Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze della Terra, della Vita e dell'Ambiente

Ciclo XXIX

Settore Concorsuale di afferenza: 05/B1
Settore Scientifico disciplinare: BIO/08

GENETIC EPIDEMIOLOGY OF TASTE PERCEPTION AND CIGARETTE USE

Presentata da: Davide Risso

Coordinatore Dottorato
Prof.ssa Barbara Mantovani

Relatore
Prof.ssa Donata Luiselli

Esame finale anno 2017

1. Introduction……………………………………………………………………………………4

1.1 Adult Tobacco Use among Ethnic Groups living in the United States …….4

1.1.1 The health consequences of smoking……………………………………………………4

1.1.2 Patterns of cigarette smoking……………………………………………………………7

1.1.3 Cigarette flavorings………………………………………………………………………9

1.2 Human taste genetics………………………………………………………………………12

1.2.1 Our sense of taste…………………………………………………………………………12

1.2.2 Molecular mechanisms of taste…………………………………………………………14

1.2.3 PTC/PROP genetics………………………………………………………………………16

1.2.4 Genetics of menthol perception…………………………………………………………18

2. Abstract…………………………………………………………………………………………19

3. Materials and Methods………………………………………………………………………20

3.1 Research Participants and collection of phenotypes…………………………………20

3.1.1 Schroeder population………………………………………………………………………20

3.1.2 Dallas Heart Study…………………………………………………………………………21

3.1.3 Dallas Biobank……………………………………………………………………………21

3.1.4 Georgia population………………………………………………………………………22

3.2 Genes selection………………………………………………………………………………22

3.3 Molecular biology……………………………………………………………………………23

3.3.1 DNA collection, purification and quantification……………………………………23

3.3.2 DNA sequencing…………………………………………………………………………24

3.4 Statistical analyses…………………………………………………………………………28

3.4.1 Descriptive statistics……………………………………………………………………28

3.4.2 Haplotypes inference 29

3.4.3 Measurements of admixture proportions 29

3.4.4 Genotype-Phenotype association analyses 30

3.4.5 Meta-analysis 30

4. Results 31

4.1 Smoking behaviors in the Georgia, DHS and Biobank cohorts 31

4.2 Menthol cigarette preference among African-Americans and European-Americans 33

4.3 Polymorphic variations at the TRPM8, TRPA1 and TAS2R38 loci 34

4.4 Admixture levels among African-American smokers 36

4.5 Associations between TAS2R38 haplotypes, PTC tasting-status and smoking 38

4.6 Meta-analysis of the association between TAS2R38 variants and menthol cigarette preference 41

4.7 Associations between TRPM8, TRPA1 variants and menthol cigarette smoking 42

5. Discussion and Conclusion 43

6. Appendix 48

6.1 Supplementary tables 48

6.1.1 Appendix Table 1 48

6.1.2 Appendix Table 2 49

6.1.3 Appendix Table 3 50

6.2 Most relevant publications during the Ph.D. course 51

7. References 133

8. Acknowledgments 140

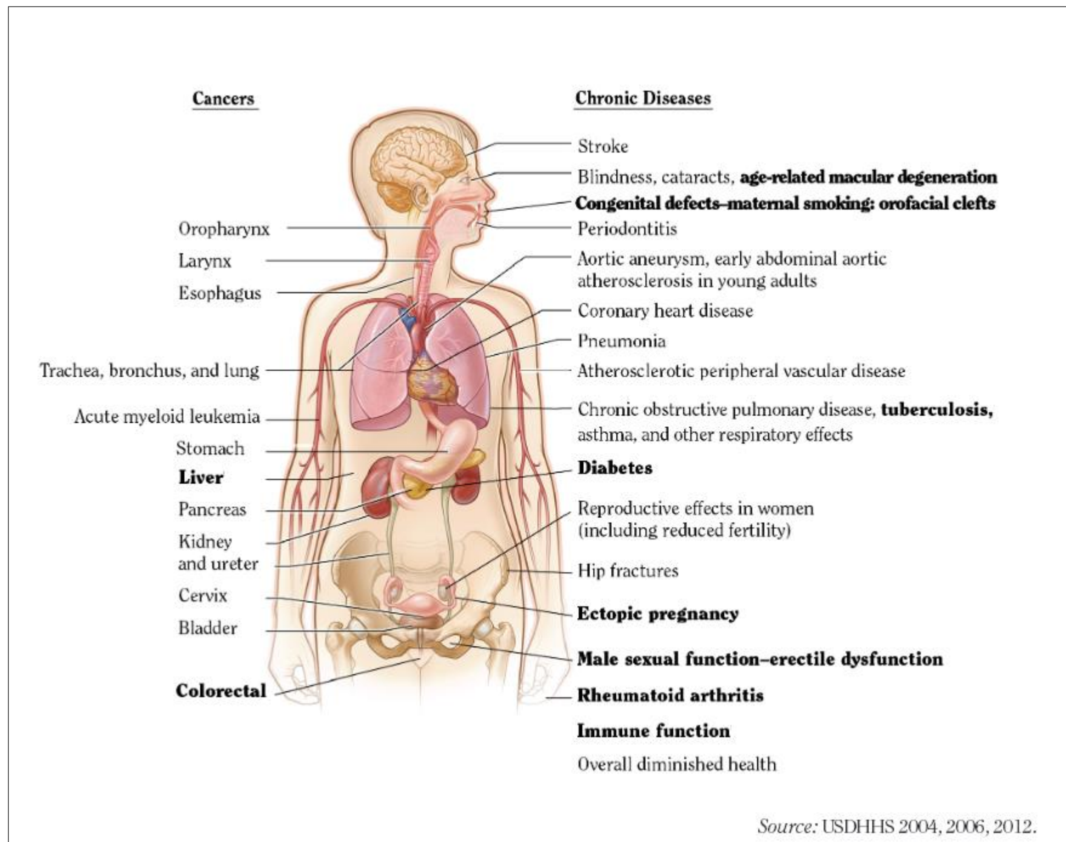
1. Introduction

1.1 Adult Tobacco Use among Ethnic Groups living in the United States

1.1.1 The health consequences of smoking

Cigarette smoking is the leading cause of preventable disease and death in the United States, resulting in more than 480,000 premature deaths every year, or one every five deaths (US Department of Health and Human Services, 2014). In addition, more than 16 million Americans live with a smoking-related disease (U.S. Department of Health and Human Services, 2014). Since the first report published in 1964 (Luther L. Terry, 1964) showing a correlation between smoking and lung cancer, thousands of scientific articles have been published documenting the diverse and serious health consequences linked to smoking (Figure 1).

Figure 1. The health consequences linked to smoking.



Smoking is in fact a risk factor for autoimmune diseases, since it compromises the immune system, making smokers more likely to have respiratory infections. Moreover, smokers are on average 30% to 40% more likely to develop type 2 diabetes when compared to non-smokers. Together with other factors, such as weight, activity level and alcohol consumption, smoking also increases the risk for osteoporosis, which is a major cause of hip fractures that have a major impact on health and longevity in the elderly (www.BeTobaccoFree.gov).

Understanding the composition of cigarettes facilitates the understanding of the correlation between smoking and its many related diseases. There are about six hundred ingredients in cigarettes which, when burned, produce more than 7,000 chemicals. Of these, at least 69 are known to have carcinogenic properties and more than 400 are known to be toxic (American Lung Association, 2010). Smoking cigarettes is the number-one risk factor for lung cancer, and it substantially raises the risk of cancer in other tissues such as trachea, bronchus, esophagus, lips, nasal cavities, bladder, liver, colon and rectum. Some of the most common chemicals found in tobacco smoke are shown in Figure 2. Surprisingly, most of these compounds are also found elsewhere. Acetone, for example, is found in nail polish removers. Acetic acid is an ingredient in hair dyes, while ammonia is a common household cleaner. Arsenic is used in rat poison and benzene is found in rubber cement. Radon and polonium-210 are both radioactive elements. However, the most dangerous compound found in cigarettes is perhaps nicotine, a neuro-active alkaloid that is highly addictive. In small doses, it acts as a brain stimulant however, in large doses, it acts as a depressant, inhibiting the flow of signals between nerve cells (Pomerleau and Rosecrans, 1989).
These chemicals, together with the other ~7000 produced when smoking, damage blood cells, impair heart function, and increase the risks for atherosclerosis, aneurysms and several cardiovascular diseases (including coronary heart diseases, heart attacks, strokes and high blood pressure). In addition, cigarette smoking damages breathing and produces scarring in the lungs, causing diseases such as pneumonia, asthma and increased risk of tuberculosis.

Lastly, secondhand smoke (the combination of smoke from the burning end of a cigarette and the smoke exhaled by smokers) causes numerous health problems in non-smokers and is especially harmful to young children (U.S. Department of Health and Human Service, 2006). Second hand smoke is responsible for between 150,000 and 300,000 lower respiratory tract infections in infants and children under 18 months of age, resulting in between 7,500 and 15,000 hospitalizations each year, and is estimated to cause 430 sudden infant death syndrome (SIDS) cases in the United States annually. It may also aggravate symptoms in 400,000 to 1,000,000 children with asthma and cause buildup of fluid in the middle ear, resulting in 790,000 physician office visits per year (California Environmental Protection Agency, 2005)

Figure 2. Some of the most common ingredients found in cigarettes (modified from https://www.cancer.ie/reduce-your-risk/smoking/health-risks/whats-in-cigarettes#sthash.QbnMb2Sn.dpbs).

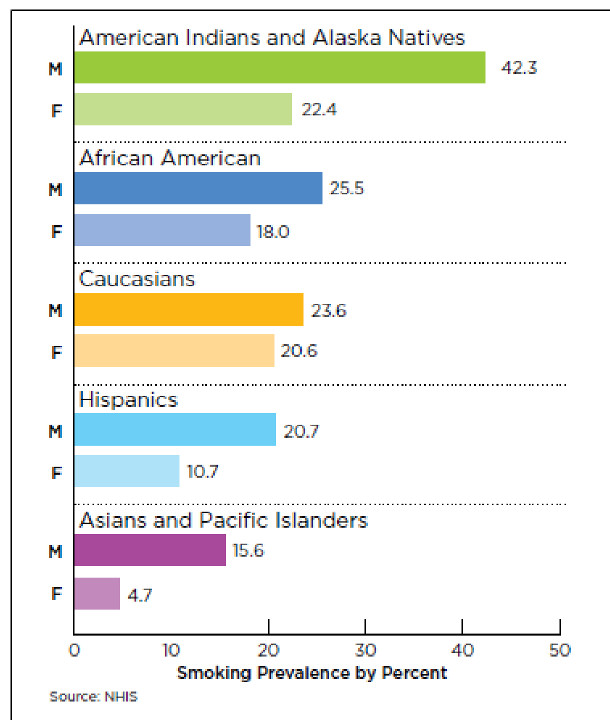
1.1.2 Patterns of cigarette smoking

Despite significant declines during the past 30 years, millions of adults in the Unites States smoke cigarettes (Agaku et al. 2014). Nearly 17 of every 100 U.S. adults (17%) currently smoke, which translates into 40 million adults in the United States who smoke cigarettes. There has been a steady decline from the smoking rate of 42% of adults in the U.S. in 1965, to 20.9% in 2005. Among U.S. adults, men are more likely to be current smokers than women: nearly 19% of adult men smoke, compared to 5% of adult women. Cigarette smoking is also higher among younger age groups compared to those aged 65 years and older. Nearly 17% of adults aged 18–24 years smoke, 20% of adults aged 25–44 years smoke, and 18% of adults aged 45–64 years smoke, however only 8.5% of adults aged 65 years and older smoke. Different patterns of cigarette smoking are also observed among different ethnicities. As shown in Figure 3, current cigarette smoking is highest among American Indian and Alaska Natives. On average, 32.35% of American Indians and Alaskan Natives are reported to smoke cigarettes. A recent study also found that American Indian and Alaska Natives have a higher percentage of smoking-related deaths from heart disease and stroke, when compared to European-American men and women (Mowery et al. 2015). One possibility that could explain this notable difference in cigarette use is that tobacco products sold on American Indian and Native Alaskan lands are not subject to state and local taxes. For this reason, cigarettes and some other tobacco products are available to American Indians and Alaska Natives at much lower prices compared to other Americans. In addition, several American tribes consider tobacco a sacred gift and use it as a traditional medicine and during religious ceremonies (Reading, 1998).

Lowest rates of cigarette smoking are currently found among Asian and Pacific Islanders, in which groups an average of 10.15% are smokers. Other ethnicities, such as African-Americans, European-Americans and Hispanics showed similar percentages of smokers (21.75%, 22.1% and 15.7% respectively) (American Lung Association, 2010). However, on average, European-American men tend to consume more (~30-40%) cigarettes per year than African-American men (Centers for Disease Control and Prevention, 2008). In addition, European-

Americans are less likely to develop lung cancers compared to African-Americans. These differences could be due to either environmental (such as disparities in access to healthcare and socio-economic status) or genetic factors. Other differences in cigarette smoking, are associated with differences such as in education. Current cigarette smoking rates are higher among people with no higher education or diploma, and lowest among those with a graduate degree. Nearly 23% of adults with 12 or fewer years of education (e.g. no higher education) are smokers, while only 5.4% of individuals with a graduate degree smoke. Cigarette smoking also varies by income status and U.S. Census Region. 26.3% of people living below the poverty level are smokers, compared to 15.2% of adults who live at or above the poverty level. Smokers are also more concentrated in the Midwest (20.7%), rather than in the West (13.1%) or in the East (16.8%) (Centers for Disease Control and Prevention, 2014).

Figure 3. Smoking prevalence by different ethnicity (Modified from American Lung Association, 2010).



1.1.3 Cigarette flavorings

On June 22nd, 2009, the Family Smoking Prevention and Tobacco Control Act gave the Food and Drug Administration (FDA) authority to regulate the manufacture, distribution and marketing of tobacco products (www.fda.gov). The result of this Act was to ban the use of all flavors other than menthol in cigarettes. However, there are currently no restrictions in terms of banning flavors in other tobacco products (such as little cigars and smokeless tobacco) (H.R. 1256—11th Congress, 2012).

Menthol is an organic compound that could be created synthetically or extracted from plants. Perhaps the most widely recognized property of menthol is its cooling effect on the mouth and throat, which arises from the fact that menthol acts as a chemical agonist of receptors that normally act to sense cold temperatures (Bautista et al. 2007). This cooling effect is especially important in terms of smoking as it reduces the harshness of cigarette smoke (Verger and McCandless, 2011). Because of this, the tobacco industry started to add menthol to cigarettes in the 1920s and 1930s, but it wasn't until the 1950s and 1960s that the practice became widespread (Tobacco Products Scientific Advisory Committee, 2011). Today menthol cigarettes represent a significant part of the U.S. cigarette market, making up to 32.0% of all cigarettes sold (Federal Trade Commission, 2013). About 20 million people in the U.S. smoke menthol cigarettes (Substance Abuse and Mental Health Services Administration, 2011).

Menthol cigarette use varies greatly depending on different factors. For example, data have shown that menthol cigarette use is higher in youth than in either young adults or adults. 56.7% of young smokers are reported to smoke menthol cigarettes, compared to 35.2% of adult smokers (Giovino et al. 2013). Several studies have shown that menthol flavoring contributes to addiction in young smokers (Sutton and Robison, 2004; Tobacco Products Scientific Advisory Committee, 2011) and that young smokers are significantly more likely to be addicted to nicotine than non-menthol cigarette smokers (Hersey et al. 2006; Nonnemaker et al. 2013). Moreover, it has been shown that adult menthol smokers are less likely to successfully quit smoking, when compared to non-menthol cigarettes smokers (Cubbin et al. 2010; Hyland et al. 2002; Fu et al.

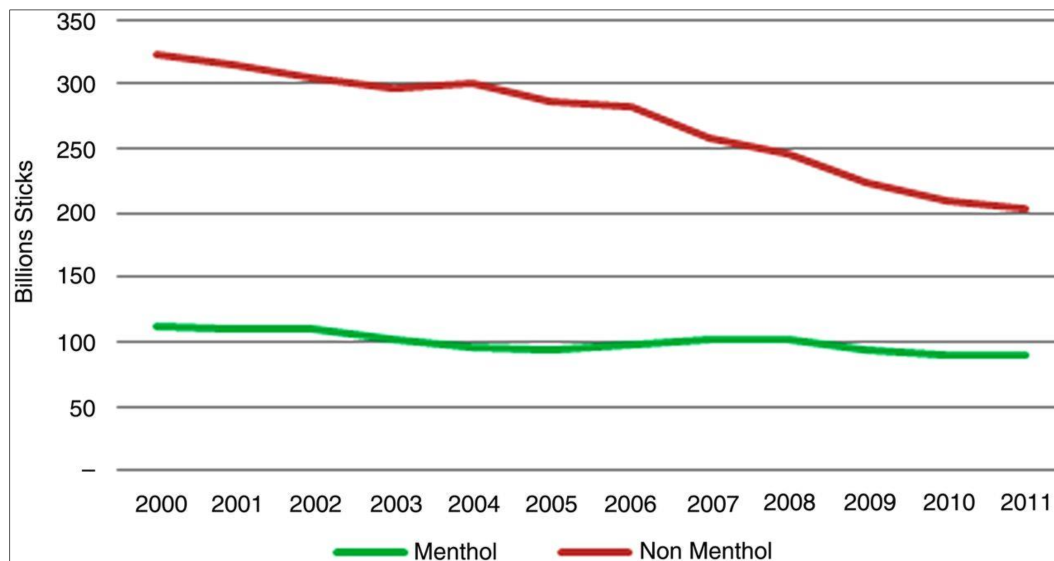
2008), suggesting that menthol cigarettes are harder to quit for everyone, regardless of age.

A remarkable difference in menthol cigarette exists among different ethnic groups. Among African-Americans who smoke, an average of 84.5% are reported to use menthol cigarettes, compared to only 26.9% of European-Americans and 33.6% and of Asian-American smokers. (Substance Abuse and Mental Health Services Administration, 2011; American Lung Association, 2010). African-American smokers are nearly 11 times more likely to use menthol cigarettes than European-American smokers. This is of importance, since there are several lines of evidence that cigarette manufacturers have specifically targeted African Americans with menthol cigarette advertising (U.S. Department of Health and Human Services, 1998; Gardiner, 2004). For this reason, around 20% of Americans believe that menthol should be banned from cigarettes (Pearson et al. 2012). Studies have shown that if menthol cigarettes were banned, around 40% of all smokers would try to quit smoking, with higher percentages (64.6%) among youth adult menthol cigarette smokers (Pearson et al. 2012). This makes studying the causes of menthol cigarette preference among smokers of great public health importance, as it is estimated that around 320,000 smoking-attributable deaths could be avoided by 2050 if a menthol ban went into effect (Levy et al. 2011).

These data are confirmed and strengthen by two recent independent reports. The first was issued by the Tobacco Products Scientific Committee (TPSAC) (2011) and the second by the Food and Drug Administration (FDA) (2013). Both these reports highlighted how menthol increases experimentation and progression to regular smoking and makes it harder for smokers (especially African-American smokers) to quit. In addition, a survey of menthol and non-menthol cigarette consumption in the USA (Delnevo et al. 2014) showed that the decline in cigarette consumption was greater among non-menthol cigarettes, compared to menthol cigarettes (37% vs. 20%) (Figure 4). An explanation that was proposed involved higher rates of initiation for menthol compared to non-menthol cigarettes, lower cessation among menthol smokers, increased consumption among current menthol smokers, and switching from non-menthol to menthol cigarettes.

Finally, it has been suggested that taste preference may play a role in cigarette smoking and/or the addiction of flavoring agents to cigarettes. This is because nicotine generally tastes unpleasantly bitter and evokes irritant sensations (Oliveira-Maia et al. 2009) that are masked and reduced by the addition of menthol to cigarettes, which elicits cooling and fresh effects.

Figure 4. Menthol and non-menthol cigarette consumption in the USA, 2000–2011 (from Delnevo et al. 2014).



1.2 Human taste genetics

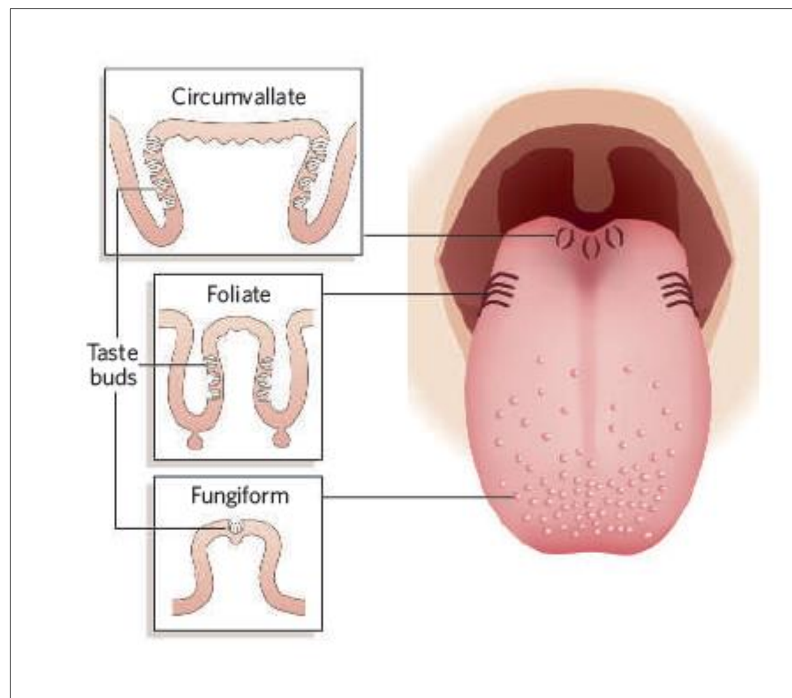
1.2.1 Our sense of taste

The sense of taste helps organisms to identify, choose and consume desirable nutrients, while avoiding harmful and potentially toxic compounds. So far, five tastes (e.g. bitter, sweet, umami, salty and sour) have been universally recognized in humans, while a sixth taste (i.e. fat) is not agreed upon by all scientists as a distinct taste perception quality (Besnard et al. 2016). Each taste has evolved in response to specific environmental stimuli. Bitter taste, for example, is believed to have evolved to prevent the ingestion of an enormous array of potentially toxic compounds, since such compounds typically taste bitter and usually evoke innate aversive behavior in different animal species. Many of these compounds are contained in plants and are produced by micro-organisms, and may generate different bitter reactions in species-specific bitter taste receptors (Bachmanov et al. 2014). On the other hand, sweet and umami taste have evolved to recognize a limited subset of nutrients. Sweet taste is universally appreciated by humans and signals the presence of carbohydrates that serve as an energy source. Although umami is lesser known than sweet taste, it has been equally important during the course of our evolution. It's the taste of L-glutamate and other L-amino acids (including aspartate and phenylalanine) which reflect a food's protein content. Salty taste is essential for maintaining body's water balance and blood circulation, governs the intake of Na^+ and other salts and the proper dietary electrolyte balance. Lastly, sour taste signals the presence of dietary acids and is generally aversive. This mechanism allows us to avoid ingesting excess acids and unbalancing the acid-base balance of the body (Bachmanov et al. 2014).

Taste-receptor cells (TRCs) represent the anatomical substrates and units of taste detection. 50 to 150 of these cells are assembled into taste buds, which are distributed across different papillae of the tongue and palate epithelium (Chandrashekar et al. 2006). Humans have three major kinds of papillae, located on different parts of the tongue. Circumvallate papillae are located at the very back of the tongue and contain thousands of taste buds. Foliate papillae are located at the posterior lateral edge of the tongue and contain a lower amount

(dozens to hundreds) of taste buds. Finally, fungiform papillae have one or a few taste buds and are found in the anterior two-thirds of the tongue (Chandrashekar et al. 2006) (Figure 5).

Figure 5. Location of human taste papillae on the tongue (Adapted from Chandrashekar et al. 2006).



The interaction between taste buds and ligands/tastants occurs through taste pores, created when taste-receptor cells converge to create a pore and project microvilli to the apical surface of the taste bud. Interaction of tastants on the surface of taste cells results in an excitation that is transmitted via afferent gustatory nerves to the brain, where it evokes taste perception which has several characteristics, including with intensity, quality and hedonics (Bachmanov et al. 2014).

1.2.2 Molecular mechanisms of taste

Taste sensitivity to bitter, sweet and umami is mediated by seven-trans-membrane domain (7TM) G-protein coupled receptors (GPCRs). GPCRs are a large protein family of receptors that sense molecules outside the cell and activate signal transduction pathways and responses inside the cell. These components are expressed in taste cells within taste buds on the tongue (Gilbertson et al. 2000; Smith and Margolskee, 2001) and in other extra-oral tissues such as gut, lungs and testis (Finger and Kinnamon 2011), where they modulate systemic functions of tastants either endogenously produced or contained in foods (Clark et al. 2015; Shaik et al. 2016). The bitter taste receptor family is designated as *TAS2R* (formerly *T2R*) and is composed of 25 functional human bitter taste receptors coded by *TAS2R* genes (Chandrashekar et al. 2000; Bufe et al. 2002). In addition, humans have 11 *TAS2R* pseudogenes (Go et al. 2005). The sweet and umami taste receptor family is much smaller, encoded by only three receptor genes. The genes encoding for these receptors are named *TAS1R1*, *TAS1R2* and *TAS1R3* (formerly designated *T1R*'s) whose products form heterodimers that serve as either sweet (*TAS1R2+TAS1R3*) or umami (*TAS1R1+TAS1R3*) receptors (Li et al. 2002). When either *TAS2R* or *TAS1R* receptors bind to taste molecules, they activate heterotrimeric GTP-binding proteins. This frees different G subunits (G α-gustducin and Gγ 13, for example) which functionally interact with phospholipase-C-beta-2 (PLCβ2), a phospholipase that stimulates the synthesis of inositol phosphate three (IP3) as a second messenger. This opens different ion channels on the endoplasmic reticulum which ultimately release Ca2+ into the cytosol of the receptor cells. The elevated intracellular Ca2+ acts to open two different ion channels found in receptor cells: *TRPM5* and a gap junction hemichannel causing a strong depolarization that permits the secretion of ATP and other taste bud transmitters into the extracellular space (Chaudhari and Roper, 2010). This leads to the transmission of neural impulses to the brain via the chorda tympani and glossopharyngeal nerves (Hellekant et al. 1997).

The molecular mechanisms of sour and salty tastes, as well as the genes encoding their receptors, are less well understood. Some lines of evidence have shown that

salt perception may be mediated by one or multiple ion channels. The ions of salt, especially sodium (Na+), can pass directly through ion channels in the tongue, leading to an action potential. One receptor that is thought to be involved in salty taste perception is the amiloride-sensitive epithelial Na channel, ENaC (Stähler et al. 2008). It is thought that an acid sensor, such one or several hydrogen ion channels, could mediate sour taste. However, despite many attempts has been made to identify the sour taste receptor, the molecular mechanisms that underlie sour perception are still uncertain (Kim et al. 2004).

1.2.3 PTC/PROP genetics

The simplest and best understood taste variation in humans is the ability to taste phenylthiocarbamide and 6-n-propylthiouracil. These compounds, otherwise known as PTC and PROP respectively, are well-known because of their ability to generate a bitter taste in some individuals but not in others. This chance discovery was made more than 70 years ago, by the chemist Arthur Fox who was pouring some powdered PTC into a bottle. Noticing that only some people started complaining about a bitter taste in the air, Fox asked different people to taste this chemical and report how it tasted. In this way, he discovered that some people could not perceive any taste at all ("non-tasters"), while others could recognize a strong bitter taste ("tasters") (Fox, 1932). Since then numerous family, twin and population studies have shown that the inability to taste PTC is inherited in a nearly Mendelian recessive manner (Blakeslee, 1932; Drayna et al. 2003; Knaapila et al. 2012). In 2003, a locus that explained approximately 75% of the variation in PTC sensitivity was identified on chromosome 7 (Drayna et al. 2003). At this locus, variation in the *TAS2R38* bitter receptor gene was subsequently found to underlie all the bimodal distribution of this phenotype and to explain >70% of the total phenotypic variance (Kim et al. 2003). A subsequent study replicated this finding and confirmed that this genetic variation also underlies the differential sensitivity to PROP (Bufe et al. 2005). *TAS2R38* is a member of the *TAS2R* bitter taste receptor gene family, which in humans consists of 25 functional genes and 11 pseudogenes, many of which show signatures of natural selection (Campbell et al. 2012, Risso et al. 2014, Risso et al. 2016). Like many other genes encoding GPCRs, *TAS2R38* has only a single coding exon of 1002 bp in length. Within this coding sequence, three single nucleotide polymorphisms (*rs714598*, *rs1726866*, *rs10246939*) at positions encoding amino acids 49, 262 and 296 represent the most common variant alleles of *TAS2R38*, and comprise the "taster" PAV (Proline, Alanine, Valine) and "non-taster" AVI (Alanine, Valine, Isoleucine) haplotypes. In addition, two uncommon (frequency < 5%) (AAV and AAI) and four rare (frequency < 1%) haplotypes (PAI, PVI, AVV and PVV) carrying different combinations of these three variant amino acids have been identified. Analysis of the population

genetics of *TAS2R38*, showed that approximately ~40% of worldwide individuals were AVI-carriers (Wooding et al. 2004; Risso et al. 2015). This represented a paradox since it is thought that bitter taste has evolved to prevent ingestion of potentially toxic or harmful compounds found in plants. How a presumably non-functional allele rose to such a high frequency in the population was not clear. Two major explanations have been presented to explain the unexpectedly high frequency of the non-taster AVI haplotype. Authors of several papers (Fisher et al. 1939; Wooding et al. 2004; Campbell et al. 2012) have hypothesized that both the taster and non-tasters form were maintained by balancing natural selection, suggesting that the AVI receptor could encode a functional receptor for some other toxic bitter compounds present in nature and not yet identified (Drayna, 2005). Another explanation is that genetic drift could explain the present-day distribution of *TAS2R38* forms, which reached their present population frequencies and distribution due to random fluctuations and could be due to demographic events, rather than selective ones (Wang et al. 2004; Risso et al. 2016).

TAS2R38 haplotypes have been hypothesized to influence smoking habits and nicotine dependence, since it has been shown that this gene has a lower expression in smokers, when compared to non-smoker individuals (Aoki et al. 2014). However, the results of previous studies have been conflicting. For example, a study examining both African-American and European-American individuals found a significant association between *TAS2R38* haplotypes and smoking, with the non-taster AVI haplotype being positively associated with smoking quantity and nicotine dependence, however this was seen only in African Americans (Mangold et al. 2008). Another study analyzed German participants and found that individuals carrying the PAV taster haplotype smoked significantly fewer cigarettes per day (Keller et al. 2008). In contrast, another study of individuals of European descent found no association between the PAV or AVI haplotypes and smoking (Cannon et al. 2005). However, this study found that the rare AAV haplotype was associated with a lower incidence of smoking.

1.2.4 Genetics of menthol perception

Another gene family codes for a different kind of temperature and ligand receptors. The products of this transient-receptor-potential (TRP) channel family can detect a wide range of changes in ambient temperature, from extremely cold to painfully hot (Jordt et al. 2003). In addition, different genes of this family encode receptors for ligands that can elicit specific psychophysical sensations, such as the cold sensation generated by menthol. In this regard, it has been shown that the eight members of the TRP M-subfamily, namely *TRPM*8, is activated by menthol and other chemical cooling agents when temperatures drop below ∼26 °C, suggesting that the menthol receptor is also the principal detector of environmental cold (Bautista et al. 2007). This gene is located on chromosome 2 and is composed by 27 exons, spanning a total genomic distance of 102,161 base pairs. Because menthol, eucalyptus oil and other similar cooling agents act as chemical agonists of a receptor that normally senses environmental cold, they provide a sensation that food containing these compound is cold. Moreover, menthol also binds to a kappa opioid receptor (Galeotti et al. 2002), which produces a numbing effect and decreases the effects of inflammation.

Interestingly, it has been discovered that *TRPA1*, another member of the TRP family, is also involved in pain and cold perception (del Camino et al. 2010). This gene is located on chromosome 8 and composed of 27 exons with a total size of 70,093 base pairs. Moreover, studies carried out in mice demonstrated that the protein encoded by this gene is a highly sensitive menthol receptor (Karashima et al. 2007). Analysis of human phenotypes and genotypes confirmed this initial hypothesis with a focus on menthol cigarette smoking. For example, a recent study showed that *TRPA1* haplotypes are highly associated with menthol preference in European-American smokers (Uhl et al. 2011).

In addition, polymorphisms in the *TAS2R38* bitter taste receptor gene have been recently associated with menthol cigarette smoking. A study examining 323 European-American pregnant smokers identified an association between the taster haplotype (i.e. PAV) of this gene and a preference for menthol cigarettes (54% vs. 30%, $P<0.001$) (Oncken et al. 2015).

2. Abstract

Tobacco use is a major worldwide health problem and is a leading cause of preventable disease. Cigarettes and other tobacco products contain bitter compounds including nicotine, which contribute to the chemosensory properties of tobacco and stimulate multiple sensory systems, including taste transduction pathways. Since bitter taste has evolved to identify potentially toxic compounds, and thus protect against harmful foods, our hypothesis is that aversion to this taste may prevent smoking and nicotine dependence.

The goal of this research was to investigate the role of inherited differences in taste perception in smoking behaviors. We sought to determine whether such genetic variation could account for the well-known differences in flavored tobacco use among different U.S. ethnic groups. For example, around 80% of African-American smokers report that they prefer menthol cigarettes, compared to only 30% of European-American smokers who express this preference. We recruited subjects from four different populations, comprising a total of 9871 individuals (8191 African-Americans and 1590 European-Americans), purified DNA's from some saliva or blood samples and used a candidate gene (e.g. Sanger sequencing) approach to sequence taste-related genes.

We identified several genetic associations between polymorphisms in taste-related genes and different smoking behaviors. We have shown that the frequency of the *TAS2R38* taster haplotype differs between smokers and non-smokers in European-American populations (37.0% vs. 44.0% in non-smokers, P=0.003). In addition, we identified two SNPs, one located in the menthol receptor *TRPM8* and one in the menthol-reactive gene *TRPA1*, that are strongly associated with menthol smoking in a study group of African-Americans (OR=6.1, P=3.12E-09, OR=2.52 OR=0.5, P=1.51E-03 and OR=0.5 and P=1.22E-05). Moreover, we found that the taster haplotype (i.e. PAV) of the *TAS2R38* bitter taste receptor gene is less common in European-American smokers (37.0% vs. 44.0% in non-smokers, P=0.003) and that the non-taster haplotype of this gene (i.e. AVI) is significantly lower in menthol smokers compared to non-menthol smokers (29.8% vs. 37.7%, P=0.008).

Overall, these findings support the hypothesis that variations in taste-related genes play a role in the choice of cigarettes when smoking. Understanding genetic differences in taste perception in tobacco use could help inform the development of more effective tobacco control policies.

3. Materials and Methods

3.1 Research Participants and collection of phenotypes

A total of 9871 individuals (8191 African-Americans and 1590 European-Americans) including both smokers (menthol and non-menthol cigarettes) and non-smokers, were analyzed. They belonged to four different cohorts, as described below. Subjects were enrolled with written informed consent under protocols approved for human subjects research protection at all participating sites.

3.1.1 Schroeder population

A total of 718 Washington DC resident smokers were recruited through the DC Tobacco Quitline (DCQL), a smoking cessation program based at The Schroeder Center in Washington, DC (https://www.tobaccocontrolresearch.org/). All subjects were African-Americans aged > 18 years and a majority (56.5%) were menthol cigarette smokers. Information on general variables, such as gender, cigarettes per day (CPD), marital status and education level was obtained from all subjects. In addition, we collected data from the brief Wisconsin Inventory of Smoking Dependence Motives (WIDSM), a 37-item scale psychometric instrument designed to evaluate the extent of smoking motivation and nicotine dependence (Piper et al. 2008). Scores for each of the 11 subscales of the WISDM scale were calculated by taking the average item scores, and the primary dependence score (PDM) and secondary dependence score (SDM) were computed. Study procedures were approved by the Western IRB and by the National Institutes of Health Combined Neurosciences IRB under protocol 01-DC-0230.

3.1.2 Dallas Heart Study (DHS)

The Dallas Heart Study is a multiethnic population-based probability sample of Dallas County, Texas residents (Victor et al. 2004). The original cohort (DHS-1) was enrolled between 2000 and 2002 and all participants, as well as their spouses or significant others, were invited for a repeat evaluation in 2007–2009 (DHS-2). During each visit, participants completed a detailed survey including questions regarding demographics, socioeconomic status, medical history, and lifestyle factors (including tobacco use), and underwent a health examination. Ethnicity was self-assigned. A total of 2,363 African-American and 1,353 European-American DHS participants with available genotype and smoking phenotype data were included in the present study. Current smokers were defined as individuals who smoked at least 100 cigarettes in their lifetime and smoked on at least some days in the previous 30 days. Smoking quantity was defined as a categorical variable in all cohorts, sub-dividing smokers into three groups (less than 6 cigarettes per day, 6–19 cigarettes per day, and 20 or more cigarettes per day). The study was approved by the Institutional Review Board of University of Texas Southwestern Medical Center and all participants provided written informed consent.

3.1.3 Dallas Biobank

The Dallas Biobank is a repository of DNA and plasma samples from individuals ascertained at various locations in north-central Texas. The present study includes a total of 4,973 African American Biobank participants for whom the genotype and smoking phenotype data were available. All participants were over 18 years of age and signed an informed consent that was approved by the University of Texas Southwestern Medical Center IRB. Current smokers were defined as people who identified themselves as smokers and said they were currently using tobacco products.

3.1.4 Georgia population

A total of 237 European-Americans was chosen based on their tobacco product usage from a longitudinal study involving young adults attending seven Georgia colleges or universities (Berg et al. 2016). Variables including sex, age and current smoking status were obtained from all participants during a web-based baseline survey in the fall of 2014. Smoking status information was also obtained again in Spring 2015. Individuals were defined as current smokers if they reported to have smoked in the past 30 days. In the spring of 2015, participants were sent a commercial taste-strip containing the bitter compound phenylthiocarbamide (PTC, Thermo Fisher Scientific Inc., Catalog Number: S85287A) via mail with instructions regarding how to complete the taste-strip test. The responses were recorded by participants and returned, and participants were then defined as "tasters" if they categorized the taste of the PTC papers-strip to be "mild or strong" and as non-tasters if they reported "no taste". Signed informed consent was approved by the Institutional Review Boards of Emory University, ICF Macro International, Albany State University, Berry College, University of North Georgia, and Valdosta State University.

3.2 Genes selection

We selected a total of 27 genes that encode products known to participate in human taste perception, designated taste-related genes, and performed a candidate-gene sequencing approach. In particular, we studied two menthol receptor genes (i.e. *TRPM8* and *TRPA1*) and 25 bitter taste receptor genes (*TAS2R*s). Of these 25 TAS2Rs, we particularly focused on one member of this family (i.e. *TAS2R38*) because of previous data in the literature suggesting possible associations between variations in this gene and different smoking behaviors (Cannon et al. 2005, Mangold et al. 2008, Uhl et al. 2011, Keller et al. 2013, Oncken et al. 2015). We mainly focused our attention on the coding regions of these genes, however for TRPM*8* and *TRPA1* we additionally selected evolutionary conserved intronic regions to assay most of the variation in these

non-coding areas. To identify them these regions, we used the comparative genomics tool of the UCSC Genome Browser (https://genome.ucsc.edu/).

3.3 Molecular biology

3.3.1 DNA collection, purification and quantification

DNA from the Schroeder and Georgia populations was collected using Oragene saliva collection kits and extracted with an ethanol precipitation protocol, according to the manufacturer's instructions (http://www.dnagenotek.com/US/pdf/PD-PR-052.pdf, Genotek Inc., Kanata, Ontario, Canada), as follows:

1. Mix the sample in the DNA Genotek kit by inversion and gentle shaking for a few seconds.
2. Incubate the sample at 50°C in a water incubator for a minimum of 1 hour or in an air incubator for a minimum of 2 hours.
3. Add 20 µL of PT-L2P to each well of the plate.
4. Transfer 5 µL of Blue Dextran (1 mg/mL) to each well of the plate.
5. Transfer 500 µL of sample to each well.
6. Cover plate with adhesive cover sheet or reusable mat. Press into place to seal. Mix manually by inversion 5 times.
7. Incubate at -20°C for 10 minutes.
8. Centrifuge plate at room temperature for 10 minutes at 4,200 x g.
9. During centrifugation, label a second 96-well plate.
10. When centrifuge has stopped, transfer 450 µL of the supernatant from the first plate to the second plate. Take care not to disrupt the pellet.
11. To the new plate containing supernatant add 350 µL of Isopropanol (room temperature) to each well.
12. Cover plate with adhesive cover sheet or reusable mat. Press into place to seal. Mix manually by slowly inverting 10 times. Incubate at room temperature for 10 minutes.
13. Centrifuge plate at room temperature for 10 minutes at 4,200 x g.
14. Carefully remove as much supernatant as possible from each well without disturbing the pellet. Discard the supernatant.
15. Add 400 µL of 70% ethanol (room temperature) to each well.
16. Cover plate with adhesive cover sheet or reusable mat. Press into place to seal. Mix by vigorous vortexing.
17. Centrifuge plate at room temperature for 10 minutes at 4,200 x g.
18. Carefully remove all the supernatant from each well without disturbing the pellet. Discard the supernatant.
19. The plate should be pulsed centrifuged for 20 seconds to collect any leftover ethanol.
20. Using a pipette carefully remove ALL residual ethanol and air-dry the plate for 5 minutes.
21. Add 50-100 µL of TE buffer to each well and cover plate with adhesive cover sheet or reusable mat.
22. Vortex vigorously to ensure any pellet on the side of the well is dislodged and re-hydrated. Place the plate on a shaker or rocker to help fully re-hydrate the DNA pellet overnight.

In the DHS and the Dallas Biobank populations, leukocytes were isolated from 40 ml of blood, and 1/3 of the cells were used to extract genomic DNA. Quantification of the DNA's was obtained through both the Quant-iT dsDNA Broad-Range Assay Kit (Invitrogen, Carlsbad, CA) and with a NanoDrop™ spectrophotometer (Thermo Fisher Scientific Inc.). The former uses an intercalating agent of DNA, PicoGreen, that is an ultra-sensitive fluorescent nucleic acid stain for quantifying in solution double-stranded DNA (dsDNA), which emits a luminous signal proportional to the length of the DNA sequence. The latter measures the level of nucleic acid absorbance at 260 and 280 nm to assess the purity of DNA.

3.3.2 DNA sequencing

Coding and intronic regions of the selected genes *TRPM8*, *TRPA1* and *TAS2R38* were completely sequenced using dideoxy Sanger sequencing (Sanger et al. 1977) in the Schroeder and Georgia populations. A polymerase chain reaction (PCR) was performed to amplify the regions of interest, using the reaction mix shown in Table 1 and the parameters shown in Table 2.

Table 1. PCR mix reagents for *TRPM8*, *TRPA1* and *TAS2R38* amplification.

Reagent	*Vol for 10 ul rxn*
Water	3.86
10 X PCR buffer	1
MgCl2 (25mM)	0.4
dNTP (10 mM)	0.2
DMSO	0.5
Forward Primer (2 uM)	1
Reverse Primer (2 uM)	1
Hotstar Taq (5U/ul)	0.04
Genomic DNA (10 ng/ul)	2
Total vol	10

Table 2. Thermal cycler setting parameters for *TRPM8*, *TRPA1* and *TAS2R38* amplification.

Step	*Time*	*Temp (°C)*	*Cycles*
Initial DNA denaturation	15 min	94	1
DNA denaturation	20 secs	94	35
Primer annealing	30 secs	56	
Extension	1 min	72	
Final extension	7 min	72	1

A complete list of the primers used to amplify the genomic regions of interest is shown in Appendix Tables 1,2,3. After this first PCR step, a PCR cleanup was performed to remove the excess of PCR primers and dNTPs, using the reaction mix shown in Table 3.

Table 3. PCR mix reagents for *TRPM8*, *TRPA1* and *TAS2R38* amplicons clean-up.

Reagent	*Vol for 10 ul*
Water	7.85
SAP buffer (10X)	2
Shrimp Alkaline Phosphatase (1 U/ul)	0.1
Exonuclease I (10 U/ul)	0.05
Total Vol	10.0

Then, forward and reverse sequencing reactions were prepared in separate tubes, using the mixture shown in Table 4, and DNA was sequenced with a BigDye® terminator dye, according to the manufacturer's protocol (https://tools.thermofisher.com/content/sfs/manuals/MAN0015666_BigDyeTerminator_V3_1_CycleSeq_QR.pdf, Thermo Fisher Scientific Inc.) and the parameters shown in Table 5.

Table 4. Reagent mix used for BigDye reactions to sequence *TRPM8*, *TRPA1* and *TAS2R38* genes.

Reagent	*Vol for 10 ul*
Water	4.3
Sequencing Buffer	2
Primer (2 uM)	1
Dimethyl Sulfoxide	0.5
Master mix	0.2
Template	2
Total Vol	10

Table 5. Thermal cycler setting parameters for *TRPM8*, *TRPA1* and *TAS2R38* sequencing.

<table>
<tr><th>Step</th><th>Time</th><th>Temp (°C)</th><th>Cycles</th></tr>
<tr><td>Initial DNA denaturation</td><td>15 min</td><td>96</td><td>1</td></tr>
<tr><td>DNA denaturation</td><td>10 secs</td><td>96</td><td rowspan="2">25</td></tr>
<tr><td>Primer annealing</td><td>4 min</td><td>60</td></tr>
</table>

A complete list of the primers used to sequence the amplified genomic regions of interest is shown in Appendix Tables 1,2,3. Products of sequencing reactions were then purified using an ethanol precipitation using the following protocol:

1. Add 2.5 ul of 125 mM EDTA
2. Add 30 ul of 100% EtOH
3. Seal the plate with aluminum foil tape and mix by vortexing
4. Incubate at room temperature for 15 min
5. Centrifuge at 3000 rpm for 30 min
6. Invert the plate over the sink to discard supernatant and spin the inverted plate at 1000 rpm
7. Add 50ul of 70 % ethanol to each well. Do not vortex
8. Centrifuge at 3000 rpm for 15 min
9. Invert the plate over the sink to discard supernatant and spin the inverted plate at 1000 rpm
10. Leave the plate for 10 min to dry it
11. Resuspend the pellet with 10 ul of Hi-Di Formamide

Finally, separation of sequencing reaction purified products based on length and fluorescent label, was performed by capillary electrophoresis on an automatic sequencer ABI 3730 DNA Analyzer (Applied Biosystems, USA).

DNA chromatograms were then analyzed and checked individually to evaluate the presence of calling errors with the Lasergene suite (DNASTAR, Madison,Wisconsin).

In the DHS and the Dallas Biobank populations, participants were previously genotyped using the Illumina Human-Exome BeadChip (http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/humanomniexpress-24/infinium_hts_assay_protocol_user_guide_15045738_a.pdf), as described in Victor et al. (2004). This genotyping produced focused coverage of exonic regions and included a total of 243,345 markers. Specific re-sequencing of candidate genes and/or variants of interests was performed in order to replicate the associations found in intronic regions in the other cohorts through Sanger sequencing.

3.4 Statistical analyses

Statistical tests focused on different smoking phenotypes. The association between *TRPM8*, *TRPA1* and *TAS2R38* gene variants and menthol cigarette smoking was tested in the Schroeder population, while associations between *TAS2R38* gene variants and cigarette smoking was examined in all the studied cohorts.

3.4.1 Descriptive statistics

R statistical analysis software (R Development Core Team 2011) was used to compare baseline characteristics of the study participants of different cohorts. Continuous variables were compared using t-tests, while categorical variables were compared using chi-square tests. In addition, estimates of some descriptive statistics were calculated using the DnaSP package (Rozas et al. 2003). In particular, we calculated the nucleotide diversity, which is defined as the probability that two randomly chosen homologous nucleotides are different in the sample, along with the average number of nucleotide differences.

As an additional measure of genetic differentiation, we calculated Wright's F_{ST} index (Wright 1969) with Arlequin v.3.5 (Excoffier et al. 2005) as follows:

$$F_{ST} = V_P/p(1-p)$$

Where p and Vp are the mean and the variance of gene frequencies between two (or more) populations, respectively. This fixation index measures the level of population differentiation due to genetic structure and ranges from 0 to 1, where values close to 0 indicate low genetic differentiations, while values approaching one suggest high genetic differentiation.

3.4.2 Haplotypes inference

The Bayesian algorithm implemented in the software PHASE v.2.1 (Stephens et al. 2001) was used to statistically infer haplotypes from genotype data, using individuals from the 1000 Genomes Phase III as a reference (1000 Genomes Project Consortium et al. 2015). Only haplotypes with posterior probability of 90% or above were considered for further analyses. In addition, we constructed haplotype networks in order to better infer and represent the evolutionary relationships among populations based on their genetic characteristics. For this purpose, we used the median joining algorithm implemented in the software Network v.4.5 (Bandelt et al. 1999).

3.4.3 Measurements of admixture proportions

To measure the admixture level of the recruited African-American individuals, 48 (24 menthol cigarette smokers and 24 non-menthol cigarette smokers) randomly chosen African-Americans from the Schroeder population were genotyped with the Illumina HumanOmni1 Chip (http://www.illumina.com/documents/products/datasheets/datasheet_human omni1_quad.pdf) which allows, on average, the assay of 1.140.419 markers, several of which are ancestry-informative. In addition, admixture levels of African-Americans from the Dallas Heart Study population (genotyped with the Illumina Human-Exome BeadChip) were also calculated. For this purpose, we used the Admixture v.1.3 software (Alexander et al. 2009) after filtering the SNPs for deviations from Hardy-Weinberg equilibrium (P>0.000001), missingness (P<0.05) and removing SNPs in high Linkage Disequilibrium (R^2>0.7). We then performed a principal component analysis (PCA) using PLINK 1.9 software (Chang et al. 2015) using populations of the 1000 Genomes population project (1000 Genomes Project Consortium et al. 2015) and of the Human Genome Diversity Project (HGDP) (Pickrell et al. 2009) as reference populations.

3.4.4 Genotype-Phenotype association analyses

PLINK 1.9 (https://www.cog-genomics.org/plink2, Chang et al. 2015) was used to perform an initial quality control of genotypes. This excluded variant with a call rate <90% or a deviation from Hardy-Weinberg equilibrium (HWE) (P<0.001). PLINK was also used to calculate measures of alleles frequencies in the examined populations and to compare minor allele frequencies (MAF) between different smoking phenotypes through the calculation of chi-square tests and resultant *p*-values. Since multiple statistical tests were performed, we adjusted the *p*-values for each SNP to reduce the possibility of false positives. We therefore applied the Bonferroni multiple testing correction, multiplying the *p*-values by the number of total SNPs tested, with the following formula:

Corrected *p*-value = *p*-value X number of tested SNPs

After this correction, we considered a corrected *p*-value < 0.05 as significant. The resultant corrected *p*-values were additionally adjusted for demographic variables such as age and sex to further reduce the possibility of false positives.

3.4.5 Meta-analysis

In order to compare our associations with previous findings, we performed a meta-analysis using PLINK 1.9 (https://www.cog-genomics.org/plink2, Chang et al. 2015) using a random effects model, to calculate *p*-values (P), odds ratios (OR) and 95% confidence intervals. The results were then shown in a forest plot, generated with R software (R Development Core Team 2011).

4. Results

4.1 Smoking behaviors in the Georgia, DHS and Biobank cohorts

Differences in smoking behavior patterns were analyzed in the Georgia, DHS and Biobank populations. Baseline and demographic characteristics of these study populations, stratified by cohort, are shown in Table 6.

Table 6. Characteristics of the study participants in the three different cohorts. DHS, Dallas Heart Study; AA, African-Americans; EA, European-Americans.

Characteristic	Georgia-EA	DHS-AA	DHS-EA	Biobank-AA
Number of participants	237	2363	1353	4973
Age, mean (SD)	20.9 (1.9)	48.2 (11.3)	50.1 (11.2)	44.8 (14.6)
Female, N (%)	121 (51.0%)`	1410 (59.7%)	724 (53.5%)	3238 (65.1%)
Smokers, N (%)	123 (51.9%)	723 (30.6%)	316 (23.4%)	1526 (30.7%)
Smoking quantity*, N (%)				
<=5 cigs/day	91 (74.0)	211 (29.5)	47 (15)	715 (53.7)
6-19 cigs/day	28 (22.7)	318 (44.5)	116 (37.1)	554 (41.6)
>=20cigs/day	4 (3.3)	186 (26)	150 (47.9)	63 (4.7)
n/r	/	8	3	194

*Smoking quantity was not available for some participants.
n/r = non-response.

The average age of the individuals of the Georgia cohort was 20.91+/-1.95. Of the 237 participants, 123 (51.9%) were current smokers and the remaining 114 (48.1%) were non-smokers. In the Georgia cohort, no differences were found in the mean age of smokers (20.6) and nonsmokers (21.2; P=0.85). A higher, but not significant, percentage of smokers than non-smokers were female (54.4% of smokers versus 47.4% of non-smokers, P=0.72). The Dallas Heart Study population was significantly older than the Georgia cohort (mean age 48.2 and 50.1 years in AA and EA participants respectively, P<0.05). The proportion of female participants was slightly higher among DHS AA than EA participants (59.7% and 53.5%, respectively, P<0.05). DHS EA participants had a lower prevalence of smoking (23.4%) than those of either the Georgia cohort (51.9%,

P<0.05) or the DHS AA participants (30.6%, P<0.05). Among DHS EA participants, smokers were on average 5 years younger than non-smokers (mean age 46.4 vs. 51.2 years, respectively, P<0.001) (Table 7). There was no difference in age between AA smokers and non-smokers. In contrast to the Georgia cohort, there was a higher proportion of women among non-smokers in both ethnicities in DHS (64.2% vs. 49.4% in AA, P<0.001; 54.6% vs. 50.0% in EA, P>0.05). The Dallas Biobank population was older than the Georgia population (mean age 44.8, P<0.001) but younger than both the DHS AA and EA participants (P<0.05). In addition, the Biobank included a higher proportion of females than either the Georgia or DHS AA/EA participants (65.1%, P<0.05). The fraction of individuals who were smokers was lower in this population compared to the Georgia cohort (30.7% vs. 51.9%, P<0.001) but similar to the DHS AA participants (P>0.05). In the Dallas Biobank population, smokers were slightly younger than non-smokers (mean age 43.8 vs. 45.3, P=0.001). In addition, smokers had a lower percentage of females than non-smokers (51.7% vs. 71.0%, P<0.001, Table 7).

Table 7. Characteristics of the Georgia, DHS and Biobank participants by smoking status.

Characteristic	Non-smokers	Smokers	*p*-value
Georgia cohort			
N	114	123	
Age, years, mean (SD)	21.2 (1.9)	20.6 (1.8)	0.85
Female, N (%)	54 (47.4%)	67 (54.5%)	0.72
DHS African American			
N	1640	723	
Age, years, mean (SD)	48.4 (11.6)	47.6 (10.4)	0.13
Female, N (%)	1053 (64.2%)	357 (49.4%)	<0.001
DHS European American			
N	1037	316	
Age, years, mean (SD)	51.2 (11.3)	46.4 (10.2)	<0.001
Female, N (%)	566 (54.6%)	158 (50.0%)	0.17
Biobank African American			
N	3447	1526	
Age, years, mean (SD)	45.3 (15.1)	43.8 (13.4)	0.001
Female, N (%)	2449 (71.0%)	789 (51.7%)	<0.001

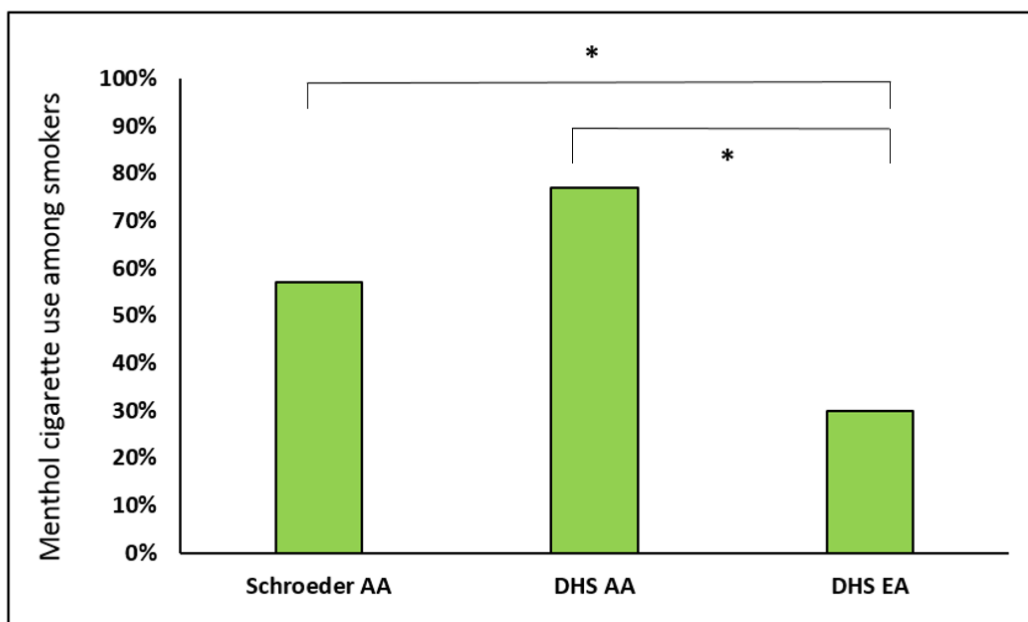
4.2 Menthol cigarette preference among African-Americans and European-Americans

Preference for menthol cigarettes was analyzed in both the Schroeder African-American population and in the DHS EA and AA populations.

The average age of the individuals belonging to the Schroeder population was 45.1 (SD=10.8) and the majority of them (406, 56.5%) were menthol smokers, with the remaining 312 (43.5%) being non-menthol smokers. A higher percentage of menthol than non-menthol smokers were female (39.6% vs. 24.2%, p=0.001), in agreement with previously reported data (Caraballo et al. 2011). No differences were found in the mean age of menthol smokers (45.4, SD=11.0) and non-menthol smokers (44.6, SD=10.6; p=0.372). Lastly, the distribution of the total WISDM items score did not differ between menthol (57.70, SD=23.36) and non-menthol (57.39, SD=23.6) smokers (p=0.86).

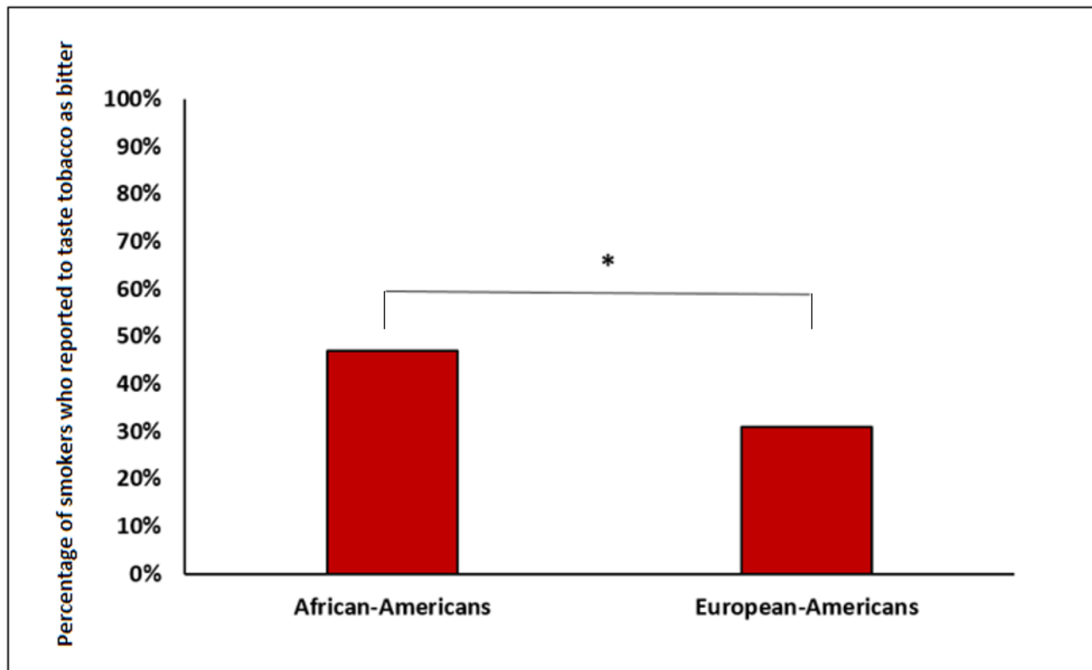
A large difference in mentholated tobacco use was found between the Schroeder and DHS African American subjects and DHS European-Americans (57% and 77% versus 30%, respectively; p=8.1e-21, Figure 6). This confirms previous findings regarding menthol cigarette use in different U.S. ethnic groups (Caraballo et al. 2011).

Figure 6. Menthol cigarette use among African-American (AA) and European-American (EA) smokers of the Schroeder and DHS populations.
* indicates significance.



Moreover, when asking how bitter tobacco tasted, African American subjects perceived tobacco products to be more bitter than non-African individuals (47% versus 31%, p=0.001, Figure 7).

Figure 6. Self-reported bitter taste sensitivity to tobacco in African-American and European-American individuals.



4.3 Polymorphic variations at the TRPM8, TRPA1 and TAS2R38 loci

TRPM8 and *TRPA1* polymorphisms were initially investigated by sequencing the entire coding regions of these genes (e.g. 27 exons each), plus evolutionary conserved intronic regions, in the Schroeder African-American individuals. As a result of this analysis, we identifed a total of 238 polymorphisms (140 at the *TRPM8* locus and 98 at *TRPA1*, respectively) with minor allele frequencies ranging from less than 0.01 to 0.50. All the identified variations were biallelic SNPs, and no insertions nor deletions were observed. In addition, 13 and 11 singletons (nucleotide changes observed only once in the total sample) were identified in these two genes, respectively. F_{ST} values indicated average low

levels of genetic differentiation at *TRPM8* and *TRPA1* loci (0.06 and 0.04, respectively), consistent with the generally low levels of nucleotide diversity (0.06 and 0.05, respectively) observed.

Eleven single nucleotide polymorphisms were found in *TAS2R38* by completely sequencing its coding region in all the analyzed populations. A complete list of identified variants, with relative minor allele frequencies, is shown in Table 8.

Table 8. Single nucleotide polymorphisms identified in the *TAS2R38* gene. Chr, chromosome; Pos, position; GRCh37, Human assembly version 37; MAF, minor allele frequency.

dbSNP ID	Chr	Pos (GRCh37)	MAF
rs10246939	7	141672604	0.48
rs114288846	7	141672670	0.02
rs1726866	7	141972905	0.34
rs140262989	7	141672752	<0.01
rs139843932	7	141673087	<0.01
rs150209521	7	141673251	<0.01
rs138869704	7	141673286	<0.01
rs115966953	7	141673299	<0.01
rs713598	7	141673345	0.47
rs141196803	7	141673384	<0.01
rs148448145	7	141673397	<0.01

Haplotype phasing of the genomic DNA sequence data predicted six major haplotypes (PAV, AVI, AAI, AAV, PVI and PAI), with frequencies of 45.47%, 33.22%, 17.90%, 2.37%, 0.76% and 0.28%, respectively. In addition, the identified rare variants (e.g. MAF<1%) contributed to form seven sub-haplotypes, illustrated in Table 9.

Table 9. Major and sub-haplotypes identified in the *TAS2R38* gene.

Major haplotype	Sub-haplotype
PAV	PAV + *rs141196803*
PAV	PAV
AAV	AAV
AAI	AAI
AAI	AAI + *rs138869704*
AAI	AAI + *rs114288846*
AAI	AAI + *rs148448145*
AAI	AAI + *rs139843932*
AAI	AAI + *rs150209521*
AVI	AVI
AVI	AVI + *rs115966953*
PVI	PVI
PAI	PAI

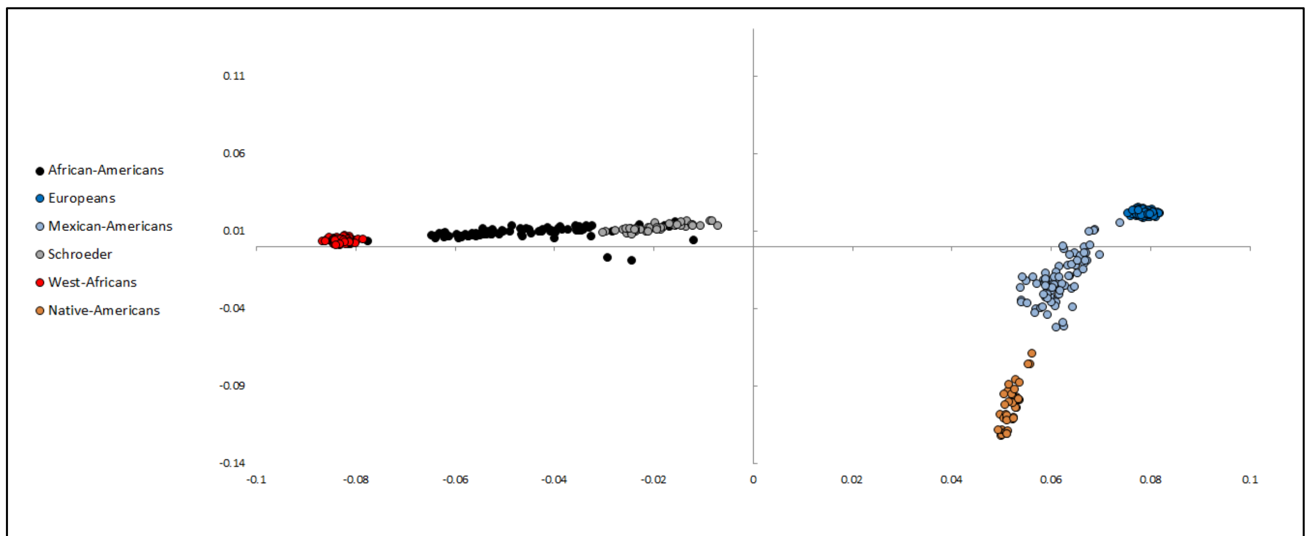
The three major *TAS2R38* SNPs (*rs714598*, *rs1726866* and *rs10246939*) showed approximately the same global F_{ST} (0.05, 0.06 and 0.04 respectively) and nucleotide diversity (0.06, 0.07 and 0.05 respectively) values.

4.4 Admixture levels among African-American smokers

Genome-wide ancestry estimates of African-Americans show, on average, proportions of ~75% African, ~24% European and ~1% Native-American ancestry, although with some differences across US states (Bryc et al. 2015). To better characterize our population sample, we calculated admixture levels in a representative sample of our African-American individuals, in order to exclude any potential ancestry differences between smokers and non-smokers that could introduce a bias in our association analyses. As shown in Figure 7, the average levels of African admixture in the Schroeder population were lower (59%) than the many previously reported levels, and these subjects carried a corresponding

higher percentage (40%) of European admixture, with an identical proportion (1%) of Native-American admixture. However, when analyzing differences between menthol cigarette smokers and non-menthol cigarette smokers, no differences were found (60% vs. 59% African ancestry, 39% vs. 40% European ancestry and 1% vs. 1% Native-American ancestry, respectively; *p*=0.98).

Figure 7. Ancestry estimates of African-Americans of the Schroeder population, and reference populations from the 1000 Genomes Project using Principal Component Analysis.



Similarly, when calculating ancestry estimates for the DHS African-Americans, no differences in ancestry distribution were noted either between smokers and non-smokers (*p*=0.78) and between menthol cigarette smokers and non-menthol cigarette smokers (*p*=0.83).

4.5 Associations between TAS2R38 haplotypes, PTC tasting-status and smoking

Sensitivity to the bitter compound PTC in the Georgia cohort was measured using PTC paper strips. Taster individuals reported the bitter taste of these papers-strip to be mild or strong" and non-taster subjects reported "no taste" at all. The frequency of *TAS2R38* haplotypes and diplotypes differed between PTC tasters and non-tasters in this cohort, where PAV (the "taster" haplotype) was the predominant haplotype in those with the phenotype of PTC-tasters (95.3%) and rarely present in PTC non-tasters (4.7%) *(p*<0.001). Most of the PAV/PAV homozygotes in this cohort were PTC-tasters (98.1%) as opposed to non-tasters (1.9%) *(p*<0.001). PTC tasting abilities also differed between smokers and nonsmokers: 71.5% of smokers were PTC tasters, while 82.5% of non-smokers were PTC tasters *(p*= 0.03), suggesting a potential connection between this particular bitter taste sensitivity and smoking behaviors.

To further explore this possibility, we examined the frequency of *TAS2R38* haplotypes in all the available cohorts. As a result, we noted a trend toward a difference between smokers (38.4%) and non-smokers (43.1%) in the Georgia cohort, although this was not significant *(p*=0.31) in this small group (N=237). We also noticed a possible trend toward a difference in the distribution of *TAS2R38* AVI (non-taster) haplotype between smokers and non-smokers (55.3% and 49.9% respectively), but again this result was not significant *(p*= 0.29) in this small sample. For this reason, we expanded the analysis to the DHS and Biobank cohorts, which comprise a total of 8689 individuals. As shown in Table 10, in the DHS European-American cohort, the frequency of the taster PAV haplotype was lower in smokers (37.0%) than in non-smokers (44.0%) *(p*=0.003). Conversely the frequency of the non-taster AVI haplotype was higher in smokers (58.7%) compared to non-smokers (51.5%) *(p*=0.002). In order to replicate this association in a sub-sample of individuals more comparable to the Georgia cohort in demographic characteristics, we repeated this analysis in DHS EA individuals<40 years of age (N=272). The observed differences in *TAS2R38* haplotypes frequencies between smokers and non-smokers in this subgroup were similar to those in the entire population (PAV haplotype frequency 35% in smokers vs 44% in nonsmokers, *p*=0.05; AVI haplotype frequency 60% in

smokers vs. 51% in non-smokers, p=0.06). Interestingly, none of the *TAS2R38* haplotypes differed in frequency between smokers and nonsmokers in African-Americans in either the DHS or Biobank populations (p's>0.05). Pooling the data for AA participants from the DHS and Biobank populations and EA participants from the DHS and Georgia populations confirmed our previous un-pooled analyses. In particular, combining the results by meta-analysis showed no association between *TAS2R38* haplotypes and current smoking in African-American individuals (p>0.05). For the European-American cohorts, we confirmed the associations found in the two independent cohorts with the PAV (p=0.001) and the AVI (p=0.001) haplotypes.

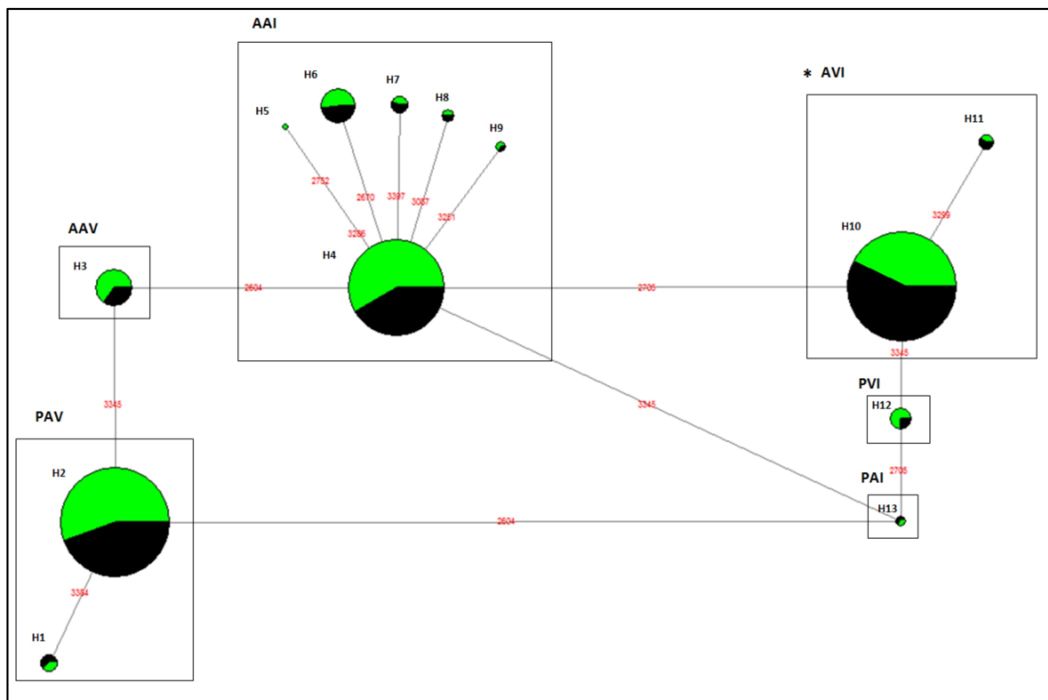
Table 10. Distribution of *TAS2R38* haplotypes between smokers and non-smokers in the DHS population.

Haplotype	Frequency Smokers	Frequency Non-Smokers	P-value
African-Americans:			
PAV	0.47	0.48	0.18
AVI	0.33	0.32	0.61
AAI	0.19	0.19	0.6
AAV	0.01	0.01	0.12
European-Americans:			
PAV	0.37	0.44	0.003
AVI	0.59	0.51	0.002
AAV	0.04	0.04	0.62

To explore potential associations between other polymorphisms in the *TAS2R38* gene and menthol cigarette preference, we sequenced the entire coding region of this gene in the Schroeder population. The *TAS2R38* major haplotypes showed a different distribution between menthol and non-menthol smokers. The frequency of the *TAS2R38* AVI haplotype was significantly higher in non-menthol smokers compared to menthol smokers, even after correction for covariates and multiple testing (OR=0.69, p-adjust=0.008). This association was replicated in both female (OR=0.89, p=0.03) and male (OR=0.72, p=0.02) individuals. The PAV

haplotype was not significantly associated with menthol smoking, although the results approached significance (OR=1.24, p=0.06). Lastly, construction of the haplotype network showed that, in addition to the major AVI haplotype, the *TAS2R38* H11 sub-haplotype (AVI + SNP rs115966953) showed a different distribution between menthol and non-menthol smokers (Figure 8). However, there were only 5 carriers of the sub-haplotype and this association did not reach significance *(p*=0.38).

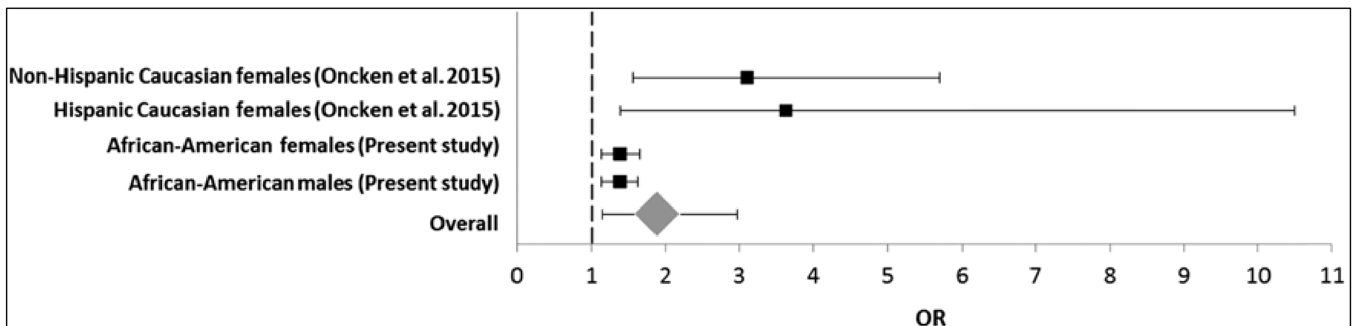
Figure 8. Haplotype network of *TAS2R38* haplotypes and sub-haplotypes. Menthol smokers are represented in light green and non-menthol smokers are represented in black. The size of each circle corresponds to the frequency of each haplotype. Red numbers indicate the site and number of nucleotide substitutions separating different haplotypes. The asterisks indicate a statistically significant difference in distribution between menthol and non-menthol smokers. H1-H13 indicate different *TAS2R38* haplotypes and sub-haplotypes.



4.6 Meta-analysis of the association between TAS2R38 variants and menthol cigarette preference

To better explore the strength and replicability of our findings, we performed a meta-analysis of the association between *TAS2R38* haplotypes and menthol cigarette smoking. This was possible because the phenotypes collected for the menthol studies were comparable, unlike those for other smoking behaviors. For this purpose, we included the results of both the present study and that of Oncken et al. (2015). As shown in Figure 9, the association between *TAS2R38* PAV haplotype and menthol cigarette smoking was still significant *(p*=0.025, using a random effects model) resulting in a common OR of 1.78 (95% confidence interval = 1.07 to 2.94), with some evidence *(p*=0.019) of expected heterogeneity, considering the differences in subject characteristics.

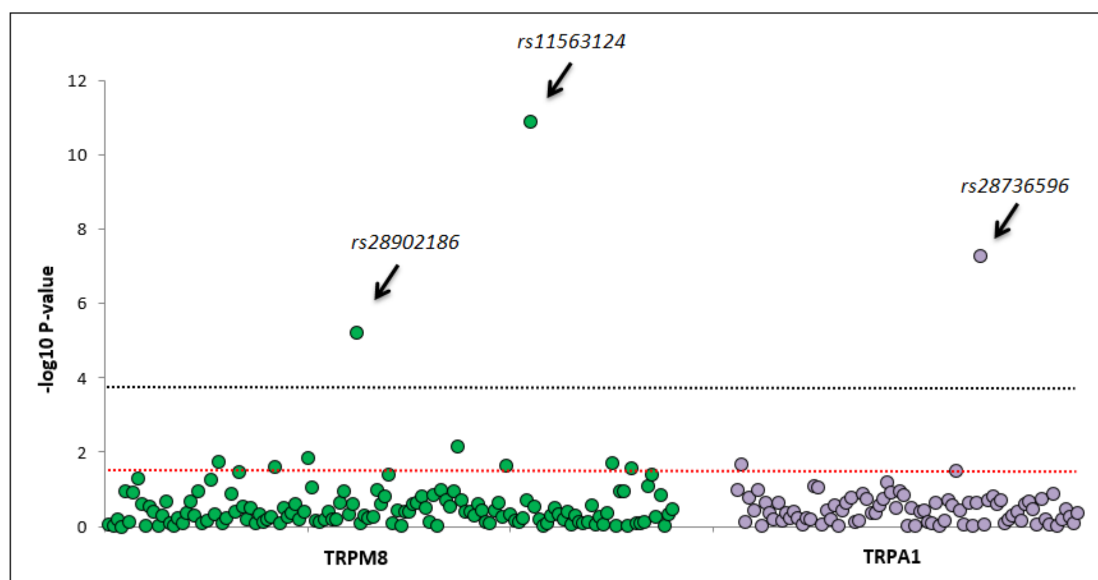
Figure 9. Forest plot illustrating the odds ratios and confidence intervals of the TAS2R38 PAV haplotype association with menthol cigarette smoking.



4.7 Associations between TRPM8, TRPA1 variants and menthol cigarette smoking

In order to further explore the genetic contributions to menthol preference in cigarettes, polymorphisms in the *TRPM8* and *TRPA1* menthol receptors were examined in both menthol and non-menthol cigarette smokers. Of the 238 analyzed SNPs in *TRPM8* and *TRPA1* (140 and 98, respectively), three SNPs were found to be strongly associated with menthol cigarette smoking in the Schroeder population, even after multiple testing correction (Figure 10). In particular, two intronic *TRPM8* SNPs (*rs28902186* and *rs11563124*) and one intronic *TRPA1* SNP (*rs35427625*) were associated with menthol cigarette preference (OR=6.1, *p*=3.12E-09; OR=2.52, *p*=1.51E-03 and OR=0.5 and *p*=1.22E-05, respectively). The frequency of the two *TRPM8* SNPs was low (MAF=0.01 and 0.03, respectively) with the derived allele of rs*28902186* (T) being only present in African populations of the 1000 Genomes populations, where it reached frequencies up to 0.09. These two SNPs were in linkage equilibrium, highlighting that these two associations were independent from each other. The frequency of the intronic *TRPA1* SNP (*rs28736596*) was high (MAF=0.43) and not in linkage disequilibrium with any SNPs in *TRPA1* coding regions. Although these two associations were intriguing, we were not able to replicate them in the DHS population (*p*>0.05).

Figure 10. Associations between *TRPM8* and *TRPA1* variants and menthol cigarette smoking. The first horizontal line (red) indicates a nominal *p*-value of 0.05, the second horizontal line (black) indicates the significance level after Bonferroni correction.



5. Discussion and Conclusion

The hypothesis that variations in taste receptor genes, in particular bitter taste, may confer protection against cigarette smoking has long been of interest. This is because the evolutionary role of bitter taste, which is generally believed to be critical for animals' survival, since it helps to avoid the ingestion of potentially harmful and poisonous compounds (Fischer et al. 2005). Cigarettes and other tobacco products contain bitter compounds, such as nicotine, that stimulate multiple sensory systems including taste transduction pathways (Ming et al. 1998). For this reason, it has been suggested that aversion to bitterness may also help prevent smoking and nicotine dependence (Enoch et al. 2001). To test this hypothesis, several previous studies produced results that suggested a connection between genetic variations in bitter taste receptor genes and smoking behaviors. However, some of these results failed to replicate in independent populations. For example, the earliest study exploring the correlation between a bitter taste receptor gene (*TAS2R38*) haplotypes and smoking behaviors examined 567 unrelated participants of European descent, comprising 384 smokers recruited from two smoking cessation trials. Although no significant associations were found between PAV/AVI (i.e. taster and non-taster, respectively) haplotypes and the odds of smoking, the analysis of current smokers revealed a correlation between these haplotypes and the importance of the taste of cigarettes as a motive for smoking (Cannon et al. 2005). A second study enrolled both European-Americans (N=197) and African-Americans (N=400) in families of heavy smokers and found a significant correlation between the non-taster AVI haplotype and smoking quantity, and concomitantly the taster PAV haplotype being associated with lower smoking quantity. This was only observed in African-Americans and no significant associations were found in European-Americans (Mangold et al. 2008). Lastly, a recent study of 1,007 German individuals comprising 330 smokers and showed that carriers of at least one taster *TAS2R38* PAV allele showed a significantly lower amount of cigarette smoking per day (Keller et al. 2013).

Such potential associations between taste perception genes and smoking behaviors is of interest to not only geneticists but also to sensory scientists, and it could have a significant impact on public health, especially when considering the well-known difference in smoking behaviors among different ethnicities. For example on average, 80% of adult African-Americans are menthol smokers, while only 25-35% of European-Americans smokers use menthol (Caraballo et al. 2011). A recent study showed how menthol improves the taste of nicotine in both cigarettes and smokeless products and could help people maintaining their smoking habits. While masking the bitter taste of nicotine and having a pleasant and cool flavor, menthol could also make it easier to consume cigarettes and similar products (Fan et al. 2016). Also bearing on this question, an association between the taster haplotype of the *TAS2R38* bitter taste receptor gene and menthol cigarette smoking in pregnant European-American women has been reported, suggesting that a genetic propensity to experience higher bitter taste perception increases the preference for menthol (Oncken et al. 2015). As the authors state, however, a replication in other ethnicities and in men is necessary to further test this hypothesis.

The goal of this research was therefore to investigate the role of inherited differences in smoking behaviors, with an emphasis on human taste perception to determine whether such genetic variation could account for the well-known differences in flavored tobacco use among different U.S. ethnic groups. For this reason, a total 9781 (8191 African-Americans and 1590 European-Americans) individuals were recruited over the course of several years. Subjects were recruited from four different populations:

1) The Dallas Heart Study (DHS), a multiethnic population-based study of 4700 adults including 2,500 African Americans enrolled at the University of Texas Southwestern Medical Center, our collaborating site on this project. These individuals have been well characterized for cardiovascular risk factors including tobacco use.

2) The Dallas Biobank (DBB), a repository of DNA and plasma samples from individuals ascertained at various locations in north-central Texas which includes a total of 4,973 African-American participants for whom the genotype and smoking phenotype data was available.

3) A population consisting of 855 African Americans from the U.S. District of Columbia, enrolled by the Schroeder Institute in Washington, DC. All of these subjects are smokers, and approximately 2/3 of them use mentholated cigarettes.

4) A population consisting of a total of 961 individuals from Emory University in Atlanta, GA. This population consists of both smokers and non-smokers. 314 of them are African Americans with the remaining 647 being Caucasians.

In the first two populations, DNA was extracted and purified from blood. General tobacco use data and DNA samples were collected from a total of 7336 African-Americans and 1353 European-American smokers and non-smokers. Regarding the Schroeder and Georgia populations, a total number of 237 European-Americans and 855 African-American saliva samples were collected and DNA purified using Oragene collection and purification kits. Tobacco use data, including flavored cigarette use, were also collected for these individuals. We completely sequenced the *TAS2R38* bitter taste receptor gene in all the examined populations and then completely sequenced the entire coding regions of *TRPM8* and *TRPA1* menthol receptors in the Schroeder population, examining potential associations between the 249 polymorphisms identified in these genes and different smoking behaviors.

In contrast to the previous studies of *TAS2R38*, which focused on relatively homogeneous populations of heavy smokers, our study included participants in which the prevalence of smoking and nicotine dependence was much lower. In addition, only a small fraction of participants recruited in our cohorts reported smoking more than 20 cigarettes per day. We fully replicated the results previously reported by Keller and colleagues (2013) in our study of two different

European-American cohorts, namely the Georgia and Dallas Heart Study populations. In these cohorts, cigarette smokers had a lower percentage of PAV-carriers (37.0% vs. 44.0% respectively, *p*=0.003). Conversely, the frequency of the non-taster haplotype was more common in smokers (58.7% vs. 51.5% respectively, *p*=0.002). In addition, this haplotype was also associated with smoking quantity. Moreover, in the Georgia population, smokers showed a lower percentage of PTC-tasters (associated with the PAV haplotype) when compared to non-smokers: 71.5% versus 82.5%, respectively *(p*=0.03). This agrees with previous studies (Hall et al. 1945, Daştan et al. 2015). We failed to replicate the results of the family-based study reported by Mangold et al. (2008) because in our two African-American cohorts, neither PAV nor AVI haplotypes showed different frequencies between smokers and non-smokers (All *p*'s>0.05). One possible explanation is that most of the individuals recruited in that study were heavy smokers, for whom nicotine dependence was a stronger motivator than taste. In addition, since the smoking data were based on self-report, it is possible that measurement error introduced a bias into our estimates.
We also confirmed and expanded the previous association between *TAS2R38* haplotypes and menthol cigarette preference in a logistic regression model. In the Schroeder population, the non-taster AVI haplotype was inversely associated with menthol cigarette smoking, even after correction for sociodemographic factors (such as age, gender, marital status, and education level) and multiple testing correction. The frequency of this haplotype was significantly lower in menthol smokers compared to non-menthol smokers (29.8% vs. 37.7% respectively, *p*=0.008). This association was replicated in both female (OR=0.89, *p*=0.01) and male *(p*=0.02) individuals and showed a gene dosage effect with 62.3%, 53.8%, and 44.0% of menthol users carrying zero, one, or two copies of this haplotype, respectively. In addition, we performed a meta-analysis including the results of both our study and of Oncken et al. (2015), finding that the association between *TAS2R38* PAV haplotype and menthol cigarette smoking was still significant *(p*=0.025).

We have also shown that polymorphisms at the *TRPM8* and *TRPA1* menthol receptor gene loci contributed to the preference for menthol cigarette smoking

in the Schroeder population. More specifically, we have identified two intronic *TRPM8* SNPs (namely *rs28902186* and *rs11563124*) and one intronic *TRPA1* SNP (*rs35427625*) were associated with menthol cigarette preference (*p*=3.12E-09, *p*=1.51E-03 and *p*=1.22E-05, respectively). Interestingly, the derived allele (T) of one of these SNPs (*rs28902186*) was only present in African populations of the 1000 Genomes populations, where it reached frequencies up to 0.09. These findings are important in light of a previous report, based on a much smaller number of subjects and phenotypes, that showed nominal associations between *TRPA1* variants and menthol cigarette smoking (Uhl et al. 2011).

Based on both previous and present data, we therefore conclude that *TAS2R38* haplotypes are factors that contribute to smoking status in European-Americans, with PAV haplotype carriers and PTC tasters less likely to be smokers. This finding has now been replicated in three independent cohorts, two cohorts in the present study and one in a previous report (Keller et al. 2013). In addition, we noted a similar trend in a recent paper studying a large cohort (N = 1,319) of individuals of Caucasian origin (Ortega et al. 2016). In contrast, *TAS2R38* haplotypes are not good predictors of smoking behaviors in African-Americans, and the lack of *TAS2R38* haplotype association with smoking in this ethnic group may be due to potentially confounding factors, such as age, gender and ascertainment of smoking status. In the previous studies, in fact, the definition of individuals as "smokers" and/or "current" smokers was different, as was the average age of individuals and the percentage of females.

Associations between *TRPM8* and *TRPA1* polymorphisms and menthol cigarette smoking further supports the hypothesis that variations in taste-related genes play a role, in certain populations, in the choice of cigarettes when smoking. Finally, understanding genetic differences in taste perceptions in tobacco use could help inform the development of more effective tobacco control policies, in particular those applicable to efforts directed at specific ethnicities and/or minorities.

6. Appendix

6.1 Supplementary tables

6.1.1 Appendix Table 1. PCR and sequencing primers used to amplify and sequence *TAS2R38* gene.

Gene	Primer name	Direction	Sequence	Reaction
TAS2R38	TAS2R38F	Forward	AGATGGGCATGCAAAACTGG	PCR, sequencing
	TAS2R38R	Reverse	ACTCACAGGCGTATTAATGAAGA	PCR, sequencing
	TAS2R38F	Forward	TCACACCTTCCTGATCTGCT	Sequencing
	TAS2R38R	Reverse	AGGCTGGGGTCACGAGAG	Sequencing

6.1.2 Appendix Table 2. PCR and sequencing primers used to amplify and sequence the *TRPM8* gene.

Gene	Primer name	Direction	Sequence	Reaction
TRPM8	TRPM8_E1F	Forward	CCTCTGCTTAAAAGAACCTCAGA	PCR, sequencing
	TRPM8_E2F	Forward	ACGTGGCAAAGGGGAGAG	PCR, sequencing
	TRPM8_E3F	Forward	CCTTTGAAGTGGGGTTACATCA	PCR, sequencing
	TRPM8_E4F	Forward	ACCAAGATGATGAAAACAGTCTG	PCR, sequencing
	TRPM8_E5F	Forward	ACCCCGCCCCTCTTCTAG	PCR, sequencing
	TRPM8_E6F	Forward	CTACAGGGATGGGGCCTTTA	PCR, sequencing
	TRPM8_E7F	Forward	TGGGAGGAGGCATAATGTGA	PCR, sequencing
	TRPM8_E8F	Forward	CAGAAGCTCTTGGTCCAACT	PCR, sequencing
	TRPM8_E9F	Forward	CGTGTGTTTCCTCCAGCTTG	PCR, sequencing
	TRPM8_E10F	Forward	AGAAGCTGATCCGGAACTCC	PCR, sequencing
	TRPM8_E11F	Forward	AGCCTGAGTTGATCCCCATT	PCR, sequencing
	TRPM8_E12F	Forward	TTACTCTGTGGCTGGAAATACA	PCR, sequencing
	TRPM8_E13F	Forward	ACAAAGCCACAGAGCCCTC	PCR, sequencing
	TRPM8_E14F	Forward	CAGTGAGCCAAGATCGTGC	PCR, sequencing
	TRPM8_E15F	Forward	TTTTGGATTCAGGGGTTGGC	PCR, sequencing
	TRPM8_E1617F	Forward	ACTGCCCCTTAGAATCCAGG	PCR, sequencing
	TRPM8_E18F	Forward	GGACCTGGAAGAAAATGAGTGG	PCR, sequencing
	TRPM8_E19F	Forward	GCTCTTGCCTGTTTCTTGAAA	PCR, sequencing
	TRPM8_E20F	Forward	GCTCTTCTCCATGGTCCACT	PCR, sequencing
	TRPM8_E21F	Forward	TTCTTGTGGCCCTGGGAAT	PCR, sequencing
	TRPM8_E22F	Forward	TCTCTCCTGTCTGTACTTAAGCT	PCR, sequencing
	TRPM8_E23F	Forward	CTGTGCAGGACGAATCTCAT	PCR, sequencing
	TRPM8_E24F	Forward	TCAGAAAGGTAAATGCAGCTGA	PCR, sequencing
	TRPM8_E25F	Forward	CATGCCACTCTCATTCTGCC	PCR, sequencing
	TRPM8_E26F	Forward	CTCTCAACTTTGGTGGTATATTTGG	PCR, sequencing
	TRPM8_E1R	Reverse	CCCTCCCAGAGACACAAACT	PCR, sequencing
	TRPM8_E2R	Reverse	ATTCAAGGCTCAGACCGGAA	PCR, sequencing
	TRPM8_E3R	Reverse	GAGCATTTGTTTGTTGTAAGGGT	PCR, sequencing
	TRPM8_E4R	Reverse	TACCCATCTCTCCTGCCTC	PCR, sequencing
	TRPM8_E5R	Reverse	TTCTCCGGAATTGCACACG	PCR, sequencing
	TRPM8_E6R	Reverse	GGCAGACTTCCCCATGATCT	PCR, sequencing
	TRPM8_E7R	Reverse	TGTTTCTCACGCATCACTCA	PCR, sequencing
	TRPM8_E8R	Reverse	AGATTAGATGAGCCCAAATCCTT	PCR, sequencing
	TRPM8_E9R	Reverse	GTGGAGTTGGGCGTTTAAGG	PCR, sequencing
	TRPM8_E10R	Reverse	TCACCAGGACAAGCGTAGTT	PCR, sequencing
	TRPM8_E11R	Reverse	CCATGCTGCAAACAAGGAAA	PCR, sequencing
	TRPM8_E12R	Reverse	GGGAAGCCTCAAAATATCCGA	PCR, sequencing
	TRPM8_E13R	Reverse	CCAAGCACTCCTTCACCAAC	PCR, sequencing
	TRPM8_E14R	Reverse	CTCAATCAGACTGGGGAGCT	PCR, sequencing
	TRPM8_E15R	Reverse	CCAAACCACTCCTTCTATGGC	PCR, sequencing
	TRPM8_E1617R	Reverse	TAGAAAGATGCCTCCCACCG	PCR, sequencing
	TRPM8_E18R	Reverse	AGACTTCCCTTGGGTTTGAAA	PCR, sequencing
	TRPM8_E19R	Reverse	GGAGAGGCCCCTGATGAAAT	PCR, sequencing
	TRPM8_E20R	Reverse	GTGCTTCTGAAGTGCGGTC	PCR, sequencing
	TRPM8_E21R	Reverse	CTCCCTCCAGTCCACCAATC	PCR, sequencing
	TRPM8_E22R	Reverse	CTAGAGTAATTCATTGGTTGGCC	PCR, sequencing
	TRPM8_E23R	Reverse	AGCTGCCTACGTTCTATGGG	PCR, sequencing
	TRPM8_E24R	Reverse	GCAATACCATTCCAGCACCC	PCR, sequencing
	TRPM8_E25R	Reverse	GGGAAGGTAATGTGAGGAGGA	PCR, sequencing
	TRPM8_E26R	Reverse	AAAGCCCCATGACACATTGG	PCR, sequencing

6.1.3 Appendix Table 3. PCR and sequencing primers used to amplify and sequence the *TRPA1* gene.

Gene	Primer name	Direction	Sequence	Reaction
TRPA1	TRPA1_E1F	Forward	GACCTGACACGCTTGACTCC	PCR, sequencing
	TRPA1_E2F	Forward	TGATCAATCTTGGGGCATTT	PCR, sequencing
	TRPA1_E3F	Forward	TCATAAAGTCACTCTCATGCTTCT	PCR, sequencing
	TRPA1_E4F	Forward	TGATTATTTCATCTAATGGGGAAA	PCR, sequencing
	TRPA1_E5F	Forward	TCCCATTGAAAAATGTGTTGA	PCR, sequencing
	TRPA1_E6F	Forward	ATTACAGGTGTGAGCCACCA	PCR, sequencing
	TRPA1_E7F	Forward	GCTACAATTTGTTTGTGGATGC	PCR, sequencing
	TRPA1_E8F	Forward	TTGAAAATGCTACCCTAATCAGAA	PCR, sequencing
	TRPA1_E9F	Forward	ACGAATTTCTTGATTTCTGATGA	PCR, sequencing
	TRPA1_E10F	Forward	CCCCACTTACACATACAGACACA	PCR, sequencing
	TRPA1_E11-12F	Forward	TGAAGACAGAGCATTCAACTTCA	PCR, sequencing
	TRPA1_E13F	Forward	CCCCAAAGTGTTTGGTAATG	PCR, sequencing
	TRPA1_E14BF	Forward	TTCCCACACATGCACACAC	PCR, sequencing
	TRPA1_E15F	Forward	GCCCACATTCCTAATGTCCA	PCR, sequencing
	TRPA1_E16F	Forward	ATTTCTGAGGCATGCAGAGG	PCR, sequencing
	TRPA1_E17F	Forward	TTCAAATCACTTAATTTTCAGCAG	PCR, sequencing
	TRPA1_E18F	Forward	AAAGAAACAGACTGAATCAAACAA	PCR, sequencing
	TRPA1_E19F	Forward	GGCTAACATCCTGGCTCTTTC	PCR, sequencing
	TRPA1_E20F	Forward	GGGCAAATTTTCCAACAGGT	PCR, sequencing
	TRPA1_E21F	Forward	ATTTCCTTAGGCGTCGGATT	PCR, sequencing
	TRPA1_E22F	Forward	GCAAGACCCTGTATCCAAAAA	PCR, sequencing
	TRPA1_E23AF	Forward	TGTTAAGCAACTTTGTCCATTG	PCR, sequencing
	TRPA1_E24F	Forward	ACCAGTTCTGCCCGAGTCTT	PCR, sequencing
	TRPA1_E25F	Forward	TGTCAGTGGGTGAGGTGAGA	PCR, sequencing
	TRPA1_E26F	Forward	TGCACTGGGGAAATAAGACA	PCR, sequencing
	TRPA1_E27F	Forward	TGCTGATTTCCTTTCTCTTGG	PCR, sequencing
	TRPA1_E1R	Reverse	ATTCGGTTGGACAAACAAGC	PCR, sequencing
	TRPA1_E2R	Reverse	GCTTGTTGTGGGGTGTCCTA	PCR, sequencing
	TRPA1_E3R	Reverse	TGAAAATGCCCGGTAAATAA	PCR, sequencing
	TRPA1_E4R	Reverse	CAGATCCTCTTAAGCAGGGAGTA	PCR, sequencing
	TRPA1_E5R	Reverse	TGAACCTGGAAGGCTGAAGT	PCR, sequencing
	TRPA1_E6R	Reverse	TTGAAACTTGGGATATTCTCTTG	PCR, sequencing
	TRPA1_E7R	Reverse	GCCTACATCCACTGGACTTGA	PCR, sequencing
	TRPA1_E8R	Reverse	TGCCAACTTCCTGCACTAAA	PCR, sequencing
	TRPA1_E9R	Reverse	ATGCTCCCTTACTGGACAGA	PCR, sequencing
	TRPA1_E10R	Reverse	ATCATCACCACCACCACCAT	PCR, sequencing
	TRPA1_E11-12R	Reverse	TGAACAAGGTTCTTCAACGCTAT	PCR, sequencing
	TRPA1_E13R	Reverse	ACAGATTTGGGGCTTCACAA	PCR, sequencing
	TRPA1_E14BR	Reverse	GATCACACCACTGCACTCCA	PCR, sequencing
	TRPA1_E15R	Reverse	TTACCAGCCAACAAACACCA	PCR, sequencing
	TRPA1_E16R	Reverse	TTCGGGGAAAACTGAGAATG	PCR, sequencing
	TRPA1_E17R	Reverse	AAAGGTTTGAGTCACATGGAAGA	PCR, sequencing
	TRPA1_E18R	Reverse	TCCCACCCTGCAAATATTTTAT	PCR, sequencing
	TRPA1_E19R	Reverse	CCTGGCAACTACAAATCTGCT	PCR, sequencing
	TRPA1_E20R	Reverse	GGCAGGGACCTTACCTAAAA	PCR, sequencing
	TRPA1_E21R	Reverse	CTGAACTCTGGAACCATGCTC	PCR, sequencing
	TRPA1_E22R	Reverse	GCTGTTGAAATGCAGTGATAAAA	PCR, sequencing
	TRPA1_E23AR	Reverse	CTTGATGTGCCTGGTGCTC	PCR, sequencing
	TRPA1_E24R	Reverse	TTGGTTTGACTCATGATTTTTGA	PCR, sequencing
	TRPA1_E25R	Reverse	TCACCCAGCTCTTTCATGTCT	PCR, sequencing
	TRPA1_E26R	Reverse	GGAAGGCAAACTGATTGGTG	PCR, sequencing
	TRPA1_E27R	Reverse	TGTAATTAACAAGCAGGAATTCAG	PCR, sequencing

6.2 Most relevant publications during the Ph.D. course

Research Articles (* Denotes Corresponding Author)

1) **Risso D***, Morini G, Pagani L, Quagliariello A, Giuliani C, De Fanti S, Sazzini M, Luiselli D, Tofanelli S. 2014. Genetic signature of differential sensitivity to stevioside in the Italian population. *Genes Nutr 9(3):401*

2) **Risso DS**, Tofanelli S, Morini G, Luiselli D, Drayna D. 2014. Genetic variation in taste receptor pseudogenes provides evidence for a dynamic role in human evolution. *BMC Evol Biol 13;14(1):198* (highly accessed article)

3) Negri R, Smarrazzo A, Galatola M, Maio A, Iaccarino Idelson P, Sticco M, Biongiovanni C, Franzese A, Greco L, **Risso DS**, Morini G. 2015. Age Variation In Bitter Taste Perception In Relation To The Tas2r38 Taste Receptor Phenotype. *International Journal of Nutrition 1 (2):2379-7835*

4) **Risso D**, Taglioli S, De Iasio S, Gueresi P, Alfani G,Nelli S, Rossi P, Paoli G, Tofanelli S. 2015. Estimating Sampling Selection Bias in Human Genetics: A Phenomenological approach. *PLoS One* Oct 9;10(10):e0140146

5) **Risso DS***, Howard L, VanWaes C, Drayna D. 2015. A potential trigger for pine mouth: a case of a homozygous PTC taster. *Nutr Res* Dec;35(12):1122-5

6) Roudnitzky N, **Risso DS**, Drayna D, Behrens M, Meyerhof W, Wooding S. 2016. Copy-number variation in *TAS2R* bitter taste receptor genes: structure, origin, and population genetics. *Chem Senses* Oct;41(8):649-59

7) **Risso DS***, Mezzavilla M, Pagani L, Robino A, Morini G, Tofanelli S, Carrai M, Campa D, Barale R, Caradonna F, Gasparini P, Luiselli D, Wooding S, Drayna D. 2016. Global diversity in the TAS2R38 bitter taste receptor: revisiting a classic evolutionary PROPosal. *Sci. Rep* 6, 25506

8) **Risso DS**, Kozlitina J, Sainz E, Gutierrez J, Wooding SP, Getachew B, Luiselli D, Berg CJ, Drayna D. Genetic Variation in the TAS2R38 Bitter Taste Receptor and Smoking Behaviors. 2016. *PLoS One* Oct 6;11(10):e0164157

9) **Risso D**, Sainz E, Gutierrez J, Kirchner T, Niaura R, Drayna D. 2016. Association of TAS2R38 Haplotypes and Menthol Cigarette Preference in an African American Cohort. *Nicotine Tob Res* Oct 12. pii: ntw275

Book Chapters

1) Morini G, Luiselli D, **Risso DS**, Tofanelli S. Il gusto degli Italiani. 2014. Cultura del cibo. Torino, Italy: *UTET Grandi Opere*

Genes Nutr (2014) 9:401
DOI 10.1007/s12263-014-0401-y

RESEARCH PAPER

Genetic signature of differential sensitivity to stevioside in the Italian population

Davide Risso · Gabriella Morini · Luca Pagani · Andrea Quagliariello · Cristina Giuliani · Sara De Fanti · Marco Sazzini · Donata Luiselli · Sergio Tofanelli

Received: 15 November 2013 / Accepted: 25 March 2014 / Published online: 6 April 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract The demand for *diet* products is continuously increasing, together with that for *natural* food ingredients. Stevioside and other steviol glycosides extracted from the leaves of the plant *Stevia rebaudiana* Bertoni are the first natural high-potency sweeteners to be approved for consumption in the United States and the European Union. However, the sweetness of these compounds is generally accompanied by aversive sensations, such as bitter and off-tastes, which may constitute a limit to their consumption. Moreover, consumers' differences in sensitivity to high-potency sweeteners are well known, as well as difficulties in characterizing their aftertaste. Recently, TAS2R4 and TAS2R14 have been identified as the receptors that mediate the bitter off-taste of steviol glycosides in vitro. In the present study, we demonstrate that *TAS2R4* gene polymorphism rs2234001 and *TAS2R14* gene polymorphism rs3741843 are functional for stevioside bitterness perception.

Keywords Stevioside · Bitter aftertaste · Genetic polymorphisms · hTAS2R4 · hTAS2R14 · hTAS2R38

Donata Luiselli and Sergio Tofanelli are co-senior authors.

Electronic supplementary material The online version of this article (doi:10.1007/s12263-014-0401-y) contains supplementary material, which is available to authorized users.

D. Risso (✉) · L. Pagani · A. Quagliariello · C. Giuliani · S. De Fanti · M. Sazzini · D. Luiselli
Laboratory of Molecular Anthropology and Centre for Genome Biology, Department of BiGeA, University of Bologna, via Selmi 3, 40126 Bologna, Italy
e-mail: davide.risso2@unibo.it

G. Morini
University of Gastronomic Sciences, Piazza Vittorio Emanuele 9, Pollenzo, 12042 Bra, Cn, Italy

L. Pagani
The Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK

S. Tofanelli
Department of Biology, University of Pisa, Via Ghini 13, 56126 Pisa, Italy

Introduction

Stevioside (Fig. 1) is the most abundant compound of a group of structurally related high-potency sweeteners, the steviol glycosides, which are secondary metabolites extracted from the leaves of the *Stevia rebaudiana* Bertoni. This plant is native to Paraguay where its leaves have a long history of use as sweetener and to treat several diseases (Kinghorn 2002; Yadav and Guleria 2012; Vega-Gálvez et al. 2012). All these compounds share a common aglycone, known as steviol (*ent*-13-hydroxykaur-16-en-18-oic acid), but differ in the number and types of sugar residues.

All steviol glycosides are high-potency sweeteners, and stevioside has been reported to have a relative sweetness, compared to sucrose, between 210 and 300, depending on the protocol used (Crammer and Ikan 1987; Kinghorn and Soejarto 1986). A detailed study of concentration/response functions for several sweeteners, including stevioside, was presented by DuBois et al. (1991). The demand for this kind of sweeteners in the production of zero and reduced-calorie food products is continuously expanding as a response to increasing health awareness. Such awareness stems from a growing body of evidence showing that overweight and obesity strongly contribute to a large proportion of non-communicable diseases.

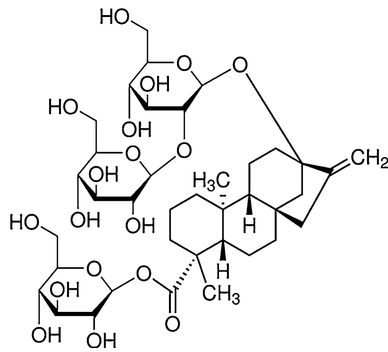


Fig. 1 Chemical structure of stevioside

As for other high-potency sweeteners (DuBois and Prakash 2012), the sweet taste of steviol glycosides presents bitter and off-tastes (Kinghorn and Soejarto 1986; Prakash et al. 2008). Moreover, consumers' differences in sweeteners sensitivity and acceptance, as well as the difficulties in characterizing their aftertaste, have been reported along with evidences suggesting that most of variation in sensitivity to chemical stimuli may have a genetic basis (Simons et al. 2008).

The recent introduction (2008 in USA, 2011 in EU) of purified steviol glycosides (at least 95 %) as the first natural high-potency sweeteners on the market opens up questions on the genetic basis of stevioside taste perception and its worldwide distribution. Particularly, understanding the differential ability to sense stevioside bitter aftertaste of various human populations has implications on its worldwide use as a natural sweetener and, consequently, on the diffusion of related healthy reduced-calorie food products. Bitter taste is detected by a set of 25 taste 2 receptors (TAS2R) (Meyerhof et al. 2010), and individual differences in the ability to taste substances like phenylthiocarbamide have been known since a long time (Fox 1932). Single nucleotide polymorphisms (SNPs) in the *TAS2R38* gene have been identified as the key determinants of this capability, as well as for that of tasting the related compound 6-*n*-propylthiouracil (PROP) (Bufe et al. 2005; Kim et al. 2003; Duffy et al. 2004). In particular, PROP status (i.e., taster or non-taster) has been proposed by some authors to have a broad effect on other oral sensations and therefore has been considered a marker for food preferences and diet selection (Tepper 2008). However, this issue is highly debated as other studies showed that PROP responsiveness had no influence on food preferences (Baranowski et al. 2011).

Recently, hTAS2R4 and hTAS2R14 have been identified as the receptors that specifically mediate the bitter off-taste of steviol glycosides in vitro (Hellfritsch et al. 2012), opening the way to investigate whether variation at the genes encoding for these receptors could explain individual differences in aftertaste perception.

To date, only some SNPs in chemosensory genes have been actually associated with differential perception of bitter compounds (Allen et al. 2013a; Campbell et al. 2013; Roudnitzky et al. 2011; Reed et al. 2010; Wooding et al. 2010; Pronin et al. 2007). Nevertheless, genetic differences in sensitivity to natural bitter compounds are an aspect of great relevance considering that bitter taste receptors are expressed also in the gastrointestinal (GI) tract (and in other extra-oral tissues). In fact, the GI tract represents the key interface between food and the human body and bitter taste receptors have been proven to play a crucial role in the control of several functions, ranging from gastric emptying to appetite modulation, to the detection of ingested harmful compounds and toxins, to emesis (Reimann et al. 2012).

That being so, in this work, we set to investigate putative associations between stevioside bitter aftertaste and SNPs on genes known to be involved in bitterness perception (i.e., *TAS2R4*, *TAS2R14* and *TAS2R38)*, in a panel of Italian subjects. Furthermore, we explore their worldwide patterns of diversity, looking for signatures of ongoing selection at these loci that could explain potential differences in the ability to sense stevioside bitter aftertaste in different human populations.

Subjects and methods

Study design and sampled population

A total of 86 healthy adult Italian donors (48 females and 38 males) were recruited. Subjects (average age of 47 ± 14 years) were not following a prescribed diet or using drugs that might interfere with taste perception, and none of them had food allergies. In addition, individuals who smoked more than ten cigarettes per week were excluded.

Experimental protocol

A training session regarding taste perception, taste qualities and how to rate perception using the Labeled Magnitude Scale (LMS, Green et al. 1993) was performed. Volunteers were asked to refrain from eating and drinking for at least 3 h before the beginning of the session and to rinse their mouth with room temperature deionized water prior to the first and between each sample. Weight (in kg) and height (in m) were collected, in order to calculate the body mass index (BMI). For the perception test, 10 mL of a solution of stevioside 1.26 mM (Nastevia, Stevia Italia s.r.l., Italy) was used. The PROP-taster status was assessed using cotton swabs dipped in 50 mM 6-*n*-propylthiouracil solution (Sigma Aldrich S.r.l.) (Caremoli 2011), modifying the

Table 1 Details of the studied SNPs in *TAS2R4*, *TAS2R14* and *TAS2R38* genes

Chr	Gene	dbSNP	cSNP	Allele	Type of mutation	Residue change
7	TAS2R4	rs2234001	G286-C	G	Missense	V [Val]
				C		L [Leu]
12	TAS2R14	rs11610105	G11088981-A	G	Silent	\
				A		
		rs3741843	G375-A	G	Cds synon	R [Arg]
				A		R [Arg]
		rs7138535	A114-T	A	Cds synon	G [Gly]
				T		G [Gly]
7	TAS2R38	rs713598	G229-C	G	Missense	A [Ala]
				C		P [Pro]
		rs10246939	A970-G	A	Missense	I [Ile]
				G		V [Val]
		rs1726866	T869-C	T	Missense	V [Val]
				C		A [Ala]

protocol for filter paper discs described in Zhao et al. (2003). Subjects were informed that they may receive stimuli eliciting more than one taste quality. They were asked to held the stevioside solution in the mouth for 5 s and the cotton PROP swab for 10 s. For both stimuli (i.e., stevioside and PROP), participants were allowed to choose among different tastes such as bitter, sweet or tasteless, as well as to rate their perceived intensity separately on multiple LMSs, one for each taste quality.

DNA collection, extraction, genotyping and quality control

DNA was collected using Oragene saliva collection kits and extracted according to the manufacturer's protocol (Genotek Inc., Kanata, Ontario, Canada). A list of genotyped SNPs in *TAS2R14* (chromosome 12), *TAS2R4* and *TAS2R38* (chromosome 7), which were selected from the literature and by surveying lists of genetic variants contained in different databases, is shown in Table 1. Genotypes for these variants were determined using Sequenom MassARRAY technology (Sequenom, San Diego, CA, USA). Genotype calls were analyzed by using SEQUENOM Typer 4.0 software, and the individual spectrograms were checked in order to evaluate the presence of calling errors. None of them showed significant deviations from Hardy–Weinberg equilibrium (Online Resource 1). We also assessed their minor allele frequency to be above 5 % and their call rate to be above 95 %.

Statistical analyses

Allelic and genotypic association analyses were performed with PLINK v. 1.07 (Purcell et al. 2007) using Fisher's exact method, and the significance threshold was adjusted according to the Bonferroni multiple testing correction (adjusted $P = P$ value $\times$ number of individual tests). Linkage disequilibrium (LD) patterns at the *TAS2R38* gene were explored by means of the solid spine approach implemented in the Haploview package (Barrett et al. 2005). The PHASE software (Stephens et al. 2001) was then used to statistically infer *TAS2R38* haplotypes within the identified high-LD blocks and to reconstruct haplotype pairs for each subject. For association analyses between individual SNPs and bitterness perception, analysis of covariance (ANCOVA) was performed using STATISTICA v. 6.0 (Stat-Soft Inc, Tulsa, OK), considering age, sex and BMI as covariates. Analysis of molecular variance (AMOVA) was carried out with Arlequin v. 3.5 (Excoffier et al. 2007). Exploration of worldwide variation patterns for the examined SNPs was performed using 1000 Genomes Project PHASE 1, HapMap and Human Genome Diversity Project datasets. Functional annotations and SIFT (http://sift.jcvi.org/) predictions (Kumar et al. 2009) for each SNP were retrieved from the Ensembl database (http://ensembl.org).

Results

Assessing taste phenotypes

Volunteers involved in the study were classified, after the perception test, in different tasting categories for both PROP and stevioside. Regarding the PROP status, the distribution of PROP sensitivity showed the classical bimodal curve (Online Resource 2), with 18 individuals (20.9 %) being classified as non-tasters and 68 (79.1 %) as

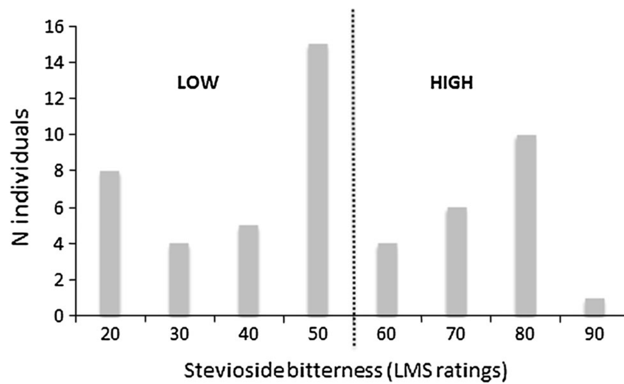


Fig. 2 Distribution of stevioside sensitivity in the Italian sample

tasters. In order to test whether the adopted deviation from the method described in Zhao et al. (2003) could affect PROP phenotype assessment, we calculated the overall genotype–phenotype concordance (90.7 %), thus inferring the correctness of our cotton swab PROP test. After stevioside tasting, 11 individuals (13 %) were able to perceive only its bitter taste. On the contrary, 22 individuals (26 %) perceived only its sweet taste, whereas the majority of samples constituted by 53 individuals (61 %) identified a sweet taste followed by a bitter after-taste.

The 53 bitter/sweet-tasters were subsequently distinguished into 32 "bitter-low" (LMS scores up to 50) and 21 "bitter-high" (LMS from 60 to 100) tasters (Fig. 2) according to LMS scale scores.

TAS2R4 SNP regulates the ability to perceive the bitter taste of stevioside

The *TAS2R4* rs2234001 (C/G) turned out to be associated with stevioside bitter status. In fact, the "bitter" and "sweet" phenotype groups showed statistically significant differences for both genotypic (adjusted $P = 0.002$, Fisher's exact test) and allelic frequencies (adjusted $P = 0.039$, Fisher's exact test). In particular, the GG genotype and the G allele were more frequent in bitter-tasters ($n = 11$), whereas genotype CC and the C allele were more frequent in sweet-tasters ($n = 22$). More specifically, 68.18 % of bitter-tasters carried the G allele, whereas 76.19 % of sweet-tasters carried the C allele at this locus (Table 2). The same segregation was observed between the sweet and the "bitter-low" tasters ($n = 32$), with the genotype GG (adjusted $P = 0.008$, Fisher's exact test) and the G (adjusted $P = 0.013$, Fisher's exact test) allele being more frequent in the "bitter-low" group. Analyses conducted with the SIFT software showed that rs2234001 caused an amino acidic substitution at residue 96, resulting in a valine–leucine change, without altering the secondary structure of the protein. However, this SNP was found to be in strong LD ($r^2 \geq 0.9$) in European populations with two other non-synonymous SNPs (rs2227264 and rs2233998), both predicted by SIFT to alter the function of TAS2R5 and TAS2R4 proteins, respectively.

Table 2 Distribution of *TAS2R4* rs2234001 and *TAS2R14* rs3741843 polymorphisms among stevioside phenotypes

rs2234001	Bitter	Bitter/sweet	Sweet
Genotype	%	%	%
CC	0	19.22	57.1
CG	63.64	50	38.1
GG	36.36	30.78	4.8
Allele			
C	31.82	44.23	76.19
G	68.18	55.77	23.81

rs3741843	Bitter-low	Bitter-high
Genotype	%	%
AA	93.55	42.85
GA	6.45	57.15
GG	0	0
Allele		
A	96.77	28.57
G	3.23	71.43

TAS2R4 SNP does not predict variations in stevioside bitterness perception

We also tested whether *TAS2R4* rs2234001 (C/G) could predict stevioside bitterness or sweetness. Differently than *TAS2R14* rs3741843, this SNP did not show variation among different levels of bitter and sweet perception (ANCOVA, $P = 0.601$ and $P = 0.623$, respectively).

TAS2R14 SNP predicts variations in stevioside bitterness perception

We found evidence that *TAS2R14* rs3741843 (A/G) has a significant impact on bitterness perception. Through comparing the "bitter-low" and "bitter-high" groups, the allelic frequency of this SNP was found to be significantly different (adjusted $P = 0.002$; Fisher's exact test). In particular, the G allele was more frequent in the "bitter-high" group ($n = 21$) compared to the "bitter-low" one ($n = 32$). Genotypic analyses confirmed this statistically significant different distribution, with genotype AA being more frequent in the "bitter-low" group (adjusted $P \leq 0.001$, Fisher's exact test). The same tests were repeated by removing the LMS modal group (LMS = 50), which is suspected to introduce a confounding effect, and obtained results confirmed our previous estimation

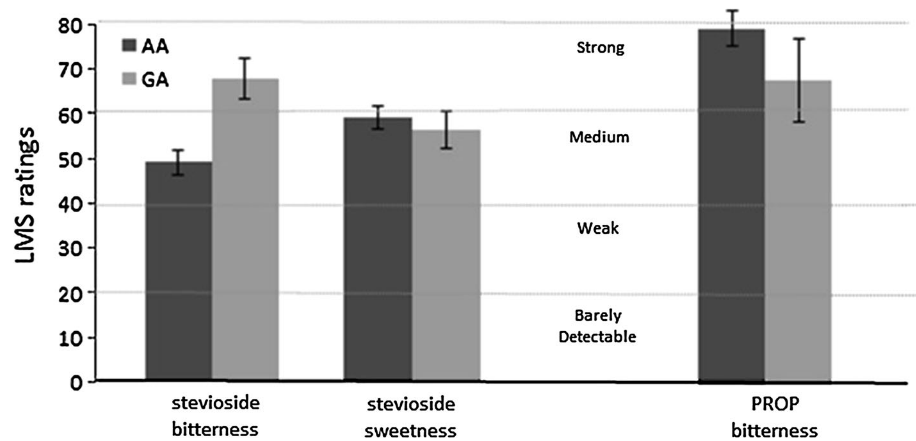


Fig. 3 Effect of the *TAS2R14* rs3741843 on bitterness and sweetness of stevioside and on bitterness of PROP

($P = 0.003$ and $P = 0.001$, Fisher's exact test, respectively).

In addition, to test whether arbitrary classifications in taste phenotypes were plausible, an ANCOVA was performed on the subset of individuals who were able to perceive stevioside bitterness ($n = 64$), considering age, sex and BMI as covariates. As shown in Fig. 3, homozygote individuals with the AA genotype of *TAS2R14* rs3741843 reported less bitterness from stevioside than heterozygote ones ($P = 0.002$). In contrast, there was no evidence that this allele predicts stevioside sweetness (ANCOVA, $P = 0.621$). We also found minimal evidence that stevioside bitterness is predictive of PROP bitterness (ANCOVA $P = 0.081$).

TAS2R38 SNPs predict variations in PROP bitterness but not in stevioside bitterness perception

As expected, PROP-tasters and non-tasters differed significantly for *TAS2R38* alleles and haplotypes. In particular, *TAS2R38* rs10246939 (T/C), rs1726866 (T/C) and rs713598 (G/C) were more frequent in PROP-tasters ($n = 68$) compared to PROP-non-tasters ($n = 18$) (adjusted $P < 0.001$, adjusted $P < 0.001$, adjusted $P < 0.001$, Fisher's exact tests). In the same way, perception of PROP bitterness varied with *TAS2R38* haplotypes. The proline–alanine–valine (PAV) homozygotes ($n = 15$) reported significantly more bitterness than the heterozygotes ($n = 42$) (ANCOVA, $P < 0.001$) or the AVI (alanine, valine, and isoleucine) homozygotes ($n = 24$) (ANCOVA, $P < 0.001$). Moreover, bitterness reported by heterozygotes was similar to that reported by PAV homozygotes (ANCOVA, $P = 0.345$). Rare haplotypes were excluded from the analyses because they are known to have intermediate phenotypes that differ from both each other and the common haplotypes. In this study, we observed both AAI (3) and AAI/AVI (2) individuals. No associations were instead found between PROP haplotypes and stevioside bitterness (ANCOVA, $P = 0.945$) or sweetness (ANCOVA, $P = 0.812$).

Comparison between stevioside bitterness and sweetness perception

Great variability in both stevioside bitterness and sweetness perception was found. Figure 4 shows this distribution of variation, with stevioside bitterness and sweetness ranging from 20 to 80 on a LMS. We therefore tried to test whether a covariation between them existed. By plotting the bitterness and sweetness of stevioside simultaneously, we did not observe a covariation between them ($R^2 = 0.007$, $P = 0.776$), with bitterness showing a decreasing trend toward the modal class and sweetness a decreasing trend followed by an increase after the modal class.

Population genetics of *TAS2R4* rs2234001 and *TAS2R14* rs3741843 SNPs

Alleles worldwide distribution

To better understand the differences observed for *TAS2R4* and *TAS2R14* SNPs, we examined their worldwide distribution. Figure 5 shows a map of allele frequencies of the two studied SNPs, based on 1000 Genomes Project data. Hardy–Weinberg test (Online Resource 3) and Heterozygosity test (Online Resource 4) were performed, confirming that all the examined populations are in Hardy–Weinberg equilibrium and no excess of heterozygosity was found ($P > 0.05$).

AMOVA

In order to identify the causes underlying the different distributions of rs3741843 and rs2234001, AMOVA based on the 1000 Genomes Project data was performed. As

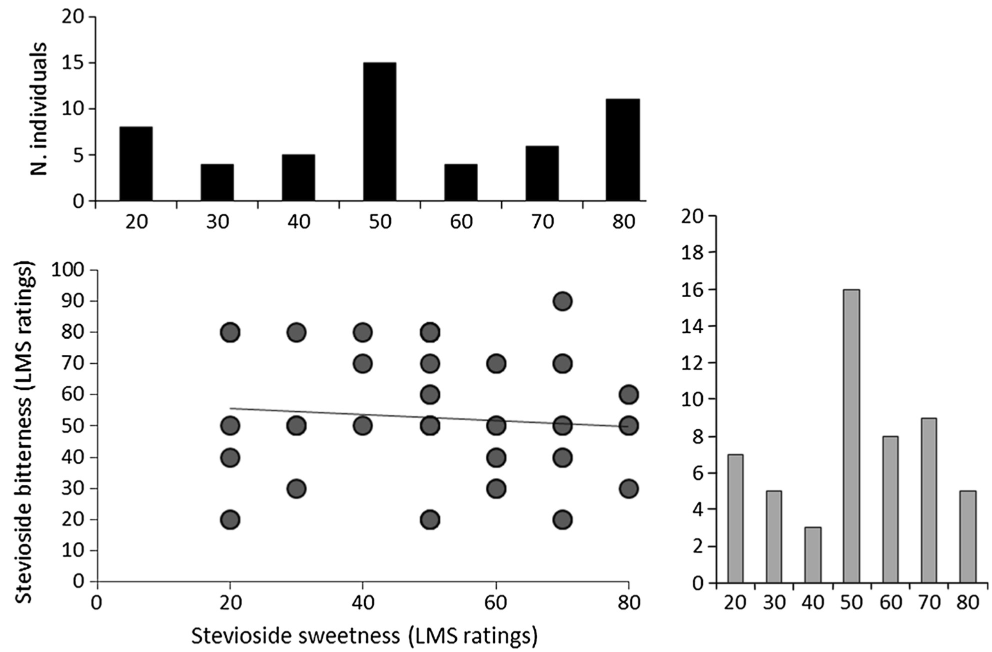


Fig. 4 *Scatter plot* showing sweetness and bitterness intensities of stevioside. *Histograms* shown along the axes illustrate the total amount of variation found across individuals

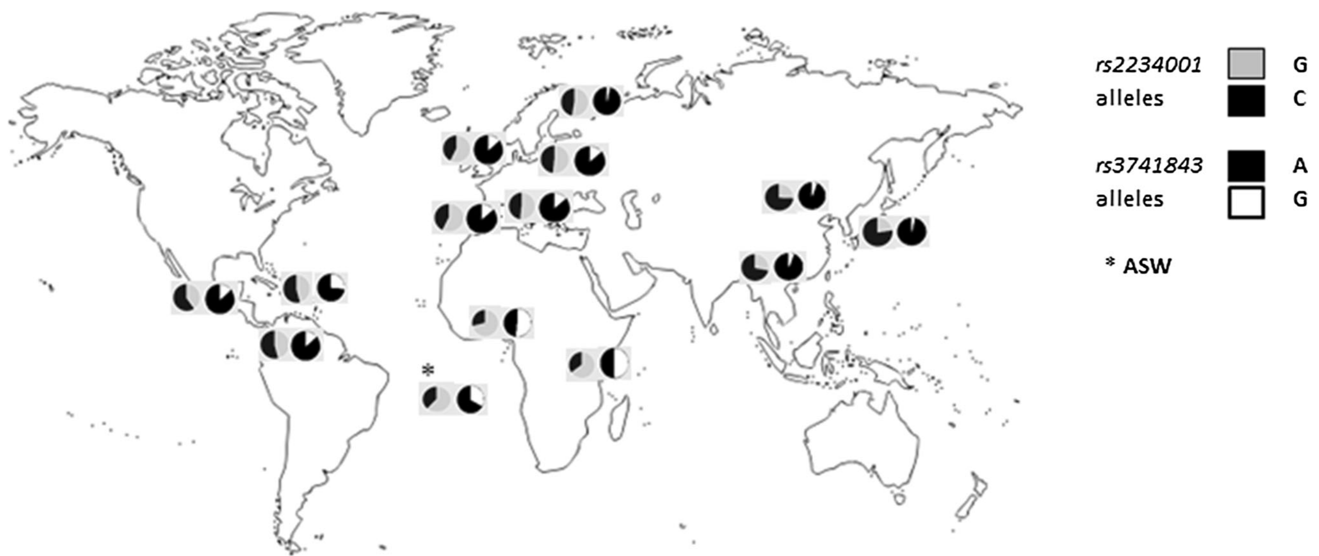


Fig. 5 Worldwide distribution of the studied *TAS2R4* rs2234001 and *TAS2R14* rs3741843 SNPs

shown in Table 3, 19.62 % of variation at rs3741843 was due to differences among population groups ($P = 0.002$), 1.32 % was due to differences among populations within groups ($P < 0.001$), and 79.06 % was accounted by differences among individuals within populations ($P < 0.001$). rs2234001 showed a similar pattern, with 11.75 % of variation which was attributed to differences among groups ($P < 0.001$), with the remaining 88.25 % accounting only for differences observable within populations ($P = 0.006$).

Genetic differentiation analyses

Genome-wide pairwise Wright's F_{ST} was calculated as a measure of genetic differentiation on 1000 Genomes populations and continental groups, and a set of relevant percentiles was extracted (Online Resource 5). F_{ST} obtained for rs3741843 and rs2234001 fell within the top percentiles (0.95) of most of the continent-wide distributions and turned out to be outlier values also for some population-level comparison.

Table 3 AMOVA performed on *TAS2R14* rs3741843 and *TAS2R4* rs2234001 SNPs

	rs3741843	*rs2234001*
Among groups[a]	19.62 %	11.75 %
Among population with groups	1.32 %	0
Within populations[b]	79.06 %	88.25 %
F_{ct}	0.19619 (*P* value 0.0019)	0.11754 (*P* value <0.001)
F_{sc}	0.01644 (*P* value <0.001)	0
F_{st}	0.20940 (*P* value <0.001)	0.1164 (*P* value 0.00587)

[a] Africans, Americans, Asian, Europeans

[b] YRI, LWK, ASW, CLM, MXL, PUR, CHB, JPT, CHS, CEU, FIN, GBR, IBS, TSI

Admixture analysis

The admixed nature of the American 1000 Genomes populations allowed us to test whether the studied SNPs could be differentially selected according to the ancestral components of these populations. The American, African and European genomic components of the 1000 Genomes American populations were extracted from the literature (1000 Genomes Project Consortium et al. 2012; Skotte et al. 2013) and used to infer the expected frequencies of the two SNPs, given the observed admixture. The African and European frequencies of the two SNPs were obtained from the relevant populations, while American frequencies were extracted from the Mexican one (MXL), after correcting for admixture. Such frequencies were then combined, according to their admixture proportions, to generate the expected frequencies of each of the American populations. Table 4 reports the observed and expected frequencies hence calculated, along with the observed excess. The difference between observed and expected frequencies is remarkable for rs3741843, where Colombians (CLM) show a defect of 22 % and Puertoricans (PUR) show an excess of 61 % on the expected frequencies.

Discussion

To our knowledge, this study represents the first attempt to explore genotype–phenotype association of stevioside bitter taste perception, as well its relationships with PROP phenotypes/genotypes. In fact, a previous study by Allen et al. (2013b) tested only for relationships between other steviol glycosides, RebA and RebD, and *TAS2R* SNPs. For this purpose, we have assayed three SNPs on *TAS2R14* (rs11610105, rs3741843 and rs7138535), one on *TAS2R4* (rs2234001) and three on *TAS2R38* (rs10246939, rs1726866 and rs713598) genes for individual associations with stevioside perception phenotypes and for their relationships with PROP genotypes/phenotypes on a sample of 86 Italian subjects.

We confirmed the associations between PROP variants and PROP-taster status at both allelic and haplotypic levels. Surprisingly, compared to other published data reporting a correlation between PROP bitterness and intensity of different tastants, including sucrose, citric acid, sodium chloride and quinine (Hayes et al. 2008), the sweetness of aspartame (Duffy et al. 2006) and Acesulfame K (Allen et al. 2013a), as well as sweet foods, such as marshmallows and iced cakes (Lanier et al. 2005), we did not find any correlation between PROP bitterness and either stevioside sweetness or stevioside bitterness.

Stevioside perception showed great variability among the candidate subjects, with the majority of the individuals perceiving both a sweet and a bitter after-taste and some individuals being not able to perceive its bitterness or sweetness. Moreover, also bitterness intensity strongly varied across subjects. Investigating polymorphisms on *TAS214* and *TAS2R4* genes, which have been recently identified as the loci encoding for receptors that mediate the bitter off-taste of steviol glycosides in vitro (Hellfritsch et al. 2012), our results suggest that two SNPs (rs2234001 and rs3741843) contribute to the perceived bitterness of stevioside in humans. After phenotype identification, we showed that *TAS2R4* rs2234001 varies among people who were able to perceive or not the bitter taste of stevioside. Specifically, our results demonstrate that the G allele of

Table 4 Ancestral proportions in American 1000 Genomes populations and observed/expected frequencies of *TAS2R4* rs2234001 and *TAS2R14* rs3741843 SNPs

Population	Ancestral proportions			rs2234001			rs3741843		
	%African	%NatAm	%European	Obs.	Exp.	%Excess	Obs.	Exp.	%Excess
ASW	0.79	0.02	0.19	0.39	0.37	4.65	0.38	0.40	−6.06
CLM	0.07	0.24	0.69	0.54	0.51	5.98	0.11	0.14	−21.75
MXL	0.02	0.60	0.38	0.60	0.60	0.00	0.11	0.11	0.00
PUR	0.13	0.13	0.74	0.55	0.48	14.08	0.26	0.16	61.41
AFR	–	–	–	0.34	–	–	0.48	–	–
EUR	–	–	–	0.47	–	–	0.12	–	–
NATAM (MXL)[a]	–	–	–	0.69	–	–	0.08	–	–

[a] % of Native Americans in the Mexican population

this locus was associated with capability to perceive bitter after-taste of stevioside, while the C allele was more abundant in subjects who could not perceive it. At the genotype level, the recessive homozygote status GG was associated with the ability to perceive the bitter taste of stevioside, while the dominant CC homozygote status was associated with the inability.

In addition, when comparing strong and low bitter stevioside perceptions, one SNP located on the *TAS2R14* gene (rs3741843) showed significant frequency differences. In particular, its G allele was associated with a stronger bitter perception. In order to confirm this, we tried to verify whether perception of stevioside bitterness varied with the genotype and, as a matter of fact, the AA genotype turned out to be strongly correlated with lower values of bitter perception. Regarding PROP status, we did not find any relationships with neither stevioside perception nor the *TAS2R4/14* studied SNPs, although minimal evidence that the rs3741843 AA genotype is predictive of PROP bitterness was observed.

While variation at rs3741843 locus was not linked with any protein modification, we found that not only rs2234001 is a non-synonymous (albeit not deleterious) SNP, but that it is also in strong LD ($r^2 > 0.9$) with two other deleterious mutations in CEU (Northern Europeans from Utah). Both such polymorphisms (rs2227264 and rs2233998) were found to be associated with eating behavior (Dotson et al. 2010), while only rs2227264 has been proved to be related to the perceived bitterness of the espresso coffee (Hayes et al. 2011), hence representing putative functional candidates for the reported genotype–phenotype association.

Patterns of worldwide variation of two SNPs related to stevioside perception, together with comparison of their pairwise population F_{st} values with empirical genome wide distributions, highlighted great variability levels within human populations. In particular, rs2234001 showed a significant differentiation between African-Asian groups ($P = 0.05$), but not between Asian-European ones ($P = 0.10$), while rs3741843 was highly differentiated between African-Asian ($P = 0.025$) and African-European ($P = 0.05$) groups (Table S5). Furthermore, considering the 10 kbp haplotypes centered on the rs3741843 locus, the modal haplotype class broadly differed between African and European populations (Online Resource 6).

The strong differentiation between African and European groups for this latter SNP allowed us to detect a 61 % excess of the African variant in Puertoricans, after accounting for their history of genetic admixture. The enrichment of the African component in Puertoricans may suggest the presence of some natural or social selective pressure acting on such variant after its arrival in their genic pool some 200 ya (Kidd et al. 2012). Such selective pressure could perhaps be related with cultural/dietary shifts caused by the introduction of African traditions at the time of their arrival in Puerto Rico. While these results might be affected by our choice of using the corrected Mexicans as proxy for the ancestral American components, we believe such choice to be conservative given that only two observed–expected pairs showed an excess higher than 20 %.

In conclusion, we found genotype–phenotype associations between stevioside perception and SNPs at the *TAS2R4* and *TAS2R14* genes, which are significantly differentiated worldwide. This finding and the strong LD of one of these variants with deleterious mutations suggest a role played by natural selection in shaping the current patterns of variation observed in different human groups.

Acknowledgments We are very grateful to the "Slow Food Communities," Silvano Zaccone, Luisella Verderi and to all the people who agreed to participate to this study. This work was supported by the Fondazione Del Monte "Taste Genetic Geography" Grant to DL and by The Wellcome Trust (098051) to LP.

Ethical standards Davide Risso, Gabriella Morini, Luca Pagani, Andrea Quagliariello, Cristina Giuliani, Sara De Fanti, Marco Sazzini, Donata Luiselli and Sergio Tofanelli declare that they have no conflict of interest. All procedures followed were in accordance with the ethical standards of the Ethics Committee of the University of Bologna and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all subjects for being included in the study.

References

Allen AL, McGeary JE, Knopik VS, Hayes JE (2013a) Bitterness of the non-nutritive acesulfame potassium varies with polymorphisms in TAS2R9 and TAS2R31. Chem Senses 38(5):379–389

Allen AL, McGeary JE, Hayes J (2013b) Rebaudioside A and Rebaudioside D bitterness do not covary with Acesulfame-K bitterness or polymorphisms in TAS2R9 and TAS2R31. Chemosens Percept 6(3):109–117

Baranowski et al (2011) 6-n-propylthiouracil taster status not related to reported cruciferous vegetable intake among ethnically diverse children. Nutr Res 31(8):594–600

Barrett JC, Fry B, Malle J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2):263–265

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

Campbell MC, Ranciaro A et al (2013) Origin and differential selection of allelic variation at TAS2R16 associated with salicin bitter taste sensitivity in Africa. Mol Biol Evol 31(2):288–302

Caremoli F (2011) Bitter taste perception: a preliminary study of the effects on food preferences and potential applications in human nutrition. Master's Thesis in Food Science and Human Nutrition, University of Milan, Italy

Crammer B, Ikan R (1987) Progress in the chemistry and properties of the rebaudiosides. In: Grenby TH (ed) Developments in sweeteners. Elsevier Applied Science, London, pp 45–64

Dotson C, Shaw H, Mitchell B, Munger S, Steinle N (2010) Variation in the gene TAS2R38 is associated with the eating behavior disinhibition in old order Amish women. Appetite 54(1):93–99

DuBois G, Prakash I (2012) Non-caloric sweeteners, sweetness modulators, and sweetener enhancers. Ann Rev Food Sci Technol 3:353–380

DuBois GE, Walters DE, Schiffman SS, Warwick ZS, Booth BJ, Pecore SD, Gibes K, Carr BT, Brands LM (1991) Concentration-response relationships of sweeteners. A systematic study. In: Walters DE, Orthoefer F, DuBois GE (eds) Sweeteners: discovery, molecular design and chemoreception, American Chemical Society Symposium Series 450. ACS Books, Washington, pp 261–276

Duffy VB, Davidson AC, Kidd JR, Kidd KK, Speed WC, Pakstis AJ, Reed DR, Snyder DJ, Bartoshuk LM (2004) Bitter receptor gene (TAS2R38), 6-n-propylthiouracil (PROP) bitterness and alcohol intake.Alcohol Clin Exp Res 28(11):1629–1637

Duffy VB, Hayes JE, Dinehart ME (2006) Genetic differences in sweet taste perception. In: Spillane WJ (ed) Optimising the sweet taste in foods. Woodhead Publishing, Cambridge, pp 30–53

Excoffier L, Laval G, Schneider S (2007) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 23(1):47–50

Fox AL (1932) The relationship between chemical constitution and taste. Proc Nat Acad Sci USA 18:115–120

Green BG, Shaffer GS, Gilmore MM (1993) Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. Chem Senses 18(6):683–702

Hayes JE, Bartoshuk LM, Kidd JR, Duffy VB (2008) Supertasting and PROP bitterness depends on more than the TAS2R38 gene. Chem Senses 33(3):255–265

Hayes JE, Wallace MR, Knopik VS, Herbstman DM, Bartoshuk LM, Duffy VB (2011) Allelic variation in TAS2R bitter receptor genes associates with variation in sensations from and ingestive behaviors toward common bitter beverages in adults. Chem Senses 36(3):311–319

Hellfritsch C, Brockhoff A, Stähler F, Meyerhof W, Hofmann T (2012) Human psychometric and taste receptor responses to steviol glycosides. J Agric Food Chem 60:6782–6793

Kidd J, Gravel S, Byrnes J et al (2012) Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. Am J Hum Genet 91(4):660–671

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(5610):1221–1225

Kinghorn AD (2002) The genus Stevia. Taylor & Francis, London

Kinghorn AD, Soejarto DD (1986) Sweetening agents of plant origin. Crit Rev Plant Sci 4:79–120

Kumar P, Henikoff S, Ng P (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081

Lanier SA, Hayes JE, Duffy VB (2005) Sweet and bitter tastes of alcoholic beverages mediate alcohol intake in of-age undergraduates. Physiol Behav 83(5):821–831

Meyerhof W, Batram C, Kuhn C, Brockhoff A, Chudoba E, Bufe B, Appendino G, Behrens M (2010) The molecular receptive ranges of human TAS2R bitter taste receptors. Chem Senses 35(2):157–170

Prakash I, DuBois GE, Clos JF, Wilkens KL, Fosdick LE (2008) Development of rebiana, a natural non-caloric sweetener. Food Chem Toxicol 46(Suppl. 7S):75–82

Pronin AN, Xu H, Tang H, Zhang L, Li Q, Li X (2007) Specific alleles of bitter receptor genes influence human sensitivity to the bitterness of aloin and saccharin. Curr Biol 17:1403–1408

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575

Reed DR, Zhu G, Breslin PA, Duke FF, Henders AK, Campbell MJ, Montgomery GW, Medland SE, Martin NG, Wright MJ (2010) The perception of quinine taste intensity is associated with common genetic variants in a bitter receptor cluster on chromosome 12. Hum Mol Genet 19(21):4278–4285

Reimann F, Tolhurst G, Gribble FM (2012) G-protein-coupled receptors in intestinal chemosensation. Cell Metab 15:421–431

Roudnitzky N, Bufe B, Thalmann S, Kuhn C, Gunn H, Xing C, Crider B, Behrens M, Meyerhof W, Wooding S (2011) Genomic, genetic and functional dissection of bitter taste responses to artificial sweeteners. Hum Mol Genet 20(17):3437–3449

Simons CT, Adam C, LeCourt G, Crawford c, Ward C, Meyerhof W, Slack JP (2008) The "bitter-sweet" truth of artificial sweeteners. In: DuBois G, Weerasinghe DK (eds) Sweetness and sweeteners. Biology, chemistry, and psychophysics. American Chemical Society, Washington, pp 335–354

Skotte L, Korneliussen TS, Albrechtsen A (2013) estimating individual admixture proportions from next generation sequencing data. Genetics 195(3):693–702

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68(4):978–989

Tepper BJ (2008) Nutritional implications of genetic taste variation: the role of PROP sensitivity and other taste phenotypes. Annu Rev Nutr 28:367–388

The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65

Vega-Gálvez A, Lemus-Mondaca R, Zura-Bravo L, Ah-Hen K (2012) *Stevia rebaudiana* Bertoni, source of a high-potency natural sweetener: a comprehensive review on the biochemical, nutritional and functional aspects. Food Chem 132:1121–1132

Wooding S, Gunn H, Ramos P, Thalmann S, Xing C, Meyerhof W (2010) Genetics and bitter taste responses to goitrin, a plant toxin found in vegetables. Chem Senses 35(8):685–692

Yadav KS, Guleria P (2012) Steviol glycosides from Stevia: biosynthesis pathway review and their application in foods and medicine. Crit Rev Food Sci Nutr 52:988–998

Zhao L, Kirkmeyera S, Tepper B (2003) A paper screening test to assess genetic taste sensitivity to 6-n-propylthiouracil. Physiol Behav 78:625–633

Risso et al. BMC Evolutionary Biology 2014, 14:198
http://www.biomedcentral.com/1471-2148/14/198

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Genetic variation in taste receptor pseudogenes provides evidence for a dynamic role in human evolution

Davide Risso[1,2], Sergio Tofanelli[3], Gabriella Morini[4], Donata Luiselli[2] and Dennis Drayna[1*]

Abstract

Background: Human bitter taste receptors are encoded by a gene family consisting of 25 functional *TAS2R* loci. In addition, humans carry 11 *TAS2R* pseudogenes, some of which display evidence for substantial diversification among species, showing lineage-specific loss of function. Since bitter taste is thought to help prevent the intake of toxic substances, diversity at *TAS2R* genes could reflect the action of natural selection on the ability to recognize some bitter compounds rather than others. Whether species-specific variation *in TAS2R* pseudogenes is solely the result of genetic drift or whether it may have been influenced by selection due to different feeding behaviors has been an open question.

Results: In this study, we analyzed patterns of variation at human *TAS2R* pseudogenes in both African and non-African populations, and compared them to those observable in nonhuman primates and archaic human species. Our results showed a similar worldwide distribution of allelic variation for most of the pseudogenes, with the exception of the *TAS2R6P* and *TAS2R18P* loci, both of which presented an unexpected higher frequency of derived alleles outside Africa. At the *TAS2R6P* locus, two SNPs were found in strong linkage disequilibrium ($r^2 > 0.9$) with variants in the functional *TAS2R5* gene, which showed signatures of selection. The human *TAS2R18P* carried a species-specific stop-codon upstream of four polymorphic insertions in the reading frame. SNPs at this locus showed significant positive values in a number of neutrality statistics, and age estimates indicated that they arose after the homo-chimp divergence.

Conclusions: The similar distribution of variation of many human bitter receptor pseudogenes among human populations suggests that they arose from the ancestral forms by a unidirectional loss of function. However we explain the higher frequency of *TAS2R6P* derived alleles outside Africa as the effect of the balancing selection acting on the closely linked *TAS2R5* gene. In contrast, *TAS2R18P* displayed a more complex history, suggesting an acquired function followed by a recent pseudogenization that predated the divergence of human modern and archaic species, which we hypothesize was associated with adaptions to dietary changes.

Keywords: Bitter taste, Evolution, Genetic polymorphisms, Pseudogenization, *TAS2Rs*

Background

The perception of bitter taste is thought to have evolved as a protection from toxic and harmful foods [1,2]. However, not all bitter compounds evoke the same aversive reaction in different species and, in particular, among primates [3]. Therefore, it has been hypothesized that consumption or rejection response depends upon the relative occurrence of bitter and potentially toxic compounds in an animal's diet [4]. The number of different bitter taste receptors varies greatly across species [5], but all of them belong to a family of seven-transmembrane G protein coupled receptors (GPCRs) known as *TAS2Rs* or *T2Rs*. In humans, this gene family is encoded by 25 functional *TAS2R* loci, which reside on chromosomes 5, 7 and 12 [6]. In addition to these genes, humans also carry 11 *TAS2R* pseudogenes [7]. Polymorphisms in these genes have been shown to modulate the taste response to different compounds, both natural and synthetic [8-10], and

* Correspondence: drayna@nidcd.nih.gov
[1]National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD 20892, USA
Full list of author information is available at the end of the article

BioMed Central

© 2014 Risso et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

showed evidence of evolutionary pressures [11-13], highlighting the importance of studying both *TAS2R* genes and pseudogenes.

Pseudogenes are considered genomic fossils, classically defined as genomic loci with sequence similarity to functional genes, but lacking coding potential [14,15], often due to disruptive mutations such as frameshifts, premature stop codons and deletions. However, recent studies have demonstrated that some pseudogenes may likely have a function, providing evidence for their plasticity and a dynamic role in evolution [16-18]. A number of authors have described these events with a "less-is-more" hypothesis, suggesting that gene loss, or pseudogenization, may serve as an engine of evolutionary change, especially in human evolution [19,20]. In fact, previous phylogenetic analyses demonstrated that remarkable variation exists in both *TAS2Rs* genes and pseudogenes in different lineages, suggesting that the ability to perceive bitterness may be mostly due to the repertoire of *TAS2R* genes [21,22]. For instance, the mouse genome contains 33 functional *TAS2Rs* loci but only three pseudogenes, and the evolutionary relationships between human and mouse genes were shown to fall into three categories, depending on their orthology: 1) one-to-one orthology; 2) one-to-multiple orthology; 3) multiple-to-one orthology [23,15,7]. In contrast, zebrafish and chicken have only four and three *TAS2R* loci, respectively [24,5]. Comparative analyses showed that primate *TAS2Rs* had a higher ratio of non-synonymous/synonymous substitutions and a lower selective pressure on this gene family compared to rodents [25,26].

To better understand the evolutionary mechanisms underlying taste receptor pseudogenization events in the human lineage, we performed a survey of the human genetic variation at nine human TAS2RP loci. We then compared these to the homologous pseudogenes present in archaic human forms (i.e. Neandertal and Denisovan), as well as in other five primate species, Chimpanzee, Gorilla, Orangutan, Gibbon and Mouse Lemur. We examined the intra- and inter-species variation patterns to search for possible footprints of natural selection at these loci, with the goal of understanding the potential adaptive role of taste receptor pseudogenes in the evolutionary history of modern humans.

Methods

Genes and data sets

Nine human *TAS2R* pseudogenes were selected from the literature and are shown in Table 1. Modern human variants located in these genes, together with their surrounding genomic regions (+/− 100 kbp), were retrieved from the 1000 Genomes Project PHASE I database [27,28], which provided data on a total number of 1,092 individuals belonging to 14 different populations. Neandertal and Denisovan variants/sequences were retrieved at the UCSC Table Browser [29,30]. Pseudogene sequences in humans (*Homo sapiens*), five other primates (*Pan troglodytes*, *Gorilla gorilla*, *Pongo abelii*, *Nomascus leucogenys* and *Microcebus murinus*) and one rodent (*Mus musculus*), were obtained from the Ensembl Genome Browser [31,32]. This project was approved by the NIH protocol 01-DC-0230, reviewed by the NIH/NINDS CNS Blue Panel IRB and complied with the Helsinki Declaration of Ethical Principles.

Table 1 *TAS2R* pseudogenes chromosomal positions (GRCh37)

Pseudogene	Chromosome	Position
TAS2R2P	7	12,530,72-12,531,630
TAS2R6P	7	141,487,614-141,488,440
TAS2R62P	7	143,134,127-143,135,066
TAS2R12P	12	11,047,542-11,048,481
TAS2R15P	12	11,117,024-11,117,951
TAS2R18P	12	11,311,384-11,312,293
TAS2R63P	12	11,200,931-11,201,855
TAS2R64P	12	11,229,368-11,231,770
TAS2R67P	12	11,332,272-11,333,061

Haplotype analyses

Haploview [33] was used to identify Linkage Disequilibrium (LD) patterns and haplotype blocks at the selected loci using the solid spine approach. Genealogical relationships among inferred haplotypes were constructed using the median-joining algorithm implemented in the Network 4.5 program [34]. Haplotypes distribution across human populations were investigated using PLINK v.1.07 [35].

Population genetics analyses

Arlequin v.3.5 [36] was used to compute summary statistics, such as nucleotide diversity (π), estimated heterozygosity (EH) and number of polymorphic sites (PS). Chi square-tests were performed to compare allele frequencies among different populations and the adopted significance threshold was adjusted using the Bonferroni correction (i.e. adjusted p = p value X number of individual tests). Metric Multidimensional Scaling (MDS) analyses based on the obtained F_{ST} values, were performed with STATISTICA v. 6.0 (Stat-Soft Inc, Tulsa, OK).

Phylogenetic studies

MUSCLE v.3.3 [37] (Edgar 2004) was used to perform multi-alignments of the examined sequences. A maximum likelihood tree was constructed with MEGA v.6.0 [38] using the Tamura-Nei substitution model. To assess the relative support for each clade, bootstrap values were calculated from 10.000 analysis replicates, and the cut-off point for bootstrap replication was 50%.

Neutrality test and age estimates

To test whether patterns of allele/haplotype frequencies and tree topology were consistent with neutral expectations, we performed three neutrality tests. For this purpose, the DNASP package [39] was used to calculate Tajima's D and Fu's FS values at each locus. Since deviations found using these tests could be caused by selection and/or demographic processes (e.g. population expansion and/or bottlenecks), we also performed the Li's MFDM test [40], which is more robust in the presence of population size changes. Finally, GENETREE [41,42] was used to infer the estimated age of selected variants and the time of the most recent common ancestor (TMRCA).

Results

Sequence variation

A total of 47 single nucleotide polymorphisms (SNPs) (Additional file 1: Table S1), annotated according to dbSNP Build 137, were observed in the nine selected pseudogenes. 32 of them had a minor allele frequency (MAF) above 5%. Unexpected patterns of genetic variation were observed for two pseudogenes, *TAS2R18P* and *TAS2R6P*, in populations from different continents. In particular, the derived alleles at four *TAS2R18P* SNPs (*rs2290318*, *rs2290319*, *rs61928604* and *rs61928603*) showed significantly increased frequency (0.5) in non-African populations compared to African populations (0.053) (Fisher's exact test, adjusted $p < 0.01$). Similarly, the derived alleles of *TAS2R6P* polymorphisms *rs1859645* and *rs11761380* showed a significantly different distribution (0.24 vs. 0.57) between Africa and other continents (Fisher's exact test, adjusted $p < 0.01$). Both Neandertal and Denisovan genomes carried the derived alleles of the *TAS2R18P* SNPs *rs2290318*, *rs2290319*, *rs61928604* and *rs61928603*. For the *TAS2R6P* SNPs *rs1859645* and *rs11761380*, the Denisovan genome showed ancestral alleles, while the Neandertal genome was heterozygous at both SNPs. In addition, derived alleles of two SNPs located in *TAS2R67P* (*rs319269* and *rs34648613*) and one in *TAS2R64P* (*rs68071847*) were more frequent in African populations with respect to non-African ones (0.68 vs. 0.11, Fisher's exact test, adjusted $p < 0.01$), and were also present in both Neandertal and Denisovan genomes. Finally, the derived allele of one SNP located in *TAS2R63P* (*rs2597986*) was present only in a few African individuals (1.58%), as well as in the two archaic species.

Haplotype structure

A total of eight haplotype blocks (Additional file 2: Figure S1 A-B) were inferred in the nine studied pseudogenes. Five of them (H1-H5) were on chromosome 7, with the remaining three (H6-H8) located on chromosome 12. H5, which contained *TAS2R6 rs11761380* and *rs1859645*, and H6, made up of *TASR18 rs2290318* and *rs2290319*, showed high level of LD ($R^2 = 1$) and differed significantly across human populations (Table 2), with African populations carrying a lower percentage of derived alleles at these loci. The other haplotype blocks (H1, H2, H3, H4, H7 and H8) showed no significant differences among continents.

To test the significance of the observed structure, the distribution of the H5 and H6 haplotypes in human populations was also investigated by means of Analysis of Molecular Variance (AMOVA). For both *TAS2R6P* and *TAS2R18P*, most of variation was accounted for by differences within populations (86.67% and 72.07%, respectively), with a smaller percentage attributed to differences among (13.2% and 27.12%) and within groups (0.13% and 0.81%). The global F_{ST} values were 0.13 for *TAS2R6P* and 0.28 for *TAS2R18P*, while those related to H5 and H6 haplotypes were 0.14 and 0.31, respectively.

We constructed median-joining networks for the *TAS2R6* and *TAS2R16* pseudogenes in order to better understand the relationships between the inferred haplotypes. The resultant topologies (Additional file 3: Figure S2 A-B) identified of two major clusters defined by the presence of either the derived (GC for *TAS2R6P*, CA for *TAS2R18P*) or ancestral (AA for *TAS2R6P*, GC for *TAS2R18P*) alleles at the H5 and H6 haplotypes.

Summary statistics and population structure

The pattern of diversity shown in most of these pseudogenes was in accordance with the usual distribution of human genetic variation [38], where the diversity is higher in African populations than in non-African ones. However, a different situation was observed in *TAS2R6P* and *TAS2R18P*. African groups (i.e. ASW, LWK and YRI) showed lower values of both nucleotide diversity (π) and estimated heterozigosity (EH) at the*TAS2R18P* locus. With the exception of the ASW group, the same pattern was observed for the *TAS2R6P* gene. The number of polymorphic sites (PS) was similar in all the studied populations (Additional file 4: Table S2 A-B).

To explore the population structure at these two outlier loci, we calculated pairwise Wright's F_{ST} indices as a measure of genetic differentiation among different groups. A multidimensional scaling (MDS) metric was then used to plot the obtained F_{ST} values. This produced a clear distinction between continental populations for the observed variation. As shown in Figure 1, *TAS2R18P* was separated from African and Asian populations along the first dimension, with a distinct cluster, containing the European and Latin American (admixed) populations, occupying an intermediate position between them. Variation at the *TAS2R6P* locus showed a similar pattern of population structure among continental clusters (Additional file 5: Figure S3).

Table 2 Comparisons of *TAS2R* pseudogene haplotype blocks among the studied populations

HBlock	Chr	PseudoGene	SNP1	SNP2	MA1	MA2	F_AFR	F_ASN	F_EUR	F_AMR	Comparison	adjusted P-value*
H5	7	TAS2R6P	rs1161380	rs1859645	C	G	0,24	0,71	0,46	0,54	AFR/EUR	1,16E-15
											AFR/AMR	1,5E-18
											AFR/ASN	1,5E-17
											EUR/ASN	2,11E-09
H6	12	TAS2R18P	rs2290318	rs2290319	C	A	0,05	0,74	0,35	0,41	AFR/EUR	2,13E-13
											AFR/AMR	1,03E-09
											AFR/ASN	1,11E-20
											EUR/ASN	1,07E-08
											ASN/AMR	2,21E-06

Hblock, Haplotype block; Chr, Chromosome; MA, Minor Allele; F, Frequence.
*after Bonferroni correction
AMR, Latin Americans; AFR, Africans, EUR, Europeans; ASN, Asians.

Inter-specific comparison

We also created a phylogenetic tree using the nucleotide sequences of *TAS2R6P* and *TAS2R18P* to understand the phylogenetic relationships between both modern and archaic humans and other primates. As shown in Figure 2, *TAS2R6P* sequences in Prosimians, Hylobatidae and Hominidae are closely related, sharing a common ancestor when compared to the outgroup sequence (i.e. *Mus musculus*) and showing short branch lengths of the tree, suggesting an ancient pseudogenization event at this locus. Longer branch lengths in Prosimians and Anthropoidea after their divergence may indicate that the pseudogenization at *TAS2R18P* locus occurred independently in the two lineages. It should be noted that these conclusions might be affected by problems with ortholog detection, considering the high dynamism and extreme complexity of this gene family.

However, DNA sequence similarities confirmed these results, showing a high percentage of *TAS2R6P* identity between Hominidae and Hylobatidae (>95%) and Prosimians, Hominidae and Hylobatidae (>75%). *TAS2R18P* sequences showed a lower degree of similarities (<70%) between primates and Prosimians (Additional file 6: Table S3 A-B).

In addition, multi-alignments of *TAS2R6P* and *TAS2R18P* produced results in agreement with the findings described above. As shown in Figure 3A, *TAS2R6P* shows a high degree of conservation among all the species examined. Some of its stop codons, such as those at positions 7:141487959 and 7:141488049, are shared with the earlier representatives of this lineage. Besides

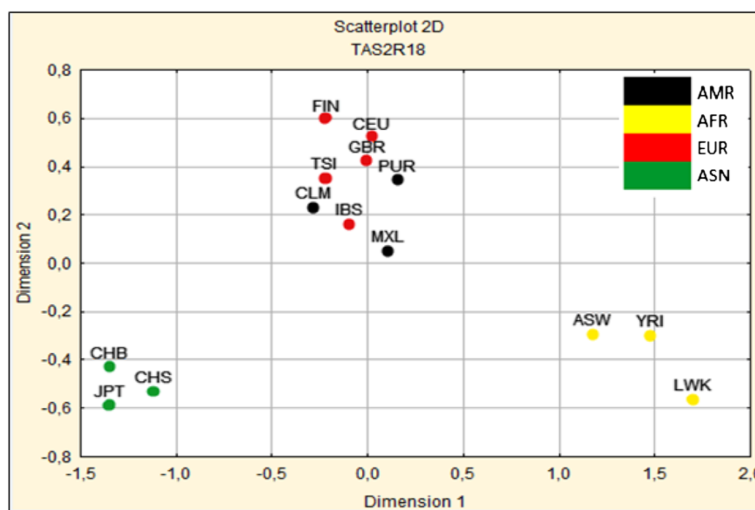


Figure 1 Two-dimensional scaling of the F_{ST} distance matrix calculated for *TAS2R18P*. AMR, Latin Americans; AFR, Africans, EUR, Europeans; ASN, Asians; YRI, Yoruba from Ibadan (Nigeria); LWK, Luhya from Webuye (Kenya); ASW, people with African ancestry from Southwest United States; IBS, Iberian populations from Spain; TSI, Tuscans from Italy; CEU, Utah residents with Northern and Western European ancestry; GBR, British from England and Scotland; FIN, Finnish; PUR, Puerto Ricans; CLM, Colombians from Medellin; MXL, people with Mexican ancestry from Los Angeles; JPT, Japanese from Tokyo; CHB, Han Chinese from Beijing; CHS, Han Chinese from Southern China.

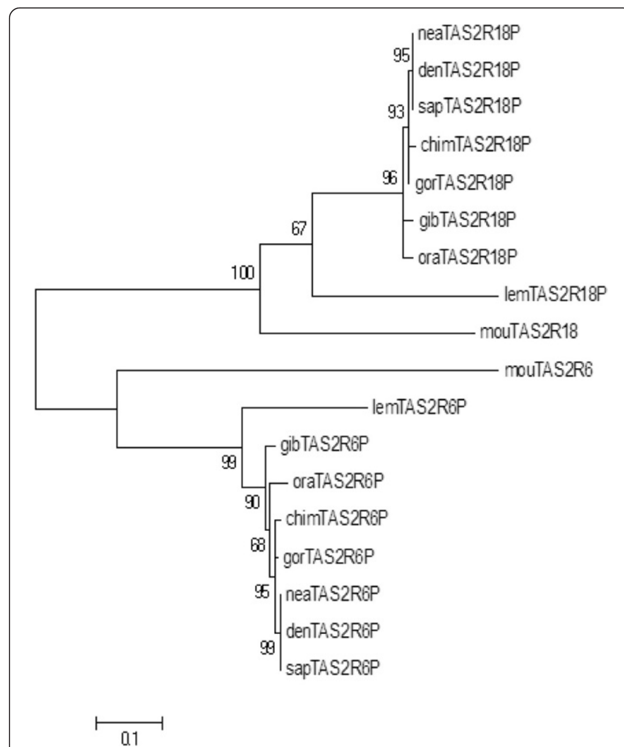


Figure 2 Maximum Likelihood tree of *TAS2R18P* and *TAS2R6P*. Bootstrap values are shown on the branch forks, the scale bar corresponds to the branch length and indicates 0.1 amino acid substitution per site. *mou*, Mouse; *lem*, Lemur; *gib*, Gibbon; *ora*, Orangutan; *gor*, Gorilla; *chi*, Chimpanzee; *nea*, Neadertal; *den*, Denisova; *sap*, Sapiens.

species-specific changes found in lemur, and in agreement with the ancient time of the Strepsirrhini-Haplorrhini divergence (64 Mya), only one unique stop-loss mutation (TAG > CAG) was found in the chimpanzee *TAS2R6P*. All the human species (i.e. H. sapiens, Neandertal and Denisova) shared a fixed stop-gain mutation (TGG > TAG) on *TAS2R18P* (Figure 3b), which was not present in other primates that carry other stop-codon mutations at this locus and show a high degree of conservation between Hominidae and Hylobatidae. The reduced conservation between Prosimians and other primates is consistent with the view that the two pseudogenization events appeared at different times during their evolutionary histories.

To better understand the species-specific changes that occurred in the *TAS2R18P* locus, the human exonic region of this gene was investigated (Figure 4). Four polymorphic insertions (*rs10619393*, *rs373807934*, *rs66547287* and *rs113657094*) were identified in the reading frame, upstream of the seven stop-codons shared across species and downstream of the human-specific stop codon. One of them (*rs113657094*) showed clear differential distribution among continental clusters of populations, with Africans displaying a significantly lower frequency of the A insertion (42%) compared to other ethnic groups (86% in Latin Americans, 88% in Europeans and 97% in Asians) (Fisher's exact test, adjusted $p < 0.01$). For comparison, the *TAS2R6P* sequence was also examined and neither insertions nor other structural variants were found.

Relationships with functional genes

A 200 kb (100 kb in both 5' and 3' directions) region surrounding both *TAS2R6P* and *TAS2R18P* was explored for linkage disequilibrium patterns. The *TAS2R6P rs11761380* and *rs1859645* polymorphisms were found to be in strong LD ($r^2 > 0.9$) with SNPs *rs62477710*, *rs10952507* and *rs6962558* that reside in the functional *TAS2R5* gene (Additional file 7: Table S4). In contrast, *TAS2R18P* showed no r^2 values above 0.3 between variants. In addition, we analyzed the sequences of all the genes found in this

A

Starting Position (GRCh37)	Sapiens	Neandertal	Denisova	Chimp	Gorilla	Orangutang	Gibbon	Mouse Lemur
7:141487731	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TG-
7:141487767	TAG	TAG	TAG	CAG	TAG	TAG	TAG	TGG
7:141487851	TAA	TAA	TAA	TAA	TAA	TAA	TAA	CAA
7:141487959	TGA	TGA	TGA	TGA	TGA	TGA	TGA	TGA
7:141487986	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAA
7:141488049	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
7:141488082	TAA	TAA	TAA	TAA	TAA	TAA	TAA	GAA
7:141488124	TGA	TGA	TGA	TGA	TGA	TGA	TGA	TGA
7:141488304	TGA	TGA	TGA	TGA	TGA	TGA	TGA	TGA

B

Starting Position (GRCh37)	Sapiens	Neandertal	Denisova	Chimp	Gorilla	Orangutang	Gibbon	Mouse Lemur
12:11312002	TAG	TAG	TAG	TGG	TGG	TGG	TGG	TGG
12:11311831	TGA	TGA	TGA	TGA	TGA	TGA	TGA	TGA
12:11311798	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAC
12:11311774	TAG	TAG	TAG	TAG	TAG	TAG	TAG	GGT
12:11311633	TAA	TAA	TAA	TAA	TAA	TAA	TAA	CAA
12:11311585	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TCA
12:11311546	TGA	TGA	TGA	TGA	TGA	TGA	TGA	TAT
12:11311414	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA

Figure 3 Multi-alignment of *TAS2R6P* (A) and *TAS2R18P* (B) stop-codon sequences. Conserved sequences are indicated in grey. Changes between humans and other primates are denoted in red, changes between Prosimians and primates are noted in blue, and the change between Chimp and other primates is noted in green.

```
>sapTAS2R18P
ATGTTCGTTGGAATTAATATTTTCTTTCTGGTGGTGGCAACAAGAGGACTTGTCTTAGGAATGCTGGGAAACGGGCTCATTGGACTGGTA
AACTGCATTGAGTGGGCCAAGAGTTGGAAGGTCTCATCAGCTGATTTCATCCTCACCAGCTTGGCTATAGTCAGAATCATTCGACTGTAT
TTAATACTATTTGATTCATTTATAATGGTATTGTCCCCTCATCTATATACCATCCGTAAACTAGTAAAACTGTTTACTATTCTTTGGGCA
TTAATTAATCAGTTAAGTATCTAGTTTGCCACCTGCCTAAGCATTTTCTACTTGCTTAAGATAGCCAATTTCTCCCGCTTAT (A) TTTT (A) T (A) GCC
TGGCTGAAGTGGAGAATGAACAGAGTGGTTCTTGTGCTTTTCCTGCGGTCTTTGTTCTTATCGTTTGTTTACCTTTTTATGTCCAATGCCA
TTAGTGAGTTGTGAAAAAAACATGACTTTGCACTCAGATACAAGTAAAATAGTCTGTCTTCAAGGCCTTAGGCTTCTCAGCTTGACATACG
TTATTCCCTTTCTTCTGACTCTGACCTCTTTGCTCCTTTTATTTATATCCTTAGTGAGACACACCAAGAATTTGCAGCTCAACTCTCTGGG
CTCAAGGGACTCCAGCACAGAGGCCC (AGTTA) ATAAAAGGGCCATGAAAATGGTGATAGCCTTCCTCCTCCTTTTTTTTATTAACTTTATTTCCA
CTTTAATAGGAGATTGGATCTTCCTTGAGGTAGAGAATTATCAGGTCATGACGTTTATTATGATGATTTACTTGCCTTTCCCTCAGGCCA
CTCATTTATTATAATTTTGGGAAACAACAAGCTAAGACAGAGCTCCTTGAGACTACTGTGGCATCTTAAATTCTCTCTGAAAAAAGCAAAA
CCTTTA
```

Figure 4 Human *TAS2R18P* sequence. Insertions are in bold, polymorphic SNPs in red and stop-codons in blue. Human-specific stop-codon is circled.

enlarged genomic interval and compared them to the pseudogene sequences, in order to document the sequence similarities between pseudogenes and functional genes. *TAS2R6P* shares a common ancestor with the functional *TAS2R5* gene (62.59% identity), while *TAS2R18P* has a high level of identity with the functional *TAS2R42* gene and with the *TAS2R67P* pseudogene (66.59 and 67.02%, respectively) (Additional file 8: Figure S4). Pseudogenes derived from a very recent pseudogenization event (i.e. *TAS2R64P*) showed a much higher sequence similarity (91.32%) with their functional forms (i.e. *TAS2R48*), indicating ancient pseudogenization events at the *TAS2R6P* and *TAS2R18P* loci.

Test of neutrality and mutation age estimates

Neutrality tests based on the analysis of frequency spectrum of variants were performed for each population. European and Latin American groups showed significantly positive values for both *TAS2R18P* Tajima's D and Fu's FS. Africans and Asians also showed positive values (1.04, 2.4 and 1.5, 2.8 respectively) for these statistics, although these did not reach significance (Figure 5). In contrast, *TAS2R6P* did not show evidence of deviation from neutral expectations in all the examined populations ($p > 0.05$). However,

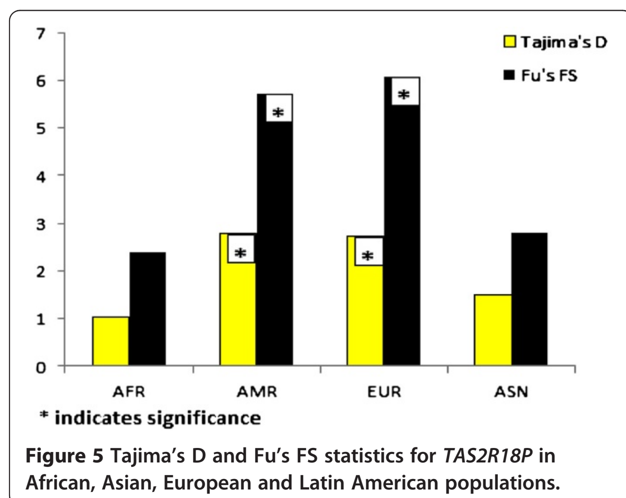


Figure 5 Tajima's D and Fu's FS statistics for *TAS2R18P* in African, Asian, European and Latin American populations.

rs62477710, *rs10952507* and *rs6962558* polymorphisms in the adjacent *TAS2R5* gene showed significantly positive values of Tajima's D and Fu's FS (Additional file 9: Figure S5). The maximum frequency of derived mutations (MFDM) test, which uses tree topology to infer selection, showed similar results. In fact, both *TAS2R18P* and *TAS2R5* showed significant p-values (p = 9.16E-4 and p = 0.025, respectively), indicating that recent selection has acted on these loci. However, the p-values for this test were not significant for *TAS2R6P* ($p > 0.05$).

Finally, the ages of these variants were estimated using a coalescent-based method implemented in the GENETREE package. Estimated ages turned out to be considerably ancient. For SNPs in *TAS2R18P*, *rs2290318* was estimated to be 1,553,750 +/− 632,500 years old, while rs2290319 was estimated to be 1,113,750 +/− 495,000 years old, with a TMRCA of 3,520,000 +/−1,198,750 years. Similar values were also obtained for the *TAS2R6P* SNPs *rs11761380* and *rs1859645*, with age estimates of 1,382,462 +/− 326,856 and 1,600,637 +/− 421,558 years, and a TMRCA of 3,457,712 +/− 1,326,856 years.

Discussion

A small number of studies have examined genes encoding functional bitter taste receptors and their corresponding pseudogenes from an evolutionary perspective [5,7,23,25,26]. In the present study, we focused on the recent evolutionary history of bitter taste receptor pseudogenes in modern humans. We found a modest level of variation and few haplotype blocks, as expected for pseudogenes in general. However, African populations showed a significantly reduced frequency of derived forms at *TAS2R6P* and *TAS2R18P* polymorphisms, at both the allelic and haplotypic level. Genomes of the archaic Neandertal and Denisova species also differed in the distribution of these derived alleles compared to modern human populations. In contrast to the usual pattern of human variation, the highest genetic diversity was found in non-African populations. Similarly, heterozygosity was higher in non-African populations. In addition, results from both population structure analyses and AMOVA demonstrated that

most of variation was due to differences within populations. In fact, Europeans and Latin Americans fell into the same cluster and did not differ in allelic distribution, suggesting a similar pattern of SNP distribution in these two populations and highlighting the admixed nature of the 1000 Genomes American populations.

Previous work [5,7,23] indicated that the pseudogenization event of *TAS2R6P* is very old, predating the divergence of Antrhopoidea from Prosimians. The *TAS2R18P* pseudogenization event occurred later in the primate genealogy, but still prior to the appearance of the Hominidae family.

Our results were consistent with these findings: *TAS2R6P* variation patterns enabled the identification of a unique cluster grouping Prosimians, Hylobatidae and Hominidae families and high level of sequence similarity among primates (>95%), as well as between primates and Prosimians (>75%). *TAS2R18P* sequence similarity was lower between the mouse lemur and other primates (<70%) and the reconstructed phylogenetic trees showed longer branch lengths after the divergence of the two lineages. In addition, sequence similarity confirmed this view, showing a lower percentage (<70%) of identity between these two pseudogenes and their functional forms, when compared to the similarities between very recent pseudogenes and their functional forms (>90%).

The analysis of the genomic regions surrounding these pseudogenes showed that *TAS2R6P* polymorphisms *rs11761380* and *rs1859645* were in strong linkage disequilibrium ($r^2 > 0.9$) with *rs62477710*, *rs6962558* and *rs6962558* SNPs of the functional *TAS2R5* gene, suggesting that evolutionary forces acting on this gene could have driven the allele differentiation at the TAS2R6P locus. To test this assumption, Tajima's D and Fu's FS tests were performed: these analyses showed that *TAS2R6P* did not show evidence of departure from neutral expectations while the neighboring *TAS2R5* gene, which shared more than 60% sequence similarity with *TAS2R6P*, showed significant positive values for both these statistics. Since positive values of these tests may indicate balancing selection or may be the result of the confounding effect due to demographic history [43,44], the maximum frequency of derived mutations (MFDM) test was applied to further investigate the evolutionary history of these regions. The results of this test indicate that *TAS2R6P* has not undergone recent selective pressures, whereas *TAS2R5* showed significant signatures of selection. This is consistent with balancing selection maintaining multiple alleles for long evolutionary times, and extending to maintain allele frequencies at closely linked neutral sites [45,46]. The application of this test to *TAS2R18* also showed significant signatures of selection and both Tajima's D and Fu's FS values were significantly positive at this locus in European and Latin American populations. This suggests that balancing selection potentially acted on the genes in these groups. In addition, these populations showed increased values of nucleotide diversity and heterozygosity in the *TAS2R18P* pseudogene compared to neighboring loci. Moreover, global F_{ST} values were unusually high compared to the typical range of 0.10 to 0.16 for estimated F_{ST} values in global populations [47,48], suggesting a high level of genetic differentiation among worldwide populations. These data are consistent with a scenario of balancing selection maintaining *TAS2R18P* alleles and enhancing genetic diversity at this locus in European and Latin American populations.

Variants at *TAS2R18P* showed no evidence of association with adjacent functional genes. This pseudogene carries a human-specific stop-codon that is shared among H. sapiens Neandertal and Denisova. In addition, the exonic region of this pseudogene carried four polymorphic insertions in the reading frame, upstream of the shared stop-codons and downstream of the human-specific one. These data suggest that these insertions may have shifted the human *TAS2R18P* reading frame, with a consequent acquired function of this gene, shortly after the homo-chimp divergence. Such an event, followed by balancing selection operating outside Africa, would produce the observed *TAS2R18P rs2290318* and *rs2290318* different allele distributions in human populations. The estimated ages of these mutations indicate that they arose after the divergence of humans from chimpanzee that occurred 7–8 million years ago [49]. We hypothesize that the human-specific stop codon located upstream of all these insertions represents a second inactivation, which happened before the split between H. sapiens, Neandertal and Denisova (from 400,000 to 800,000 years ago).

Conclusions

Our results provide evidence for a dynamic role for *TAS2R18P* in primate evolution, suggesting that this locus may have acquired its function during the evolution of the human lineage, shortly after the homo-chimp divergence. This was followed by a much more recent deactivation due to the stop-gain mutation which was shared among modern humans, Neandertals and Denisovans. We speculate that this event may have been due to the disappearance of some bitter compound only found in Africa, which was specifically recognized by the product of this pseudogene, and that therefore the functionality of *TAS2R18P* was useful only in a given stage of human evolution.

Availability of supporting data

The article does not report new empirical data since the analyzed sequences were already deposited at public databases, including the 1000 Genomes Project, the UCSC Genome Browser and the Ensemble Genome Browser (see the Methods section for further details).

Additional files

Additional file 1: Table S1. Identified SNPs in the examined pseudogenes.

Additional file 2: Figure S1. A) Haplotype blocks found on chromosome 7. B) Haplotype blocks found on chromosome 12.

Additional file 3: Figure S2. Median-joining network of inferred haplotypes on A) *TAS2R6P* (H5) and B) *TAS2R18P* (H6) genes.

Additional file 4: Table S2. Number of Polymorphic Sites (PS), Nucleotide Diversity (π) and Estimated Heterozigosity (EH) in the examined populations in A) *TAS2R6P* and B) *TAS2R18P*. YRI, Yoruba from Ibadan (Nigeria); LWK, Luhya from Webuye (Kenya); ASW, people with African ancestry from Southwest United States; IBS, Iberian populations from Spain; TSI, Tuscans from Italy; CEU, Utah residents with Northern and Western European ancestry; GBR, British from England and Scotland; FIN, Finnish; PUR, Puerto Ricans; CLM, Colombians from Medellin; MXL, people with Mexican ancestry from Los Angeles; JPT, Japanese from Tokyo; CHB, Han Chinese from Beijing; CHS, Han Chinese from Southern China.

Additional file 5: Figure S3. Two dimensional scaling of FST distance matrix calculated for *TAS2R6P* in the examined populations.

Additional file 6: Table S3. Sequence similarity among the studied species in A) *TAS2R6P* and B) *TAS2R18P* genes.

Additional file 7: Table S4. LD values in *TAS2R6P* surrounding regions.

Additional file 8: Figure S4. Cladograms based on comparisons between *TAS2R6P* and *TAS2R18P* and neighboring genes.

Additional file 9: Figure S5. Tajima's D and Fu's FS statistic in African, Asian, European and Latin American populations for *TAS2R5 rs62477710, rs10952507* and *rs6962558* polymorphisms.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
DR and DD designed the study, performed the initial analyses, constructed the figures and tables, and wrote the first draft of the manuscript. DL and GM drafted the manuscript and guided all aspects of the study. ST contributed to design and implementation of statistical tests. All authors read and approved the final manuscript.

Acknowledgements
This research was supported by the National Institute on Deafness and Other Communication Disorders Intramural grant number Z1A-000046-14. We would like to thank Dr. Marco Sazzini for help in preparing the manuscript and Drs. Igor Rogozin and Rob Morell for their helpful suggestions.

Author details
[1]National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD 20892, USA. [2]Department of BiGeA, Laboratory of Molecular Anthropology and Centre for Genome Biology, University of Bologna, via Selmi 3, 40126 Bologna, Italy. [3]Department of Biology, University of Pisa, Via Ghini 13, 56126 Pisa, Italy. [4]University of Gastronomic Sciences, Piazza Vittorio Emanuele 9, Bra, Pollenzo 12042, CN, Italy.

Received: 20 May 2014 Accepted: 1 September 2014
Published online: 13 September 2014

References
1. Garcia J, Hankins WG: The evolution of bitter and the acquisition of toxiphobia. *Olfaction and Taste* 1975, 5:39–45.
2. Chandrashekar J, Mueller KL, Hoon MA, Adler E, Feng L, Guo W, Zuker CS, Ryba NJ: T2Rs function as bitter taste receptors. *Cell* 2000, 100(6):703–711.
3. Imai H, Suzuki N, Ishimaru Y, Sakurai T, Yin L, Pan W, Abe K, Misaka T, Hirai H: Functional diversity of bitter taste receptor TAS2R16 in primates. *Biol Lett* 2012, 8(4):652–656.
4. Glendinning JI: Is the bitter rejection response always adaptive? *Physiol Behav* 1994, 56(6):1217–1227.
5. Shi P, Zhang J: Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol Biol Evol* 2006, 23(2):292–300.
6. Adler E, Hoon MA, Mueller KL, Chandrashekar J, Ryba NJ, Zuker CS: A novel family of mammalian taste receptors. *Cell* 2000, 100(6):693–702.
7. Go Y, Satta Y, Takenaka O, Takahata N: Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates. *Genetics* 2005, 170(1):313–326.
8. Bufe B, Breslin PA, Kuhn C, Reed DR, Tharp CD, Slack JP, Kim UK, Drayna D, Meyerhof W: The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. *Curr Biol* 2005, 15(4):322–327.
9. Wooding S, Gunn H, Ramos P, Thalmann S, Xing C, Meyerhof W: Genetics and bitter taste responses to goitrin, a plant toxin found in vegetables. *Chem Senses* 2010, 35(8):685–692.
10. Risso D, Morini G, Pagani L, Quagliariello A, Giuliani C, De Fanti S, Sazzini M, Luiselli D, Tofanelli S: Genetic signature of differential sensitivity to stevioside in the Italian population. *Genes Nutr* 2014, 9(3):401.
11. Wooding S, Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D: Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. *Am J Hum Genet* 2004, 74(4):637–646.
12. Campbell MC, Ranciaro A, Froment A, Hirbo J, Omar S, Bodo JM, Nyambo T, Lema G, Zinshteyn D, Drayna D, Breslin PA, Tishkoff SA: Evolution of functionally diverse alleles associated with PTC bitter taste sensitivity in Africa. *Mol Biol Evol* 2012, 29(4):1141–1153.
13. Campbell MC, Ranciaro A, Zinshteyn D, Rawlings-Goss R, Hirbo J, Thompson S, Woldemeskel D, Froment A, Rucker JB, Omar SA, Bodo JM, Nyambo T, Belay G, Drayna D, Breslin PA, Tishkoff SA: Origin and differential selection of allelic variation at TAS2R16 associated with salicin bitter taste sensitivity in Africa. *Mol Biol Evol* 2014, 31(2):288–302.
14. Balakirev ES, Ayala FJ: Pseudogenes: are they "junk" or functional DNA? *Annu Rev Genet* 2003, 37:123–151.
15. Zhang Z, Gerstein M: Large-scale analysis of pseudogenes in the human genome. *Curr Opin Genet Dev* 2004, 14(4):328–335.
16. Bekpen C, Marques-Bonet T, Alkan C, Antonacci F, Leogrande MB, Ventura M, Kidd JM, Siswara P, Howard JC, Eichler EE: Death and resurrection of the human IRGM gene. *PLoS Genet* 2009, 5(3):e1000403.
17. Korrodi-Gregório L, Abrantes J, Muller T, Melo-Ferreira J, Marcus K, da Cruz e Silva QA, Fardilha M, Esteves PJ: Not so pseudo: the evolutionary history of protein phosphatase 1 regulatory subunit 2 and related pseudogenes. *BMC Evol Biol* 2013, 13:242.
18. Zhang Q: Using pseudogene database to identify lineage-specific genes and pseudogenes in humans and chimpanzees. *J Hered* 2014, 105(3):436–443.
19. Olson MV: When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet* 1999, 64(1):18–23.
20. Wang X, Grus WE, Zhang J: Gene losses during human origins. *PLoS Biol* 2006, 4(3):e52.
21. Shi P, Zhang J, Yang H, Zhang YP: Adaptive diversification of bitter taste receptor genes in Mammalian evolution. *Mol Biol Evol* 2003, 20(5):805–814.
22. Dong D, Jones G, Zhang S: Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evol Biol* 2009, 9:12.
23. Conte C, Ebeling M, Marcuz A, Nef P, Andres-Barquin PJ: Evolutionary relationships of the Tas2r receptor gene families in mouse and human. *Physiol Genomics* 2003, 14(1):73–82.
24. Go Y, SMBE Tri-National Young Investigators: Proceedings of the SMBE Tri-National Young Investigators' Workshop. Lineage-specific expansions and contractions of the bitter taste receptor gene repertoire in vertebrates. *Mol Biol Evol* 2005, 23(5):964–972.
25. Wang X, Thomas SD, Zhang J: Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum Mol Genet* 2004, 13(21):2671–2678.
26. Fischer A, Gilad Y, Man O, Pääbo S: Evolution of bitter taste receptors in humans and apes. *Mol Biol Evol* 2005, 22(3):432–436.
27. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012, 491(7422):56–65.
28. 1000 Genomes Browser. [http://browser.1000genomes.org/index.html]
29. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ: The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 2004, 32:D493–D496.

30. UCSC Table Browser. [http://genome.ucsc.edu/cgi-bin/hgTables]
31. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, Gil L, García-Girón C, Gordon L, Hourlier T, Hunt S, Juettemann T, Kähäri AK, Keenan S, Komorowska M, Kulesha E, Longden I, Maurel T, McLaren WM, Muffato M, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, *et al*: **Ensembl 2013.** *Nucleic Acids Res* **41**(D1):D48–D55.
32. Ensembl Genome Browser. [http://www.ensembl.org/index.html]
33. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**(2):263–265.
34. Bandelt HJ, Forster P, Röhl A: **Median-joining networks for inferring intraspecific phylogenies.** *Mol Biol Evol* 1999, **16**(1):37–48.
35. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**(3):559–575.
36. Excoffier L, Lischer HE: **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.** *Mol Ecol Resour* 2010, **10**(3):564–567.
37. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**(5):1792–1797.
38. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.** *Mol Biol Evol* 2013, **30**(12):2725–2729.
39. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25**(11):1451–1452.
40. Li H: **A new test for detecting recent positive selection that is free from the confounding impacts of demography.** *Mol Biol Evol* 2011, **28**(1):365–375.
41. Bahlo M, Griffiths RC: **Inference from gene trees in a subdivided population.** *Theor Popul Biol* 2000, **57**(2):79–95.
42. GENETREE software. [http://www.stats.ox.ac.uk/~griff/software.html]
43. Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123**(3):585–595.
44. Fu YX, Li WH: **Statistical tests of neutrality of mutations.** *Genetics* 1993, **133**(3):693–709.
45. Schierup MH, Charlesworth D, Vekemans X: **The effect of hitch-hiking on genes linked to a balanced polymorphism in a subdivided population.** *Genet Res* 2000, **76**(1):63–73.
46. Navarro A, Barton NH: **The effects of multilocus balancing selection on neutral variability.** *Genetics* 2002, **161**(2):849–863.
47. Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG: **Measures of human population structure show heterogeneity among genomic regions.** *Genome Res* 2005, **15**(11):1468–1476.
48. Cheng F, Chen W, Richards E, Deng L, Zeng C: **SNP@Evolution: a hierarchical database of positive selection on the human genome.** *BMC Evol Biol* 2009, **9**:221.
49. Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, Robbins MM, Schubert G, Stoinski TS, Viola B, Watts D, Wittig RM, Wrangham RW, Zuberbühler K, Pääbo S, Vigilant L: **Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution.** *Proc Natl Acad Sci U S A* 2012, **109**(39):15716–15721.

doi:10.1186/s12862-014-0198-8
Cite this article as: Risso *et al.*: Genetic variation in taste receptor pseudogenes provides evidence for a dynamic role in human evolution. *BMC Evolutionary Biology* 2014 **14**:198.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

OPEN ACCESS
Freely Available Online

OpenAccessPub

INTERNATIONAL JOURNAL OF NUTRITION

ISSN NO: 0000-0000

RESEARCH ARTICLE

DOI : COMING SOON

Age Variation in Bitter Taste Perception in Relation to the Tas2r38 Taste Receptor Phenotype

Rossella Negri[1], Andrea Smarrazzo[1], Martina Galatola[1], Antonietta Maio[1], Paola Iaccarino Idelson[1] Maura Sticco[1] , Carmen Biongiovanni[1] , Adriana Franzese[1] , Luigi Greco[1*], Davide Risso[2], Gabriella Morini[3]

[1] Department of Translational Medicine, Pediatrics, European Laboratory for Food Induced Diseases, University of Naples Federico II, Italy.
[2] Lab. of Molecular Anthropology & Centre for Genome Biology, Dept. BIGEA, University of Bologna, Italy
[3] University of Gastronomic Sciences, Pollenzo, Bra (CN), Italy.

Abstract

Objectives

Taste sensitivity is an important determinant of food choice and differs between children and adults. This difference is probably due to several factors that constitute an individual's phenotype. The aim of this study was to explore taste perception in relation to the TAS2R38 and CAVI (gustin) genotypes in age classes from infancy to adulthood in a Mediterranean population.

Methods

In this cross-sectional study we evaluated the TAS2R38 and gustin genotypes and administered a standardized PROP taste test in 705 individuals (435 adults, 270 children); the sample included 224 mother-child dyads. We also explored the acceptance and consumption of bitter and non-bitter vegetables.

Results

Sensitivity to bitterness was strongly related to the TAS2R38 haplotype, and we observed an intriguing relationship with age. In fact, children were more sensitive than adults with the same TAS2R38 haplotype also within mother-child dyads. The mother-child tasting differences decreased with age and became minimal when children reached adolescence. Variations in the gustin gene did not contribute significantly to the overall taste phenotype, but helped to differentiate among non-tasters.

Conclusions

The genetic profile of the bitter-taste receptor TAS2R38 explains most of the variance in bitter taste perception, but the related phenotype is also strongly influenced by age, also in mother-child dyads that share the same genotype. This finding is likely to have a significant impact on the complex feeding relationship between mother and child.

Corresponding author:
e-mail: rosnegri@unina.it (Rossella Negri) tel. +39 0817463260, fax +39 0817462375
e-mail. ydongre@unina.it (Luigi Greco) tel. +39 0817463275, fax +39 0817462375
Key words:
Bitter perception; Food choices; TAS2R38 polymorphism; Gustin polymorphism; Mother-child dyads.

INTRODUCTION

Recent discoveries related to the molecular domain of human taste have opened new avenues for the study of the relationship between inherited factors and food behaviour [1,2]. This is particularly important in children because an unprecedented change in food choices is modifying the shape of generations of children in affluent societies and in the upper class of developing countries as well as their health status. The ongoing epidemic of childhood obesity, which is particularly notable in the Campania Region (southern Italy), cannot be attributed to the advent of industrial foods alone. Indeed, current genomic profiles developed over thousands of years, while environmental changes are only one-generation old. Consequently, the relationship between genomic profile and food choice should be explored in relation to age and family environment.

Food acceptance in children is clearly related to the inherited genomic profile of taste receptors, much more than in adults, since environmental and cultural factors experienced during the life have a strong relevance in shaping taste preferences.

In fact, they much prefer sweet (energy) and reject bitter (potentially toxic) foods; consequently, they tend to avoid low energy-density foods like vegetables and fruit, and favour high-energy density-rich food such as sugars and fats [3]

To date, the most studied genotype-phenotype association related to taste is the one of Phenylthiocarbamide (PTC) and the related compound 6-n-propylthiouracil (PROP), both not found in nature, which sensitivity closely related to common genetic variants in the TAS2R38 gene located on chromosome [7].

PROP bitter taste sensitivity appears to be a marker of a wide variety of factors that condition food choices. Indeed, hypersensitivity to propylthiouracil (PROP) is associated to heightened responses to sweeteners [5] and salt [6], as well as to such chemesthetic [7,8] and somatosensory [9,10] sensations as the feel of astringency and the hotness of fats in the mouth [11].

Therefore, the PROP bitter taste phenotype is considered a general marker of oral sensitivity that is able to influence food selection and body weight [12-17]; however, these results were not confirmed by other studies [18,19].

Individuals are usually classified as being PROP bitter insensitive, sensitive or super sensitive according to the perceived intensity of the PROP solutions [4]. In this context, the salivary protein gustin (CAVI) that has been implicated in taste bud growth and maintenance was implicated in the control of PROP phenotype [21,22]. We previously confirmed the relationship between the TAS2R38 genotype and the PROP taste phenotype in children and adults, and observed a difference in this relationship between children and adults carrying the same genotype [20]. The aim of the present study was to examine systematically in a cross-section study how TAS2R38 and CAVI polymorphisms affect bitter taste perception and food preference in relation to age and sex. We also explored the feeding behaviour of children in relation to their mother's food preferences in order to stratify for familial environment and shared or unshared genomic profile.

MATERIALS AND METHODS

Study Population

705 healthy individuals from southern Italy (435 adults and 270 children; the sample included 224 mother-child dyads), were enrolled in the study. All were Caucasian, and resided in the same geographic area (Campania) for at least two generations. The average age of adults was 31.17 years (± 12.83 S.D.). One-hundred children were aged between 1 and 6 years, 151 between 7and 12 years and 23 between 13 and 8 years; their body mass

index ranged between the 5th and the 85th percentiles. This study was conducted according to the guidelines of the Declaration of Helsinki 2008 and all procedures involving human subjects and patients were approved by the Ethics Committee of the University of Naples Federico II. Mothers gave written informed consent for themselves and their children.

PROP sensitivity assessment

Adults and children above the age of 6 were requested to refrain from eating and drinking for at least 1 hour before the bitter taste test. In ascending order, subjects tasted two suprathreshold PROP solutions, namely 280 mM and 560 mM (Aldrich Chemical, Milwaukee, WI, USA) in distilled water, rinsing their mouth with water before and after each test solution as reported elsewhere [18]. Children were instructed not to swallow the test solutions and, if required, underwent brief training to sip and spit water. Taste sensation was assessed on a 4-point scale in which the labels "no taste", "weakly unpleasant" (bitter, barely perceptible), "unpleasant" (bitter), and "very unpleasant" (extremely bitter) corresponded to values ranging from 0 to 3. We used a 4-point scale to assess the phenotype because individuals tend to select the middle number in a span of 3 numbers.

To assess taste sensation in very young children (< 6 years old) we used a hedonic scale rating of four facial movements made by the child when exposed to the taste test. A neutral expression = no taste (0); depression of mouth corners = a weakly unpleasant taste (1); frown and depression of mouth corners = unpleasant taste (2); frown and grimace = a very unpleasant taste (3). In fact, corner elevation of lip and mouth, grimace and frowning are landmarks of the reaction to bitterness [23,24],. Facial expressions were recorded and analyzed by two independent observers. In case of dubious expressions, the child was retested and his/her mother helped to interpret the child's expression. Based on the score, subjects were classified as non-tasters (score 0–2), medium tasters (score 3–4), and super tasters (score 5–6). The score is based on the two PROP solutions.

For quality control, a random sample of 30 adults was tested for threshold sensitivity using a standard forced-choice procedure [25] . Six PROP solutions, (0.032; 0.1; 0.32; 0.56; 1; 3.2 mM) were used and threshold values were identified as the first concentration correctly chosen in two subsequent presentations. Subjects were classified as non-tasters if the threshold was ≥ 0.15mM, and tasters if the threshold was ≤ 0.1 mM PROP. The phenotype assessment was validated by comparing the means of the two methods (threshold and suprathreshold) with a significant correlation (Cohen's k=0.439, P=0.011).

In addition, a random sample of 20 children (age < 6 years) and 58 adults (age 24-60 year) underwent a second supra-threshold test to assess the consistency of PROP taste assessment. We found a strong correlation between the first and second test (Pearson r = 0,627, Spearman r=0.617). and the correlation was better for small children (Pearson r =0.806; Spearman r=0.808) than for adults (Pearson r = 0.515; Spearman r=0.503). The average of the differences (absolute values) between the first and the second test was 0.53 (95% C.I. 0.39-0.69). The coefficient of variation was equal to 22.3%. Most of the retests were within ±0.5 on the 4 points scale.

Genotyping

Genetic analyses were conducted on genomic DNA obtained from saliva with the phenol-chloroform extraction method, following a protocol developed in our laboratory. TAS2R38 gene C145G (rs713598), C785T (rs1726866), G886A (rs10246939) and CAVI A268G (rs2274333) polymorphisms were determined by means

of RT-PCR. Allelic Discrimination Assay with Applied Biosystems 7900HT fast thermal cycler, using allele-specific probes (TAS2R38rs713598: C__8876467_10; TAS2R38rs1726866:C__9506827_10; TAS2R38rs10246939: C__9506826_10, CAVIrs2274333: C__1739329_10) and primers from Applied Biosystems (Life Technologies Corporation CA, USA), according to standard Taqman SNP Genotyping assay protocol. All subjects were typed for the three polymorphic sites of TAS2R38, corresponding to the amino acid substitutions A49P, V262I, I296V, that give rise to the most common haplotypes PAV and AVI. A total of 265 (198 adults and 67 children) were also typed for the gustin/CAVI gene polymorphism rs2274333. Children of the mother-child dyads were excluded from the analysis of population prevalence because their data were not independent of those of their mothers.

Food acceptance

We used a pictorial food frequency questionnaire to estimate the total weekly consumption of bitter–tasting vegetables (cabbage, broccoli, cauliflower, spinach, rocket, radicchio) and non-bitter vegetables (lettuce, roman salad, escarole, tomato, zucchini, eggplant). Pictures of each vegetable were shown to the mothers of children < 8 years old to estimate their child's food choices. Preference scores for legumes, bitter and non-bitter tasting greens were estimated by a 4-point scale ranging from "highly liked" (score 4) to "disliked most" (score 0) for each single item. A global estimate was obtained by summing the scores of each food item within the class of legumes, bitter vegetables and non bitter vegetables. Similarly the weekly consumption of each vegetable was recorded and a global score of weekly food consumption was obtained by summing the reported weekly consumption of each vegetable within the class of legumes, bitter and non bitter vegetables. Within the 224 mother-child dyads, the difference in the global score of food preferences and consumption was obtained by subtracting child's score from his/her mother's score, stratified for shared/unshared genotype.

Retrospective food analysis questionnaires are open to significant random variation due to inaccuracy of the mother's report. We attempted to control for this variance by: (i) using a pictorial questionnaire that shows the food item to be questioned; (ii) an experienced dietician who verified the accuracy of the food item questionnaires; and (iii) 22 food questionnaires were repeated by two independent operators. The difference between the two sets of questionnaires was evaluated with the Wilcoxon rank sum test. No significant differences were found between the two observers regarding the liking and consumption questionnaires. Indeed, the greatest difference in liking was recorded for the bitter vegetables (mean 0.27), which corresponds to a 3% coefficient of variation of the grand mean of the estimated liking. Regarding the frequency of food consumption, the greatest difference was observed for bitter vegetables, the mean of which being 0.068 that corresponds to a 3.13% coefficient of variation.

Statistical analysis

Variables were screened for their distribution, and parametric or non parametric test adopted accordingly. Analysis of variance (ANOVA) was performed to compare groups with a significance level of 0.05, using sex and age as covariates, after controlling for normality of distribution. Concordances between the different methods used to assess PROP bitterness were evaluated using Cohen's kappa. Data were analyzed using SPSS 19.0 (SPSS Inc, Chicago, IL) and PLINK software (http://pngu.mgh.harvard.edu/~purcell/plink/) [26].

RESULTS

PROP phenotypes

PROP sensitivity was measured using the suprathreshold method in 705 subjects, 435 adults and 270 children. There were 214 (30.4%) non-tasters, 276 (39.1%) medium-tasters, and 215 (30.5 %) supertasters. Sensitivity to bitterness differed significantly between children and adults, with more supertasters among children (113/270; 41.9%) and more non tasters among adults (152/435; 34.9%), chi square = 28,1 p < 0.0001. No sex differences were observed for children, but significantly fewer adult females were non tasters and more were supertasters compared to adult males (chi square = 6.6 p < 0.037; Table 1). Sensitivity to bitterness increased constantly with age in children (Figure 1): 50% of 3-year-old children were non tasters versus only 26% of teenagers (14-18 years old); similarly the per cent of super-tasters increased from 21% to 52% in the latter group (chi square for trend 9.08 p = 0.002).

TAS2R38 haplotypes

Haplotype analysis performed in 517 individuals revealed that about 25% of the sample were homozygous for the sensitive haplotype PAV, while 18.4 % were homozygous for the non functional haplotype AVI and 47% were PAV/AVI heterozygous. Other variants accounted for less than 10% (Table 2). Because of the presence of mother and child pairs, we recalculated the frequencies excluding randomly one member of the couple: but the results were completely overlapping with those shown. These frequencies were obviously not different by sex and age.

Gustin genotypes

The results of the genotyping of the gustin gene polymorphism A286G performed in 340 individuals, showed a higher frequency of the A allele (71.3%) compared to G allele (28.7%). At genotypic level, 170 (50%) were homozygous AA, 139 (40.9%) heterozygous AG, while 31 (9.1%) were homozygous GG (Table 3). Given the rarity of the gustin GG genotype we could not explore its effect in detail; however, the GG genotype was significantly less frequent (2/64 = 3.1%) in non-tasters carrying the expected AVI haplotype than in non-tasters carrying the functional PAV haplotype (5/38 = 13.2%). Conversely, the gustin AA genotype was significantly more frequent (34/64 = 53.1%) in AVI non-tasters than in PAV non-tasters (14/38 = 36.8%, chi square = 4.826 p = 0.0028).

Genotype-phenotype association

Sensitivity to PROP was strongly related to the TAS2R38 haplotype in children and adults (chi square = 150 p < 0.00001). However, as shown in Table 4, 75% of carriers of the AVI haplotype (AVI/AVI, AVI/AAV) were non-tasters, and 83.7% of PAV homo- or heterozygous were tasters or supertasters. Interestingly, PAV children were as sensitive to bitterness (126/143; 88.1%) as PAV adults (222/249; 81.1%), but significantly fewer AVI children were non-tasters (23/38; 60.5%) than AVI adults (63/77; 81.8%, chi square = 6.11 p = 0.013). Similar differences were observed in AVI or PAV homozygous : 95.5% (43/45) of adults and 92.8% (65/70) of children were tasters/supertasters while 53/63 (84.1%) AVI/AVI adults were non-tasters versus 20/31 (64.5%) AVI/AVI children.

Food preferences, age and PROP status

Bitter and non-bitter tasting vegetables preferences (consumption and liking) were calculated for all age groups. As shown in Figure 2, both bitter and non-bitter tasting greens intake increased significantly with age (ANOVA; F=20.17, p<0.0001 and F=13,93 , p <

Table 1 : Sensitivity to bitter taste by sex in children and adults

Adults	Status	Male	Female	Total
	Non taster	31 (49.2%)	121 (35.2%)	152 (34.9%)
	Taster	20 (31.7%)	161 (43.3%)	181 (41.6%)
	Super taster	12 (19%)	90 (24.2%)	102 (23.4%)
	Total	63	372	435
Children	**Status**	**Male**	**Female**	**Total**
	Non taster	38 (24.8%)	24 (20.5%)	62 (23%)
	Taster	56 (36.6%)	39 (33.3%)	95 (35.2%)
	Super taster	59 (38.6%)	54 (46.2%)	113 (41.9%)
	Total	153	117	270

Sensitivity to bitter differed significantly between children and adults, with more supertasters among children and more non tasters among adults. No sex differences were observed for children, but among adults significantly less females were non tasters and more were supertasters

Table 2. Frequency of the TAS2R38 diplotypes

TAS2R38 Diplotype	FREQUENCY	PERCENTAGE
AVI/AVI	95	18.4
PAV/AVI	243	47.0
PAV/PAV	129	25.0
PAV/AAV	22	4.3
AVI/AAV	23	4.4
OTHERS	5	1.0
Total	517	100.0

The most common haplotypes in the Caucasian population, PAV and AVI recurred as common diplotypes PAV/PAV, PAV/AVI, AVI/AVI in 90.4 % of the subjects, the less common haplotype AAV was present in 8.7 % of the sample, the rare haplotypes (AVV, PVV, AAI) in 1% of the population, as expected.

Table 3. Distribution of *gustin* genotypes in unrelated individuals.

Allele	*%*	*Genotype*	*Frequency*	*%*
A	71.3	AA	170	50.0
Total	100	Total	340	100

Three hundred and forty subjects were typed for the single polymorphism A286G of the Gustin gene, resulting in the substitution of serine with glycine at amino acid position 90.

The population prevalence was: 0.713 for the A allele and 0.287 for the G allele. Two diplotypes, the homozygous AA and the heterozygous AG, accounted for the 90.9% of the population, whereas the homozygous GG accounted for only 9.1%, as expected by the Hardy-Weinberg equilibrium.

Table 4 Genotype-phenotype correlation

T2R38	**STATUS**	**TOTAL**		
	NON TASTER	*TASTER*	*SUPER TASTER*	
AVI[a]	86 (74.8%)	24 (20.9%)	5 (4.3%)	115
PAV[b]	64 (16.3%)	178 (45.4%)	150 (38.3%)	392
RARE GENOTYPES	2 (40.4%)	1 (20%)	2 (40%)	5
TOTAL	152 (29.7%)	203 (39.6%)	157 (30.7%)	512

[a] AVI/AVI, AVI/AAV

[b] PAV/PAV, PAV/AVI, PAV/AAV

Sensitivity to the bitter compound (tasting phenotype) was strongly related to the TAS2R38 haplotype, but the association between genotype and phenotype was more stringent for the PAV haplotype.

Table 5. Differences (Δ) in vegetables preferences between mothers and AVI or PAV children.

Children's haplotype	**Mother/child difference for non bitter greens**	**Mother/child difference for bitter greens**	**p-value**	
AVI	Mean	5.25	8	0.003
	S.D.	7.85	7.55	
PAV	Mean	10.23	11.13	0.0001
	S.D.	8.38	6.25	

TASR38 haplotypes affected the differences between children and mothers global preference score (consumption + fondness) for non-bitter and bitter tasting vegetables.

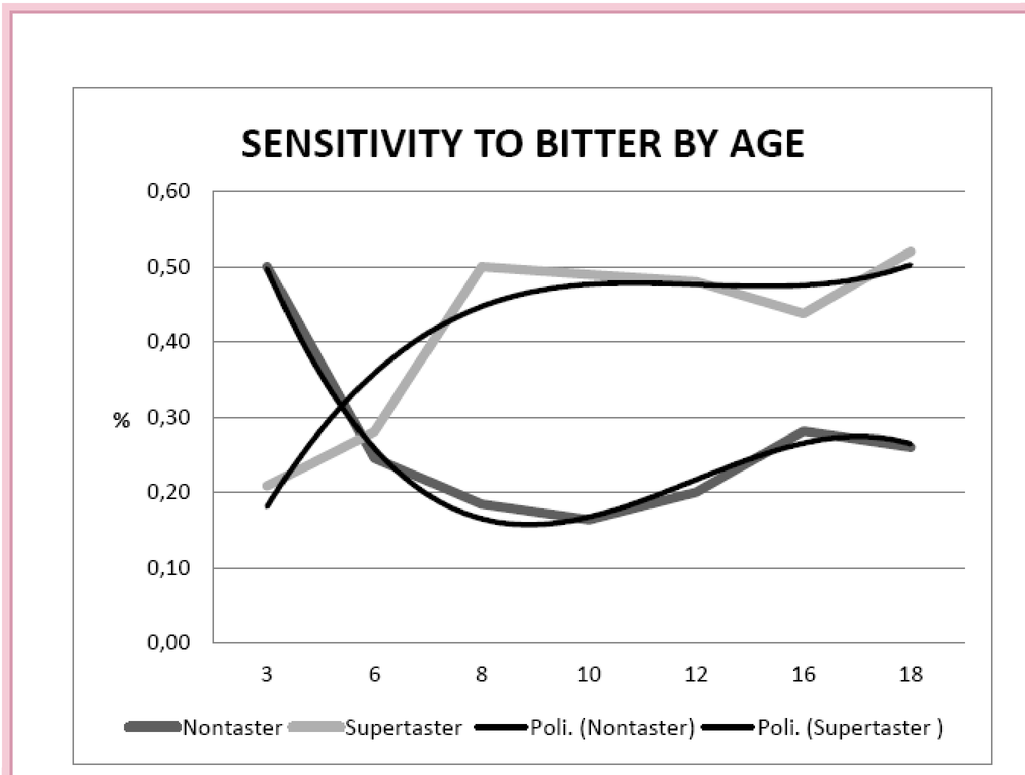


Figure 1. Sensitivity to bitterness by age. Mean percentages of Supertaster and Non Tasters is shown by age. A third degree polynomial was fitted to the data.

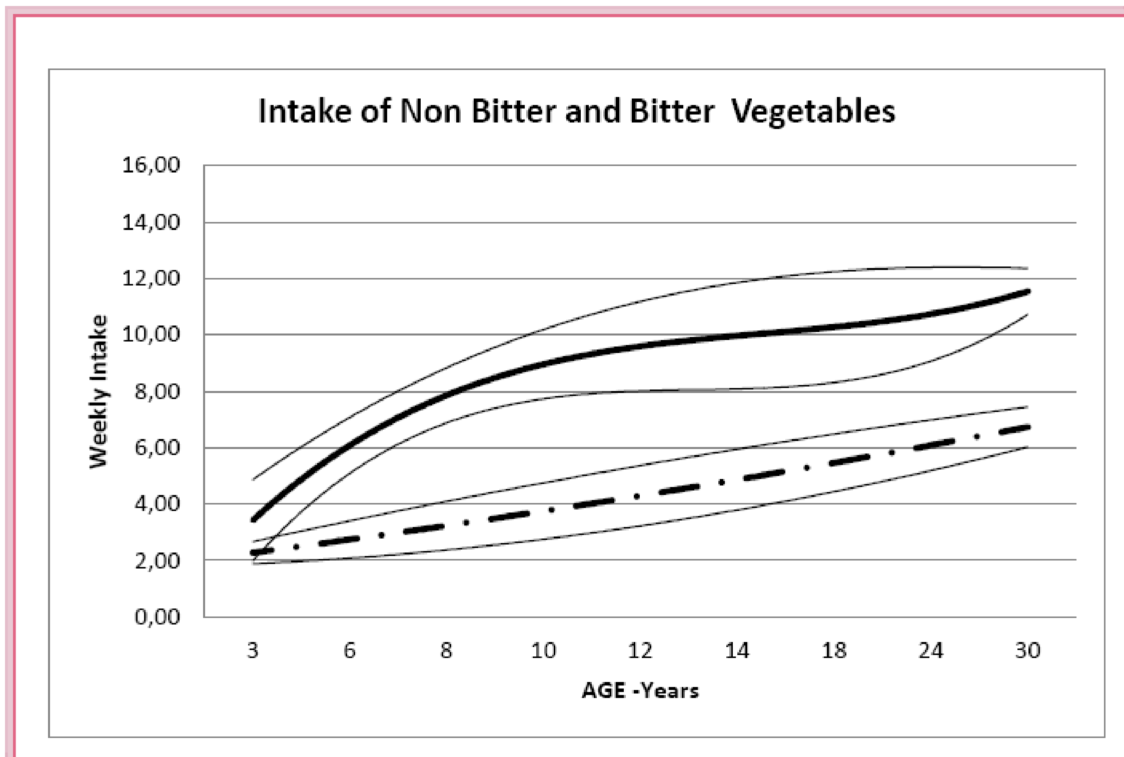


Figure 2. Mean (+/-95% C.I.) weekly intake of non bitter (continuous line) and bitter (dashed line) by age. A second and third degree polynomial was fitted to the raw data.

0,0001). As previously reported, no association was found between bitter and non-bitter tasting vegetables neither with PROP phenotypes nor haplotypes.

Children versus Mothers

In the mother-child dyads, the child was generally more sensitive to PROP than the mother. In fact, in 32% of the pairs the number of medium/supertaster children exceeded that of mothers, while in only 15% of the pairs the mother was more sensitive than the child. In the remaining 53% of dyads, mothers and children shared phenotype irrespective of genotype. A comparison of the global preference score (consumption + fondness) of both bitter and non bitter tasting vegetables within the dyads revealed striking differences between the child and the mother (mean preference score= 9.7) for non-bitter tasting greens, and particularly for the bitter tasting greens (mean preferences score= 11.2). These differences decreased with the age of the child: adolescents showed a trend "more compatible" with that of their mothers (Pearson correlation r=-0.308 for non-bitter vegetables; r=-0.174 for bitter tasting ones; both P <0.05).

In addition, the TASR38 haplotypes affected the differences between children and mothers for non-bitter (ANOVA; F=6.28, P=0.003) and bitter tasting vegetables (ANOVA; F= 11.3; P<0.001). Children with the PAV haplotype, in fact, had almost a double distance from their mothers compared to those carrying the AVI haplotype (10.23 vs. 5.25 ANOVA; F=6.3, P=0.003 for non-bitter vegetables; 11.13 vs. 8.0 points, ANOVA; F=11.4, P<0.001 for bitter vegetables) (Table 5). At ANOVA, the child's haplotype was the best contributor to the variance (F= 5.05, P=0.05 for non-bitter and F=5.68, P=0.005 for bitter tasting vegetables), while the maternal haplotype did not contribute significantly to the intra-couple difference in the consumption of the foods investigated (F=0.25 and F=0.20, with both P>0.05).

DISCUSSION

The main aim of this study was to evaluate in a Mediterranean population whether PROP genotypes/ phenotypes influence feeding behaviour in children and adults, and in mother–child dyads. Here we confirm our previous finding, obtained in a smaller sample [20], that PROP sensitivity differs between children and adults. In fact, the frequency of supertasters was higher in children, even in mother-child dyads sharing the same TAS2R38 haplotype. Sensitivity to PROP bitterness was strongly influenced by the TAS2R38 genotype and only marginally influenced by gustin gene polymorphisms. Concordance between PROP bitter sensitivity and the expected tasting haplotype (PAV) was gradually reached as children approached adolescence. Age-related changes in PROP sensitivity was particularly evident in genetic non tasters (AVI homozygous) in whom the genotype-phenotype concordance decreased from 84% in adults to 66% in children. The taster PAV diplotype was less affected by age: the proportions of tasters and supertasters were similar in adults and children with the PAV haplotype. This is in accordance with Mennella et al. [27] who reported that AVI/PAV children, but not adults, perceived a bitter taste at low PROP concentrations, whereas no such effect occurred in AVI or PAV homozygous.

Other factors besides age have been implicated in PROP sensitivity. Calò et al. [21] reported that, in a genetically homogeneous cohort, bitter sensitivity might be influenced by the polymorphic site A286G in the taste bud growth factor gustin/CAVI gene, as well by salivary zinc ion concentration. However, Melis et al. [22] reported that, among PAV homo- and heterozygous, the supertaster status is not more frequently associated to the AA genotype. In line with the findings of Feeney et al. [28], and Tomassini Barbarossa et al., [29] we did not find a correlation between CAVI polymorphisms and the PROP phenotype. In fact, the frequency of the gustin genotypes was not associated with the TAS2R38

genotypes or with the bitter sensitivity phenotypes. It is a taste-modifying gene, with no overall implication in producing the tasting phenotype.

PROP responsiveness is also linked to the secretion of salivary peptides from the basic proline-rich protein family. In fact, these peptides were found to be more abundant in supertasters than in non-tasters and their concentration in saliva was increased by PROP stimulation [30] ,which suggests that PROP sensitivity could be more complex than hitherto thought.

Genetic and/or environmental modifiers could contribute, albeit it to different degrees, to the definition of the phenotype throughout life. Changes in gene expression in the development phase or hormonal influences around the time of puberty may account for a different penetrance of the TAS2R38 genotypes at different ages. Furthermore, individual differences in the expression of the PAV haplotype among heterozygous may account for the variation in bitter taste perception [31] . Consequently, PROP sensitivity should be considered a quantitative rather than a qualitative trait, and a continuum of intermediate levels of responsiveness probably separate the insensitive phenotype from the hypersensitive phenotype [32].

Although several studies support the relationships between PROP taster status and food preferences [33-37] others do not [18,19,38]. In our population, the consumption of bitter and non-bitter tasting vegetables did not appeared to be related to PROP sensitivity or to TAS2R38 haplotype. However, we found that the acceptance and consumption of both kinds of greens increased with age irrespective of the PROP taster phenotype. These findings are not surprising given the evidence that other taste genes contribute to determining food selection [39-43] and to the supertaster phenotype [44]. Indeed, enhanced global sensory acuity may be uncoupled from genetic sensitivity to PROP [45-46]. This concept suggested a new kind of phenotype, and the term "general supertaster" was coined to describe individuals more responsive to all or most oral stimuli, regardless of their TAS2R38 genotype [47]. Sociocultural factors rather than PROP status could influence cruciferous vegetable consumption as shown by Baranowski et al. [48] in children from the USA; however, the burden of cultural differences is very limited in our cohort because it is constituted totally by individuals of Caucasian origin resident in Campania (south Italy) for at least two generations. On the other hand, foods are complex mixtures of different components, and bitterness may result from chemically different compounds that could interact with more than one taste receptor. Bitter sensitivity in humans is linked to 25 different receptors of the T2R family thereby providing a receptacle to a wide range of molecules [49-50].

Regarding food preferences within the mother-child dyads, we show that the difference between mothers and children in terms of acceptance and consumption of greens decreased with the age of children, and it was significantly related to the children's TAS2R38 haplotype. Because only 68.9% of the mothers shared the same TAS2R38 haplotype with their children, these differences could explain a significant proportion of the discrepancy in food consumption between mothers and children. Consumption of vegetables depends largely on age and less on the specific sensing genotype. Within the mother -child dyads, the differences in food preference were related to age, and to the specific genotype of the child. Because mothers are not aware of the tasting genotype of their children, they may feel that their child's refusal to eat vegetables is a problem of relationship and behaviour and not of taste. In fact, as shown in this and other studies [12,14,51,52], food preferences are also strongly influenced by cultural and environmental factors as well as by genetic factors. Further studies on the expression of taste receptors are going on in our

laboratory in the attempt to better understand "when" taste sensitivity are changing and "what" may affect them, both key points to unravel the biological bases of food choices. Therefore, there is ample scope to continue to explore this domain given the impact of food choices on the health and quality of life of our species.

Conflict of interest:

No conflict of interest can be attributed to any of the authors.

Sources of funding: The study was supported by the European Laboratory for Food Induced Diseases (ELFID), University of Naples Federico II, Italy.

References

1. Grimm ER, Steinle NI (2011). Genetics of eating behavior: established and emerging concepts. Nutr Rev. 69(1):52-60.

2. Reed DR, Knaapila A.(2010). Genetics of taste and smell: poisons and pleasures. Prog Mol Biol Transl Sci.94:213-40.

3. Mennella JA, Pepino MY and Reed DR (2005) Genetic and environmental determinants of bitter perception and sweet preferences. Pediatrics 115: e216-22.

4. Kim UK, Jorgenson E, Coon H, Leppert M, Risch N et al. (2003) Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. Science. 299: 1221-5.

5. Duffy VB, Hayes JE, Dinehart ME. (2006). Genetic differences in sweet taste perception. Optimising the sweet taste in foods. Spillane WJ, editor. Cambridge: Woodhead Publ.30-53.

6. Bartoshuk LM, Duffy VB, Lucchina LA, Prutkin J, Fast K (1998) . PROP (6-n-propylthiouracil) supertasters and the saltiness of NaCl. Ann N Y Acad Sci. 855:793-6.

7. Prescott J, Swain-Campbell N. (2000). Responses to repeated oral irritation by capsaicin, cinnamaldehyde and ethanol in PROP tasters and non-tasters. Chem Senses. 25:239–246.

8. Pickering GJ, Gordon R. (2006). Perception of mouthfeel sensations elicited by red wine are associated with sensitivity to 6-N-propylthiouracil. J Sens Stud. 21:249–265.

9. Prutkin J, Fisher EM, Etter L, Fast K, Gardner E et al. (2000). Genetic variation and inferences about perceived taste intensity in mice and men. Physiol Behav. 69:161–173

10. Hayes JE, Duffy VB. (2007). Revisiting sugar-fat mixtures: sweetness and creaminess vary with phenotypic markers of oral sensation. Chem Senses. 32:225–236.

11. Yackinous CA and Guinard JX (2002) Relation between PROP taster status and fat perception, touch, and olfaction. Physiol. Behav. 72: 427-37

12. Robino A, Mezzavilla M, Pirastu N, Dognini M, Tepper BJ et al. (2014) A population-based approach to study the impact of PROP perception on food liking in populations along the Silk Road. PLoS One.9(3):e91716.

13. Tepper BJ, White EA, Koelliker Y, Lanzara C, d'Adamo P,et al.(2009) Genetic variation in taste sensitivity to 6-n-propylthiouracil and its relationship to taste perception and food selection. Ann N Y Acad Sci.1170:126-39

14. Hayes JE, Feeney EL, Allen AL (2013) Do polymorphisms in chemosensory genes matter for human ingestive behavior? Food Qual Prefer. 30:202-216

15. Keller KL, Olsen A, Cravener TL, Bloom R, Chung WK et al.(2014) Bitter taste phenotype and body weight predict children's selection of sweet and savory foods at a palatable test-meal. Appetite. 77:113-21.

16. 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). Ann Hum Biol. 41:29-39.

17. Feeney E, O'Brien S, Scannell A, Markey A, Gibney ER. (2011) Genetic variation in taste perception: does it have a role in healthy eating? Proc Nutr Soc. 70:135-43.

18. Mattes, R.D. (2004) . 6-n-propylthiouracil taster status: dietary modifier, marker or misleader? Genetic Variation in Taste Sensitivity. J. Prescott & B.J. Tepper, Eds.: 229–250. Marcel Dekker, New York.

19. Drewnowski A, Henderson SA, Cockroft JE (2007). Genetic sensitivity to 6-n-propylthiouracil has no influence on dietary patterns, body mass indexes , or plasma lipid profiles of women. J. Am. Diet. Assoc.107: 1340–1348.

20. Negri R, Di Feola M, Di Domenico S, Scala MG, Artesi G, et al. (2012) Taste perception and food choices. J Pediatr Gastroenterol Nutr. 54: 624-9.

21. Calò C, Padiglia A, Zonza A, Corrias L, Contu P et al. (2011) Polymorphisms in TAS2R38 and the taste bud trophic factor, gustin gene co-operate in modulating PROP taste phenotype. Phys & Behavior 104: 1065-71

22. Melis M, Atzori E, Cabras S, Zonza A, Calò C et al. (2013) The Gustin (CAVI) Gene Polymorphism, rs2274333 (A/G), as a Mechanistic Link between PROP Tasting and Fungiform Taste Papilla Density and Maintenance. PLoS One 8(9):e74151.

23. Chen AW , Resurreccion AV, Paguio LP (1996) Age appropriate hedonic scales to measure food preferences of young children. J. Sensory Studies 11:141-163

24. Steiner JE, Glaser D, Hawilo ME, Berridge KC (2001) Comparative expression of hedonic impact: affective reactions to taste by human infants and other primates. Neurosc.Behav.Rev.25:53-74

25. Anliker JA, Bartoshuk L, Ferris AM, Hooks LD. (1991). Children's food preferences and genetic sensitivity to the bitter taste of 6-n-propylthiouracil (PROP). Am J Clin Nutr. 54:316-20

26. Purcell S, Neale B, Todd-Brown K, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81: 559-75

27. Mennella JA, Pepino MY, Duke FF, Reed DR (2010). Age modifies the genotype-phenotype relationship for the bitter receptor TAS2R38. BMC Genet. 1:11-60.

28. Feeney EL, Hayes JE (2014). Exploring associations between taste perception, oral anatomy and polymorphisms in the carbonic anhydrase (gustin) gene CAVI. Physiol Behav.128:148-54.

29. Tomassini Barbarossa I, Melis M, Mattes M Z, Calò C, Muroni P,et al.. (2015). The gustin (CAVI) gene polymorphism, rs2274333 (A/G), is associated with fungiform papilla density, whereas PROP bitterness is mostly due to TAS2R38 in an ethnically-mixed population. Physiology & Behavior 138:6–12

30. Cabras T, Melis M, Castagnola M, Padiglia A, Tepper BJ et al. (2012) Responsiveness to 6-n-propylthiouracil (PROP) is associated with salivary levels of two specific basic proline-rich proteins in humans. PLoS One 7: e30962

31. Lipchock SV, Mennella JA, Spielman AI, Reed DR (2013) Human bitter perception correlates with bitter receptor messenger RNA expression in taste cells. Am J Clin Nutr. Oct;98:1136-43

32. Hayes JE and Keast RS (2011). Two decades of supertasting: where do we stand? Physiol Behav. 104:1072-4.

33. Drewnowski A, Henderson SA, Levine A, Hann C (1999). Taste and food preferences as predictors of dietary practices in young women. Pub. Health Nutr.2: 513–519.

34. Dinehart, M.E., J.E. Hayes, L.M. Bartoshuk, Lanier SL, Duffy VB (2006). Bitter taste markers explain variability in vegetable sweetness, bitterness, and intake. Physiol. Behav. 87: 304–313.

35. Keller KL, Steinmann L, Nurse RJ, Tepper BJ (2002). Genetic taste sensitivity to 6-n-propylthiouracil influences food preference and reported intake in preschool children. Appetite 38: 3–12.

36. Turnbull, B and Matisoo-Smith E (2002). Taste sensitivity to 6-n-propylthiouracil predicts acceptance of bitter-tasting spinach in 3–6-y-old children. Am. J.Clin. Nutr. 76: 1101–1105.

37. Bell, KI and Tepper BJ (2006). Short-term vegetable intake by young children classified by 6-npropylthoiuracil bitter-taste phenotype. Am. J. Clin.Nutr. 84: 245–251.

38. Drewnowski A, Henderson SA, Cockroft JE (2007). Genetic sensitivity to 6-n-propylthiouracil has no influence on dietary patterns, body mass indexes , or plasma lipid profiles of women. J. Am. Diet. Assoc.107: 1340–1348.

39. Pronin, A.N., Xu H, Tang H, Zang L, Li Q et al. (2007). Specific alleles of bitter receptor genes influence human sensitivity to the bitterness of aloin and saccharin. Curr. Biol. 17: 1403–1408.

40. Wang, J.C., A.L. Hinrichs, S. Bertelsen, Stock H, Budde JP et al. (2007). Functional variants in TAS2R38 and TAS2R16 influence alcohol consumption in high-risk families of African-American origin. Alcohol Clin. Exp. Res. 31:209–215.

41. Stewart JE, Newman LP, Keast RS. (2011).Oral sensitivity to oleic acid is associated with fat intake and body mass index. Clin Nutr. 30:838-44.

42. Keller KL, Liang LC, Sakimura J, May D, van Belle C et al. (2012). Common Variants in the CD36 Gene Are Associated With Oral Fat Perception, Fat Preferences, and Obesity in African Americans. Obesity 20: 1066-1073,

43. Martínez-Ruiz NR, López-Díaz JA, Wall-Medrano A, Jiménez-Castro JA, Angulo O (2014). Oral fat perception is related with body mass index, preference and consumption of high-fat foods. Physiol Behav. 129:36-42.

44. Hayes JE, Bartoshuk LM, Kidd JR, Duffy VB (2008). Supertasting and PROP bitterness depends on more than the TAS2R38 gene. Chem Senses.33:255-65.

45. Lim J, Urban L, Green BG. (2008). Measures of individual differences in taste and creaminess perception. Chem. Senses 33: 493–501.

46. Bajec MR. and Pickering GJ (2008). Thermal taste, PROP responsiveness, and perception of oral sensations. Physiol. Behav. 95: 581–590.

47. Reed DR (2008). Birth of a new breed of supertaster. Chem. Senses 33: 489–491

48. Baranowski T, Baranowski JC, Watson KB, Jago R, Islam N et al.(2011). 6-n-propylthiouracil taster status not related to reported cruciferous vegetable intake among ethnically diverse children. Nutr Res. 31:594-600

49. Meyerhof W, Batram C, Kuhn C, Brockhoff A, Chudoba E et al. (2010). The molecular receptive ranges of human TAS2R bitter taste receptors. Chem Senses. 35:157-70.

50. Tepper BJ. (2008) Nutritional implications of genetic taste variation: the role of PROP sensitivity and other taste phenotypes. Annu Rev Nutr.28:367-88.

51. Larson N, Story M (2009). A review of environmental influences on food choices. Ann Behav Med.38 Suppl 1:S56-73.

52. Franchi M (2012). Food choice: beyond the chemical content. Int J Food Sci Nutr. 63 Suppl 1:17-28.

RESEARCH ARTICLE

Estimating Sampling Selection Bias in Human Genetics: A Phenomenological Approach

Davide Risso[1,2], Luca Taglioli[3], Sergio De Iasio[4], Paola Gueresi[5], Guido Alfani[6], Sergio Nelli[7], Paolo Rossi[8], Giorgio Paoli[3], Sergio Tofanelli[3]*

1 National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD 20854, United States of America, **2** Laboratory of Molecular Anthropology and Centre for Genome Biology, Department of BiGeA, University of Bologna, via Selmi 3, 40126 Bologna, Italy, **3** Dipartimento di Biologia, University of Pisa, Via Ghini 13, 56126 Pisa, Italy, **4** Dipartimento di Genetica Biologia dei Microrganismi Antropologia Evoluzione, University of Parma, Parco Area delle Scienze 11/a, 43124 Parma, Italy, **5** Dipartimento di Scienze Statistiche, University of Bologna, Via Belle Arti 41, 40126 Bologna, Italy, **6** Bocconi University, Dondena Centre and IGIER, Milan, Italy, **7** Archivio di Stato, Lucca, Italy, **8** Dipartimento di Fisica, University of Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy

* sergio.tofanelli@unipi.it

OPEN ACCESS

Citation: Risso D, Taglioli L, De Iasio S, Gueresi P, Alfani G, Nelli S, et al. (2015) Estimating Sampling Selection Bias in Human Genetics: A Phenomenological Approach. PLoS ONE 10(10): e0140146. doi:10.1371/journal.pone.0140146

Editor: Gyaneshwer Chaubey, Estonian Biocentre, ESTONIA

Received: June 19, 2015

Accepted: September 21, 2015

Published: October 9, 2015

Copyright: © 2015 Risso et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

This research is the first empirical attempt to calculate the various components of the hidden bias associated with the sampling strategies routinely-used in human genetics, with special reference to surname-based strategies. We reconstructed surname distributions of 26 Italian communities with different demographic features across the last six centuries (years 1447–2001). The degree of overlapping between "reference founding core" distributions and the distributions obtained from sampling the present day communities by probabilistic and selective methods was quantified under different conditions and models. When taking into account only one individual per surname (low kinship model), the average discrepancy was 59.5%, with a peak of 84% by random sampling. When multiple individuals per surname were considered (high kinship model), the discrepancy decreased by 8–30% at the cost of a larger variance. Criteria aimed at maximizing locally-spread patrilineages and long-term residency appeared to be affected by recent gene flows much more than expected. Selection of the more frequent family names following low kinship criteria proved to be a suitable approach only for historically stable communities. In any other case true random sampling, despite its high variance, did not return more biased estimates than other selective methods. Our results indicate that the sampling of individuals bearing historically documented surnames (founders' method) should be applied, especially when studying the male-specific genome, to prevent an over-stratification of ancient and recent genetic components that heavily biases inferences and statistics.

Introduction

The question of how representative is a sample of a given population is particularly critical in human genetic research, where data collection is strongly affected by ethical and social factors.

The choice of the sampling strategy can affect the estimation of a number of relevant parameters and, accordingly, the inferences one can make about either the groups or the genetic variants under study. The ideal goal is to achieve the best compromise among recruitment costs, statistical power and representativeness. To make a sample an unbiased proxy of the underlying population, either for genetic screening, genotype-phenotype association studies or evolutionary reconstructions, its composition should minimize close kinship, ambiguous individual ancestry assignments, and the side effects of population stratification.

Most of present-day human populations have experienced recent admixture to some extent [1] and genetic flows make it difficult to discern ancestral genotypes and alleles, stabilized by many generations of historical residency and reinforced by endogamy, from genotypes and alleles introduced by recent migration. In population-based association studies, methodological corrections have been widely used to overcome spurious associations driven by undetected genetic structure. All of these methods have restrictions and drawbacks [2, 3]. In human evolutionary genetics, especially when inferring past demography upon Y chromosome diversity, direct methods based on either self-declared ancestry or a number of selective criteria based on surnames are routinely adopted, but the reliability of these methods has not been assessed so far. Owing to the co-segregation of surnames and Y markers in the Occidental practice, the conclusions inferred from samplings based on family names should be intended as specifically referred to the male-specific genome with the corollary that the higher the parents' co-ancestry (marital kinship) of a community, the more they could be extended to whole genome studies. The use of donors with surnames documented in the early historical records of a given community allows obtaining genetic profiles not affected by recent shuffling. In addition, the adhesion to the oldest surname distribution of a community can provide the null hypothesis, in order to test how fully a sample reflects the ancestral genetic composition. Unfortunately, this method is rarely adopted due to difficulties in retrieving historical records. The so-called "grandparents", or "two-generation residency", criterion gives a picture of the state of a population about a century earlier than the sampling time and generally depends on the accuracy of self-declarations. Selecting only surnames whose distribution is limited to a relatively small area of origin (geographically restricted surnames), with or without the aid of neural network algorithms [4, 5], helps discarding potentially polyphyletic lineages. The selection of donors with the most frequent surnames, assuming that the most common surnames represent lineages of long residency, and the random drawing from the current distributions of family names, are two other widely used methods [6, 7, 8].

Using surnames to select research subjects has the advantage that family names are cultural markers typically transmitted from fathers to sons in most human societies. The distributions and mode of inheritance of surnames mirror the non-recombinant markers of the male specific regions of the Y chromosome (MSY). The main difference between surnames and MSY sequences is the nature of the transmitted units which are predominantly, if not exclusively, identical by descent for MSY markers and identical by state (polyphyly) other than by descent (monophyly) for surnames. In addition, the depth of the genealogical reconstruction is largely unbounded for MSY markers, while for surnames it is limited to the time since such family names have been stabilized as markers of inheritance (around 1200–1800 AD in Western societies). Such differences have been invoked to explain deviations from neutral expectations when surnames were used to estimate consanguinity or gene propagation [9, 10]. Nonetheless, the long-term dynamics of both natural and social phenomena tend to scale frequency spectra and are subject to the same laws of parameterization [11]. With few exceptions [12], within the last 20–30 generations surnames and MSY markers can be confidently considered as neutral alleles evolving under the same stochastic model with negligible effects from mutation and selection.

Hence, last names can be considered as cost-efficient estimators of recent demographic changes occurred by migration and size fluctuations.

A number of reasons makes the Italian population an ideal case-study for surnames-related research: Italians show one of the highest surname diversity in the world [8] with a mean number of bearers per surname approaching twenty. In addition, historical documents are available since the Middle Ages for many communities, and standardized religious population registers, with barely the same geographic relevancy than present-day Italy, have been maintained since the Council of Trent (1545–1563 AD). Lastly, genetically and culturally closed communities still survive in proximity to open populations [13].

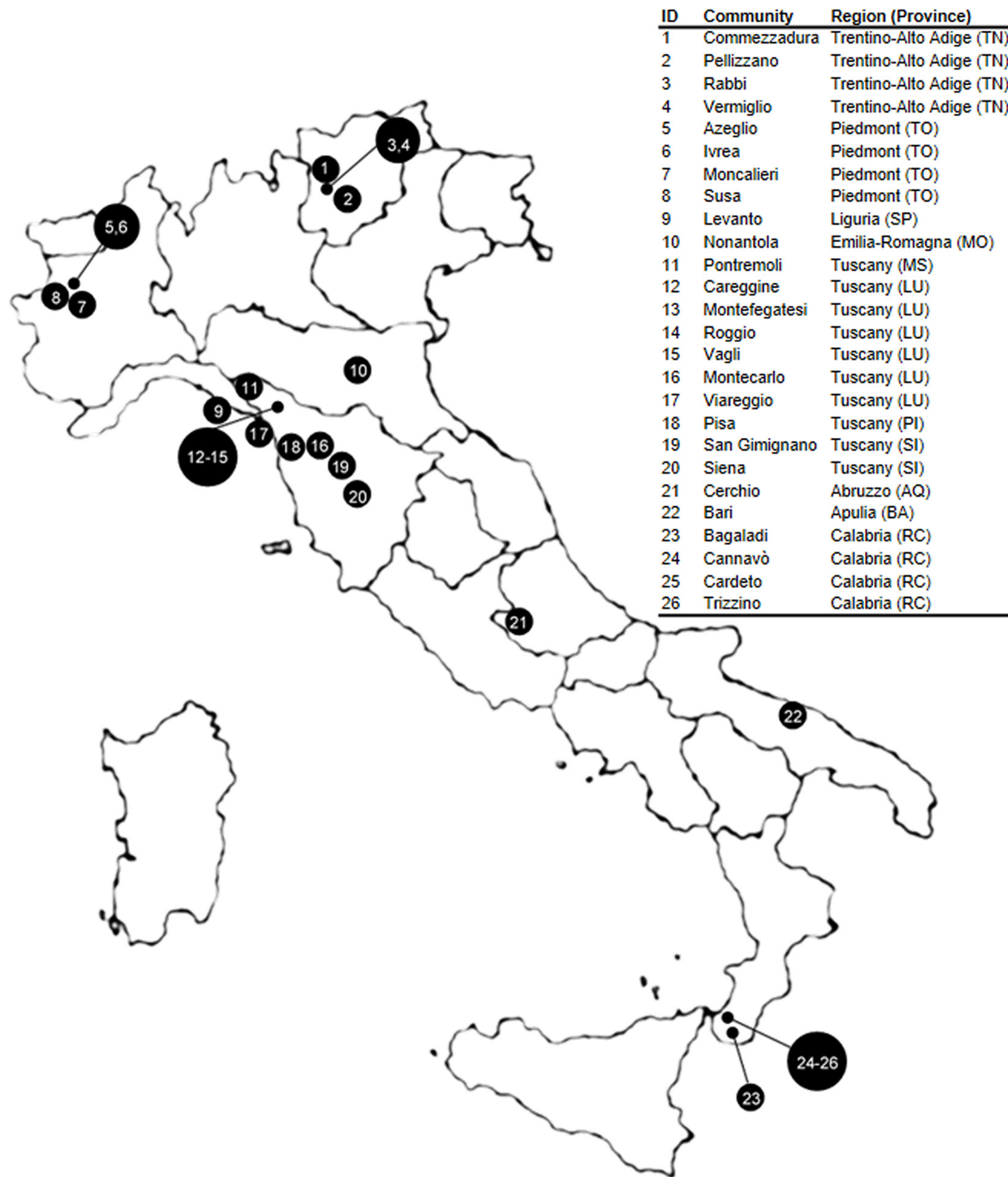
The present work aims at first comparing different surname-based sampling strategies, to estimate the bias associated with each method. In order to do this, we reconstructed the surname distributions of 26 Italian communities with different demography in the time interval encompassing the 1447–2001 AD. The degree of overlapping between "reference core" surname distributions, obtained from analyzing the oldest historical frames, and the distributions obtained from sampling present-day surnames by probabilistic (random) and selective (based on historically resident, locally spread and more frequent surnames) methods was quantified at different geographic scales (municipality, region, Country). Whenever possible, jackknife resampling and comparisons between different data sources and models, with low and high levels of kinship, were applied to weigh the various sources of the sampling bias.

Materials and Methods

Populations and records

Twenty-six Italian communities belonging to 11 provinces and 8 regions were analyzed (Fig 1): four of them are located in Trentino-Alto Adige [Rabbi (TN), Pellizzano (TN), Commezzadura (TN), Vermiglio (TN)], four in Piedmont [Ivrea (TO), Azeglio (TO), Moncalieri (TO), Susa (TO)], one in Liguria [Levanto (SP)], one in Emilia-Romagna [Nonantola (MO)], ten in Tuscany [Viareggio (LU), Pisa (PI), Siena (SI), San Gimignano (SI), Roggio (LU), Careggine (LU), Vagli (LU), Montefegatesi (LU), Pontremoli (MS), Montecarlo (LU)], one in Abruzzo [Cerchio (AQ)], one in Apulia [Bari (BA)] and four in Calabria [Bagaladi (RC), Cannavò (RC), Cardeto (RC), Trizzino (RC)]. Table 1 recapitulates the relevant data for each of the investigated communities, which are widely dispersed over the peninsula at different altitudes, differing in size, growth rate, and time since the oldest repertoire of surnames.

We reconstructed the earlier distributions of surnames from either three kinds of historical records: census (*Stato delle anime*), baptismal (*Atti battesimali*), and marriage (*Atti matrimoniali*) sacramental registers. The present-day distributions of surnames from the same communities were extracted from the digitalized versions of the 1993 complete national phone directory [14] and/or from the 1991 and 2001 National Institute of Statistics (ISTAT) [15] records. All the original historical sources of data are public and can be accessed upon request from the relevant institutional archives: diocesan and /or civil anagraphical archives of Azeglio, Bagaladi, Bari, Cannavo', Cardeto, Careggine, Cerchio, Commezzadura, Ivrea, Levanto, Moncalieri, Montecarlo, Montefegatesi, Nonantola, Pellizzano, Pisa, Pontremoli, Rabbi, Roggio, San Gimignano, Siena, Susa, Trizzino, Vagli, Vermiglio, Viareggio. From each record we extracted the following information: year, surnames, birthplace and, whenever possible, surnames and birthplaces of both parents. ISTAT original data were provided by the municipal offices of Montefegatesi (1993 records), Roggio (1993 records), Cannavo' (2001 records) and Trizzino (2001 records). Original data from 1993 phone directories were kindly provided by SEAT in the form of magnetic tapes. Commercial data were subsequently filtered out. Each of present-day areas of jurisdiction (SEAT phone districts, ISTAT municipalities), largely overlap with the



ID	Community	Region (Province)
1	Commezzadura	Trentino-Alto Adige (TN)
2	Pellizzano	Trentino-Alto Adige (TN)
3	Rabbi	Trentino-Alto Adige (TN)
4	Vermiglio	Trentino-Alto Adige (TN)
5	Azeglio	Piedmont (TO)
6	Ivrea	Piedmont (TO)
7	Moncalieri	Piedmont (TO)
8	Susa	Piedmont (TO)
9	Levanto	Liguria (SP)
10	Nonantola	Emilia-Romagna (MO)
11	Pontremoli	Tuscany (MS)
12	Careggine	Tuscany (LU)
13	Montefegatesi	Tuscany (LU)
14	Roggio	Tuscany (LU)
15	Vagli	Tuscany (LU)
16	Montecarlo	Tuscany (LU)
17	Viareggio	Tuscany (LU)
18	Pisa	Tuscany (PI)
19	San Gimignano	Tuscany (SI)
20	Siena	Tuscany (SI)
21	Cerchio	Abruzzo (AQ)
22	Bari	Apulia (BA)
23	Bagaladi	Calabria (RC)
24	Cannavò	Calabria (RC)
25	Cardeto	Calabria (RC)
26	Trizzino	Calabria (RC)

Fig 1. Map of Italy showing the location of the 26 investigated communities. This Fig is similar, although not identical, to the original image, and is therefore for illustrative purposes only.

doi:10.1371/journal.pone.0140146.g001

geographic extension pertained to the corresponding sacramental archive and comprise no more than one founding parish. In no case either population replacements or massive migrations occurred since the time of the historical source.

Data analysis

An accurate context-specific transliteration of historical characters was done from original sources. Subsequently, a lemmatization of both "founding" and recent surnames was carried out. Namely, ancillary historical sources were used to correct for editing mistakes (i.e."Bevlacqua = Bevilacqua"), single-digit mismatches (i.e. "Marino/Marini") and double

Table 1. Details of the studied communities/municipalities. N_0, number of individuals of the oldest list of surnames; N_t, number of individuals of the present-day list of surnames; altitude is indicated in meters above sea level; relative growth rate calculated as N_t-N_0/N_t.

Community	Region (Province)	Oldest Historical Source	Recent Source	N_o	N_t	S_o	S_t	Altitude	Growth rate
Commezzadura	Trentino-Alto Adige (TN)	Marriage acts 1700	SEAT 1993	297	350	44	166	850	0.15
Pellizzano	Trentino-Alto Adige (TN)	Marriage acts 1700	SEAT 1993	548	324	85	134	925	-0.69
Rabbi	Trentino-Alto Adige (TN)	Marriage acts 1566	SEAT 1993	292	482	88	103	1,095	0.39
Vermiglio	Trentino-Alto Adige (TN)	Marriage acts 1714	SEAT 1993	341	462	28	68	1,261	0.26
Azeglio	Piedmont (TO)	Baptismal acts 1543	SEAT 1993	895	423	100	217	260	-1.12
Ivrea	Piedmont (TO)	Census Paper 1613	SEAT 1993	3,835	9,816	568	4,861	253	0.61
Moncalieri	Piedmont (TO)	Census Paper 1613	SEAT 1993	6,129	20,436	776	9,382	219	0.7
Susa	Piedmont (TO)	Census Paper 1613	SEAT 1993	4,447	2,283	341	1,276	503	-0.95
Levanto	Liguria (SP)	Census Paper 1662	SEAT 1993	1,728	5,716	329	1,151	3	0.7
Nonantola	Emilia-Romagna (MO)	Census Paper 1629	SEAT 1993	3,451	3,407	181	1,092	24	-0.01
Careggine	Tuscany (LU)	Marriage acts 1566	SEAT 1993	243	206	115	70	882	-0.18
Montecarlo	Tuscany (LU)	Baptismal acts 1527	SEAT 1993	3,913	1,226	283	545	162	-2.09
Montefegatesi	Tuscany (LU)	Marriage acts 1600	ISTAT 1991	398	270	108	55	842	-0.47
Pisa	Tuscany (LU)	Baptismal acts 1447	SEAT 1993	17,504	35,921	1,830	10,913	4	0.51
Roggio	Tuscany (LU)	Marriage acts 1775	ISTAT 1991	115	175	30	41	858	0.34
San Gimignano	Tuscany (LU)	Marriage acts 1700	SEAT 1993	290	2,357	73	1,013	324	0.88
Siena	Tuscany (LU)	Census Paper 1767	SEAT 1993	2,941	56,956	1,373	5,626	322	0.95
Vagli	Tuscany (LU)	Marriage acts 1700	SEAT 1993	553	379	96	100	575	-0.46
Viareggio	Tuscany (LU)	Census Paper 1705	SEAT 1993	290	57,514	86	7,263	2	0.99
Pontremoli	Tuscany (MS)	Marriage acts 1559	SEAT 1993	249	3,400	61	976	236	0.93
Cerchio	Abruzzo (AQ)	Census Paper 1700	SEAT 1993	932	1,735	144	146	834	0.46
Bari	Puglia (BA)	Census Paper 1598	SEAT 1993	8,872	111,221	1,065	12,993	5	0.92
Bagaladi	Calabria (RC)	Baptismal acts 1657	SEAT 1993	125	399	31	120	460	0.69
Cannavò	Calabria (RC)	Baptismal acts 1601	ISTAT 2001	994	3,935	274	577	147	0.75
Cardeto	Calabria (RC)	Baptismal acts 1670	SEAT 1993	924	695	129	129	700	-0.33
Trizzino	Calabria (RC)	Baptismal acts 1706	ISTAT 2001	137	104	34	28	551	-0.32

TN, Trento; TO, Turin; SP, La Spezia; MO, Modena; LU, Lucca; MS, Massa and Carrara; AQ, L'Aquila; BA, Bari; RC, Reggio Calabria.

doi:10.1371/journal.pone.0140146.t001

surnames (i.e. "Albertini Marchesani = Albertini"). In presence of doubtful cases, a stringent criterium was adopted in order to minimize the bias due to mutational events (i.e. "Benvenuti = Benvenutini", "Rosso/Rossi"). Overall, such a method tends to inflate the degree of overlapping between "founding" and present-day distributions and, hence, to underestimate the sampling bias. The lemmatization process was applied separately to each community. The complete database is provided in S1 Dataset and counts 67,071 surnames and 310,571 individuals.

In order to validate the use of different sources, as either historical or recent references, we recovered and compared the surname distributions from SEAT and ISTAT records in the community of Cerchio (Abruzzo, Central Italy), as well the surname distributions from marriage and baptismal acts in the community of Montecarlo (Tuscany, Central Italy).

Simulating sampling strategies

For each strategy we measured the sampling-dependent bias (SDB) according to the formula:

$$SDB = 1 - \frac{\sum_k^1 x}{N_k},$$

where x is the number of the k surnames which are in common between the historical and current repository and N_k is the size of the current repository. SDB ranges from 1, no surname shared, to 0, all the surnames of the current repository are shared with the historical repository. A "low-kinship" (LK) model, namely when only one individual per surname was considered, and a "high-kinship" model (HK), when using the entire set of surnames and their bearers, were applied to each pairwise comparison. Under HK, k was intended as the total number of surnames, allowing multiple counts for the same type; under LK, k was intended as the number of different surnames, one for type. Being our null hypothesis the adhesion of the sampled surnames to the oldest surname distribution of each community, as occurs when collecting according to the founding surnames method, the SDB rate expectation is zero under both models.

To estimate how far the departures of SDB from zero are exclusively due to genetic drift we simulated the evolution of the 26 studied communities under both a growth (observed relative growth rates) and a stationary model (null growth rate) by the Markov chain Monte Carlo method implemented in the software ASHEs [16, 17]. For each model, 100 iterations were performed and averaged SDB values were compared with the observed values.

To estimate how far SDB is affected by sample size, we performed a total of 201 jackknife re-samplings of twenty, fifty and, when possible, one hundred individuals out of every dataset, using PopTools version 3.2.5 [18].

STATISTICA v. 6.0 (Stat-Soft Inc, Tulsa, OK) and XLSTAT (Addinsoft) were utilized to perform the Chi-square tests, the correlation analyses and to create graphics.

When simulating by the random sampling criterion (R), the entire list of surnames of the recent repository (SEAT or ISTAT) was considered as starting set. When selecting samples by a spatial distribution criterion ("locally spread" surnames or LS), we chose only those surnames whose geographic origin was unambiguously centered within a single province (the metropolises Milan and Rome excluded), as shown in the SEAT 1993-derived GENS database [19]. The surnames fitting such a criterion showed to be often, but not always, the rarest, with a number of bearers usually lower than 5,000. It is important to stress that local surnames satisfying this condition are not necessarily monophyletic, that is derived from a single ancestor. The frequency-based method (FQ) was applied by selecting the most frequent surnames, those falling above the upper quartile of the present-day distribution. The availability of marriage acts between 1845 and 1915, along with important information such as the place and date of birth of both the spouses and their parents, gave us the opportunity to select individuals by the grandparents' criterion (GP), since this time interval represents the period when the grandparents of the 1991–1993 population lived. Simulating the "founding" surnames method (FS) meant to refer to the most ancient distribution found in the historical records. We actually used this sampling strategy as a template with SDB = 0 to compare the efficiency of other methods.

Results

Source-dependent bias

Bivariate analyses of present-day surname lists obtained from ISTAT (2001) and SEAT (1993) records in the community of Cerchio (S1A Fig) and of "founder" surname lists obtained from baptismal (A.D. 1527) and marriage (A.D. 1576) acts in the community of Montecarlo (S1B Fig) were performed. The two distributions, albeit differing as total number of surnames (SEAT = 168, ISTAT = 150; baptismal acts = 242, marriage acts = 224) and individuals (SEAT = 557, ISTAT = 481; baptismal acts = 2,310, marriage acts = 2,164), strongly correlate with each other ($R^2 = 0.87$, $P<0.01$ and $R^2 = 0.82$, $P<0.01$ respectively) and did not differ

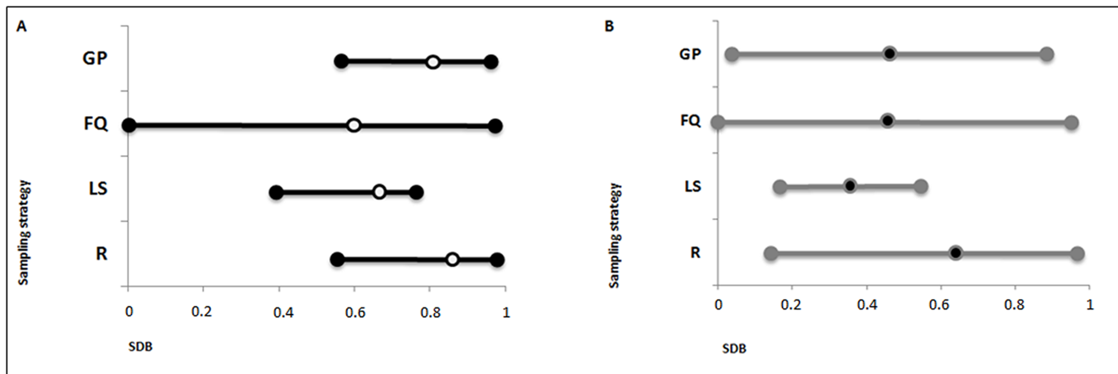


Fig 2. Minimum, maximum and mean values of the sampling-dependent bias (SDB) calculated after sampling by random (R), locally spread (LS), first quartile (FQ) and grandparents (GP) strategies, under the low-kinship (A) and the high-kinship (B) models in the 26 investigated communities.

doi:10.1371/journal.pone.0140146.g002

significantly (Chi-square test, P = 0.71 and P = 0.66 respectively). The fact that Cerchio and Montecarlo are communities with intermediate size and growth rate would suggest that, for most of the investigated communities, the availability of data sources exerted a negligible effect on the estimate of the intrinsic bias (SDB).

Strategy-dependent bias

As a rule, the comparisons between "founder" and present-day lists of surnames showed a low level of concordance independently of the applied sampling strategy and the kinship model (Fig 2A and 2B). Regarding the LK model, the observed average SDB was never below 0.595 (FQ strategy) with peaks of 0.770 and 0.846 (GP and R strategies, respectively). When multiple individuals per surname were considered (HK model), the average bias always decreased (by 8–30%) but at the cost of a larger variance.

The percentage of shared surnames sampled with a random strategy (R) was predictably low under a LK model (from 1 to 48%, SDB = 0.52–0.99) as well as under a HK model (from 2 to 83%, SDB = 0.17–0.98). Instead, it was somewhat surprising to observe the same features when applying the grandparents' criterion (GP): surnames sharing ranged from 3 to 47% (SDB = 0.53–0.97) under the LK and from 6 to 95% (SDB = 0.05–0.94) under the HK.

The frequency-based criterion (FQ) showed the highest degree of uncertainty: surnames sharing, in fact, ranged from 4 to 100% (SDB = 0.00–0.96) under both the LK and the HK. The criterion based on locally-spread surnames (LS), on the other hand, showed the lowest bias: the overlapping degree varying from 22 to 65% under the LK (SDB = 0.35–0.78) and from 41 to 86% under the HK (SDB = 0.14–0.59). In this case preserved patrilineages approach the estimated proportion of "autochthony" of Italian surnames at the province level (from 22.8% to 77.9%, [5]). The selection of locally spread or rarer surnames, therefore, ensures a more similar albeit overall low degree of representativeness across communities. The historical residence of individuals bearing rare and localized family names, in fact, is not ensured because a focused spatial distribution may be the consequence of founder effects due to family renaming and/or migrations in the last few centuries.

In order to quantify how much the estimation of single population parameters is affected by the choice of the sampling method, we measured both the isonymy (according to [20]) and the value S/N, the relative number of different surnames S in present-day lists of surnames of size N, applying each strategy of sample selection in all the 26 communities. Such values differed with respect to the reference value calculated upon the FS method (sampling only "founding"

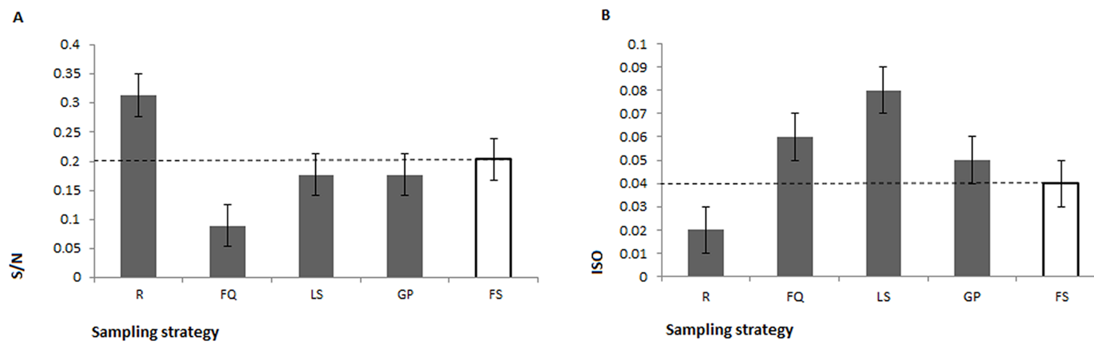


Fig 3. Average values and standard deviations of the S/N parameter (A) and isonymy (B) calculated with different sampling strategies in the present-day communities. S/N, relative number of surnames; ISO, isonymy. Sampling strategies: R, random; FQ, first quartile; LS, locally spread; GP, grandparents; FS, founder surnames.

doi:10.1371/journal.pone.0140146.g003

surnames) depending on the strategy adopted (Fig 3A and 3B). Interestingly, the deviations observed were of inverse sign for the R and the other strategies, respectively inflating and deflating the ancestral diversity.

Drift-dependent bias

Simulations performed under growth and stationary models allowed us to measure the expected bias accumulated by lineage sorting alone (genetic drift), from foundation to present times. As shown in S2 Fig, SDB rates observed by sampling randomly one individual per surname (R criterion under a LK model) gave a good estimation of the drift-dependent bias only in those small communities which underwent negligible migration (i.e. Careggine). The data in our possession (not shown) suggest that positive departures from this value are proportional to the cumulative impact of outwards and inwards migratory flows in all the other investigated communities.

Community-dependent bias

A more accurate analysis was done by evaluating the dependence of SDB by the type of the community (S3–S6 Figs). A first observation is that the lowest SDB values were obtained in the alpine valleys of Trentino (North Italy). Here, the four communities investigated are known to show high levels of both consanguinity and emigration rates [21, 22]. For these communities, geographic factors likely reinforced isolation from the neighboring demes more than in other geographic contexts, thus facilitating the retention of surname-based legacies.

A second observation is the generally lower SDB values scored using HK compared to LK sampling criteria, with a more pronounced trend in small communities. The drop of SDB was significantly lower in communities with present-day size (N_t) > 500 (Chi-square test, $P<0.05$). Similarly, we found that the higher the difference between the intrinsic bias calculated under the HK and LK methods (ΔSDB), the lower the value of S/N in the "founding" list, here used as a crude estimate of early population diversity [23]. This finding is somewhat expected because, in a stable population, the number of individuals is large when compared to the number of surnames and the effect is accentuated when the size is small because of drift. However, as shown in S7 Fig, a functional relation with S/N was found for the R ($R^2 = 0.35$, P = 0.02), the FQ ($R^2 = 0.36$, P = 0.02) and the GP ($R^2 = 0.78$, $P<0.001$) sampling strategies but not for the LS criterion ($R^2 = 0.005$, P = 0.96). The same tendency was observed between ΔSDB and isonymy, here intended as the random component of inbreeding ($R^2>0.35$, $P<0.05$) (data not shown).

Interestingly, Nonantola (Po river plain, North Italy), despite being not geographically isolated and with a size in the order of thousands, showed the highest ΔSDB when sampling by the FQ criterion. In this case, the retention of frequency-based surname legacies is more easily explained by its hosting a "Partecipanza", a social-economic entity where rights over common lands have been long transmitted strictly following residency and patrilineal genealogies [24].

Lastly, we tested whether there could be any geographic or demographic trait of the community that would allow us to predict the sampling bias. We created a scatter plot showing the relationships between the intrinsic bias and altitude, present-day population size, growth rate and the time since the oldest repertoire of surnames (S8 Fig). Of the four variables tested, only altitude showed significant lines of tendency, even after multiple testing correction, under both low and high kinship models and for all the sampling strategies with the exception of the LS criterion (Table 2).

Scale-dependent bias

When the sampling bias was calculated at the level of mid (region) and macro (Country) area, the overall patterns observed at the municipality level held for the R and FQ sampling methods under both the LK and HK models (data not shown). Deviations are mainly due to the proportion of closed *vs* open communities in each region, which underestimates SDB in regions over-represented by mountain villages (Trentino) and, *vice versa*, overestimates SDB in regions represented by large open cities (Apulia). Missing or overabundance of data prevented us to extend the analysis to higher levels than municipality, respectively for the GP and LS method. Nonetheless, a breakdown of SDB is expected when locally spread surnames are assigned at the regional and, of course, at the Country level.

Size-dependent bias

The mean values of the percentage of "founder" surnames (1-SDB), calculated after randomly re-sampling twenty, fifty and one hundred individuals out of the complete list of present-day surnames, were not appreciably different under the LK (Chi-square test, P = 0.87) and HK (Chi-square test, P = 0.88) models. However, the range of values decreased geometrically by increasing the sample size.

The outputs of random Jackknife resampling for the S/N parameter are shown in S9 Fig, taking Bagaladi (South Italy) as an example of the general tendency in the other twenty six communities (LK, Chi-square test, P = 0.91; HK, Chi-square test, P = 0.93). This tendency shows that only samples approaching N = 100 ensure a reliable estimate of diversity.

Discussion

The identification of the most reliable sampling strategy, among the many available for human genetic studies, could be of great utility. Conversely, an improper collection of samples from the population may introduce hidden confounding factors. With this perspective, we calculated the various sources of bias associated with common methods of sample recruitment: those employing family names as ancestry markers. Sampling by surnames is in fact one of the most accurate and cost-efficient ways to assess the genetic composition of historical populations, despite the fact that family names behave as a single locus transmitted along only one parental lineage, and the fact that mismatches are expected in Y-DNAs due to illegitimate paternities and adoptions [25].

A main result of our analyses is that the fidelity to the ancestral composition of the population (measured from 1447–1775) by samples selected from current lists of surnames is variable but generally low whatever is the size of the sample, the selection strategy, the primary source

Table 2. Detailed list of R2 values and correlation signs (S) with nominal and Bonferroni adjusted P-values for different sampling strategies and models. Sampling strategies: R, random; FQ, first quartile; LS, locally spread; GP, grandparents; FS, founder surnames. Models: LK, low-kinship; HK, high-kinship.

Sampling strategy	Model	Covariate	R2	S	P-value	Adjusted P-value
R	LK	Altitude	0.42	-	0.00	0.008
	HK	Altitude	0.58	-	0.01	0.036
LS	LK	Altitude	0.01	-	0.98	1.000
	HK	Altitude	0.12	-	0.74	1.000
GP	LK	Altitude	0.36	-	0.01	0.036
	HK	Altitude	0.41	-	0.00	0.012
FQ	LK	Altitude	0.42	-	0.00	0.012
	HK	Altitude	0.43	-	0.00	0.008
R	LK	Present-day N	0.13	+	0.41	1.000
	HK	Present-day N	0.16	+	0.28	1.000
LS	LK	Present-day N	0.01	+	0.96	1.000
	HK	Present-day N	0.01	+	0.96	1.000
GP	LK	Present-day N	0.21	+	0.16	0.640
	HK	Present-day N	0.23	+	0.11	0.440
FQ	LK	Present-day N	0.11	+	0.46	1.000
	HK	Present-day N	0.12	+	0.42	1.000
R	LK	Foundation year	0.11	-	0.46	1.000
	HK	Foundation year	0.12	-	0.42	1.000
LS	LK	Foundation year	0.01	-	0.99	1.000
	HK	Foundation year	0.01	-	0.99	1.000
GP	LK	Foundation year	0.11	-	0.46	1.000
	HK	Foundation year	0.11	-	0.46	1.000
FQ	LK	Foundation year	0.02	-	0.89	1.000
	HK	Foundation year	0.02	-	0.89	1.000
R	LK	Growth rate	0.03	+	0.84	1.000
	HK	Growth rate	0.11	+	0.46	1.000
LS	LK	Growth rate	0.01	+	0.99	1.000
	HK	Growth rate	0.01	+	0.99	1.000
GP	LK	Growth rate	0.01	+	0.99	1.000
	HK	Growth rate	0.12	+	0.42	1.000
FQ	LK	Growth rate	0.01	+	0.99	1.000
	HK	Growth rate	0.09	+	0.54	1.000

doi:10.1371/journal.pone.0140146.t002

of data, the geographic scale, or type of the community. As expected, lineage sorting accounts for the deviations from fidelity only partially and in very closed communities.

Despite the use of a relaxed lemmatization, biases lower than 40%, intending this term as the complement to one of the degree of overlapping between "founding" and present-day lists of family names (SDB), have been observed only in small mountainous hamlets as far as a close parenthood among participants was admitted (high-kinship model). The selection of unrelated donors is usually associated with a substantial increase of the bias other than to higher costs of recruitment and, in remote areas, a lower sample size. Low kinship in most cases means rates of bias much higher than 50%, unless the first quartile method is applied in stable communities where cultural (socio-economic rules, language) or geographic (mountain terrain) barriers preserved family names with high frequency over long periods of time.

Population stratification due to recent fluctuations in allele frequencies over time thus appears a foremost level of apportionment of the overall genetic variance. While typically not recognized because of the lack of contribution of bio-demographic data to genetic sampling designs, the diachronic variance would play a crucial role in the representativeness of a population sample and, hence, to its utility in biological research.

The proportion of ancestry from ancient and recent demography can vary widely between individuals, leading to step-ups of the internal structure of the population which, in turn, may easily introduce either spurious results (type I and II errors) in genetic association studies or distortions in historical and evolutionary reconstructions. The latter case has been demonstrated by Manni et al. [4] who found that, sampling Dutch Y chromosomes without a prior selection of donors on the base of surnames' origin, led to skewed population structures. In addition, Bowden et al. [26] demonstrated that independent samples ascertained on the basis of residency (GP method) and on the possession of medieval surnames (FS) showed significantly different Y-haplotypes. Similarly, Calò et al. [27] showed how different sampling strategies, founding surnames (FS) and the grandparents' (GP) methods, can lead to contrasting population affinities on the basis of mitochondrial HVRI haplotypes.

Here, we demonstrated that sampling criteria could easily affect population parameters resulting in inflated (random methods) or deflated (selective methods) estimates of diversity.

The most efficient method to correct population parameters for the diachronic bias is to select individuals bearing historically documented surnames ("founders" method). Alternative but equally time-demanding methods, such as those based on surname distribution (LS) and residency (GP), are heavily influenced by recent gene flows and appeared to be efficient only in specific contexts, as macro geographic scales and alpine villages. Moreover, they generally result in a reduction of the maximum gatherable sample size down to a hundred or even less. Consequently, their use in population genetics is limited by the poor statistical power, and is very limited or impracticable in genotype-phenotype association studies. In these latter cases a truly random sampling under a HK model, despite it carrying a high SDB, is able to give samples of suitable size at affordable costs.

Among the various features used to define a human community both geographically and demographically, altitude was found the best proxy of the bias inherent to most of the sampling procedures in the modern Italian population.

We can conclude that sampling individuals bearing historically documented surnames should be the method of choice to buffer the stratification of ancient and recent genetic components of the sampled population. Nonetheless, the selection of "founders", while excluding misleading results from markers of the male specific fraction of the Y chromosome, does not prevent that recent population reshufflings heavily driven by women mobility could bias inferences and statistics based on the variability at autosomal, X-chromosome and mitochondrial markers.

Supporting Information

S1 Fig. Number of individuals (N) per surname, comparing data sources from ISTAT and SEAT records (Fig A) and from baptismal and marriage acts records (Fig B).
(TIF)

S2 Fig. Observed (red dots) and simulated values of SDB sampling randomly under a low-kinship model in the 26 investigated communities. Yellow lines, simulated SDB values with population growth; yellow diamonds, simulated SDB values with constant population size.
(TIF)

S3 Fig. Sampling-dependent bias (SDB) calculated after sampling by the random strategy in each of the 26 investigated communities. HK, high-kinship model; LK, low-kinship model.
(TIF)

S4 Fig. Sampling-dependent bias (SDB) calculated after sampling by the first quartile strategy. HK, high-kinship model; LK, low-kinship model.
(TIF)

S5 Fig. Sampling-dependent bias (SDB) calculated after sampling by the grandparents strategy. HK, high-kinship model; LK, low-kinship model.
(TIF)

S6 Fig. Sampling-dependent bias (SDB) calculated after sampling by the locally spread strategy in the investigated communities. HK, high-kinship model; LK, low-kinship model.
(TIF)

S7 Fig. Correlation between ΔSDB and S/N values under the random (A), first quartile (B), locally spread (C) and grandparents (D) methods.
(TIF)

S8 Fig. Plots showing correlation between communities' features (altitude, present-day population size, year of the oldest list of surnames, growth rate) and sampling-dependent bias (SDB), calculated sampling by random (R), locally spread (LS), first quartile (FQ), grandparents (GP) strategies under low-kinship (LK) (A) and high-kinship (HK) (B) models.
(TIF)

S9 Fig. Variation of the S/N parameter with different Jackknife re-sampled population sizes in the present-day population of Bagaladi, following the high-kinship (A) and low-kinship (B) models.
(TIF)

S1 Dataset. Complete surnames database.
(XLSX)

Acknowledgments

Authors would like to acknowledge the staff of the Demography Offices of all the studied communities for historical information and, in particular, the voluntaries of the association "Terra di Viareggio" for the precious help in studying the historic marriage records. We would like to thank Dr. Dennis Drayna for help in revising the manuscript and for his helpful suggestions.

Author Contributions

Conceived and designed the experiments: DR ST. Performed the experiments: DR ST. Analyzed the data: DR ST PG GA GP LT. Contributed reagents/materials/analysis tools: SD PG GA SN PR GP. Wrote the paper: DR ST GP PG GA.

References

1. Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, et al. A Genetic Atlas of Human Admixture History. Science. 2014; 14; 343(6172): 747–751. doi: 10.1126/science.1243518 PMID: 24531965

2. Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol. 2001; 20(1):4–16. PMID: 11119293

3. Shringarpure S, Xing EP. Effects of sample selection bias on the accuracy of population structure and ancestry inference. G3 (Bethesda). 2014; 4(5): 901–911.
4. Manni F, Toupance B, Sabbagh A, Heyer E. New method for surname studies of ancient patrilineal population structures and possible application to improvement of Y-chromosome sampling. Am J Phys Anthr. 2005; 126: 214–228.
5. Boattini A, Lisa A, Fiorani O, Zei G, Pettener D, Manni F. General method to unravel ancient population structures through surnames, final validation on Italian data. Hum Biol. 2012; 84(3): 235–270. PMID: 23020096
6. King TE, Ballereau SJ, Schürer KE, Jobling MA. Genetic signatures of coancestry within surnames. Curr Biol. 2006; 16(4): 384–348. PMID: 16488872
7. King TE, Jobling MA. Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. Mol Biol Evol. 2009; 26(5): 1093–1102. doi: 10.1093/molbev/msp022 PMID: 19204044
8. King TE, Jobling MA. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. Trends Genet. 2009; 25(8): 351–360. doi: 10.1016/j.tig.2009.06.003 PMID: 19665817
9. Crow JF, Mange AP. Measurement of inbreeding from the frequency of marriages between persons of the same surname. Eugen Q. 1965; 12: 199–203. PMID: 5839742
10. Wijsman E, Zei G, Moroni A, Cavalli-Sforza LL. Surnames in Sardinia II. Computation of migration matrices from surname distributions in different periods, Ann Hum Genet. 1984; 48: 65–78. PMID: 6712156
11. Rossi P. Surname distribution in population genetics and in statistical physics. Phys Life Rev. 2013; 10 (4): 395–415 doi: 10.1016/j.plrev.2013.06.005 PMID: 23830614
12. McEvoy B, Bradley DG. Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. Hum Genet. 2006; 119(1–2): 212–219. PMID: 16408222
13. Capocasa M, Anagnostou P, Bachis V, Battaggia C, Bertoncini S, Biondi G, et al. Linguistic, geographic and genetic isolation: a collaborative study of Italian populations. J Anthropol Sci. 2014; 92: 201–231. doi: 10.4436/JASS.92001 PMID: 24607994
14. SEAT Pagine Bianche. Elenco abbonati al telefono anno1993.
15. ISTAT—Istituto nazionale di statistica. 14° Censimento della popolazione e delle abitazioni 2001. Available: http://www.istat.it/it/censimento-popolazione/censimento-popolazione-2001
16. Merlitti D, Tofanelli S. An Application to Simulate Haplotype Evolutions. Available: http://ashes.codeplex.com/.
17. Tofanelli S, Taglioli L, Merlitti D, Paoli G. Tools which simulate the evolution of uni-parentally transmitted elements of the human genome. J Anthropol Sci. 2011; 89: 201–219. doi: 10.4436/jass.89017 PMID: 21911915
18. Hood, G. M. PopTools version 3.2.5. Available: http://www.poptools.org
19. GENS database. Available: www.gens.info
20. Scapoli C, Rodriguez-Larralde A, Beretta M, Nesti C, Lucchetti A, Barrai I. Correlations between isonymy parameters. Int J Anthropol. 1997; 12: 17–37.
21. Boattini A, Calboli FC, Blanco Villegas MJ, Gueresi P, Franceschi MG, Paoli G, et al. Migration matrices and surnames in populations with different isolation patterns: Val di Lima (Italian Apennines), Val di Sole (Italian Alps), and La Cabrera (Spain). Am J Hum Biol. 2006; 18: 676–690. PMID: 16917890
22. Martuzzi Veronesi F, Gueresi P, Pettener D. Biodemographic analysis of Italian Alpine communities (Upper Sole Valley, 1725–1923). Rivista di Antropologia. 1996; 74: 55–75. (Volume dedicated to Gabriel Ward Lasker).
23. Zei G, Guglielmino CR, Siri E, Moroni A, Cavalli-Sforza LL. Surnames as neutral alleles: observations in Sardinia. Hum Biol. 1983; 55(2): 357–365. PMID: 6873922
24. Alfani G, Munno C. Godparenthood and social networks in an Italian rural community. Nonantola, sixteenth-seventeenth centuries. In: Alfani G, Gourdon V, editors. Spiritual Kinship in Europe, 1500–1900, London, Palgrave; 1991. pp 96–123.
25. Jobling MA (2001) In the name of the father: surnames and genetics. Trends Genet 17:353–357. PMID: 11377798
26. Bowden GR, Balaresque P, King TE, Hansen Z, Lee AC, Pergl-Wilson G, et al. Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England. Mol Biol Evol. 2008; 25(2): 301–309. PMID: 18032405
27. Calò CM, Corrias L, Vona G, Bachis V, Robledo R. Sampling strategies in a linguistic isolate: results from mtDNA analysis. Am J Hum Biol. 2012; 24(2): 192–194. doi: 10.1002/ajhb.22215 PMID: 22287111

NUTRITION RESEARCH 35 (2015) 1122–1125

Available online at www.sciencedirect.com

ScienceDirect

www.nrjournal.com

Communication

A potential trigger for pine mouth: a case of a homozygous phenylthiocarbamide taster

Davide S. Risso [a,b,*], Louisa Howard [c], Carter VanWaes [d], Dennis Drayna [a]

[a] National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892
[b] Department of BiGeA, Laboratory of Molecular Anthropology and Centre for Genome Biology, University of Bologna, 40126 Bologna, Italy
[c] National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, MD 20892
[d] Clinical Genomics Unit and Tumor Biology Section, Head and Neck Surgery Branch, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892

ARTICLE INFO

Article history:
Received 10 June 2015
Revised 11 September 2015
Accepted 17 September 2015

Keywords:
Dysgeusia
Pine nuts
PTC
TAS2R38
Taste disturbance

ABSTRACT

Pine mouth, also known as pine nut syndrome, is an uncommon dysgeusia that generally begins 12 to 48 hours after consuming pine nuts. It is characterized by a bitter metallic taste, usually amplified by the consumption of other foods, which lasts 2 to 4 weeks. Recent findings have correlated this disorder with the consumption of nuts of the species *Pinus armandii*, but no potential triggers or common underlying medical causes have been identified in individuals affected by this syndrome. We report a 23-year-old patient affected by pine mouth who also underwent a phenylthiocarbamide taste test and was found to be a taster for this compound. *TAS2R38* genotyping demonstrated that this subject was a homozygous carrier of the proline-alanine-valine taster haplotype.
We, therefore, hypothesize that homozygous phenylthiocarbamide taster status may be a potential contributor for pine mouth events. Although based on a single observation, this research suggests a connection between genetically determined bitter taste perception and the occurrence of pine nut dysgeusia events.

Published by Elsevier Inc.

1. Introduction

In the past 5 years, studies have reported that thousands of cases of individuals in Europe and the United States have experienced a persistent bitter taste reaction to pine nuts [1-3]. The classical symptoms of this disorder, also known as pine nut syndrome (PNS), are a delayed and constant bitter or metallic taste, occurring 1 to 3 days after the ingestion of pine nuts and lasting up to several weeks [4]. In addition, occasional cases of diarrhea, headache, and nausea have been reported to be associated with this syndrome [1]. Only a few studies of PNS have been reported [3-5], and 1 study has described the symptoms of pine nut–related dysgeusia in a total of 6 subjects [6]. A single species of pine nuts (*Pinus armandii*) has previously been associated with this syndrome [1,7,8], although recent studies have reported taste disturbances from samples containing nuts from a mixture of different *Pinus* species [2,9,10]. However, the syndrome occurs in only a small fraction of individuals who consume pine nuts, and efforts to identify a common trigger or medical cause responsible for the occurrence of PNS have not been successful [4]. The rate of

Abbreviations: LMS, Labeled Magnitude Scale; PNS, pine nut syndrome; PROP, propylthiouracil; PTC, phenylthiocarbamide.

* *Corresponding author at:* National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892.
E-mail address: davide.risso@nih.gov (D.S. Risso).

http://dx.doi.org/10.1016/j.nutres.2015.09.011
0271-5317/Published by Elsevier Inc.

resolution of PNS symptoms has been noted to be similar to the turnover rate of taste cells of the tongue, and PNS has been hypothesized to act through taste perception pathways [4]. A widespread genetic difference in taste perception is specified by the *TAS2R38* gene, which encodes a bitter taste receptor that occurs in 2 common forms, designated phenylthiocarbamide (PTC) taster and nontaster [11]. Taster individuals have been shown to be more sensitive to bitter taste, and genetic responsiveness to PTC may affect eating habits and food choices [12,13]. For this reason, we completely sequenced the *TAS2R38* gene in a 23-year-old woman affected by pine mouth, to investigate potential correlations between *TAS2R38* haplotypes and the occurrence of this syndrome. In addition, a suprathreshold PTC solution was used to test the taster status tasted in this subject. The objective of our research was to investigate the hypothesis that PTC taster status may be a potential trigger for pine mouth events.

2. Methods and materials

2.1. Case presentation

A white 23-year-old American woman participated in a PTC taste perception study and subsequently reported a persistent metallic/bitter taste in her mouth which began approximately 24 hours after eating a salad containing pine nuts from *Pinus sibirica* species. These symptoms worsened during the following 4 days and progressively improved without medication. Other symptoms (ie, nausea, diarrhea, and headache) were not reported by the subject. The patient was a nonsmoker and did not have any history of medication, infections or surgery associated with dysgeusia, or any neurologic conditions.

2.2. Experimental protocol

Suprathreshold measurements of both PTC and propylthiouracil (PROP) have been shown to provide strong correlations between TAS2R38 genotypes and phenotypes [14-16]. A single PTC concentration of 256 μmol/L [11] was, therefore, chosen as the best single discriminant between taster and nontaster status. The subject was asked to hold this solution in her mouth for 10 seconds and rate its bitterness on a Labeled Magnitude Scale (LMS) (Fig. 1) [17].

2.3. DNA collection, extraction, and genotyping

A saliva sample was collected using an Oragene collection kit (Genotek, Inc, Kanata, Ontario, Canada), and genomic DNA was purified according to the manufacturer's protocol [18]. *TAS2R38*, the gene encoding the bitter taste receptor responsible for PTC perception, was completely sequenced using dideoxy Sanger sequencing [19], with a dedicated set of primers modified from Kim et al [11] (Table). DNA chromatograms were analyzed with the Lasergene suite (DNASTAR, Madison, WI, USA) [20].

2.4. Ethical standards

This study has been approved by the National Institutes of Health Combined Neurosciences/Blue Panel Institutional Review Board (National Institutes of Health protocol 01-DC-0230), and all procedures were performed in accordance with the Helsinki Declaration of 1975, as revised in 2000. Written informed consent form was obtained from the patient included in this study.

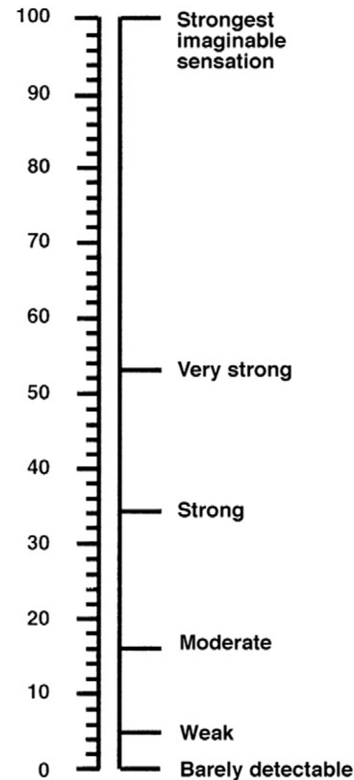


Fig. 1 – The LMS used to evaluate PTC bitterness perception.

3. Results

Physical examination performed on the subject reported no significant findings: ears, nose, and oral cavity were clear. There were no neck adenopathy or masses and no dermal lesions.

Regarding the PTC tasting, the subject defined the perceived bitterness to be "between strong and very strong," marking a value of 40 on the LMS, consistent with her being a PTC taster. We, therefore, explored and confirmed the taster status of this individual at the genetic level: she was found to carry a proline, alanine, and valine at amino acid positions 49, 262, and 295, respectively. This demonstrated that she is homozygous for the proline-alanine-valine (PTC taster) haplotype (Fig. 2).

4. Discussion

A recent survey of pine mouth events in the United States has shown that the delayed dysgeusia after pine nut ingestion is an emerging phenomenon particularly related to the ingestion of nuts from *P armandii*[4]. Removal of these nuts in food has been associated with a lower number of PNS events [4]. The specific agent responsible for these symptoms, however, remains unknown, and researchers have found no clear association between PNS and underlying medical conditions, age, or

Table – PCR and sequencing primers used to amplify and sequence *TAS2R38* gene

Primer name	Direction	Sequence	Reaction
TAS2R38F	Forward	AGATGGGCATGCAAAACTGG	PCR, sequencing
TAS2R38R	Reverse	ACTCACAGGCGTATTAATGAAGA	PCR, sequencing
TAS2R38F1	Forward	TCACACCTTCCTGATCTGCT	Sequencing
TAS2R38R1	Reverse	AGGCTGGGGTCACGAGAG	Sequencing

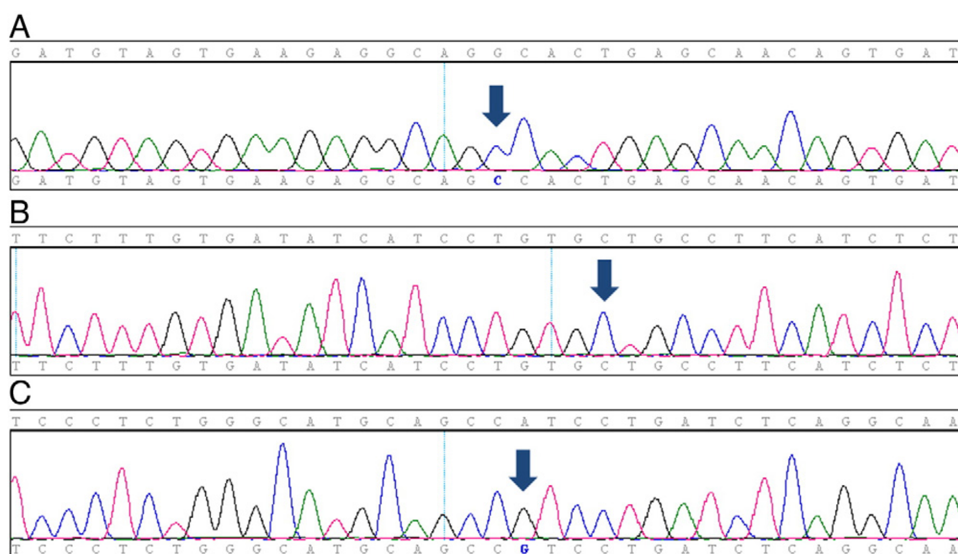


Fig. 2 – Single nucleotide polymorphisms identified at amino acid positions 49 (A), 262 (B), and 295 (C) of *TAS2R38* gene. The sequence above the chromatogram indicates the reference sequence; the presence/absence of the single nucleotide polymorphism–derived allele is indicated with an arrow.

tobacco use. Here, we hypothesize that the PTC taster status, encoded by the *TAS2R38* bitter receptor gene, could be a common feature of people who experience PNS. Multiple anecdotal reports have noted the co-occurrence of PNS and the ability to taste PTC [21]. In addition to these, we have here shown that the homozygous taster *TAS2R38* genotype, which is associated with higher sensitivity to PTC, PROP, and many other compounds [11,22,23], is associated with PNS in a 23-year-old woman. This subject was affected by PNS after ingesting pine nuts from *P sibirica*, which have previously not been associated with PNS [1,7,8]. However, recent reports of taste disturbances from samples containing nuts from a mixture of *P sibirica* and *P armandii* have been reported [2,9,10]. As a study based on a single individual, it is intended to be a hypothesis-generating research with limited generalizability. Confirmation of this hypothesis in a larger number of individuals is needed, possibly by introducing a simple PTC/PROP taste test in toxicological investigations and clinical evaluation of PNS. At this time, the only current available intervention is avoidance of recurrent exposure.

Acknowledgment

DR, CVW, and DD were supported by the National Institute on Deafness and Other Communication Disorders intramural grant numbers Z1A-000046-15 and ZIA-DC-000075.

The authors report no conflicts of interest or financial relationships to disclose. The authors alone are responsible for the content and writing of the manuscript.

REFERENCES

[1] Flesch F, Rigaux-Barry F, Saviuc P, Garnier R, Daoudi J, Blanc I, et al. Dysgeusia following consumption of pine nuts: more than 3000 cases in France. Clin Toxicol (Phila) 2011;49(7): 668–70.

[2] Handy SM, Parks MB, Deeds JR, Liston A, de Jager LS, Luccioli S, et al. Use of the chloroplast gene ycf1 for the genetic differentiation of pine nuts obtained from consumers experiencing dysgeusia. J Agric Food Chem 2011;59(20): 10995–1002.

[3] Mostin M. Taste disturbances after pine nut ingestion. Eur J Emerg Med 2001;8:76.

[4] Kwegyir-Afful EE, Dejager LS, Handy SM, Wong J, Begley TH, Luccioli S. An investigational report into the causes of pine mouth events in US consumers. Food Chem Toxicol 2013;60: 181–7.

[5] Munk MD. "Pine mouth" syndrome: cacogeusia following ingestion of pine nuts (genus: *Pinus*). An emerging problem? J Med Toxicol 2010;6(2):158–9.

[6] Ballin NZ. A trial investigating the symptoms related to pine nut syndrome. J Med Toxicol 2012;8(3):278–80.

[7] Destaillats F, Cruz-Hernandez C, Giuffrida F, Dionisi F. Identification of the botanical origin of pine nuts found in

food products by gas-liquid chromatography analysis of fatty acid profile. J Agric Food Chem 2010;58(4):2082–7.
[8] Destaillats F, Cruz-Hernandez C, Giuffrida F, Dionisi F, Mostin M, Verstegen G. Identification of the botanical origin of commercial pine nuts responsible for dysgeusia by gas-liquid chromatography analysis of fatty acid profile. J Toxicol 2011; 2011:316789.
[9] Fardin-Kia AR, Handy SM, Rader JI. Characterization of pine nuts in the U.S. market, including those associated with “pine mouth”, by GC-FID. J Agric Food Chem 2012;60(10): 2701–11.
[10] Handy SM, Timme RE, Jacob SM, Deeds JR. Development of a locked nucleic acid real-time polymerase chain reaction assay for the detection of *Pinus armandii* in mixed species pine nut samples associated with dysgeusia. J Agric Food Chem 2013;61(5):1060–6.
[11] Kim UK, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. Science 2003;299(5610):1221–5.
[12] Drewnowski A, Rock CL. The influence of genetic taste markers on food acceptance. Am J Clin Nutr 1995;62(3):506–11.
[13] Drewnowski A, Henderson SA, Barratt-Fornell A. Genetic taste markers and food preferences. Drug Metab Dispos 2001; 29(4 Pt 2):535–8.
[14] Drewnowski A, Henderson SA, Shore AB. Genetic sensitivity to 6-n-propylthiouracil (PROP) and hedonic responses to bitter and sweet tastes. Chem Senses 1997;22(1):27–37.
[15] Bufe B, Breslin PA, Kuhn C, Reed DR, Tharp CD, Slack JP, et al. The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. Curr Biol 2005;15(4): 322–7.
[16] Risso D, Morini G, Pagani L, Quagliariello A, Giuliani C, De Fanti S, et al. Genetic signature of differential sensitivity to stevioside in the Italian population. Genes Nutr 2014;9(3):401.
[17] Green BG, Shaffer GS, Gilmore MM. Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. Chem Senses 1993;18(6):683–702.
[18] http://www.dnagenotek.com/US/pdf/PD-PR-006.pdf.
[19] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977; 74(12):5463–7.
[20] http://www.dnastar.com/t-allproducts.aspx.
[21] https://pinenutsyndrome.wordpress.com/pinenutsyndrome/.
[22] Bartoshuk LM, Duffy VB, Miller IJ. PTC/PROP tasting: anatomy, psychophysics, and sex effects. Physiol Behav 1994;56(6): 1165–71.
[23] Harris H, Kalmus H. Chemical specificity in genetical differences of taste sensitivity. Ann Eugen 1949;15(1):32–45.

Chemical Senses, 2016, Vol 41, 649–659
doi:10.1093/chemse/bjw067
Original Article
Advance Access publication June 23, 2016

OXFORD

Original Article

Copy Number Variation in *TAS2R* Bitter Taste Receptor Genes: Structure, Origin, and Population Genetics

Natacha Roudnitzky[1], Davide Risso[2], Dennis Drayna[2], Maik Behrens[1], Wolfgang Meyerhof[1] and Stephen P. Wooding[3]

[1]Department of Molecular Genetics, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114–116, 14558 Nuthetal, Germany, [2]National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892, USA and [3]Health Sciences Research Institute, University of California, Merced, 5200 North Lake Road, Merced, CA 95343, USA

Correspondence to be sent to: Stephen P. Wooding, Health Sciences Research Institute, University of California, Merced, 5200 North Lake Road, Merced, CA 95343, USA. e-mail: swooding@ucmerced.edu

Abstract

Bitter taste receptor genes (*TAS2R*s) harbor extensive diversity, which is broadly distributed across human populations and strongly associated with taste response phenotypes. The majority of *TAS2R* variation is composed of single-nucleotide polymorphisms. However, 2 closely positioned loci at 12p13, *TAS2R43* and *-45*, harbor high-frequency deletion (Δ) alleles in which genomic segments are absent, resulting in copy number variation (CNV). To resolve their chromosomal structure and organization, we generated maps using long-range contig alignments and local sequencing across the *TAS2R43–45* region. These revealed that the deletion alleles (*43*Δ and *45*Δ) are 37.8 and 32.2 kb in length, respectively and span the complete coding region of each gene (~1 kb) along with extensive up- and downstream flanking sequence, producing separate CNVs at the 2 loci. Comparisons with a chimpanzee genome, which contained intact homologs of *TAS2R43*, *-45*, and nearby *TAS2R*s, indicated that the deletions evolved recently, through unequal recombination in a cluster of closely related loci. Population genetic analyses in 946 subjects from 52 worldwide populations revealed that copy number ranged from 0 to 2 at both *TAS2R43* and *TAS2R45*, with *43*Δ and *45*Δ occurring at high global frequencies (0.33 and 0.18). Estimated recombination rates between the loci were low ($\rho = 2.7 \times 10^{-4}$; $r = 6.6 \times 10^{-9}$) and linkage disequilibrium was high ($D' = 1.0$), consistent with their adjacent genomic positioning and recent origin. Geographic variation pointed to an African origin for the deletions. However, no signatures of natural selection were found in population structure or integrated haplotype scores spanning the region, suggesting that patterns of diversity at *TAS2R43* and *-45* are primarily due to genetic drift.

Key words: bitter, evolution, genetic, genomic, natural selection, taste receptor

Introduction

Bitter taste perception is mediated in its earliest stages by TAS2Rs, a family of ~25 G protein-coupled receptors expressed in the apical microvilli of taste bud cells (Adler *et al*. 2000). When stimulated, these receptors trigger a transductional cascade culminating in depolarization, leading to sensation (Munger and Meyerhof 2015). TAS2Rs are responsive to numerous compounds at low concentrations and exhibit overlapping specificities, with most receptors responsive to multiple compounds and many compounds capable of stimulating multiple receptors (Meyerhof *et al*. 2010). A striking

© The Author 2016. Published by Oxford University Press. All rights reserved.
For permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

proportion of their agonists are toxins originating in plants, suggesting that the native biological role of TAS2Rs is to detect noxious substances in potential foods, preventing overexposure (Sandell and Breslin 2006). However, they retain their importance in modern populations through their influence on food likes, dislikes, and consumption (Duffy *et al.* 2004; Dinehart *et al.* 2006; Duffy 2007; Hayes *et al.* 2011; Hayes *et al.* 2013).

*TAS2R*s harbor extensive mutational polymorphism associated with receptor function and phenotypic variability. To date, more than 150 single-nucleotide polymorphisms (SNPs) have been identified in *TAS2R* coding regions, approximately 75% of which result in amino acid substitutions (Kim *et al.* 2005). These drive profound variation in receptor functionality. For instance, 3 amino acid replacements in *TAS2R38* distinguish functional variants underlying nearly 10 000-fold variation in taste sensitivity to phenylthiocarbamide (PTC) among subjects (Kim *et al.* 2003). Similar associations are found at other *TAS2R*s and are predicted to exist for most loci (Pronin *et al.* 2007; Roudnitzky *et al.* 2011, 2015; Campbell *et al.* 2014). Variation in *TAS2R*s is also associated with downstream health-connected phenotypes such as food and drink preferences, metabolic traits, and physical measures such as body mass index (Duffy 2007; Dotson *et al.* 2008).

The *TAS2R*s harboring the highest levels of genetic diversity occur in a region designated the *TAS2R30-31* cluster, which includes *TAS2R30*, *-31*, *-43*, *-45* and *-46* (Kim *et al.* 2005; Roudnitzky *et al.* 2015) (Figure 1). Genes in this cluster contain numerous SNPs associated with taste phenotypes such as sensitivity to bitterness of artificial sweeteners and phytotoxins (Pronin *et al.* 2007; Roudnitzky *et al.* 2011). Two genes in the cluster, *TAS2R43* and *-45*, contain structural variation as well as SNPs, including common deletion alleles in which at least part of the coding sequence of each gene is absent (Pronin *et al.* 2007; Roudnitzky *et al.* 2011; Wooding *et al.* 2012). In addition, coarse genomic maps support the presence of large genomic duplications and deletions in the region, including variants spanning *TAS2R43* (Redon *et al.* 2006; McCarroll *et al.* 2008; Perry *et al.* 2008a; Conrad *et al.* 2010; Sudmant *et al.* 2010; Sudmant *et al.* 2015a). Together, these findings suggest that *TAS2R43* and *-45* reside in regions exhibiting copy number variation (CNV), major polymorphisms in which genomic segments thousands or millions of nucleotides in length are present or absent in alternate alleles.

CNVs are a common feature of chemosensory genes and have been implicated as an important source of phenotypic variation beyond that accounted for by nucleotide substitutions (Nei *et al.* 2008; Nozawa and Nei 2008). Odorant receptor genes (ORs) are highly enriched with CNVs, which show complex patterns of genomic structure, population distribution, and evidence of natural selection (Hasin *et al.* 2008; Young *et al.* 2008). These are an emerging focal point in olfactory genetics because they are predicted to underlie variance in perception. Evidence that CNVs affect *TAS2R*s suggests that they too vary in structure, and may display differences in population frequency. To test the hypothesis that *TAS2R43* and *-45* reside in CNVs and delineate their genomic organization, we used computational and molecular techniques to establish the presence of structural variants spanning *TAS2R43* and *-45*, map their fine-scale genomic architectures, and determine their presence worldwide.

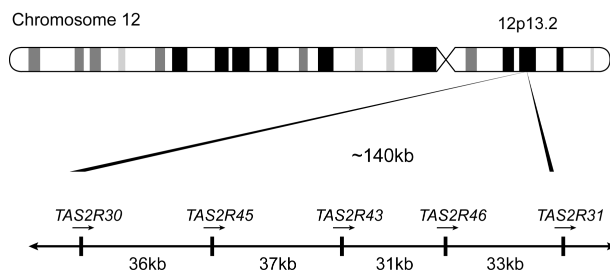


Figure 1. Gene locations. The *TAS2R30-31* cluster is located at 12p13.2. In the longest contig spanning the region (NW_001838055.2), their order in forward orientation is *TAS2R30*, *-45*, *-43*, *-46*, *-31*, with a mean separation of 34 kb.

Materials and methods

Subjects

CNV mapping was performed in 48 unrelated Caucasian subjects recruited at the German Institute for Human Nutrition Potsdam-Rehbruecke (Germany) as part of a previous study (Roudnitzky *et al.* 2011). The sample was collected and analyzed following the guidelines of the Declaration of Helsinki on Biomedical Research Involving Human Subjects and the study was approved by the Ethics Committee of the University of Potsdam (Germany) through decision 10/27. Session/2009 (Roudnitzky *et al.* 2011). All participants gave written informed consent. Geographic and population distributions of variation were analyzed in the CEPH H952 panel, a set of 952 nonrelated subjects identified by Rosenberg (2006) in the CEPH-HGDP diversity panel, an anthropologically focused collection of 1064 subjects representing 52 worldwide groups (Table 1) (Cann *et al.* 2002).

Structural mapping

Genomic alignments spanning the 5 *TAS2R*s at the 12p13.2 locus (the *TAS2R30-31* cluster) in Genome Reference Consortium Human builds 37 and 38 (GRCh37 and -38) and HuRef assembly were examined to ascertain the presence or absence of gaps and other inconsistencies potentially indicative of insertions and deletions. Alignments were generated using the FSA (Fast Statistical Alignment) software package with default settings (Bradley *et al.* 2009). To provide a basis of comparison with nonhumans, alignments were also performed against the homologous genomic region in chimpanzee (*Pan troglodytes*), which was identified in the CSAC 2.1.4/panTro4 genomic assembly using BLAST. The boundaries of structural variants discovered in these alignments were confirmed in human subjects using targeted fragment PCR amplification with primer sets positioned upstream, downstream, and spanning putative deleted regions. Multiplex PCR genotyping assays (Figure 3) were performed using Advantage 2 polymerase (Clontech Laboratories, Inc.) and standard PCR conditions. Primers and annealing temperature were specific for *TAS2R43* (43A_forward/43C_forward: 5′-AGCAACAGACAAGTTACTATTCAAAAGAAGC-3′; 43A_reverse: 5′-CCAATGTCAAACAGGAAAGCATCTCAAT-3′; 43B_forward: 5′-TCACGGATAGGGATTAAATGGTGGAAT-3′; 43B_reverse/43C_reverse: 5′-ACAATGCTTCTGGCCCATTTTCTCT-3′; annealing temperature: 68 °C) and for *TAS2R45* (45A_forward/45C_forward: 5′-GTCAGGATATTCAAGCAATCACAACCAG-3′; 45A_reverse: 5′-TCTTAAAACTCCAAACTGATATTATTACAGACACA-3′; 45B_forward: 5′-AGAGTTTTTGCTGAATAAAGGAGAATAGAACA-3′; 45B_reverse/45C_reverse: 5′-CCTTCTAACTCCATCATCACTCACTCAA-3′; annealing temperature: 67 °C).

Copy-number estimates

Copy-number estimates for *TAS2R43* and *-45* were obtained using the real time PCR protocol of Roudnitzky *et al.* (2011). This method utilizes ABI TaqMan Gene Copy Number Assays to compare

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

Table 1. H952 population sample

Population	N
Americas (N = 64)	
Colombian	7
Karitiana	14
Maya	21
Pima	14
Surui	8
Asia (N = 430)	
Balochi	24
Brahui	25
Burusho	23
Cambodian	10
Dai	10
Daur	10
Han	44
Hazara	22
Hezhen	9
Japanese	29
Kalash	23
Lahu	8
Makrani	25
Miaozu	10
Mongolian	10
Naxi	7
Oroqen	9
Pathan	24
She	10
Sindhi	24
Tu	10
Tujia	10
Uygur	10
Xibo	9
Yakut	25
Yizu	10
Europe (N = 158)	
Adyegi	22
French	28
French Basque	24
North Italian	13
Orcadian	15
Russian	25
Sardinian	28
Tuscan	8
Middle East (N = 133)	
Bedouin	46
Druze	41
Palestinian	46
North Africa (N = 29)	
Mozabite	29
Oceania (N = 29)	
Melanesian	11
Papuan	17
Subsaharan Africa (N = 104)	
Bantu (North)	11
Bantu (South)	8
Biaka Pygmy	22
Mandenka	22
Mbuti Pygmy	13
San	6
Yoruba	22

simultaneously the rate of PCR amplification of *TAS2R43* and *-45* with that of *RNaseP*, a stable control product, revealing their relative abundance. To avoid non-specific amplification, which could occur because *TAS2R43* and *-45* have a high level of sequence identity (92.7%), probes were targeted at regions constant across all known published sequences for each gene. Each assay was performed in triplicate for each subject, and only subjects producing positive control results and identical estimates across replicates were included in further analyses. Under this criterion, 6 H952 subjects were excluded from the study, resulting in a final HGDP sample size of 946.

Population genetic analysis

CNV genotypes in the H952 subjects were analyzed to ascertain allele frequencies, identify haplotypes, and determine levels of differentiation across worldwide populations. Tests for Hardy–Weinberg equilibrium (HWE), which detect bias in genotype frequencies due to factors such as nonrandom mating (e.g., inbreeding), population substructure, and genotyping error were performed in each of the 52 study populations. The PHASE software package was used to identify haplotypes (allelic combinations on individual chromosomes), and their pairings in subjects (Stephens *et al.* 2001). These provided measures of linkage disequilibrium (LD), D' and r^2, which estimate correlations between alleles at different genomic positions. PHASE was also used to establish recombination rates, r and ρ, which indicate long-term trends in genetic crossover in a region. Signatures of natural selection were analyzed using the HGDP selection browser (http://hgdp.uchicago.edu) to examine 2 indicators of pressure across a 1.5 Mb region centered on the *TAS2R30-31* segment: iHS (integrated haplotype score), which detects changes in haplotype lengths due to selection, and F_{ST}, which detects selection-caused changes in population divergence (Pickrell *et al.* 2009).

Results

Genomic structure

Alignments of human genomic contigs spanning the *TAS2R30-31* region revealed the presence of CNVs characterized by the complete absence of *-43* and *-45* on some genetic backgrounds. Thus, each locus was represented by 2 alleles: deleted (Δ) and nondeleted (n). In the GRCh38.7 release of the genome reference assembly, both genes were present in NW_003571050.1. Three representative examples were GRCh NW_001838055.2, NW_003571047.1, and NT_009714.17. While *TAS2R43* and *-45* were both present in NW_001838055.2, *TAS2R43* was absent from NW_003571047.1 and *TAS2R45* was absent from NT_009714.17 (Figure 2 and Supplementary Data). Thus, 3 haplotypes were observed: *43n/45n* (both nondeleted), *43n/45Δ* (*TAS2R43* non-deleted and *-45* deleted), and *43Δ/45n* (*-43* deleted and *-45* nondeleted). The fourth possible configuration, *43Δ/45Δ* (both deleted) was not detected at any point in the study. An aligned chimpanzee contig most resembled that of human contig NW_003571050.1, the variant containing intact copies of both *TAS2R43* and *-45*, and harbored orthologs of all 5 genes in the *TAS2R30-31* cluster (Figure 2).

In addition to spanning the *TAS2R43* and *-45* coding regions, the deleted regions in NW_003571047.1 and NT_009714.17 extended far up- and downstream. While *-43* and *-45* are 930 and 900 bp in length, respectively, the deletions encompassing them were 37787 and 32247 bp. They were also asymmetrical, with the *TAS2R43* deletion extending from 11.2 kb upstream to 25.7 kb downstream of the coding region and the *TAS2R45* deletion extending from 0.7 kb upstream to 30.1 kb downstream. Further, the genomic region deleted in NW_003571047.1 overlapped with that deleted in NT_009714.17, producing a 5.2 kb intergenic segment absent in both relative to NW_003571050.1.

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

The boundaries of the *TAS2R43* and *-45* deletions were confirmed by targeted PCR amplifications in the 48 subject CNV mapping panel (Figure 3A). Three defining amplification products were identified at each locus: 2 crossing the up and downstream boundaries of each CNV (*TAS2R43* PCRs 43A and 43B; *TAS2R45* PCRs 45A and 45B), and 1 crossing the complete CNV (*TAS2R43* PCR

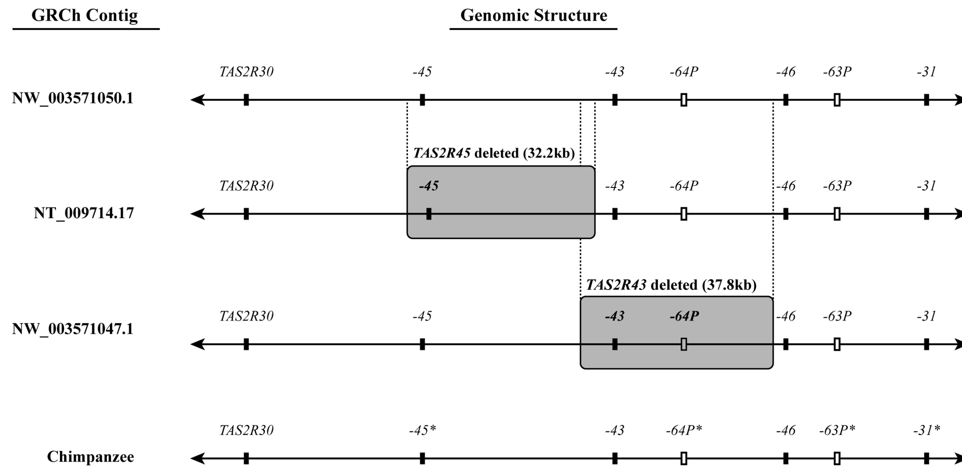
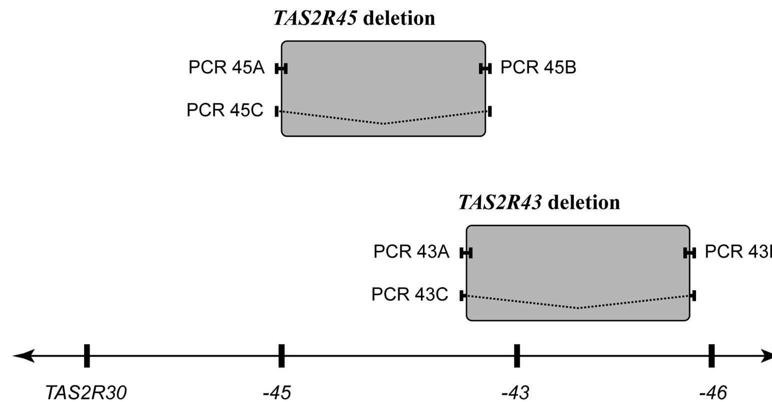


Figure 2. Genomic alignments. Alignments of three representative GRCh contigs are shown. Alignment gaps (highlighted), indicate deletion with respect to the reference contig, NW_003571050.1. The chimpanzee genome (bottom) was most similar to the reference contig, harboring homologs of all genes and pseudogenes in the human *TAS2R30-31* cluster. Homologs present but not currently annotated are denoted by the * symbol.



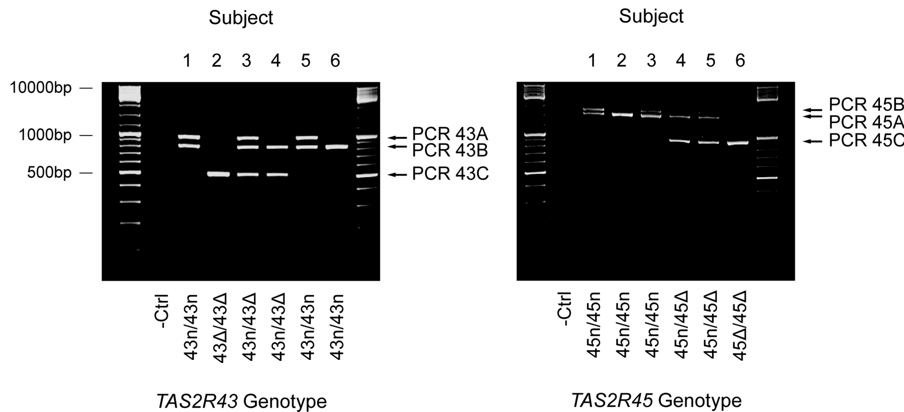


Figure 3. Amplification maps. (**A**) Iteratively amplifying targets in the *TAS2R43-45* region confirmed the boundaries of the deletions discovered in contig alignments. Three defining amplifications were established for each: one crossing the upstream limit of the deletion (43A and 45A), one crossing the downstream limit (43B and 45B), and one spanning the deletion (43C and 45C). (**B**) Multiplex PCR reactions in 6 representative subjects from the CNV mapping panel illustrate patterns in individuals with zero, one or two copies of each gene. For example, the sole presence of C products (which span the deletion) indicated a copy number of 0, while the sole presence of A and B products indicated a copy number of 2. The simultaneous presence of all products indicated a copy number of 1.

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

43C; *TAS2R45* PCR 45C). Thus, the pattern of products produced depended on the CNV allele. In nondeleted alleles products, A and B were produced, but C was not produced due to its excessive length. In deleted alleles products, A and B were not produced due to the absence of annealing regions, which lie in the deleted region, but C was produced because its length was short. In aggregate, these results confirmed the genotype of subjects (Figure 3B). For example, samples homozygous for the *TAS2R43* nondeleted allele (e.g., Subject 1 in Figure 3B) successfully produced products 43A and 43B, which only amplify in the presence of the nondeleted *TAS2R43* allele, but not 43C, which was too long to amplify (~38 kb). Conversely, samples homozygous for the deletion allele of *TAS2R43* (e.g., Subject 2 in Figure 3B) failed to produce products 43A and 43B, but successfully produced 43C, whose annealing positions up- and downstream of the deletion were in sufficient proximity to amplify. Analogous patterns were produced by primer sets surrounding *TAS2R45*.

Variation in CNV breakpoints, the up- and downstream boundaries of inserted or deleted regions, is a common phenomenon that includes both large and small scale differences among alleles (Perry *et al.* 2008b; Conrad *et al.* 2010). These have been well documented in human *OR* genes (Hasin *et al.* 2008). Our analyses revealed no large scale polymorphism in breakpoints, but small scale polymorphism was not ruled out. Genotypes ascertained using PCR amplification in the 48-subject mapping panel were consistent both internally and with TaqMan CNV assays, which were composed of 3 independent assays. Evidence of polymorphism in break point within PCR products, such as change in fragment size or the unexpected presence or absence of fragments, was also absent. In addition, genotypes in the H952 populations were in HWE, suggesting that large variants such as recombinant alleles with break points shifted from upstream of a gene to downstream, were rare or absent. However, small changes in breakpoints were beyond the resolution of our PCR and TaqMan analyses and may have been present but undetected. For instance, changes smaller than ~100 bp could be undetected in comparisons of PCR products.

A hallmark signature of the presence of CNVs in SNP datasets is elevated homozygosity, a preponderance of individuals carrying 2 identical alleles. The pattern occurs because standard genotyping methods incorrectly identify all SNP genotypes in CNV regions as homozygous in individuals with one gene copy. This provides the opportunity to establish approximate breakpoint positions. In population samples, CNV boundaries reside at the up- and downstream extent of genotypes out of HWE. In individuals they reside at the up- and downstream extent of homozygous SNPs. Current H952 datasets provide genomic coverage at a low resolution relative to the size of the *TAS2R43* and *-45* CNVs, averaging 6 kb genomewide, and are sparse across the *TAS2R43-45* region. Therefore, we did not utilize the HGDP SNP genotypes to obtain breakpoint estimates. However, we found that the presence and location of CNVs in the *TAS2R43-45* region was supported by SNP data from the 1000 Genomes Project (1000GP), an inventory of variation in 2504 subjects from 26 worldwide populations (1000 Genomes Project Consortium 2015; Sudmant *et al.* 2015b). In the 1000GP, SNP homozygosity levels were elevated in the region spanning *TAS2R43*, resulting in departures from HWE at sites extending from rs34847625 to rs372854040 (Supplementary Figure). Data for the *TAS2R45* region were not available.

Population genetics

Analysis of allele frequencies in 946 subjects from 52 worldwide populations in the CEPH Human Genetic Diversity Panel revealed differences among geographic regions, with different patterns at *TAS2R43* and *-45* (Figures 4 and 5). At *TAS2R43*, the deletion allele ranged in frequency from 0.21 (in Oceania) to 0.46 (in the Americas), with a global frequency of 0.33. At *TAS2R45* the deletion had a greater frequency range but was less common overall, with a minimum of 0.02 (in Oceania), a maximum of 0.64 (in Subsaharan Africa), and a global frequency of 0.18. Analogous patterns were found on a population level (Supplementary Data). Across populations, the *43*Δ allele had a mean frequency of 0.33, ranging from 0.00 (Mbuti Pygmy) to 0.75 (in San). Again, the *45*Δ allele had a slightly greater frequency range but was less common overall, with a minimum of 0.00 (in several populations in Asia and the Americas), a maximum of 1.00 (in Bantu and Mbuti Pygmy), and a mean of 0.17. Thus, both functional and deleted copies of *TAS2R43* and *-45* were found in all global regions, and nearly all populations. The only population sample fixed with respect to a deletion allele was Mbuti, which harbored no intact copies of *TAS2R45*.

CNV genotyping indicated the presence of individuals carrying 0–2 gene copies at both *TAS2R43* and *TAS2R45*, corresponding to the Δ/Δ, n/Δ, and n/n genotypes at each locus. Genotypes were in HWE in all populations when multiple testing was taken into account using a Bonferroni correction, which specified a significance cutoff of $P = 0.0005$. At the less conservative $P = 0.05$ level, three populations showed significant departures from expectations. At *TAS2R43*, excess heterozygosity was observed in the Northern Bantu ($N = 11$; $P = 0.02$) and Tuscan ($N = 8$; $P = 0.005$) samples, while an excess of homozygosity was observed in She ($N = 10$; $P = 0.002$). At *TAS2R45*, Northern Bantu exhibited a deficit of heterozygosity ($P = 0.006$). However, the number of departures (4) was roughly that expected by chance when performing 104 tests (5). Thus, potentially confounding effects from hidden population structure and nonrandom mating were not detected. This finding also supported the accuracy of copy number assays, which should manifest departures from HWE if systematic errors occur.

Although most populations carried both deleted and nondeleted alleles at *TAS2R43* and *-45*, the frequencies of deleted alleles were high (0.33 and 0.18 worldwide, respectively) and numerous subjects were homozygous with respect to the deletion alleles (Figure 4). Thus, these individuals completely lacked 1 of the 2 genes. The highest frequency of homozygotes for the *TAS2R43* deletion was found in Europe, where more than 20% of subjects were missing the gene (Figure 4A). The pattern was less pronounced at *TAS2R45*, which had a lower deletion allele frequency (0.18 globally), and the prevalence of *TAS2R45* deletion homozygotes was below 2% for most regions (Figure 4B). However, it was high among Subsaharan Africans, 41% of whom carried no copy of the gene.

Analyses using PHASE indicated that the H952 panel carried 3 haplotypes, which corresponded to the 3 found in contig alignments: *43n/45n*, *43n/45Δ*, and *43Δ/45n* (Figure 5A). Linkage disequilibrium between the 2 loci was high ($D' = 1.0$; $r^2 = 0.11$), and the estimated population recombination rate ρ, was low, 2.7×10^{-4}, consistent with their adjacent genomic locations. Under a conservative estimate of effective population size (N_e) in humans, ~10 000, it implied a recombination rate ($r = \rho/4N_e$) of ~6.6×10^{-9} per generation. The double-deletion haplotype (*43Δ/45Δ*) was not detected at any point in the study although recombination between *43n/45Δ*, and *43Δ/45n* would be expected to produce the double-deletion type. This was explained by the low level of recombination between *TAS2R43* and *-45*, which should generate double-deletion alleles at very low rates. The production of *43n/45Δ*, and *43Δ/45n* through recombination between the double deletion allele and *43n/45n*

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

A

Region	*TAS2R43* Alleles		*TAS2R43* Genotypes		
	43n	43Δ	43n/43n	43n/43Δ	43Δ/43Δ
Americas	69 (0.54)	59 (0.46)	18 (0.28)	33 (0.52)	13 (0.20)
Asia	619 (0.72)	241 (0.28)	222 (0.52)	175 (0.40)	33 (0.08)
Europe	173 (0.55)	143 (0.45)	50 (0.32)	73 (0.46)	35 (0.22)
Middle East	180 (0.68)	86 (0.32)	64 (0.48)	52 (0.39)	17 (0.13)
North Africa	39 (0.67)	19 (0.33)	14 (0.48)	11 (0.38)	4 (0.14)
Oceania	44 (0.79)	12 (0.21)	17 (0.60)	10 (0.36)	1 (0.04)
Subsaharan Africa	139 (0.67)	69 (0.33)	44 (0.42)	51 (0.49)	9 (0.09)
Worldwide	1263 (0.67)	629 (0.33)	429 (0.45)	405 (0.43)	112 (0.12)

B

Region	*TAS2R45* Alleles		*TAS2R45* Genotypes		
	45n	45Δ	45n/45n	45n/45Δ	45Δ/45Δ
Americas	122 (0.95)	6 (0.05)	59 (0.92)	4 (0.06)	1 (0.02)
Asia	781 (0.91)	79 (0.09)	354 (0.82)	73 (0.17)	3 (0.01)
Europe	261 (0.83)	55 (0.17)	106 (0.67)	49 (0.31)	3 (0.02)
Middle East	212 (0.80)	54 (0.20)	88 (0.66)	36 (0.27)	9 (0.07)
North Africa	47 (0.81)	11 (0.19)	18 (0.62)	11 (0.38)	0 (0.00)
Oceania	55 (0.98)	1 (0.02)	27 (0.96)	1 (0.04)	0 (0.00)
Subsaharan Africa	75 (0.36)	133 (0.64)	13 (0.12)	49 (0.47)	42 (0.41)
Worldwide	1553 (0.82)	339 (0.18)	665 (0.70)	223 (0.24)	58 (0.06)

Figure 4. Allele and genotype frequencies. Deleted and nondeleted allele and genotype frequencies are shown for major geographical regions. Occurrences are given, with frequencies in parentheses.

A

Haplotype	Contig	Chromosomal Configuration
43n/45n	NW_003571050.1	
43n/45Δ	NT_009714.17	
43Δ/45n	NW_003571047.1	

B

Region	Occurrences (frequency)		
	43n/45n	43n/45Δ	43Δ/45n
Americas	63 (0.49)	6 (0.05)	59 (0.46)
Asia	540 (0.63)	79 (0.09)	241 (0.28)
Europe	118 (0.37)	55 (0.17)	143 (0.45)
Middle East	126 (0.47)	54 (0.20)	86 (0.32)
North Africa	28 (0.48)	11 (0.19)	19 (0.33)
Oceania	43 (0.77)	1 (0.02)	12 (0.21)
Subsaharan Africa	6 (0.03)	133 (0.64)	69 (0.33)
Worldwide	924 (0.49)	339 (0.18)	629 (0.33)

Figure 5. Haplotype configurations and frequencies. (**A**) Three haplotypes were observed with respect to the *TAS2R43* and *-45* deletion alleles, consistent with 3 chromosomal configurations: 43n/45n, 43n/45Δ, and 43Δ/45n. (**B**) Haplotype occurrences in major geographical regions. *N* indicates number of subjects. Occurrences are shown, with frequencies in parentheses.

should be similarly rare. Therefore, we hypothesize that *43Δ/45Δ* is absent from our sample, but is likely present at low frequencies in human populations and might even occur at high frequencies in populations outside our study.

Like single-locus alleles, haplotypes varied in frequency across both regions and populations (Figures 5B and 6). Haplotype *43n/45n* was the most common globally, with a mean frequency of 0.49. It was least common in Subsaharan Africa, where it was found at a frequency of 0.03, and most common in Oceania, where it occurred at a frequency of 0.77. Haplotype *43n/45Δ* was found at frequencies ranging from 0.02 in Oceania to 0.64 in Subsaharan Africa, with a global mean of 0.18. *43Δ/45n* was found at intermediate frequencies, ranging from 0.21 (in Oceania) to 0.46 (in the Americas) with a mean of 0.33. Similar patterns were observed on a population level (Supplementary Data). Again, haplotype *43n/45n* was most common, ranging in frequency from 0.00 (in Bantu and Mbuti Pygmy) to >0.85 in 5 Asian populations, with a mean of 0.50. Haplotype *43n/45Δ* was the least common at frequencies ranging from 0.00 in 10 Asian and American populations to 1.00 in Mbuti Pygmy, with a mean of 0.17. Haplotype *43Δ/45n* was found at intermediate frequencies, ranging from 0.00 (in Bantu and Mbuti) to 0.75 (in San) with a mean of 0.33.

Patterns of population differentiation and haplotype structure in the H952 panel showed no evidence of selective pressure across the 1.5 Mb region spanning the *TAS2R30-31* segment

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

(Voight *et al.* 2006). F_{ST} values both within the *TAS2R30-31* cluster and across broader flanking regions lay predominantly between the 5th and 95th percentiles among empirical genome wide measures, indicating that populations are not exceptionally divergent or homogeneous with respect to SNPs in the region. While high and low F_{ST} values were observed they were not extreme and values were scattered, not consistent as would be expected if driven by population structure. Likewise, the decay of linkage disequilibrium with distance as measured by iHS, which can be indicative of positive or balancing selection, was low locally but was not extreme relative to other regions analyzed in the HGDP. For instance, scores in established examples of selection typically exceed 3.0 and often exceed 4.0 (Voight *et al.* 2006), but no iHS score exceeding 2.5 was observed in our sample and most were lower, averaging ~1.5. Thus, both the geographic distributions of haplotypes and genomic structure across the *TAS2R43-45* region must be primarily due to genetic drift.

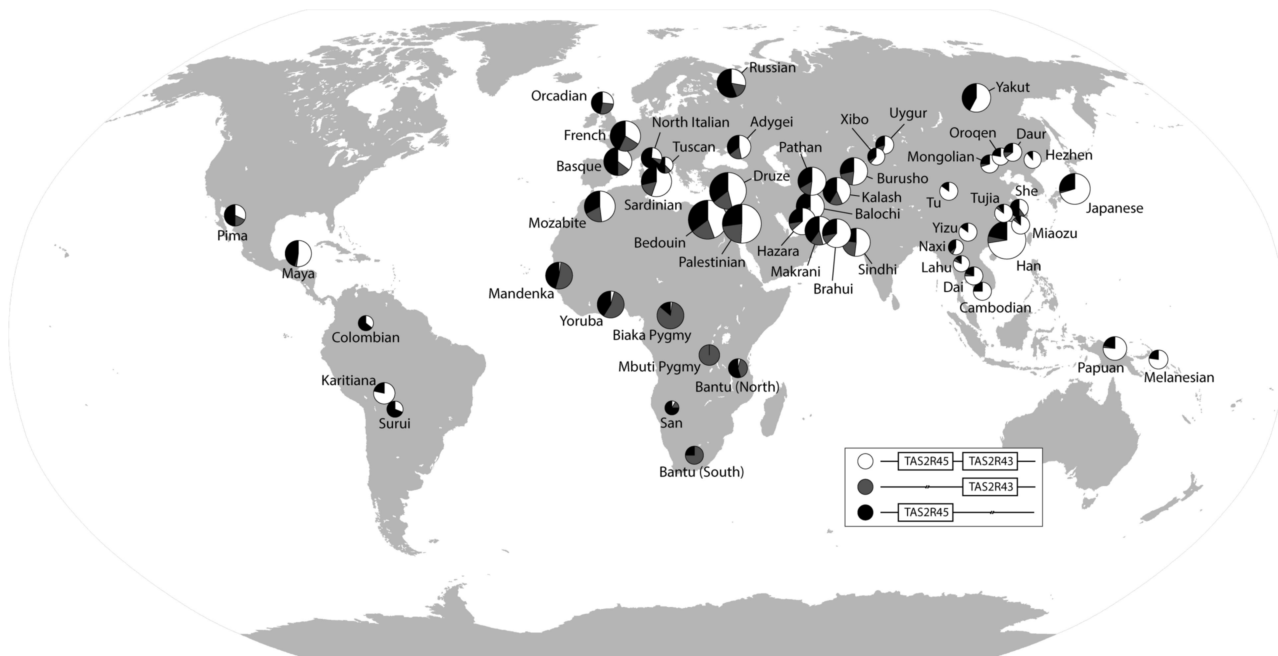


Figure 6. Global haplotype frequencies. Each point represents a sampled H952 population. Point size is proportional to sample size. Sections indicate the relative haplotype frequency within populations.

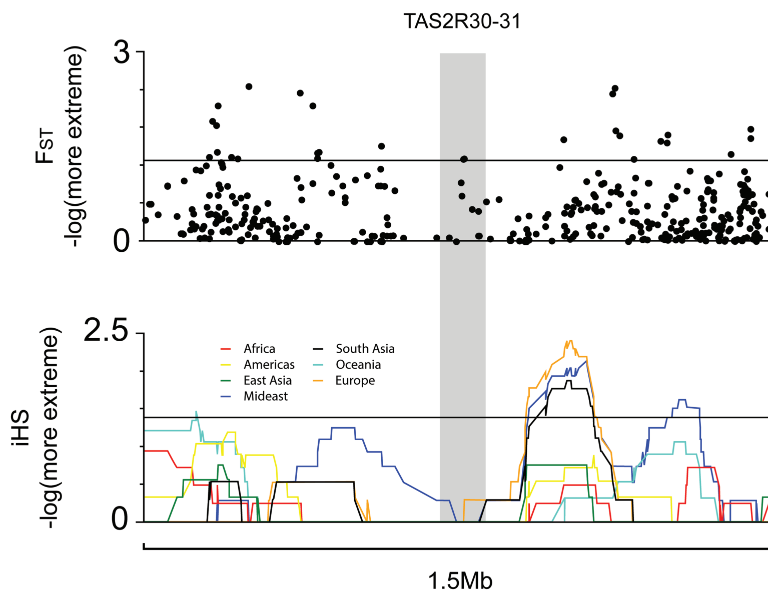


Figure 7. Tests for natural selection. Patterns across a 1.5 Mb region centered on the *TAS2R30-31* cluster in the HGDP are shown. (**A**) Plot of SNP F_{ST} values relative to the distribution across all SNPs in the HGDP. Y-axis indicates the negative log of the fraction of HGDP SNPs with more extreme F_{ST} values than the plotted SNP. Horizontal line indicates 95th percentile. (**B**) Thirty SNP sliding window plot of iHS scores (integrated haplotype scores) relative to the genome-wide distribution in the HGDP. Y-axis indicates the negative log of the fraction of windows with more extreme iHS values than the plotted window. Horizontal line indicates 95th percentile.

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

Discussion

The role of TAS2Rs as toxin sensors suggests that their fitness relevance varies across environments, resulting in differences in evolutionary constraint among loci and species (Sandell and Breslin 2006; Lei *et al.* 2015; Sandau *et al.* 2015). These are well documented in both population genetic and phylogenetic studies (Wang *et al.* 2004; Wooding *et al.* 2004; Kim *et al.* 2005; Soranzo *et al.* 2005; Wooding 2011; Campbell et al. 2012, 2014). Diversity patterns consistent with variation in selective pressure are found at multiple loci in humans, indicating that some are evolving neutrally while others are under positive or balancing selection, which cause deficits and excesses of genetic diversity in affected genes (Wang *et al.* 2004; Wooding *et al.* 2004; Kim *et al.* 2005; Soranzo *et al.* 2005; Nozawa *et al.* 2007; Campbell et al. 2012, 2014). The number of *TAS2R* loci also varies substantially among taxa, suggesting that CNV could be tolerated or even actively maintained by selective pressure (Shi *et al.* 2003; Fischer *et al.* 2005; Go *et al.* 2005; Shi and Zhang 2006; Nozawa and Nei 2008; Lei *et al.* 2015; Sandau *et al.* 2015). A key trend is that *TAS2R* repertoires vary systematically among species differing in dietary exposure to phytotoxins. For example, carnivores, which face relatively little exposure to TAS2R agonists, carry smaller *TAS2R* repertoires than do herbivores (Suzuki *et al.* 2010; Wooding 2011; Feng *et al.* 2014; Li and Zhang 2014; Lei *et al.* 2015; Sandau *et al.* 2015; Suzuki-Hashido *et al.* 2015). However, this pattern is not universal, suggesting that factors beyond diet are also important drivers of *TAS2R* evolution (Lei *et al.* 2015; Sandau *et al.* 2015).

Numerous genome-scale studies have reported the presence of structural variation in the *TAS2R30-31* region but size estimates have been inconsistent, ranging from <1 kb to >500 kb (Redon *et al.* 2006; Wong *et al.* 2007). Our findings indicate that CNVs are present at both *TAS2R43* and *-45* and establish their size and genomic organization. The deletions we identified are most similar to those reported by Sudmant *et al.* (2015a), who detected a 29 kb copy-number variant spanning *TAS2R43*. However, we obtained a slightly larger estimate, 37.8 kb, and defined a deletion allele at *TAS2R45* that has not been described previously. We also found that the orientation of the deletions is asymmetrical with respect to their corresponding genes. In the case of *TAS2R43* the deletion extends from 11.2 kb upstream to 25.7 kb downstream (totaling 37.8 kb including the coding region) and the *TAS2R45* deletion extends from 0.7 kb upstream to 30.1 kb downstream (totaling 32.2 kb), and no evidence of large-scale variation in the up- and downstream boundaries of deleted regions was detected. Further, none of the haplotypes identified in genomic contigs or inferred in subjects supported the presence of chromosomes deleted with respect to *TAS2R43* and *-45* simultaneously. All harbored *TAS2R43*, *TAS2R45*, or both.

Structural maps of the *TAS2R43* and *-45* deletions reveal their mutational origins. First, comparisons with the chimpanzee genome, which contains intact copies of all genes in the *TAS2R30-31* cluster including *-43* and *-45*, suggest that the haplotype containing intact copies of both was present in the common ancestor of humans and chimpanzees 5–7 million years ago (Kumar *et al.* 2005). Thus, while it cannot be ruled out that the *-43* and *-45* copy number variants have been segregating throughout that interval, the simplest scenario is that the ancestral haplotype contained both genes. Thus, the copy number variants most likely evolved through deletion from an intact ancestral chromosome, not through duplication and insertion, and must have emerged after the evolutionary divergence of humans and chimpanzees. Second, the localization of the two deletions to different haplotype backgrounds and the presence of the partial genomic overlap (5.3 kb) indicate that they are products of separate deletion events, not a single event spanning both loci, and represent distinct CNVs. Finally, levels of nucleotide identity among loci in the *TAS2R30-31* cluster exceed 80%, with a mean intergene distance of 34.3 kb, similar to the lengths of the *TAS2R43* and *-45* deletions (37.8 and 32.2 kb, respectively). This suggests that they are the result of unequal recombination among tandem loci, the most common mechanism of expansion in the *OR* gene family and the primary source of CNV genome wide (Young *et al.* 2008; Zhang *et al.* 2009). If so, it introduces likelihood that parallel deletion processes may have been occurring in chimpanzees, resulting in separate but similar patterns of CNV in the 2 species. Such parallels have been reported previously, with independent losses of function at *TAS2R38* resulting in shared patterns of bitter perception of PTC (Wooding *et al.* 2006).

In addition to supporting their emergence through a deletion process, as opposed to an insertion process, haplotype frequencies in our data point to Africa as the geographic origin of *43Δ* and *45Δ*. It has long been observed that allele frequencies provide an indication of population origins. For instance, a consistent pattern in DNA sequence data is that the relative abundance of ancestral alleles is highest in ancestral populations (Rogers *et al.* 2007). These patterns are shaped by mutation rate (Rogers *et al.* 2007). When mutation rates are high, the frequency of ancestral alleles is highest in ancestral populations; when mutation rates are low, the frequency of ancestral alleles is lowest in ancestral populations. *De novo* CNVs evolve at low rates, suggesting that they follow the latter pattern (Itsara *et al.* 2010). Thus, diversity in our sample, in which the ancestral allele (*43n/45n*) is least common in Africa and most common in Oceania, is most consistent with African origins of *43Δ/45Δ* followed by dispersal. This finding agrees with numerous lines of evidence that modern humans originated in Africa and migrated outward to other continents (Li *et al.* 2008)

The absence of strong positive or balancing selective pressures in the *TAS2R30-31* cluster raises questions about the fitness relevance of polymorphism at *TAS2R43* and *-45*. Because feeding behaviors and diet choice are central to human growth and development, mutations causing variation in taste would seem to be subject to strong pressures from natural selection, yet signs of selection are absent in measures of both haplotype structure and geographic distribution in our sample. Thus, patterns of variation found at *TAS2R43* and *-45* are most likely the result of neutral processes, such as genetic drift. Evidence of neutral processes has been reported in studies of other *TAS2R*s, as well (Wang *et al.* 2004). This may be explained by aspects of the functional contributions of *TAS2R43* and *-45* to phenotype. First, Meyerhof *et al.* (2010) found that while TAS2R43 is responsive to a range of agonists, most are agonists for other TAS2Rs as well, such that TAS2R43's responses are redundant. For instance, amarogentin is a TAS2R43 agonist, but it is also an agonist of 6 other TAS2Rs. Thus, loss of function at TAS2R43 may be compensated for by other loci, reducing selective pressures. Second, despite extensive screening, no agonist of TAS2R45 has been discovered (Meyerhof *et al.* 2010; Thalmann *et al.* 2013). Thus, it appears to be completely dysfunctional (Figure 8).

Similarities and differences in allele frequency among human populations have complex implications for phenotypic variation. TAS2R43 has numerous known agonists, including both natural compounds (such as caffeine, falcarindiol, and grosheimin) and synthetic ones (such as the sweeteners saccharin and acesulfame K, and the deterrent denatonium benzoate), and polymorphism in *TAS2R43* shapes perception of these compounds (Kuhn *et al.* 2004;

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

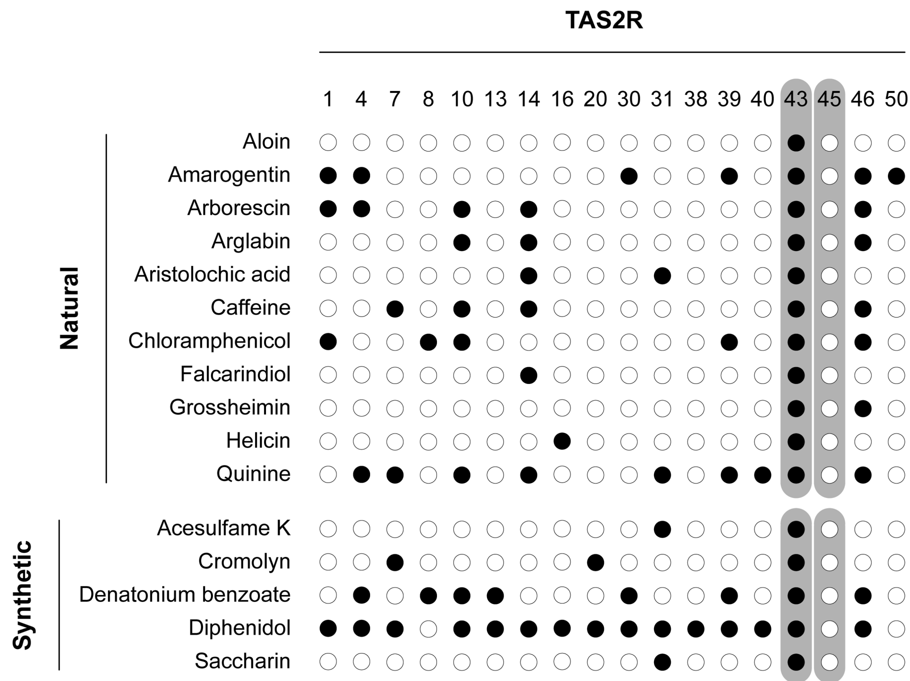


Figure 8. TAS2Rs and agonists. Responses of 18 TAS2Rs to each of the 16 known agonists of TAS2R43, reported by Meyerhof *et al.* (2010), are shown. TAS2R45 has no known agonists (Meyerhof *et al.* 2010; Thalmann *et al.* 2013). Filled circles indicate responsiveness. Of TAS2R43's 16 agonists, 15 are agonists of at least one other TAS2R, such that their responses are redundant.

Pronin *et al.* 2007; Meyerhof *et al.* 2010; Roudnitzky *et al.* 2011). Thus, the patterns we observe suggest that populations vary phenotypically as the result of frequency differences in *43Δ*, for example, for Grosheimin (Roudnitzky *et al.*, 2015). In contrast, evidence that TAS2R45 is nonfunctional suggests that differences in the frequency of *45Δ* account for no phenotypic variance among populations despite complete removal of the gene from the genome. However, if TAS2R45 is functionally responsive to agonists not yet identified, then *45Δ* could account for individual and population differences. Further, high LD between *TAS2R43* and *-45* suggests that if both contribute to phenotype, then correlations between phenotypes driven by the two genes will likely be strong, and negative.

CNV has been reported at *TAS2R* loci in addition to *TAS2R43* and *-45*, suggesting that *TAS2R* copy number polymorphisms may be an underappreciated contributor to bitter taste phenotypes. For example, genome-wide scans have consistently detected structural variation spanning *TAS2R39* (Itsara *et al.* 2010; Cooper *et al.* 2011; Sudmant *et al.* 2015a), which is responsive to a range of compounds and is the only *TAS2R* responsive to the analgesic, acetaminophen (paracetamol). The precise boundaries of the CNVs found at *TAS2R39* are not known, but a simple prediction is that these mediate taste responses to the drug. Evidence for CNV has also been reported at *TAS2R19*, *-20*, *-42*, and *-50* (Redon *et al.* 2006; Ahn *et al.* 2009; Coe *et al.* 2014). Taken together these findings indicate that while most variance in bitter perception is likely accounted for by nonsynonymous nucleotide variants, which are ubiquitous and known to exert strong effects on taste sensitivity, CNV is an important additional contributor. These findings also raise questions about population-level patterns of diversity of CNV at *TAS2R* loci, which are not yet known but could shed light on broader evolutionary pressures on structural evolution in the family.

Supplementary material

Supplementary material can be found at http://www.chemse.oxfordjournals.org/

Acknowledgments

The authors thank Stefanie Schultz for technical assistance.

Conflict of Interest

W.M., co-author, is Editor-in-Chief of Chemical Senses.

References

1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature*. 526:68–74.

Adler E, Hoon MA, Mueller KL, Chandrashekar J, Ryba NJ, Zuker CS. 2000. A novel family of mammalian taste receptors. *Cell*. 100(6):693–702.

Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C, *et al*. 2009. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res*. 19:1622–1629.

Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, Holmes I, Pachter L. 2009. Fast statistical alignment. *PLoS Comput Biol*. 5:e1000392.

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*. 2014. Origin and differential selection of allelic variation at *TAS2R16* associated with salicin bitter taste sensitivity in Africa. *Mol Biol Evol*. 31:288–302.

Cann H, de Toma C, Cazes L, Legrand M, Morel V, Piouffre L, Bodmer J, Bodmer W, Bonne-Tamir B, Cambon-Thomsen A, *et al*. 2002. A human genome diversity cell line panel. *Science*. 296:261–262.

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, Friend KL, Baker C, Buono S, Vissers LE, *et al*. 2014. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet*. 46:1063–1071.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, *et al*. 2010. Origins and functional impact of copy number variation in the human genome. *Nature*. 464:704–712.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, *et al*. 2011. A copy number variation morbidity map of developmental delay. *Nat Genet*. 43:838–846.

Dinehart M, Hayes J, Bartoshuk L, Lanier S, Duffy V. 2006. Bitter taste markers explain variability in vegetable sweetness, bitterness, and intake. *Physiol Behav*. 87:304–313.

Dotson CD, Zhang L, Xu H, Shin YK, Vigues S, Ott SH, Elson AE, Choi HJ, Shaw H, Egan JM, *et al*. 2008. Bitter taste receptors influence glucose homeostasis. *PLoS One*. 3:e3974.

Duffy VB. 2007. Variation in oral sensation: implications for diet and health. *Curr Opin Gastroenterol*. 23(2):171–177.

Duffy VB, Davidson AC, Kidd JR, Kidd KK, Speed WC, Pakstis AJ, Reed DR, Snyder DJ, Bartoshuk LM. 2004. Bitter receptor gene (*TAS2R38*), 6-n-propylthiouracil (PROP) bitterness and alcohol intake. *Alcohol Clin Exp Res*. 28(11):1629–1637.

Feng P, Zheng J, Rossiter SJ, Wang D, Zhao H. 2014. Massive losses of taste receptor genes in toothed and baleen whales. *Genome Biol Evol*. 6(6):1254–1265.

Fischer A, Gilad Y, Man O, Pääbo S. 2005. Evolution of bitter taste receptors in humans and apes. *Mol Biol Evol*. 22(3):432–436.

Go Y, Satta Y, Takenaka O, Takahata N. 2005. Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates. *Genetics*. 170(1):313–326.

Hasin Y, Olender T, Khen M, Gonzaga-Jauregui C, Kim PM, Urban AE, Snyder M, Gerstein MB, Lancet D, Korbel JO. 2008. High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genet*. 4(11):e1000249.

Hayes JE, Feeney EL, Allen AL. 2013. Do polymorphisms in chemosensory genes matter for human ingestive behavior? *Food Qual Pref*. 30:202–216.

Hayes JE, Wallace MR, Knopik VS, Herbstman DM, Bartoshuk LM, Duffy VB. 2011. Allelic variation in *TAS2R* bitter receptor genes associates with variation in sensations from and ingestive behaviors toward common bitter beverages in adults. *Chem Senses*. 36(3):311–319.

Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE. 2010. De novo rates and selection of large copy number variation. *Genome Res*. 20(11):1469–1481.

Kim U, Wooding S, Ricci D, Jorde LB, Drayna D. 2005. Worldwide haplotype diversity and coding sequence variation at human bitter taste receptor loci. *Hum Mutat*. 26(3):199–204.

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(5610):1221–1225.

Kuhn C, Bufe B, Winnig M, Hofmann T, Frank O, Behrens M, Lewtschenko T, Slack JP, Ward CD, Meyerhof W. 2004. Bitter taste receptors for saccharin and acesulfame K. *J Neurosci*. 24(45):10260–10265.

Kumar S, Filipski A, Swarna V, Walker A, Hedges SB. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc Natl Acad Sci U S A*. 102(52):18842–18847.

Lei W, Ravoninjohary A, Li X, Margolskee RF, Reed DR, Beauchamp GK, Jiang P. 2015. Functional analyses of bitter taste receptors in domestic cats (*Felis catus*). *PLoS One*. 10(10):e0139670.

Li D, Zhang J. 2014. Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Mol Biol Evol*. 31(2):303–309.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, *et al*. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 319:1100–1104.

McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, *et al*. 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet*. 40:1166–1174.

Meyerhof W, Batram C, Kuhn C, Brockhoff A, Chudoba E, Bufe B, Appendino G, Behrens M. 2010. The molecular receptive ranges of human TAS2R bitter taste receptors. *Chem Senses*. 35(2):157–170.

Munger SD, Meyerhof W. 2015. The molecular basis of gustatory transduction. In: Doty RL, editor. *Handbook of olfaction and gustation*. Oxford: Wiley-Blackwell. p. 685–700.

Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet*. 9(12):951–963.

Nozawa M, Kawahara Y, Nei M. 2007. Genomic drift and copy number variation of sensory receptor genes in humans. *Proc Natl Acad Sci U S A*. 104:20421–20426.

Nozawa M, Nei M. 2008. Genomic drift and copy number variation of chemosensory receptor genes in humans and mice. *Cytogenet Genome Res*. 123:263–269.

Perry G, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee A, Hyland C, Stone A, Hurles M, Tyler-Smith C, *et al*. 2008a. Copy number variation and evolution in humans and chimpanzees. *Genome Res*. 18:1698–1710.

Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, *et al*. 2008b. The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet*. 82:685–695.

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.

Pronin AN, Xu H, Tang H, Zhang L, Li Q, Li X. 2007. Specific alleles of bitter receptor genes influence human sensitivity to the bitterness of aloin and saccharin. *Curr Biol*. 17(16):1403–1408.

Redon R, Ishikawa S, Fitch K, Feuk L, Perry G, Andrews T, Fiegler H, Shapero M, Carson A, Chen W, *et al*. 2006. Global variation in copy number in the human genome. *Nature*. 444:444–454.

Rogers AR, Wooding S, Huff CD, Batzer MA, Jorde LB. 2007. Ancestral alleles and population origins: inferences depend on mutation rate. *Mol Biol Evol*. 24(4):990–997.

Rosenberg NA. 2006. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet*. 70:841–847.

Roudnitzky N, Behrens M, Engel A, Kohl S, Thalmann S, Hübner S, Lossow K, Wooding SP, Meyerhof W. 2015. Receptor polymorphism and genomic structure interact to shape bitter taste perception. *PLoS Genet*. 11(9):e1005530.

Roudnitzky N, Bufe B, Thalmann S, Kuhn C, Gunn HC, Xing C, Crider BP, Behrens M, Meyerhof W, Wooding SP. 2011. Genomic, genetic and functional dissection of bitter taste responses to artificial sweeteners. *Hum Mol Genet*. 20(17):3437–3449.

Sandau MM, Goodman JR, Thomas A, Rucker JB, Rawson NE. 2015. A functional comparison of the domestic cat bitter receptors Tas2r38 and Tas2r43 with their human orthologs. *BMC Neurosci*. 16:33.

Sandell MA, Breslin PA. 2006. Variability in a taste-receptor gene determines whether we taste toxins in food. *Curr Biol*. 16(18):R792–R794.

Shi P, Zhang J. 2006. Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol Biol Evol*. 23(2):292–300.

Shi P, Zhang J, Yang H, Zhang YP. 2003. Adaptive diversification of bitter taste receptor genes in Mammalian evolution. *Mol Biol Evol*. 20(5):805–814.

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(14):1257–1265.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 68(4):978–989.

Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE; 1000 Genomes Project. 2010. Diversity of human copy number variation and multicopy genes. *Science*. 330(6004):641–646.

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, Coe BP, Baker C, Nordenfelt S, Bamshad M, *et al*. 2015a. Global diversity, population stratification, and selection of human copy number variation. *Science*. 349:aab3761.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Hsi-Yang Fritz M, *et al*. 2015b. An integrated map of structural variation in 2,504 human genomes. *Nature*. 526:75–81.

Suzuki N, Sugawara T, Matsui A, Go Y, Hirai H, Imai H. 2010. Identification of non-taster Japanese macaques for a specific bitter taste. *Primates*. 51(4):285–289.

Suzuki-Hashido N, Hayakawa T, Matsui A, Go Y, Ishimaru Y, Misaka T, Abe K, Hirai H, Satta Y, Imai H. 2015. Rapid expansion of phenylthiocarbamide non-tasters among Japanese macaques. *PLoS One*. 10(7):e0132016.

Thalmann S, Behrens M, Meyerhof W. 2013. Major haplotypes of the human bitter taste receptor *TAS2R41* encode functional receptors for chloramphenicol. *Biochem Biophys Res Commun*. 435:267–273.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol*. 4(3):e72.

Wang X, Thomas SD, Zhang J. 2004. Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum Mol Genet*. 13(21):2671–2678.

Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, *et al*. 2007. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet*. 80(1):91–104.

Wooding S. 2011. Signatures of natural selection in a primate bitter taste receptor. *J Mol Evol*. 73(5–6):257–265.

Wooding S, Bufe B, Grassi C, Howard M, Stone A, Vazquez M, Dunn D, Meyerhof W, Weiss R, Bamshad M. 2006. Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature*. 440:930–934.

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(4):637–646.

Wooding SP, Atanasova S, Gunn HC, Staneva R, Dimova I, Toncheva D. 2012. Association of a bitter taste receptor mutation with Balkan Endemic Nephropathy (BEN). *BMC Med Genet*. 13:96.

Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM, Trask BJ. 2008. Extensive copy-number variation of the human olfactory receptor gene family. *Am J Hum Genet*. 83(2):228–242.

Zhang F, Gu W, Hurles ME, Lupski JR. 2009. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 10:451–481.

Downloaded from http://chemse.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

SCIENTIFIC REPORTS

OPEN

Global diversity in the *TAS2R38* bitter taste receptor: revisiting a classic evolutionary PROPosal

Received: 31 December 2015
Accepted: 18 April 2016
Published: 03 May 2016

Davide S. Risso[1,2], Massimo Mezzavilla[3,4], Luca Pagani[2,5], Antonietta Robino[3], Gabriella Morini[6], Sergio Tofanelli[7], Maura Carrai[7], Daniele Campa[7], Roberto Barale[7], Fabio Caradonna[8], Paolo Gasparini[3], Donata Luiselli[2], Stephen Wooding[9] & Dennis Drayna[1]

The ability to taste phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP) is a polymorphic trait mediated by the *TAS2R38* bitter taste receptor gene. It has long been hypothesized that global genetic diversity at this locus evolved under pervasive pressures from balancing natural selection. However, recent high-resolution population genetic studies of *TAS2Rs* suggest that demographic events have played a critical role in the evolution of these genes. We here utilized the largest *TAS2R38* database yet analyzed, consisting of 5,589 individuals from 105 populations, to examine natural selection, haplotype frequencies and linkage disequilibrium to estimate the effects of both selection and demography on contemporary patterns of variation at this locus. We found signs of an ancient balancing selection acting on this gene but no post Out-Of-Africa departures from neutrality, implying that the current observed patterns of variation can be predominantly explained by demographic, rather than selective events. In addition, we found signatures of ancient selective forces acting on different African *TAS2R38* haplotypes. Collectively our results provide evidence for a relaxation of recent selective forces acting on this gene and a revised hypothesis for the origins of the present-day worldwide distribution of *TAS2R38* haplotypes.

More than 70 years ago A.L. Fox reported that phenylthiocarbamide (PTC) tastes extremely bitter to some people (defined as "tasters") but not bitter at all to others ("non-tasters")[1,2]. Since then, numerous family, twin and population studies have shown that the inability to taste PTC is inherited in a nearly Mendelian recessive manner[3–5]. In more recent studies, the use of the chemically similar 6-n-propylthiouracil (PROP) has often been substituted for PTC because of its ability to generate similar taste responses in humans and its better known toxicity profile[6]. In 2003, a locus that explained approximately 75% of the variation in PTC sensitivity was identified on chromosome 7[4]. At this locus, variation in the *TAS2R38* bitter receptor gene was subsequently found to underlie all of the bimodal distribution of this phenotype and to explain >70% of the total phenotypic variance[7]. *TAS2R38* is a member of the TAS2R bitter taste receptor gene family, which in humans consists of 25 functional genes and 11 pseudogenes, many of which show signatures of natural selection[8–11]. Three single nucleotide polymorphisms (*rs714598*, *rs1726866*, *rs10246939*) at positions encoding amino acids 49, 262 and 296 represent the most common variant alleles of *TAS2R38*, and comprise the "taster" PAV (Proline, Alanine, Valine) and "non-taster" AVI (Alanine, Valine, Isoleucine) haplotypes. In addition, two rare (frequency <5%) (AAV and AAI) and two extremely rare (frequency <1%) (PAI and PVI) haplotypes have been identified. The other two possible haplotypes (AVV and PVV) have been individually reported in two studies[12,13] but not otherwise observed.

[1]National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD 20892, USA. [2]Department of Biological, Geological and Environmental Sciences BiGeA, Laboratory of Molecular Anthropology and Centre for Genome Biology, University of Bologna, via Selmi 3, 40126 Bologna, Italy. [3]Institute for Maternal and Child Health, IRCCS "Burlo Garofolo", University of Trieste, 34137 Trieste, Italy. [4]Division of Experimental Genetics, Sidra Medical and Research Center, Doha, Qatar. [5]Division of Biological Anthropology, University of Cambridge, CB2 1QH, Cambridge, UK. [6]University of Gastronomic Sciences, Piazza Vittorio Emanuele 9, Bra, Pollenzo 12042, CN, Italy. [7]Department of Biology, University of Pisa, Via Ghini 13, 56126 Pisa, Italy. [8]Biological, Chemical and Pharmaceutical Sciences and Technologies Department, STEBICEF, Università degli Studi di Palermo, V.le delle Scienze, Edificio 16, 90128 Palermo, Italy. [9]Health Sciences Research Institute, University of California at Merced, 5200 North Lake Road, Merced, CA 95343, USA. Correspondence and requests for materials should be addressed to D.S.R. (email: davide.risso@nih.gov)

Population	PAV	AVI	AAV	AVV	PAI	PVI	AAI	PVV
All	50.76%	42.70%	2.48%	0.32%	0.18%	0.07%	3.39%	0.10%
Africans	50.76%	35.18%	0.61%	0.08%	0.00%	0.15%	13.22%	0.00%
Asians	64.51%	35.31%	0.00%	0.17%	0.00%	0.00%	0.00%	0.00%
Europeans	45.66%	49.22%	3.56%	0.49%	0.32%	0.03%	0.55%	0.17%
Americans	68.61%	26.69%	2.26%	0.00%	0.00%	0.19%	2.26%	0.00%

Table 1. Detailed distributions of *TAS2R38* haplotypes in the studied populations.

A longstanding question has been the reason for the presence of two high-frequency haplotypes in worldwide populations. Because the perception of bitter taste is thought to protect us from the ingestion of toxic substances, how could the presumably non-functional AVI haplotype come to high frequency in populations worldwide? Analyses of the frequency distribution of *TAS2R38* haplotypes in different populations showed significant positive values for Tajima's D statistic, low F_{ST} values (0.001–0.05) and a deep coalescent time (TMRCA ~1 million years old) for this locus[9,14], providing evidence that balancing selection maintained both the taster and non-taster alleles at high frequency. However, a study of inter-specific variations of bitter taste receptor genes showed that TAS2R genes, and in particular *TAS2R38*, have undergone relaxation of selection in humans when compared with many other mammals[15].

Balancing selection hypotheses have suggested the possibility that the AVI non-taster allele encodes a fully functional receptor for another hypothetical bitter substance[8]. It has also been suggested that pathogens may have been the real targets of natural selection, since bitter receptors are expressed in the respiratory and enteric system[16,17] and one study suggested that common polymorphisms in the *TAS2R38* gene were linked to significant differences in the ability of the upper respiratory cells to clear and kill bacteria[18]. In addition, a recent study showed that *TAS2R38* genotypes regulate innate immune responses to oral bacteria[19].

Here we studied the distribution of *TAS2R38* haplotypes in a large number of human populations and available archaic hominid genomes to provide a fine-scale view of worldwide *TAS2R38* diversity, and we applied selection tests to evaluate the processes underlying the evolution of *TAS2R38* haplotypes.

Results

Population Genetics of *TAS2R38*. The analysis of the *TAS2R38* gene and surrounding regions (500 bp 5′ + 500 bp 3′) in the 1000 Genomes database showed that 96.07% of the variation in this locus was due to differences within populations, with little existing among continental sub-groups (4.1%) or populations (0.83%). However, geographically diverse populations showed differences at the nucleotide level, and African populations, in particular, carried more polymorphic sites when compared to populations of European, Asian and American ancestry. In addition, both gene and nucleotide diversity were higher in African populations (Supplementary Table S1). This is confirmed by the Wright's fixation indices calculated between all the analyzed populations, divided in continental groups ($F_{SC} = 0.01$, $F_{CT} = 0.04$ and $F_{ST} = 0.05$; all tests with $P < 0.001$), indicating a moderate to low level of differentiation at this locus between populations. The three *TAS2R38* SNPs (*rs714598*, *rs1726866* and *rs10246939*) showed approximately the same global F_{ST} values (0.05, 0.06 and 0.04 with P = 0.01, 0.01 and 0.002 respectively). However, when comparing these values to those obtained for SNPs of similar minor allele frequency across the genome, *TAS2R38* F_{ST} values were not outliers (Supplementary Figure S1), and the probability of observing a SNP with $F_{ST} < 0.05$ in this distribution is 0.34. Taking into account the presence of three SNPs in the same gene with high LD levels (average R^2 of 0.87), the probability of drawing three linked SNPs showing $F_{ST} < 0.05$ is 0.29 (0.34*0.87), not statistically lower than that of SNPs in the 5th percentile genome-wide.

We then calculated the global frequency of the *TAS2R38* haplotypes in three population datasets for which we had sequence data for the three common *TAS2R38* variants: 1000 Genomes (2N = 2722), our Italian populations (2N = 2878) and AGVP chromosomes (2N = 456) for which whole-genome sequences were available ($2N_{tot} = 6{,}056$). In the combined dataset, the PAV and AVI haplotypes were predominant (50.76% and 42.70%, respectively), followed by AAI (3.39%) and AAV (2.48%). Other haplotypes occurred at very low frequencies, AVV (0.32%), PAI (0.18%), PVV (0.10%) and PVI (0.07%) (Table 1). All the identified haplotypes were in Hardy-Weinberg equilibrium in each analyzed population ($P > 0.05$, data not shown).

The worldwide distribution of the PAV, AVI and AAI haplotypes calculated in our entire database ($2N_{tot} = 11{,}178$) is shown in Fig. 1. Other than a few exceptions with low frequencies in Central-South Italy (≤1%) and in the Near East (≤5%), the AAI haplotype is uniquely present in Africa, where it occurs at moderate-high frequencies (7–33%). It should be noted that, given the lack of data for *rs714598* in both the Silk Road and the HGDP datasets and of *rs10246939* in some individuals of the AGVP, the percentage of AAI could be slightly overestimated in these populations compared to that of AAV. However, the frequency of AAV in our African and Asian populations proved to be extremely low (0.61% and 0% respectively, Table 1), indicating that any overestimation effect is likely negligible.

Supplementary Figure S2 shows details of the distribution of AAI diplotypes in Africa. The PAV/AAI combination is the most common (ranging from 25 to 100%), with AAI homozygotes being present exclusively in the sub-tropical parts of Africa, where they displayed frequencies of 3–17%. In addition to AAI, the two major haplotypes (i.e. PAV and AVI) showed a different degree of distribution among continents, with AVI being less prevalent in both the African and American continental populations (Chi-square test, $P < 0.001$).

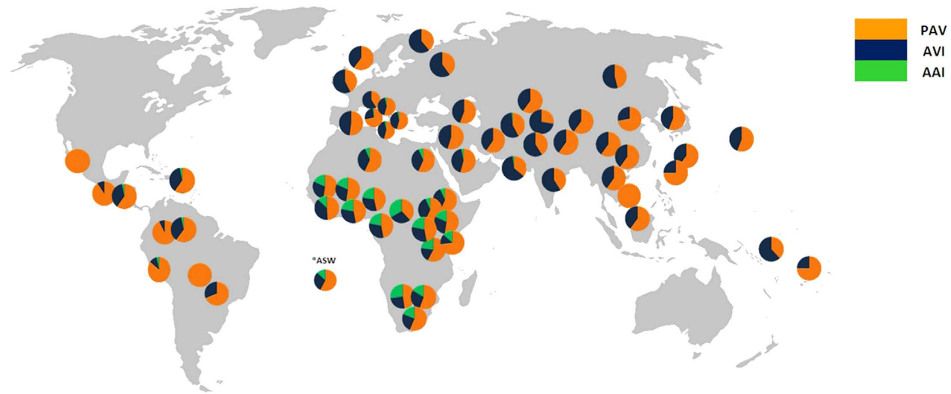


Figure 1. Worldwide distribution of *TAS2R38* PAV, AVI and AAI haplotypes in the studied populations. This map has been modified from its original version (https://commons.wikimedia.org/wiki/File:BlankMap-World-noborders.png).

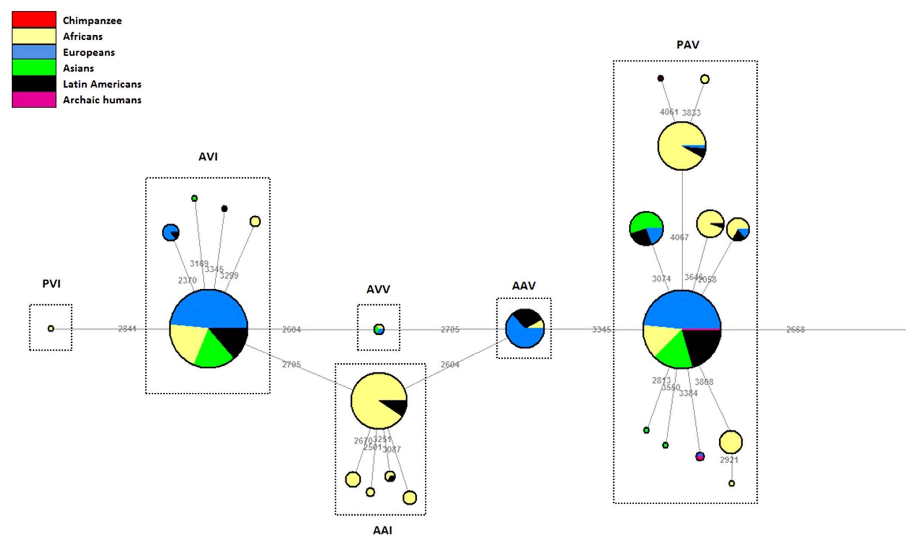


Figure 2. Neighbor-joining haplotype network illustrating the genealogical relationships between *TAS2R38* haplotypes in archaic hominids and modern African, European, Asian and Latin American populations. The position of the chimpanzee (outgroup) is highlighted in red.

Haplotype and Linkage Disequilibrium Analysis. The distribution of *TAS2R38* haplotypes, based on nucleotide data from the 1000 Genomes dataset, is represented in the median-joining network shown in Fig. 2. Six different clusters could be observed, with the PAV clade containing the highest number of different sub-haplotypes (N = 11). Our data indicate that the extremely rare (frequency $<1\%$) AVV haplotype occupies an intermediate position between the AVI and the rare (frequency $<5\%$) AAV haplotype. We also find that the archaic hominins (i.e. Neanderthal and Denisovan), lie in the PAV clade, suggesting an ancestral state of this haplotype.

Linkage Disequilibrium (LD) analyses performed on the *TAS2R38* gene and surrounding regions in the 1000 Genomes dataset, confirmed that *TAS2R38 rs714598*, *rs1726866* and *rs10246939* variants are in strong LD ($R^2 = 0.80-0.93$, $D' = 0.997-0.998$). In addition, a fourth SNP (*rs4726481*) situated in a nearby gene (*MGAM*, maltase-glucoamylase) was found to lie in the same haplotype block and in moderately strong LD ($R^2 = 0.63-0.81$, $D' = 0.992-0.994$) with the *TAS2R38* SNPs in all populations analyzed. In addition, Asian populations showed substantial LD values ($R^2 = 0.66-0.67$, $D' = 0.812-0.833$) with another SNP (*rs17162635*) in the same gene. Both the *MGAM* SNPs are located in intronic regions and display different distributions across continents, with rs4726481 having a derived allele frequency (DAF) approaching 0.50 in Africa, Europe, Asia, Latin America and *rs17162635* showing a DAF much higher in Asia (0.29) compared to Europe (0.13), Latin America (0.10) and Africa (0.04).

Testing neutrality and selection. We performed several tests to measure the deviation of the genetic differences at the *TAS2R38* locus from neutral expectations using the 1000 Genomes dataset. First, we computed Tajima's D statistics in different worldwide populations. The coding region of *TAS2R38* gene showed positive,

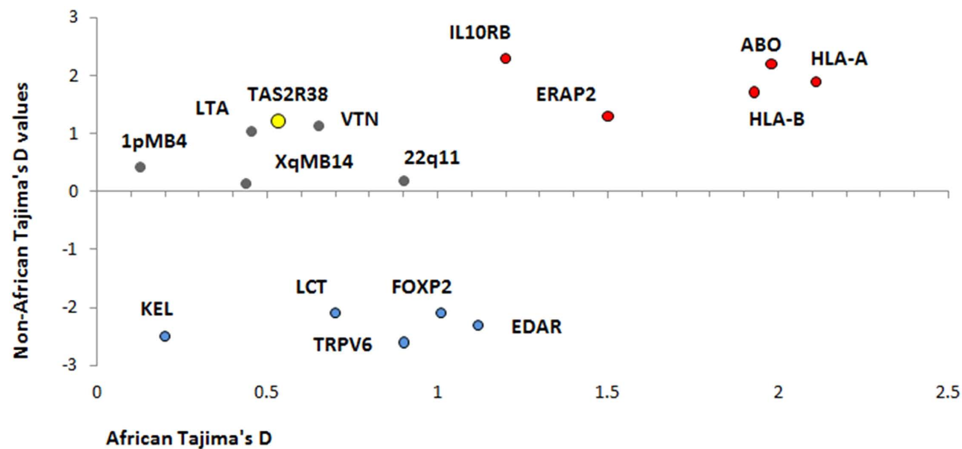


Figure 3. Comparison between Tajima's D values calculated in African and non-Africans for genes under different selective pressures. Genes under positive selection are represented with blue circles, genes under balancing selection with red circles and neutral regions with grey circles. The position of *TAS2R38* is highlighted in yellow.

although not significant (P > 0.05) values in all populations examined. Both the flanking 5′ and 3′ regions showed negative and not significant (P > 0.05) values. Moreover, when compared to Tajima's D values obtained across the genome for coding loci of similar size (1,143 +/− 500 bp) the *TAS2R38* values resided between the 5th and 95th percentiles (Supplementary Figure S3). Expanding the analyzed region to 10,000 bases around these selected loci showed similar results (data not shown). In addition, when compared to Tajima's D values calculated for genes for which a clear adaptive significance has been suggested, *TAS2R38* clustered with genes considered to be evolutionarily neutral (Fig. 3).

We further explored this using Li's MFDM, which has been shown to be more robust in the presence of confounding effects such as population size fluctuations and other demographic events. This analysis failed to detect any signs of deviation from neutral expectations in *TAS2R38* (P = 0.63). The same results were obtained when using the HKA test (P = 0.35). We also performed a Bayesian analysis in an effort to detect evidence of balancing selection. The calculated alpha values for *TAS2R38* detected no evidence of selection at *TAS2R38* (−0.5 < Alpha < 0).

Finally, Supplementary Table S2 shows the window containing *TAS2R38* did not differ from the average heterozygosity levels and did not fall below the 5th or above the 95th percentile

Simulations of haplotype evolution. We simulated the evolution of the two common *TAS2R38* haplotypes (PAV and AVI), the less common AAI, the rare (frequency < 5%) AAV and the rarest (frequency < 1%) AVV and PVI haplotypes, under several conditions (see Materials and Methods and Supplementary Information for more details). The only scenario that produced simulated haplotype frequencies similar to current frequencies involved an ancient (e.g. before Out of Africa) balancing selection (s = 0.001) acting on PAV/AVI individuals (Supplementary Figure S4). Under this scenario, the distribution of simulated global haplotype frequencies for the PAV, AVI, AAV, AVV and PVI (0.50 +/− 0.13, 0.47 +/− 0.14, 0.01 +/− 0.05, 0.003 +/− 0.005 and 0.0001 +/− 0.005 respectively) did not differ significantly (P = 0.39) from the distribution of the observed frequencies (0.51 +/− 0.13, 0.43 +/− 0.14, 0.02 +/− 0.07, 0.003 +/− 0.006 and 0.0001 +/− 0.0005 respectively), regardless of the degree of population expansion. However, the simulated frequencies obtained for AAI in African individuals did not match the empirical data (0.02 +/− 0.04 vs 0.13 +/− 0.05, P = 0.001), although the distribution of this haplotype outside Africa was consistent with the frequencies (P = 0.42) obtained from this model (0.02 +/− 0.03 vs 0.03 +/− 0.01). We therefore performed simulations involving different selecting forces acting on AAI in African individuals, both before and after the Out Of Africa (OOA) event. The only scenario that produced results similar to the observed AAI distribution in Africa (0.13 +/− 0.11 vs 0.13 +/− 0.05, P = 0.11) involved a weak (s = 0.001) directional selection force acting on this locus before the OOA event (Supplementary Figure S5).

Discussion

Although the presence of PTC or PROP has not been documented in nature, chemically similar compounds, known as glucosinolates, are found in common bitter foods, such as brussels sprouts, cabbage and broccoli[20]. For this reason, it has been suggested that a correlation may exist between PROP taster status and dietary intake[21], which could have important evolutionary consequences, a hypothesis supported by the results of a number of previous studies[22–24]. Other studies have identified relationships between the intensity of PTC/PROP bitterness and the perception of several natural compounds[25–28]. These correlations generally support the hypothesis that natural selection has been acting on *TAS2R38*, maintaining the AVI haplotype at roughly the same worldwide frequency as the PAV haplotype. In addition to this hypothesis, it has been suggested that the non-taster AVI haplotype may be a functional receptor for another bitter compound[8] and that pathogens may have been the real targets of natural selection[17,18].

Using the largest sample available, we have confirmed the global predominance of the PAV and AVI haplotypes of *TAS2R38*. We also found that the AVI form is less common in Africa and that the AAI haplotype is primarily present in this continent, as previously reported[9]. The three common *TAS2R38* SNPs (i.e. *rs714598*, *rs1726866* and *rs10246939*) showed low F_{ST} values, indicating little genetic differentiation at these sites. This was previously interpreted[9] as a footprint of balancing selection that maintained similar frequencies for the alternative alleles at these sites. However in contrast to previous studies[8,9], we failed to detect any recent departures from neutral expectations for the variation at the *TAS2R38* locus and surrounding regions. Although we found positive Tajima's D values in the *TAS2R38* coding region, these did not reach significance and where similar to those observed at evolutionary neutral loci. When correcting for population stratification and taking into account demographic events with Li's MFDM test, *TAS2R38* did not show significant P-values, suggesting that these variables may contribute to the observed positive Tajima's D values. Another study, although based on a much smaller number of individuals (N = 22), previously pointed out this possibility[15]. Other approaches that have previously successfully detected signs of balancing natural selection, including the HKA and Bayescan tests, failed to find any departures from neutrality in our data set. We also note that two other studies applying new methods to detect signatures of balancing selection failed to identify such signatures in *TAS2R38*[29,30].

Here we propose a modified hypothesis for the evolution of *TAS2R38* haplotypes in humans. We suggest ancient balancing selection acting during the early stages of hominin evolution, before the Out-Of-Africa event, that maintained both PAV and AVI alleles at roughly the same frequency. We speculate that both haplotypes were in fact important for detecting potentially toxic bitter compounds found uniquely in the African continent.

Since *TAS2R38* shows no signs of recent departure from neutral expectations, we hypothesize that the modern frequency distribution of the non-taster AVI allele in non-Africans is largely due to recent demographic and population stratification events. This is also supported by our simulations, which confirmed that the existing high frequencies of the PAV and AVI haplotypes outside Africa could have arisen by a series of bottlenecks and population expansion with a relaxation of the selective forces acting on this gene. It has been shown that bottleneck episodes after the out-of-Africa event have strongly contributed to a modification of the genetic structure and of the selective sweeps acting on different human populations[31,32].

The AAI haplotype seems to have a more complex evolutionary history. Its distribution is unique in Africa, aside from a few exceptions in Latin America presumably due to the African admixture in these populations[33] and in the Near East and southern Italy, regions with strong flow from Africa in historical times[34,35]. From our simulations we hypothesize that the AAI haplotype has undergone weak but directional selection, in addition to the balancing selection that previously acted on PAV/AVI. This view is strengthened by a recent work studying a large collection of African populations, where the authors found that African *TAS2R38* haplotypes evolved under a more complex scenario that includes a combination of balancing and directional selective pressures[9].

Finally, the high LD values between *TAS2R38* and *MGAM* variants suggest that this gene may also have had an effect on the current distribution of the PAV and AVI haplotypes. This gene is involved in pathways and interacts with genes involved in starch and sucrose metabolism, carbohydrate digestion and absorption and galactose and lactose, suggesting a potentially important role for human nutrition.

Materials and Methods

Sources of samples and genetic data. Archaic hominids studied in this project consisted of the Altai Neanderthal individual[36] and the Denisovan hominin from the same site[37]. The modern human subjects consisted of both published databases and newly recruited samples. For the former, we included 1,361 individuals belonging to 17 worldwide populations from the 1000 Genomes Project Phase 3[38], 942 individuals from 53 different populations from the Human Genome Diversity Project (HGDP)[39] and 1,428 African individuals from 15 populations of the African Genome Variation Project (AGVP)[40,41]. In addition, 419 individuals belonging to 20 communities living along the Silk Road in Caucasus and Central Asia[42,43] and 1,439 adult healthy subjects from 9 different regions within Italy (this study) were included in the analysis. Individuals from different databases but belonging to the same populations were pooled together. A total of 11,178 chromosomes from 105 different worldwide populations were analyzed (Supplementary Table S3). All samples were anonymized, coded identifiers were assigned to them and each donor was given an information sheet. All experimental protocols were approved by the ethical committee of IRCCS-Burlo Garofolo Hospital, by the bioethics committee of Pisa University and by the National Institutes of Health (NIDCD protocol 01-DC-0230) combined Neurosciences/Blue Panel Institutional Review Board. Written informed consent was obtained from all subjects included in the study. All experiments and methods were performed in accordance with the approved guidelines and regulations.

DNA collection, extraction and genotyping. Samples of saliva from subjects of the Silk Road populations were collected using the Oragene kits and extracted according to the manufacturer's protocol (Genotek Inc., Kanata, Ontario, Canada). Genotyping was carried out using the Illumina 700 K high density SNP array. Samples of saliva from Italians were collected using buccal swabs rinsed in a 0.9% NaCl solution, using specific purification kits (Invisorb Spin Swab Kit and Genomed; Celbio, Milan, Italy, respectively). Genotyping of the known variant sites (*rs714598*, *rs1726866* and *rs10246939*) was carried out using the KASPar SNP Genotyping Assay (Kbioscience, Heddesdon, UK) and the reads were performed with the ABI PRISM 7900HT instrument (Applied Biosystems).

Data analysis. Three *TAS2R38* SNPs (*rs714598*, *rs1726866*, *rs10246939*) were analyzed in the HGDP, AGVP, Silk Road and Italian samples. The program PHASE[44] was used to infer *TAS2R38* haplotypes in the Italian and Silk Road subjects. Individuals from the 1000 Genomes Project Phase 1 were used as a reference to impute *rs714598* in both the Silk Road and HGDP populations and *rs10246939* in some individuals of the AGVP population (N = 1,200), since these SNPs were not present in the genotyping arrays used to type these

individuals. Only haplotypes with posterior probability of 0.9 or above were considered. VCF files containing sequence data at the *TAS2R38* locus and its 500 bp upstream and downstream regions (7:141972131-141974273, GRCh38) were retrieved from both the 1000 Genomes Phase 3 Database (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/) and from the 50 and 30-fold coverage Neanderthal and Denisovan genomes aligned to the hg19 human reference sequence, after obtaining the genomic coordinates as appropriate (http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/VCF/ and http://cdna.eva.mpg.de/denisova/VCF/hg19_1000g/) using VCFtools[45].

The software Arlequin v.3.5[46] was used to calculate basic population genetics statistics, such as nucleotide diversity, number of polymorphic sites and estimated heterozygosity, and to perform the Analysis of Molecular Variance (AMOVA). Allele frequencies between populations were compared performing Chi-square tests, and the Bonferroni correction was used to correct the nominal P-values (i.e. adjusted P = P value x number of individual tests), with the software PLINK v 1.07[47]. This software was also used to calculate pairwise LD measures and to estimate haplotype blocks which were also confirmed with Haploview[48]. For this analysis, we expanded the analyzed region to 50,000 base pairs in both directions. The haplotype network was constructed with Network 4.5[49] using a median-joining algorithm.

Testing natural selection. Departures from neutrality were tested using several approaches: we calculated Tajima's D values using the DNASP package[50] in *TAS2R38* and surrounding regions. In addition, we compared these values to the ones calculated across the genome for coding loci with sizes similar to *TAS2R38* (e.g. 1,143 bases) and to those calculated in candidate genes selected from literature known to have undergone some kind of selective pressure. Since it has been shown that Tajima's test may not be powerful enough to detect departures from neutrality in regions shorter than 5,000 bases[51], Tajima's D values were also calculated in a 10,000 bases region encompassing the selected loci. Pairwise F_{ST} values for all pairs of populations and within the same sub-population for *TAS2R38* and surrounding regions were calculated with the software Arlequin v.3.5[46] and the significance of these statistics was tested using a coalescent simulation adapted from Hudson[52]. In addition, we calculated F_{ST} values for the three *TAS2R38* common SNPs. These values were then compared to the ones calculated across the genome for SNPs of similar frequency (e.g. MAF ranging from 0.42 to 0.47) and to the genome-wide F_{ST} distribution. In order to further explore the signatures of natural selection at *TAS2R38*, we performed Li's MFDM test[53], which has been shown to be very robust in distinguishing selection from demography, even in presence of balancing selection. In addition, the Bayesian regression method implemented in Bayescan v2.1[54] and the HKA test[55] were applied to our dataset. Finally, simulations of the evolution *TAS2R38* haplotypes outside Africa were performed with the python library simuPOP v1.1.4[56]. We simulated several scenarios for the six *TAS2R38* haplotypes with increasing level of expansions (i.e. 10 fold, 20 fold and 40 fold) and different natural selection pressures (i.e. s = 0.05, s = 0.01 and s = 0.001). Each scenario was replicated 1000 times to generate a prediction of the resulting haplotype frequencies. Finally, sliding-window analyses (100 kb) of heterozygosity were calculated on the entire chromosome 7 in the the 1000 Genome Project populations. More details on the analytical methods can be found in the Supplementary Information.

References

1. Anonymous. Six in ten "tasteblind" to bitter chemical. *Sci. News Lett*. **9,** 249 (1931).
2. Fox, A. L. The Relationship between chemical constitution and taste. *Proc Natl Acad Sci USA* **18,** 115–120 (1932).
3. Blakeslee, A. F. Genetics of sensory thresholds: taste for phenyl thio carbamide. *Proc Natl Acad Sci USA* **18,** 120–130 (1932).
4. Drayna, D. *et al.* Genetic analysis of a complex trait in the Utah Genetic Reference Project: a major locus for PTC taste ability on chromosome 7q and a secondary locus on chromosome 16p. *Hum Genet*. **112,** 567–72 (2003).
5. Knaapila, A. *et al.* Genetic analysis of chemosensory traits in human twins. *Chem Senses*. **37,** 869–81.
6. Fischer, R. Gustatory behavioral, and pharmacological manifestations of chemoreception in man. In Ohloff, G. & Thomas, A. F. (Ed.), *Gustation and Olfaction*. New York: Academic Press. pp. 187–237 (1971).
7. Kim, U. K. *et al.* Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science*. **299,** 1221–5 (2003).
8. Wooding, S. *et al.* Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. *Am J Hum Genet*. **74,** 637–46 (2004).
9. Campbell, M. C. *et al.* Evolution of functionally diverse alleles associated with PTC bitter taste sensitivity in Africa. *Mol Biol Evol*. **29,** 1141–53 (2012).
10. Risso, D., Tofanelli, S., Morini, G., Luiselli, D. & Drayna, D. Genetic variation in taste receptor pseudogenes provides evidence for a dynamic role in human evolution. *BMC Evol Biol*. **13,** 14: 198 (2014).
11. Dong, D., Jones, G. & Zhang, S. Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evol Biol*. **15,** 9: 12 (2009).
12. Pemberton, T. J. Prevalence of common disease-associated variants in Asian Indians. *BMC Genet*. **4,** 9: 13 (2008).
13. Carrai, M. *et al.* Association between *TAS2R38* gene polymorphisms and colorectal cancer risk: a case-control study in two independent populations of Caucasian origin. *Plos One*. **6,** e20464 (2011).
14. Wooding, S. *et al.* Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature*. **440,** 930–4 (2006).
15. Wang, X., Thomas, S. D. & Zhang, J. Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum Mol Genet*. **13,** 2671–8 (2004).
16. Wu, S. V. *et al.* Expression of bitter taste receptors of the T2R family in the gastrointestinal tract and enteroendocrine STC-1 cells. *Proc Natl Acad Sci USA* **99,** 2392–2397 (2002).
17. Lee, R. J. & Cohen, N. A. The emerging role of the bitter taste receptor T2R38 in upper respiratory infection and chronic rhinosinusitis. *Am J Rhinol Allergy*. **27,** 283–6 (2012).
18. Lee, R. J. *et al.* T2R38 taste receptor polymorphisms underlie susceptibility to upper respiratory infection. *J Clin Invest*. **122,** 4145–59 (2012).
19. Gil, S. *et al.* Genotype-specific regulation of oral innate immunity by T2R38 taste receptor. *Mol Immunol*. **68,** 663–670 (2015).
20. Carlson, D., Daxenbichler, M. E. & VanEtten, C. H. Glucosinolates in crucifer vegetables: broccoli, Brussel sprouts, cauliflower, collards, kale, mustard greens, and Kohlrabi. *J. Amer. Soc. Sci*. **112,** 173–178 (1987).
21. Tepper, B. J. Nutritional implications of genetic taste variation: the role of PROP sensitivity and other taste phenotypes. *Annu Rev Nutr*. **28,** 367–88 (2008).

22. Turnbull, B. & Matisoo-Smith, E. Taste sensitivity to 6-n-propylthiouracil predicts acceptance of bitter-tasting spinach in 3-6-y-old children. *Am J Clin Nutr*. **76,** 1101–5 (2002).
23. Sandell, M. A. & Breslin, P. A. Variability in a taste-receptor gene determines whether we taste toxins in food. *Curr Biol*. **16,** R792-4, 9 (2006).
24. Wooding, S. *et al.* Genetics and bitter taste responses to goitrin, a plant toxin found in vegetables. *Chem Senses*. **35,** 685–92 (2010).
25. Risso, D. S., Howard, L., VanWaes, C. & Drayna, D. A potential trigger for pine mouth: a case of a homozygous phenylthiocarbamide taster. *Nutr Res*. **35,** 1122–5 (2015).
26. Hayes, J. E., Bartoshuk, L. M., Kidd, J. R. & Duffy, V. B. Supertasting and PROP bitterness depends on more than the *TAS2R38* gene. *Chem Senses*. **33,** 255–265 (2008).
27. Allen, A. L., McGeary, J. E., Hayes, J. & Rebaudioside, A. and Rebaudioside D bitterness do not covary with Acesulfame-K bitterness or polymorphisms in TAS2R9 and TAS2R31. *Chemosens Percept*. **6,** 109–117 (2013).
28. Risso, D. *et al.* Genetic signature of differential sensitivity to stevioside in the Italian population. *Genes Nutr*. **9,** 401 (2014).
29. DeGiorgio, M., Lohmueller, K. E. & Nielsen, R. A model-based approach for identifying signatures of ancient balancing selection in genetic data. *Plos Genet*. **10,** e1004561 (2014).
30. Andrés, A. M. *et al.* Targets of balancing selection in the human genome. *Mol Biol Evol*. **26,** 2755–64 (2009).
31. Amos, W. & Hoffman, J. I. Evidence that two main bottleneck events shaped modern human genetic diversity. *Proc Biol Sci*. **277,** 131–7 (2010).
32. Messer, P. W. & Petrov, D. A. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol*. **28,** 659–69 (2013).
33. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature*. **491,** 56–65 (2012).
34. Moorjani, P. *et al.* The history of African gene flow into Southern Europeans, Levantines, and Jews. *Plos Genet*. **7,** e1001373 (2011).
35. Botigué, L. R. *et al.* Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci USA* **110,** 11791–6 (2013).
36. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. **505,** 43–9 (2014).
37. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. **468,** 1053–60 (2010).
38. 1000 Genomes Project, Consortium et al. A global reference for human genetic variation. *Nature*. **526,** 68–74 (2015).
39. Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res*. **19,** 826–37 (2009).
40. Gurdasani, D. *et al.* The African Genome Variation Project shapes medical genetics in Africa. *Nature*. **517,** 327–32 (2015).
41. Pagani, L. *et al.* Tracing the route of modern humans out of Africa using 225 human genome sequences from Ethiopians and Egyptians. *Am J Hum Genet*. **96,** 986–91 (2015).
42. Robino, A. *et al.* A population-based approach to study the impact of PROP perception on food liking in populations along the Silk Road. *Plos One*. **9,** e91716 (2014).
43. Mezzavilla, M. *et al.* Genetic landscape of populations along the Silk Road: admixture and migration patterns. *BMC genetics*. **15,** 131 (2014).
44. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. **68,** 978–89 (2011).
45. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics*. **27,** 2156–8 (2011).
46. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. **10,** 564–7 (2010).
47. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. **81,** 559–75 (2007).
48. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. **21,** 263–265 (2005).
49. Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. **16,** 37–48 (1999).
50. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. **25,** 1451–2 (2009).
51. Hu, M. *et al.* Exploration of signals of positive selection derived from genotype-based human genome scans using re-sequencing data. *Hum Genet*. **131,** 665–74 (2012).
52. Hudson, R. Gene genealogies and the coalescent process. p. 1-44 in *Oxford Surveys in Evolutionary Biology*, edited by Futuyama & J. D. Antonovics. Oxford University Press, New York (1990).
53. Li, H. A new test for detecting recent positive selection that is free from the confounding impacts of demography. *Mol Biol Evol*. **28,** 365–375 (2011).
54. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*. **180,** 977–93 (2008).
55. Hey, J. Heylab Software, 2010, Date of access: 15/10/2015, https://bio.cst.temple.edu/~hey/software/software.htm#HKA.
56. Peng, B. & Kimmel, M. simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*. **18,** 3686–7 (2005).

Acknowledgements

DR and DD were supported by the National Institute on Deafness and Other Communication Disorders Intramural grant number Z1A-000046-14. LP was funded by the ERC Grant FP7 – 261213. MM, AR and PG were funded by the sponsors of the scientific expedition Marcopolo 2010, by Ministry of Health (IRCCS-Burlo Garofolo - Ricerca Corrente) and by Italian Ministry of Health RC 35/09. DL was supported by the PRIN 2010EL8TXP_006 grant and, by the European Research Council [ERC-2011-AdG 295733 (Langelin). ST, DC and RB were supported by a University of Pisa PRA 2016 Grant.

Author Contributions

D.S.R. conceived the study, performed the population genetics analyses and wrote the first draft of the paper. M.M. and A.R. performed the simulations of haplotype evolution and contributed to the manuscript drafting. L.P. helped in performing the experiments and drafting the manuscript. G.M., S.T., D.L., D.D. and S.W. helped in designing the study and contributed to the statistical analyses. M.C., D.C., R.B., F.C. and P.G. collected the samples and helped in performing the experiments. All authors have read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Risso, D. S. *et al.* Global diversity in the *TAS2R38* bitter taste receptor: revisiting a classic evolutionary PROPosal. *Sci. Rep.* **6**, 25506; doi: 10.1038/srep25506 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

SCIENTIFIC REPORTS

OPEN

Corrigendum: Global diversity in the *TAS2R38* bitter taste receptor: revisiting a classic evolutionary PROPosal

Davide S. Risso, Massimo Mezzavilla, Luca Pagani, Antonietta Robino, Gabriella Morini, Sergio Tofanelli, Maura Carrai, Daniele Campa, Roberto Barale, Fabio Caradonna, Paolo Gasparini, Donata Luiselli, Stephen Wooding & Dennis Drayna

Scientific Reports **6**:25506; doi: 10.1038/srep25506; published online 03 May 2016; updated 27 June 2016

In this Article, all instances of "*rs714598*" should read "*rs713598*".

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

PLOS | ONE

RESEARCH ARTICLE

Genetic Variation in the *TAS2R38* Bitter Taste Receptor and Smoking Behaviors

Davide S. Risso[1,2], Julia Kozlitina[3], Eduardo Sainz[1], Joanne Gutierrez[1], Stephen Wooding[4], Betelihem Getachew[5], Donata Luiselli[2], Carla J. Berg[5], Dennis Drayna[1]*

1 Laboratory of Communication Disorders, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Laboratory of Molecular Anthropology and Centre for Genome Biology, Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy, **3** McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **4** Health Sciences Research Institute, University of California Merced, Merced, United States of America, **5** Department of Behavioral Sciences and Health Education, Emory University Woodruff Health Sciences Center, Atlanta, GA, United States of America

* drayna@nidcd.nih.gov

OPEN ACCESS

Citation: Risso DS, Kozlitina J, Sainz E, Gutierrez J, Wooding S, Getachew B, et al. (2016) Genetic Variation in the *TAS2R38* Bitter Taste Receptor and Smoking Behaviors. PLoS ONE 11(10): e0164157. doi:10.1371/journal.pone.0164157

Editor: Sergio Tofanelli, Universita degli Studi di Pisa, ITALY

Received: July 15, 2016

Accepted: September 20, 2016

Published: October 6, 2016

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was supported by the National Institute on Deafness and Other Communication Disorders Intramural grant number Z1A-000046-16 to DD, by the National Institutes of Health grant (HHSN263201300011C-03; OLAO) to JK, and by National Cancer Institute grant number 1R01CA179422-01 to CB, as well as a Georgia Cancer Coalition grant to CB. The Dallas Heart Study is supported by the National Center for Advancing Translational Sciences award number

Abstract

Common *TAS2R38* taste receptor gene variants specify the ability to taste phenylthiocarbamide (PTC), 6-n-propylthiouracil (PROP) and structurally related compounds. Tobacco smoke contains a complex mixture of chemical substances of varying structure and functionality, some of which activate different taste receptors. Accordingly, it has been suggested that non-taster individuals may be more likely to smoke because of their inability to taste bitter compounds present in tobacco smoke, but results to date have been conflicting. We studied three cohorts: 237 European-Americans from the state of Georgia, 1,353 European-Americans and 2,363 African-Americans from the Dallas Heart Study (DHS), and 4,973 African-Americans from the Dallas Biobank. Tobacco use data was collected and *TAS2R38* polymorphisms were genotyped for all participants, and PTC taste sensitivity was assessed in the Georgia population. In the Georgia group, PTC tasters were less common among those who smoke: 71.5% of smokers were PTC tasters while 82.5% of non-smokers were PTC tasters (P = 0.03). The frequency of the *TAS2R38* PAV taster haplotype showed a trend toward being lower in smokers (38.4%) than in non-smokers (43.1%), although this was not statistically significant (P = 0.31). In the DHS European-Americans, the taster haplotype was less common in smokers (37.0% vs. 44.0% in non-smokers, P = 0.003), and conversely the frequency of the non-taster haplotype was more common in smokers (58.7% vs. 51.5% in non-smokers, P = 0.002). No difference in the frequency of these haplotypes was observed in African Americans in either the Dallas Heart Study or the Dallas Biobank. We conclude that *TAS2R38* haplotypes are associated with smoking status in European-Americans but not in African-American populations. PTC taster status may play a role in protecting individuals from cigarette smoking in specific populations.

UL1TR001105, and the Dallas Biobank was funded by the National Institutes of Health. Research reported in this publication was supported by the NIH intramural research program and the FDA through funds obtained under the Family Smoking Prevention and Tobacco Control Act. The content was not reviewed by the Food and Drug Administration, but underwent the standard manuscript clearance process for scientific papers published from the NIH intramural research program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Tobacco smoking is a major worldwide health problem and is a leading cause of preventable disease [1–2].

Cigarettes and other tobacco products contain bitter compounds including nicotine, which contribute to the chemosensory properties of tobacco [3] and stimulate multiple sensory systems, including taste transduction pathways [4]. Since bitter taste has evolved to identify potentially toxic compounds [5], and thus protect against harmful foods, aversion to this taste may prevent smoking and nicotine dependence [6].

Receptors for human bitter taste are encoded by the *TAS2R* gene family which comprises 25 functional genes [7] and 11 pseudogenes [8] that have been subject to evolutionary forces [9–10–11–12]. The most studied gene in this family is *TAS2R38*, which encodes a receptor that mediates the ability to taste the bitter compounds phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP) [13–14]. Two common forms of this gene exist worldwide, defined by amino acids at positions 49, 262 and 296 that constitute the PAV (Proline, Alanine, Valine, "PTC taster") and AVI (Alanine, Valine, Isoleucine, "PTC non-taster") haplotypes.

TAS2R38 haplotypes have been hypothesized to influence smoking habits and nicotine dependence, since it has been shown that this gene has a lower expression in smokers, when compared to non-smoker individuals [15]. However, the results of previous studies have been conflicting. For example, a study examining both African-American (AA) and European-American (EA) individuals found a significant association between *TAS2R38* haplotypes and smoking, with the non-taster AVI haplotype being positively associated with smoking quantity and nicotine dependence. This was seen only in AA [16]. Another study analyzed German participants and found that individuals carrying the PAV taster haplotype smoked significantly fewer cigarettes per day [17]. In contrast, another study of individuals of European descent found no association between the PAV or AVI haplotypes and smoking. Moreover, this study found that the rare AAV haplotype was associated with a lower incidence of smoking [18]. In addition, a recent study investigated the relationship between *TAS2R38* haplotypes and menthol cigarette smoking and found that the PAV haplotype was associated with menthol cigarette use in pregnant female Caucasian smokers [19].

These mixed findings motivated the current study, which examined the association between *TAS2R38* PAV, AVI and rarer haplotypes and cigarette smoking in a larger number of individuals from three independent cohorts of both EA and AA individuals.

Materials and Methods

Research Participants

Georgia population. A total of 237 EA were chosen based on their tobacco product usage from a longitudinal study involving young adults attending seven Georgia colleges or universities [20]. Variables including sex, age and current smoking status were obtained from all participants during the web-based baseline survey in the fall of 2014; smoking status was also obtained again in Spring 2015. Individuals were defined as current smokers if they reported to have smoked in the past 30 days, as previously described [20–21]. In the spring of 2015, participants were sent an Oragene kit and a commercial taste-strip containing PTC (Thermo Fisher Scientific Inc., Catalog Number: S85287A) via mail with instructions regarding how to complete saliva provision and the taste-strip test; the responses for the latter were recorded by participants and sent back with the Oragene kit. Participants were defined as tasters if they categorized the taste of the PTC papers-strip to be "mild or strong" and as non-tasters if they reported "no taste".

Dallas Heart Study. The Dallas Heart Study is a multiethnic population-based probability sample of Dallas County, Texas residents. The study design and recruitment procedures have been previously described in detail [22]. The original cohort was enrolled between 2000 and 2002, and all participants, as well as their spouses or significant others, were invited for a repeat evaluation in 2007–2009 (DHS-2). During each visit, participants completed a detailed survey including questions regarding demographics, socioeconomic status, medical history, and lifestyle factors (including tobacco use), and underwent a health examination. Ethnicity was self-assigned. A total of 2,363 AA and 1,353 EA DHS participants with available genotype and smoking phenotype data were included in the present study. Current smokers were defined as individuals who smoked at least 100 cigarettes in their lifetime and smoked on at least some days in the previous 30 days. Smoking quantity was defined as a categorical variable in all cohorts, sub-dividing smokers in three groups (less than 6 cigarettes per day, 6–19 cigarettes per day and 20 or more cigarettes per day).

Dallas Biobank. The Dallas Biobank is a repository of DNA and plasma samples from individuals ascertained at various locations in north-central Texas. The present study includes a total of 4,973 AA Biobank participants for whom the genotype and smoking phenotype data were available. Current smokers were defined as people who identified themselves as smokers and said they were currently using tobacco products.

Ethical Statement

All participants were over 18 years of age and were enrolled with written informed consent. For the Georgia population, the study was approved by the Institutional Review Boards of Emory University, ICF Macro International, Albany State University, Berry College, University of North Georgia, and Valdosta State University. For the DHS and Biobank populations, the study was approved by the Institutional Review Board of the University of Texas Southwestern Medical Center.

DNA Collection, Purification, and Sequencing

DNA from the Georgia population was collected using Oragene saliva collection kits and extracted according to the manufacturer's protocol (Genotek Inc., Kanata, Ontario, Canada). The single coding exon of the *TAS2R38* gene was completely sequenced using dideoxy Sanger sequencing [23]. A dedicated set of primers modified from Kim et al., 2003, was adopted (Table A in S1 File) as previously published [24]. DNA chromatograms were analyzed and checked individually in order to evaluate the presence of calling errors with the Lasergene suite (DNASTAR, Madison, Wisconsin) [25]. In the DHS and the Dallas Biobank, genomic DNA was extracted from circulating leukocytes. A total of 4,597 DHS and 4,973 Biobank participants were previously genotyped using the Illumina Human-Exome BeadChip, which assayed the *TAS2R38 rs713598*, *rs1726866*, and *rs10246939* variants, residing in the codons for amino acid positions 49, 262, and 296 within the *TAS2R38* coding sequence. Genetic ancestry was estimated using EIGENSTRAT software [26].

Statistical Analyses

Statistical analysis was performed using the R statistical analysis software [27]. Baseline characteristics of the study participants were compared using t-tests for continuous variables and chi-square tests for categorical variables. We used PLINK [28] to perform an initial quality control of genotypes and excluded variants with a call rate $<90\%$ or a deviation from Hardy-Weinberg equilibrium (HWE) ($P<0.001$). PHASE [29] was used to statistically infer *TAS2R38* haplotypes, using individuals from the 1000 Genomes Phase 1 [30] as a reference. Only haplotypes

with posterior probability of 0.9 or above were considered for further analyses. Differences in *TAS2R38* haplotype distributions between smokers and non-smokers were explored using logistic regression in PLINK, with adjustments for demographic variables such as age, sex, the leading principal components of ancestry and the study indicator. An additive model was assumed for the effect of haplotypes. The significance levels of the association tests were adjusted using the Bonferroni correction in the Georgia cohort (adjusted P = P value X number of individual tests) and $P<0.05$ was considered statistically significant. For replication of the results in the DHS and Biobank populations, we reported the nominal p-values.

Results

Subject Cohort Demographics and Smoking Behaviors

Baseline and demographic characteristics of our study populations, stratified by cohort, are shown in Table 1. The average age of the individuals of the Georgia cohort was 20.91 +/- 1.95. Of the 237 participants, 123 (51.9%) were current smokers and the remaining 114 (48.1%) were non-smokers. No differences were found in the mean age of smokers (20.6) and non-smokers (21.2; P = 0.85). A higher, but not significant, percentage of smokers than non-smokers were female (54.4% of smokers versus 47.4% of non-smokers, P = 0.72). PTC sensitivity showed the classical bimodal distribution among participants, with 182 (76.8%) individuals classified as tasters and 55 (23.2%) as non-tasters. No significant age (P = 0.61) or gender (P = 0.34) differences were observed between PTC-taster and non-tasters.

The Dallas Heart Study population was significantly older than the Georgia cohort (mean age 48.2 and 50.1 years in AA and EA participants respectively, $P<0.05$). The proportion of female participants was slightly higher among DHS AA than EA participants (59.7% and 53.5%, respectively, $P<0.05$). DHS EA participants had a lower prevalence of smoking (23.4%) than those of either the Georgia cohort (51.9%, $p<0.05$) or the DHS AA participants (30.6%, $P<0.05$). Among DHS EA participants, smokers were on average 5 years younger than non-smokers (mean age 46.4 vs. 51.2 years, respectively, $P<0.001$, Table A in S1 File). There was no difference in age between AA smokers and non-smokers. In contrast to the Georgia cohort, we found a higher proportion of women among non-smokers in both ethnicities in DHS (64.2% vs. 49.4% in AA, $P<0.001$; 54.6% vs. 50.0% in EA, $P>0.05$).

The Dallas Biobank population was older than the Georgia population (mean age 44.8, $P<0.001$) but younger than both the DHS AA and EA participants ($P<0.05$). In addition, the

Table 1. Characteristics of the study participants in the three different cohorts. DHS, Dallas Heart Study; AA, African-Americans; EA, European-Americans.

Characteristic	Georgia—EA	DHS—AA	DHS—EA	Biobank—AA
Number of participants	237	2363	1353	4973
Age, mean (SD)	20.9 (1.9)	48.2 (11.3)	50.1 (11.2)	44.8 (14.6)
Female, N (%)	121 (51.0%)	1410 (59.7%)	724 (53.5%)	3238 (65.1%)
Smokers, N (%)	123 (51.9%)	723 (30.6%)	316 (23.4%)	1526 (30.7%)
Smoking quantity*, N (%)				
≤5 cigs/day	91 (74.0)	211 (29.5)	47 (15)	715 (53.7)
6–19 cigs/day	28 (22.7)	318 (44.5)	116 (37.1)	554 (41.6)
≥20 cigs/day	4 (3.3)	186 (26)	150 (47.9)	63 (4.7)
n/r	/	8	3	194

*Smoking quantity was not available for some participants.

n/r–non-response.

doi:10.1371/journal.pone.0164157.t001

Biobank included a higher proportion of females than either the Georgia or DHS AA/EA participants (65.1%, P<0.05). The fraction of individuals who were smokers was lower in this population compared to the Georgia cohort (30.7% vs. 51.9%, P<0.001) but similar to the DHS AA participants (P>0.05). In the Dallas Biobank population, smokers were slightly younger than non-smokers (mean age 43.8 vs. 45.3, P = 0.001). In addition, smokers had a lower percentage of females than non-smokers (51.7% vs. 71.0%, P<0.001, Table B in S1 File).

Associations between *TAS2R38* haplotypes, PTC Tasting-Status and Smoking Behaviors

Haplotype phasing of the genomic DNA sequence data revealed five major haplotypes (AVI, PAV, AAV, AAI, and PVI), with frequencies of 53.4%, 41.6%, 4.4%, 0.4% and 0.2%, respectively, in the Georgia sample. In the DHS EA population, the frequencies of the three major haplotypes (PAV, AVI, and AAV) were similar to those in Georgia (53.2%, 42.5%, and 4.1%, respectively). The DHS AA carried a greater number of haplotypes: PAV (47.0%), AVI (32.6%), AAI (19.1%), AAV (1.1%) and PVI (0.3%). Similarly, the Dallas Biobank population showed higher frequencies of rare haplotypes: PAV (47.6%), AVI (32.3%), AAI (19.0%), AAV (1.0%), and PVI (0.1%). The distribution of *TAS2R38* haplotypes did not show any statistical differences between Biobank and DHS AA individuals and between DHS EA and Georgia EA individuals (Table B in S1 File).

As expected, the frequency of *TAS2R38* haplotypes and diplotypes differed between PTC-tasters and non-tasters in the Georgia cohort, where PAV was the predominant haplotype in PTC-tasters (95.3%) and rarely present in PTC non-tasters (4.7%) (P<0.001). Most of the PAV/PAV homozygotes in this cohort were PTC-tasters (98.1%) as opposed to non-tasters (1.9%) (P<0.001). In this EA cohort, PTC tasting abilities differed between smokers and non-smokers: 71.5% of smokers were PTC tasters, while 82.5% of non-smokers were PTC tasters (P = 0.03). The frequency of the *TAS2R38* PAV haplotype showed a trend toward a difference between smokers (38.4%) and non-smokers (43.1%), although this was not significant (P = 0.31) in this small group. We also noticed a possible trend toward a difference in the distribution of *TAS2R38* AVI haplotype between smokers and non-smokers (55.3% and 49.9% respectively, P = 0.29), but again this result was not significant in this small sample.

In the DHS EA cohort, the frequency of the taster PAV haplotype was lower in smokers (37.0%) than in non-smokers (44.0%) (P = 0.003). Conversely the frequency of the non-taster AVI haplotype was higher in smokers (58.7%) compared to non-smokers (51.5%) (P = 0.002). We did not find a difference in the frequency of the AAV haplotype between smokers and non-smokers (Table 2). In order to replicate this association in a sub-sample of individuals

Table 2. Distribution of *TAS2R38* haplotypes between smokers and non-smokers in the DHS population.

Haplotype	Frequency Smokers	Frequency Non-Smokers	P-value
African Americans:			
PAV	0.47	0.48	0.18
AVI	0.33	0.32	0.61
AAI	0.19	0.19	0.60
AAV	0.01	0.01	0.12
European Americans:			
PAV	0.37	0.44	0.003
AVI	0.59	0.51	0.002
AAV	0.04	0.04	0.620

doi:10.1371/journal.pone.0164157.t002

more comparable to the Georgia cohort in demographic characteristics, we repeated this analysis in DHS EA individuals <40 years of age (N = 272). The observed differences in *TAS2R38* haplotypes frequencies between smokers and non-smokers in this subgroup were very similar to those in the entire population (PAV haplotype frequency 35% in smokers vs 44% in non-smokers, P = 0.05; AVI haplotype frequency 60% in smokers vs. 51% in non-smokers, P = 0.06). None of the *TAS2R38* haplotypes differed in frequency between smokers and non-smokers in African Americans in either the DHS or Biobank populations (P's>0.05).

Pooling the data for AA participants from the DHS and Biobank populations and EA participants from the DHS and Georgia populations confirmed our previous un-pooled analyses. In particular, combining the results by meta-analysis showed no association between *TAS2R38* haplotypes and current smoking in AA individuals (P>0.05). For the EA cohorts, we confirmed the associations found in the two independent cohorts with the PAV (P = 0.001) and the AVI (P = 0.001) haplotypes.

Lastly, a significant association was observed between the AVI haplotype and the prevalence of heavy smoking (>20 cigarettes per day) in DHS EA, in the same direction as with smoking status. This haplotype was in fact associated with higher prevalence of heavy smoking (P = 0.009). We also noted an opposite, although not significant, trend for the PAV haplotype (P = 0.08). No significant association was found in DHS AA or in the combined AA cohorts.

Discussion

Although the hypothesis that variations in bitter taste receptor genes confer protection against cigarette smoking has long been of interest, previous findings have been conflicting [16–17–18]. We have therefore recruited a larger number of individuals from three independent cohorts to further explore this question.

Our results show a significant association between common *TAS2R38* haplotypes and smoking in EA: carriers of the taster PAV haplotype, and PTC tasters, were significantly less likely to be current smokers. Conversely, carriers of the non-taster AVI haplotype and PTC non-tasters were significantly more likely to be regular smokers. In contrast, in two large samples of AA including a total of more than 7,000 participants, we found no association between the major *TAS2R38* haplotypes and smoking status. These findings support the hypothesis that *TAS2R38* haplotypes play a role in modulating smoking behaviors, although the effects may be population-specific.

The reasons for the lack of consistency among the previous findings and across ethnic groups are not completely clear. One possibility is that taste plays a differential role as a motivation for smoking in individuals who are heavy tobacco users compared to those who smoke only occasionally. Indeed, the characteristics of participants included in previous reports were quite varied and different from individuals included in the current study. The earliest study exploring the correlation between *TAS2R38* haplotypes and smoking behaviors [18] examined 567 unrelated participants of European descent, comprising 384 smokers recruited from two smoking cessation trials. Although no significant associations were found between PAV/AVI haplotypes and the odds of smoking, the analysis of current smokers revealed a correlation between these haplotypes and the importance of the taste of cigarettes as a motive for smoking, as measured by the WISDM-68 taste/sensory processes scale [31].

A second study [16] enrolled both EA (N = 197) and AA (400) families of heavy smokers (defined as individuals who have smoked for at least the previous 5 years, and have consumed at least 20 cigarettes per day for the preceding 12 months). A significant correlation between the non-taster AVI haplotype and smoking quantity (cigarettes per day) was reported in AA, and the taster PAV haplotype was associated with lower smoking quantity. No significant

associations, however, were observed in their EA participants. Lastly, a recent study [17] recruited 1,007 German individuals comprising 330 smokers with 10.9 mean cigarettes per day and showed that carriers of at least one PAV allele showed significantly lower cigarette smoking per day.

In contrast to these previous studies, which focused on relatively homogeneous populations of heavy smokers, the current study included participants from three demographically diverse cohorts (the Georgia cohort, the DHS cohort, and the Dallas Biobank cohort), in which the prevalence of smoking and nicotine dependence was much lower. In addition, only a small fraction of participants recruited in our cohorts reported smoking more than 20 cigarettes per day. Together, our cohorts contained a total of 1,590 EA and 7,336 AA participants. We fully replicated the results previously reported by Keller and colleagues [17] in our study of two different EA cohorts, the Georgia and DHS populations. In these cohorts, cigarette smokers had a lower percentage of PAV-carriers. In addition, this haplotype was also associated with smoking quantity (data not shown). Moreover, in the Georgia population, smokers showed a lower percentage of PTC-tasters (associated with the PAV haplotype) when compared to non-smokers. This also agrees with previous findings [32–33].

We failed to replicate the results of the family-based study reported by Mangold et al. 2008 In our two AA cohorts, neither PAV nor AVI haplotypes showed different frequencies between smokers and non-smokers. One possible explanation is that most of the individuals recruited in that study were heavy smokers, for whom nicotine dependence was a stronger motivator than taste. Finally, since the smoking data were based on self-report, it is possible that measurement error introduced a bias in our estimates. Nevertheless, this is the largest study to date to investigate the relationship between *TAS2R38* haplotypes and smoking in an ethnically diverse cohort.

Based on both previous and present data, we conclude that *TAS2R38* haplotypes appear to be factors contributing to smoking status in EA, with PAV haplotype carriers and PTC tasters less likely to be smokers. This finding has now been replicated in three independent cohorts, two cohorts in the present study and one in a previous report [17]. In addition, we noted a similar trend in a recent paper studying a large cohort (N = 1,319) of individuals of Caucasian origin [34]. In contrast, *TAS2R38* haplotypes are not good predictors of smoking behaviors in AA. The lack of *TAS2R38* haplotype association with smoking in AA may be due to potentially confounding factors, such as age, gender and ascertainment of smoking status. In the previous studies, in fact, the definition of individuals as "smokers" and/or "current" smokers was different, as was the average age of individuals and the percentage of females. Our study has several possible limitations, including different recruitment mechanisms and baseline characteristics of the cohorts and the fact that tobacco use was self-reported resulting in possible misclassification of smoking status. In addition, the number of cigarettes smoked is quite different in the examined cohorts, with the Georgia population being mainly composed of light smokers. This may have attenuated the association between TAS2R38 haplotypes and smoking in this cohort. However, our results support the hypothesis that *TAS2R38* haplotypes and the related ability to taste specific bitter compounds (such as PTC and PROP) influence smoking behaviors in EA. This does not appear to be true in AA populations. Future studies will need to address potentially confounding variables such as ascertainment of smoking status, population stratification, ethnicity, age and sex.

Supporting Information

S1 File. Table A. Characteristics of the Georgia, DHS and Biobank participants by smoking status. Table B. TAS2R38 Haplotype Frequencies in the Study Cohorts.
(DOCX)

Acknowledgments

The authors thank Drs. Robert Morell and Carter Van Waes for helpful comments on the manuscript, Dr. Jonathan Cohen for the Biobank data set accession, Dr. Helen Hobbs for DHS data accession, and Drs. Thomas Kirchner and Raymond Niaura for helpful discussions.

Author Contributions

Conceptualization: DD SW JK.

Data curation: DR DD JK.

Formal analysis: DR JK.

Funding acquisition: DD JK CB.

Investigation: DR JK ES JG BG.

Methodology: DR JK DD.

Project administration: ES JG DD DL.

Resources: CB BG DD.

Software: DR JK.

Supervision: DD DL.

Validation: CB JK DR.

Visualization: DR JK.

Writing – original draft: DR JK.

Writing – review & editing: DD DR JK CB SW DL.

References

1. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014. PMID: 24455788
2. Centers for Disease Control and Prevention. Current Cigarette Smoking Among Adults—United States, 2005–2013. Morbidity and Mortality Weekly Report 2014; 63(47):1108–12. PMID: 25426653
3. Hummel T, Livermore A, Hummel C, Kobal G. Chemosensory event-related potentials in man: relation to olfactory and painful sensations elicited by nicotine. Electroencephalogr Clin Neurophysiol. 1992 Mar-Apr; 84(2):192–5. doi: 10.1016/0168-5597(92)90025-7 PMID: 1372235
4. Ming D, Ruiz-Avila L, Margolskee RF. Characterization and solubilization of bitter-responsive receptors that couple to gustducin. Proc Natl Acad Sci U S A. 1998 Jul 21; 95(15):8933–8. doi: 10.1073/pnas.95.15.8933 PMID: 9671782
5. Behrens M, Meyerhof W. Mammalian bitter taste perception. Results Probl Cell Differ. 2009; 47:203–20. doi: 10.1007/400_2008_5 PMID: 19145413
6. Enoch MA, Harris CR, Goldman D. Does a reduced sensitivity to bitter taste increase the risk of becoming nicotine addicted? Addict Behav. 2001; 26(3):399–404. doi: 10.1016/S0306-4603(00)00117-9 PMID: 11436931
7. Chandrashekar J, Mueller KL, Hoon MA, Adler E, Feng L, Guo W et al. T2Rs function as bitter taste receptors. Cell. 2000; 100(6):703–11. doi: 10.1016/S0092-8674(00)80706-0 PMID: 10761935
8. Go Y, Satta Y, Takenaka O, Takahata N. Lineage-Specific Loss of Function of Bitter Taste Receptor Genes in Humans and Nonhuman Primates. Genetics. 2005; 170(1): 313–326. doi: 10.1534/genetics.104.037523 PMID: 15744053

9. Wooding S., Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. Am J Hum Genet. 2004; 74(4): 637–646. doi: 10.1086/383092 PMID: 14997422
10. Risso D, Tofanelli S, Morini G, Luiselli D, Drayna D. Genetic variation in taste receptor pseudogenes provides evidence for a dynamic role in human evolution. BMC Evol Biol. 2014; 14:198. doi: 10.1186/s12862-014-0198-8 PMID: 25216916
11. Campbell MC, Ranciaro A, Froment A, Hirbo J, Omar S, Bodo JM et al. Evolution of functionally diverse alleles associated with PTC bitter taste sensitivity in Africa. Mol Biol Evol. 2012; 29(4):1141–53. doi: 10.1093/molbev/msr293 PMID: 22130969
12. Risso D, Mezzavilla M, Pagani L, Robino A, Morini G, Tofanelli S et al. Global diversity in the TAS2R38 bitter taste receptor: revisiting a classic evolutionary PROPosal. Sci. Rep 2016; 6, 25506. doi: 10.1038/srep25506 PMID: 27138342
13. Kim UK, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. Science. 2003; 299(5610):1221–5 doi: 10.1126/science.1080190 PMID: 12595690
14. Bufe B, Breslin PA, Kuhn C, Reed DR, Tharp CD, Slack JP et al. The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. Curr Biol. 2005; 15 (4):322–7. doi: 10.1016/j.cub.2005.01.047 PMID: 15723792
15. Aoki M, Takao T, Takao K, Koike F, Suganuma N. Lower expressions of the human bitter taste receptor TAS2R in smokers: reverse transcriptase-polymerase chain reaction analysis. Tob Induc Dis. 2014; 12(1): 12. doi: 10.1186/1617-9625-12-12 PMID: 25152706
16. Mangold JE, Payne TJ, Ma JZ, Chen G, Li MD. Bitter taste receptor gene polymorphisms are an important factor in the development of nicotine dependence in African Americans. J Med Genet. 2008; 45 (9):578–82. doi: 10.1136/jmg.2008.057844 PMID: 18524836
17. Keller M, Liu X, Wohland T, Rohde K, Gast MT, Stumvoll M et al. TAS2R38 and its influence on smoking behavior and glucose homeostasis in the German Sorbs. PLoS One. 2013; 8(12):e80512. doi: 10.1371/journal.pone.0080512 PMID: 24312479
18. Cannon DS, Baker TB, Piper ME, Scholand MB, Lawrence DL, Drayna DT et al. Associations between phenylthiocarbamide gene polymorphisms and cigarette smoking. Nicotine Tob Res. 2005; 7(6):853–8. doi: 10.1080/14622200500330209 PMID: 16298720
19. Oncken C, Feinn R, Covault J, Duffy V, Dornelas E, Kranzler HR et al. Genetic Vulnerability to Menthol Cigarette Preference in Women. Nicotine Tob Res 2015; 17(12):1416–20. doi: 10.1093/ntr/ntv042 PMID: 25832883
20. Berg CJ, Haardörfer R, Lewis M, Getachew B, Lloyd SA, Thomas SF et al. DECOY: Documenting Experiences with Cigarettes and Other Tobacco in Young Adults. Am J Health Behav. 2016; 40 (3):310–21. doi: 10.5993/AJHB.40.3.3 PMID: 27103410
21. Berg CJ, Parelkar PP, Lessard L, Escoffery C, Kegler MC, Sterling KL et al. Defining “smoker”: college student attitudes and related smoking characteristics. Nicotine Tob Res. 2010; 12(9):963–969. doi: 10.1093/ntr/ntq123 PMID: 20675365
22. Victor RG, Haley RW, Willett DL, Peshock RM, Vaeth PC, Leonard D et al. The Dallas Heart Study: a population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. Am J Cardiol. 2004; 93(12):1473–80. doi: 10.1016/j.amjcard.2004.02.058 PMID: 15194016
23. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977; 74(12): 5463–5467. doi: 10.1073/pnas.74.12.5463 PMID: 271968
24. Risso DS, Howard L, VanWaes C, Drayna D. A potential trigger for pine mouth: a case of a homozygous PTC taster. Nutr Res. 2015; 35(12):1122–5. doi: 10.1016/j.nutres.2015.09.011 PMID: 26463018
25. Available: http://www.dnastar.com/t-allproducts.aspx.
26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D.Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug; 38(8):904–9. doi: 10.1038/ng1847 PMID: 16862161
27. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing. 2011. ISBN: 3-900051-07-0. Available: http://www.R-project.org/.
28. Purcell S, Neale B, Todd-Brown at al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007; 81(3): 559–575. doi: 10.1086/519795 PMID: 17701901
29. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001; 68(4):978–89. doi: 10.1086/319501 PMID: 11254454

30. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422):56–65. doi: 10.1038/nature11632 PMID: 23128226

31. Smith S, Piper M, Bolt D, Fiore MC, Wetter DW, Cinciripini PM et al. Development of the Brief Wisconsin Inventory of Smoking Dependence Motives. Nicotine Tob Res. 2010; 12(5): 489–499. doi: 10.1093/ntr/ntq032 PMID: 20231242

32. Hall AR, Blakeslee AF. Effect of Smoking on Taste Thresholds for Phenyl-Thio-Carbamide (PTC). Proc Natl Acad Sci U S A. 1945; 31(12): 390–396. doi: 10.1073/pnas.31.12.390 PMID: 16588708

33. Daştan S, Durna YM, Daştan T. The relationships between phenylthiocarbamide taste perception and smoking, work out habits and susceptibility to depression. Turkish Journal of Agriculture—Food Science and Technology. 2015; 3(6): 418–424.

34. Ortega FJ, Agüera Z, Sabater M, Moreno-Navarrete JM, Alonso-Ledesma I, Xifra G et al. Genetic variations of the bitter taste receptor TAS2R38 are associated with obesity and impact on single immune traits. Mol Nutr Food Res. 2016 Jul; 60(7):1673–83. doi: 10.1002/mnfr.201500804 PMID: 27059147

Nicotine & Tobacco Research Advance Access published October 25, 2016

Nicotine & Tobacco Research, 2016, 1–2
doi:10.1093/ntr/ntw275
Letter

OXFORD

Letter

Association of *TAS2R38* Haplotypes and Menthol Cigarette Preference in an African American Cohort

Davide Risso MS[1,2], Eduardo Sainz MS[1], Joanne Gutierrez MS[1], Thomas Kirchner PhD[3], Raymond Niaura PhD[4], Dennis Drayna PhD[1]

[1]Laboratory of Communication Disorders, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD; [2]Laboratory of Molecular Anthropology and Centre for Genome Biology, Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy; [3]NYU College of Global Public Health, New York, NY; [4]Schroeder Institute for Tobacco Research and Policy Studies, Washington, DC

Corresponding Author: Dennis Drayna, PhD, Laboratory of Communication Disorders, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD, USA. Telephone: 301-402-4930; Fax: 301-827-9637; E-mail: drayna@nidcd.nih.gov

In a recent publication in *Nicotine & Tobacco Research*, Oncken *et al.*[1] examined 323 pregnant female Caucasian cigarette smokers, including menthol and nonmenthol users, and genotyped three *TAS2R38* bitter taste receptor gene polymorphisms (*rs713598*, *rs1726866*, and *rs10246939*). These polymorphisms specify whether an individual is a taster (associated with the PAV haplotype) or a nontaster (associated with the AVI haplotype) for a variety of bitter compounds, including the well-known phenylthiocarbamide and propylthiouracil.[2–3] The rationale behind this study was to test whether variations in the well-studied *TAS2R38* bitter taste receptor gene could contribute to the preference of smokers for menthol cigarettes, because menthol could mask the bitter taste of nicotine or other components of cigarette smoke. Oncken *et al.* reported the frequency of the PAV taster haplotype to be greater in menthol smokers than in nonmenthol smokers in both non-Hispanic (54% vs. 30%, respectively, $p < 0.001$) and Hispanic (53% vs. 25%, respectively, $p = 0.016$) women, confirming this initial hypothesis.

While this research reported an intriguing association between variations in a bitter taste receptor gene and preference for menthol cigarettes, the authors pointed out a number of weaknesses in their study, including the relatively small sample size and the fact that they only studied pregnant Caucasian women. The authors noted that it was not clear if these results generalize to other ethnicities, nonpregnant women or men. In addition, genotype–phenotype association studies typically require replication in an independent group in order to exclude any potential confounding population stratification and increase the confidence in the findings.[4] Moreover, focusing the attention on Caucasian individuals precludes the possibility of studying rarer alleles and haplotypes, since individuals of African descent carry greater levels of genetic diversity and more polymorphic sites when compared to non-African populations.[5] This is of particular importance when studying menthol cigarette smoking, since an average of 80% of adult African Americans are menthol smokers, while only 25–35% of Caucasian and Hispanic smokers use menthol.[6]

In order to overcome these limitations, we recruited 718 African Americans smokers (236 females and 482 males) in the Washington DC area with written informed consent under protocols reviewed and approved by the National Institutes of Health Combined Neurosciences IRB and the Western IRB (National Institutes of Health protocol 01-DC-0230). All procedures were performed in accordance with the Helsinki Declaration of 1975, as revised in 2000. The average age of these individuals was 45.1 (SD = 10.8) and 406 (56.5%) of them were menthol smokers, with the remaining 312 (43.5%) being nonmenthol smokers. A higher percentage of menthol than nonmenthol smokers were female (39.6% vs. 24.2%, $p = 0.001$), in agreement with previously reported data.[6] No differences were found in the mean age of menthol smokers (45.4, SD = 11.0) and nonmenthol smokers (44.6, SD = 10.6; $p = 0.372$). We collected saliva samples from all participants using Oragene saliva collection kits (Genotek Inc, Kanata, ON, Canada) and we purified genomic DNA following the manufacturer's protocol. A dedicated set of primers was designed to completely sequence the single coding exon of the TAS2R38 gene using dideoxy Sanger sequencing, as previously published.[7] We constructed and calculated the frequencies of the taster PAV (45.47%), nontaster AVI (33.22%), and the rarer AAI (17.90%), AAV (2.37%), PVI (0.76%), and PAI (0.28%) haplotypes, and we examined associations between their frequency and menthol cigarette preference in the group. The nontaster AVI haplotype was inversely associated with menthol cigarette smoking, even after correction for sociodemographic factors (such as age, gender, marital status, and education level) and multiple testing correction, implemented in a logistic regression model.

Downloaded from http://ntr.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

Published by Oxford University Press on behalf of the Society for Research on Nicotine and Tobacco 2016.
This work is written by (a) US Government employee(s) and is in the public domain in the US.

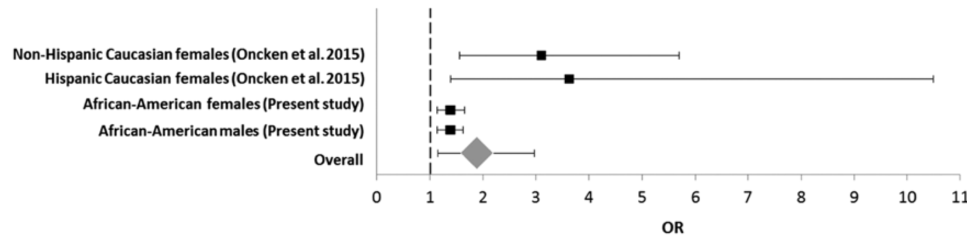


Figure 1. Forest plot illustrating the odds ratios and confidence intervals of the TAS2R38 PAV haplotype association with menthol cigarette smoking in both African American and Caucasian individuals.

The frequency of this haplotype was significantly lower in menthol smokers (29.8%) compared to nonmenthol smokers [37.7%; odds ratio (OR) = 0.70, p = 0.008]. This association was replicated in both female (OR = 0.89, p = 0.01) and male (OR = 0.72, p = 0.02) individuals and showed a gene dosage effect with 62.3%, 53.8%, and 44.0% of menthol users carrying zero, one, or two copies of this haplotype, respectively. Conversely, the taster PAV haplotype was higher in menthol smokers (48.2%), compared to nonmenthol smokers (42.1%; OR = 1.24, p = 0.04). The fact that the association between the PAV haplotype and menthol smoking was not as strong as previously found[1] could be due to the relatively greater number of intermediate sensitivity haplotypes (AAI, AAV, PVI, and PAI haplotypes) in our subject sample that did not show a significant difference in frequency between menthol and nonmenthol smokers (all p's > 0.1). In addition, since our population displayed a high prevalence of African-derived haplotypes, there was not a strong and complementary relation between the common nontaster AVI and taster PAV haplotypes (r = −0. 64), in comparison to the one previously shown in Caucasians[1] (r = −0.87). Lastly, when performing a meta-analysis including the results of both the present study and of Oncken *et al.*,[1] the association between *TAS2R38* PAV haplotype and menthol cigarette smoking was still significant (p = 0.025, using a random effects model) resulting in a common OR of 1.78 (95% confidence interval = 1.07 to 2.94; Fig.1), with some evidence (p = 0.019) of expected heterogeneity, considering the differences in subject characteristics.

Our results confirm the previous work by Oncken *et al.*,[1] stressing the validity and importance of their findings which we have now extended to men and to different ethnic groups. Together, these data suggest that genetic variations that modify the ability of tasting bitter compounds could explain the observed differences and preference toward mentholated tobacco use. This is of particular importance, considering the fact that the potential risks associated with adding menthol to cigarettes have been a subject of considerable recent study by the Food and Drug Administration (FDA)[8] and risks to specific racial and ethnic minorities, such as African Americans and Hispanics have been raised in this context.[9]

Funding

This research was supported by the Intramural Research Program of the NIH, NIDCD (grant numbers Z1A-000046-15 and Z1C-000086), by NIH, Office of the Director and the National Institute on Drug Abuse Grant RC1-DA028710, and by the FDA through funds obtained under the Family Smoking Prevention and Tobacco Control Act. The content was not reviewed by the Food and Drug Administration, but underwent the standard manuscript clearance process for scientific papers published from the NIH intramural research program.

Declaration of Interests

None declared.

Acknowledgments

The authors thank Carter VanWaes and Thomas Friedman for helpful comments and Julia Kozlitina for critical revision of the manuscript.

References

1. Oncken C, Feinn R, Covault J, et al. Genetic vulnerability to menthol cigarette preference in women. *Nicotine Tob Res*. 2015;17(12):1416–1420.
2. Kim UK, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science*. 2003;299(5610):1221–1225.
3. Harris H, Kalmus H. Chemical specificity in genetical differences of taste sensitivity. *Ann Eugen*. 1949;15(1):32–45.
4. NCI-NHGRI Working Group on Replication in Association Studies; Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature*. 2007;447(7145):655–660.
5. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*. 2008;9:403–433. doi:10.1146/annurev.genom.9.081307.164258
6. Caraballo RS, Asman K. Epidemiology of menthol cigarette use in the United States. *Tob Induc Dis*. 2011;9(suppl 1):S1. doi:10.1186/1617-9625-9-S1-S1
7. Risso DS, Howard L, VanWaes C, Drayna D. A potential trigger for pine mouth: a case of a homozygous PTC taster. *Nutr Res*. 2015; 35(12):1122–1125.
8. Food and Drug Administration. Preliminary Scientific Evaluation of the Possible Public Health Effects of Menthol Versus Nonmenthol Cigarettes. 2013. http://www.fda.gov/downloads/UCM361598.pdf. Accessed September 16, 2016.
9. Tobacco Products Scientific Advisory Committee. *Menthol Cigarettes and Public Health: Review of the Scientific Evidence and Recommendations*. Rockville, MD: Food and Drug administration; 2011. http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/tobaccoproductsscientificadvisorycommittee/ucm269697.pdf. Accessed September 16, 2016.

Downloaded from http://ntr.oxfordjournals.org/ at National Institutes of Health Library on December 15, 2016

7. References

1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68-74

Agaku IT, King BA, Husten CG, et al. Tobacco product use among adults—United States, 2012–2013. MMWR Morb Mortal Wkly Rep 2014;63:542–7

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep;19(9):1655-64

American Lung Association. State of Lung Disease in Diverse Communities, 2010. Available at: http://action.lung.org/site/DocServer/state-of-lung-disease-in-diverse-communities-2010.pdf?docID=8744

Aoki M, Takao T, Takao K, Koike F, Suganuma N. Lower expressions of the human bitter taste receptor TAS2R in smokers: reverse transcriptase-polymerase chain reaction analysis. Tob Induc Dis. 2014; 12(1): 12

Bachmanov AA, Bosak NP, Lin C, Matsumoto I, Ohmoto M, Reed DR, Nelson TM. Genetics of Taste Receptors. Curr Pharm Des. 2014 ; 20(16): 2669–2683

Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. 1999. Mol Biol Evol 16(1):37–48

Bautista DM et al. The menthol receptor TRPM8 is the principal detector of environmental cold. Nature. 2007 Jul 12;448(7150):204-8

Berg CJ, HaardoÈ rfer R, Lewis M, Getachew B, Lloyd SA, Thomas SF et al. DECOY: Documenting Experiences with Cigarettes and Other Tobacco in Young Adults. Am J Health Behav. 2016; 40(3):310-21

Besnard P, Passilly-Degrace P, Khan NA. Taste of Fat: A Sixth Taste Modality? Physiol Rev. 2016 Jan;96(1):151-76

Blakeslee, A. F. Genetics of sensory thresholds: taste for phenyl thio carbamide. Proc Natl Acad Sci USA 18, 120–130 (1932)

Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am J Hum Genet. 2015 Jan 8;96(1):37-53

Bufe B, Breslin P, Kuhn C, Reed D, Tharp C, et al. 2005. The molecular basis of individual differences in phyenythiocarbamide and propylthiouracil bitterness perception. Curr. Biol. 15:1–20

California Environmental Protection Agency. Health Effects of Exposure to Environmental Tobacco Smoke: Final Report. 2005. Available at: http://oehha.ca.gov/air/report/health-effects-exposure-environmental-tobacco-smoke-final-report

Campbell et al. Evolution of functionally diverse alleles associated with PTC bitter taste sensitivity in Africa. Mol Biol Evol. 2012 Apr;29(4):1141-53

Cannon DS, Baker TB, Piper ME, Scholand MB, Lawrence DL, Drayna DT et al. Associations between phenylthiocarbamide gene polymorphisms and cigarette smoking. Nicotine Tob Res. 2005; 7(6):853-8

Caraballo RS, Asman K. Epidemiology of menthol cigarette use in the United States. Tob Induc Dis. 2011;9(suppl 1):S1

Centers for Disease Control and Prevention. Health, United States, with special feature on the health of young adults. 2008. Available at: http://www.cdc.gov/nchs/data/hus/hus08.pdf

Centers for Disease Control and Prevention . State Tobacco Activities Tracking & Evaluation (STATE) System. 2014. Available at: http://www.cdc.gov/statesystem/cigaretteuseadult.html

Chandrashekar J, Mueller K, Hoon M, Adler E, Feng L, et al. 2000. T2Rs function as bitter taste receptors. Cell 100:703–11

Chandrashekar J, Hoon MA, Ryba NJ, Zuker CS. The receptors and cells for mammalian taste. Nature. 2006 Nov 16;444(7117):288-94

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015 Feb 25;4:7

Chaudhari N., Roper S.D. 2010. The cell biology of taste. J. Cell Biol. Vol. 190 No. 3 285–296

Clark, A.A. Dotson, C.D. Elson, A.E.T. Voigt, A. Boehm, U. Meyerhof, W. Steinle N.I. Munger, S.D. 2015. TAS2R bitter taste receptors regulate thyroid function. FASEB J. Jan;29(1):164-72

Cubbin C, Soobader MJ, LeClere FB. The intersection of gender and race/ethnicity in smoking behaviors among menthol and non-menthol smokers in the United States. Addiction. 2010 Dec;105 Suppl 1:32-8. doi: 10.1111/j.1360-0443.2010.03191.x

Daştan S, Durna YM, Daştan T. The relationships between phenylthiocarbamide taste perception and smoking, work out habits and susceptibility to depression. Turkish Journal of AgricultureÐFood Science and Technology. 2015; 3(6): 418-424

del Camino D, Murphy S, Heiry M, Barrett LB, Earley TJ, Cook CA, Petrus MJ, Zhao M, D'Amours M, Deering N, Brenner GJ, Costigan M, Hayward NJ, Chong JA, Fanger CM, Woolf CJ, Patapoutian A, Moran MM. TRPA1 contributes to cold hypersensitivity. J Neurosci. 2010 Nov 10;30(45):15165-74

Delnevo CD, Villanti AC, Giovino GA. 2014. Trends in menthol and non-menthol cigarette consumption in the USA: 2000–2011. Tob Control 2014;23:e154-e155 doi:10.1136/tobaccocontrol-2013-051385

Drayna D, Coon H, Kim UK, Elsner T, Cromer K, Otterud B, Baird L, Peiffer AP, Leppert M; Utah Genetic Reference Project. Genetic analysis of a complex trait in the Utah Genetic Reference Project: a major locus for PTC taste ability on chromosome 7q and a secondary locus on chromosome 16p. Hum Genet. 112, 567–72 (2003)

Drayna, D. Human taste genetics. 2005. Annu. Rev. Genomics Hum. Genet. 6;217-35

Enoch MA, Harris CR, Goldman D. Does a reduced sensitivity to bitter taste increase the risk of becoming nicotine addicted? Addict Behav. 2001; 26(3):399-404

Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online. 2007 Feb 23;1:47-50

Fan L, Balakrishna S, Jabba SV, Bonner PE, Taylor SR, Picciotto MR, Jordt SE. Menthol decreases oral nicotine aversion in C57BL/6 mice through a TRPM8-dependent mechanism. Tob Control. 2016 Nov;25(Suppl 2):ii50-ii54

Federal Trade Commission Cigarette Report. 2013. Available at: https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-cigarette-report-2013/2013cigaretterpt.pdf

Finger T.E. and Kinnamon S.C. 2011. Taste isn't just for taste buds anymore. F1000 Biol Rep. 3: 20

Fischer A, Gilad Y, Man O, Pääbo S. 2005. Evolution of bitter taste receptors in humans and apes. Mol Biol Evol. 2005 Mar;22(3):432-6

Fisher R, Ford E, Huxley J. 1939. Taste-testing in Anthropoid apes. Nature 144.750

Fox, A. L. The Relationship between chemical constitution and taste. Proc Natl Acad Sci USA 18, 115–120 (1932)

Fu SS, Kodl MM, Joseph AM, Hatsukami DK, Johnson EO, Breslau N, Wu B, Bierut L. Racial/Ethnic disparities in the use of nicotine replacement therapy and quit ratios in lifetime smokers ages 25 to 44 years. Cancer Epidemiology, Biomarkers and Prevention. 2008;17:1640–1647

Galeotti N, Di Cesare Mannelli L, Mazzanti G, Bartolini A, Ghelardini C. Menthol: a natural analgesic compound. Neurosci Lett. 2002 Apr 12;322(3):145-8

Gardiner PS. The African Americanization of menthol cigarette use in the United States. Nicotine Tob Res. 2004 Feb;6 Suppl 1:S55-65

Gilbertson T, Damak S, Margolskee R. 2000. The molecular physiology of taste transduction. Curr. Opin. Neurobiol. 10: 519–27

Giovino GA, Villanti AC, Mowery PD, Sevilimedu V, Niaura RS, VAllone DM, Abrams DB. Differential trends in cigarette smoking in the USA: is menthol slowing progress? Tob Control doi:10.1136/tobaccocontrol-2013-051159

Go Y, Satta Y, Takenaka O, Takahata N. Lineage-Specific Loss of Function of Bitter Taste Receptor Genes in Humans and Nonhuman Primates. Genetics. 2005 May; 170(1): 313–326

Hall AR, Blakeslee AF. Effect of Smoking on Taste Thresholds for Phenyl-Thio-Carbamide (PTC). Proc Natl Acad Sci U S A. 1945; 31(12): 390-396

Hellekant G, Danilova B, Ninomiya Y. 1997. Primate sense of taste: behavioral and single chroda tympani and glossopharyngeal nerve fiber recordings in the rhesus monkey, Macca mulatta. J. Neurophysiol. 77:978–93

Hersey JC, Ng SW, Nonnemaker JM, Mowery P, Thomas KY, Vilsaint MC, et al. Are menthol cigarettes a starter product for youth? Nicotine & Tobacco Research. 2006;8:403–413. doi:10.1080/14622200600670389

Hyland A, Garten S, Giovino GA. Mentholated cigarettes and smoking cessation: Findings from COMMIT. Tobacco Control. 2002;11:135–139. doi:10.1136/tc.11.2.135

H.R. 1256—11th Congress, 2012. Available at: https://www.govtrack.us/congress/bills/111/hr1256/text

Jordt SE, McKemy DD, Julius D. Lessons from peppers and peppermint: the molecular logic of thermosensation. Current Opinion in Neurobiology. 2003;(13)

Karashima Y, Damann N, Prenen J, Talavera K, Segal A, Voets T, Nilius B. Bimodal action of menthol on the transient receptor potential channel TRPA1. Neurosci. 2007 Sep 12;27(37):9874-84

Keller M, Liu X, Wohland T, Rohde K, Gast MT, Stumvoll M et al. TAS2R38 and its influence on smoking behavior and glucose homeostasis in the German Sorbs. PLoS One. 2013; 8(12):e80512

Kim U-K, Breslin P, Reed D, Drayna D. 2004. Genetics of human taste perception. J. Dental Res. 83:448–53

Knaapila, A. et al. Genetic analysis of chemosensory traits in human twins. Chem Senses. 2012. 37, 869–81

Levy DT, Pearson JL, Villanti AC, Blackman K, Vallone DM, Niaura RS, Abrams DB. Modeling the future effects of a menthol ban on smoking prevalence and smoking-attributable deaths in the United States. Am J Public Health. 2011 Jul;101(7):1236-40. doi: 10.2105/AJPH.2011.300179

Li X, Staszewski L, Xu H, Durick K, Zoller M, Adler E. 2002. Human receptors for sweet and umami taste. Proc. Natl. Acad. Sci. USA 99:4692–96

Luther L Terry. Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service, 1964

Mangold JE, Payne TJ, Ma JZ, Chen G, Li MD. Bitter taste receptor gene polymorphisms are an important factor in the development of nicotine dependence in African Americans. J Med Genet. 2008; 45 (9):578-82

Ming D, Ruiz-Avila L, Margolskee RF. Characterization and solubilization of bitter-responsive receptors that couple to gustducin. Proc Natl Acad Sci U S A. 1998 Jul 21;95(15):8933-8

Mowery PD, Dube SR, Thorne SL, Garrett BE, Homa DM, Nez Henderson P. Disparities in Smoking-Related Mortality Among American Indians/Alaska Natives. Am J Prev Med. 2015 Nov;49(5):738-44

Nonnemaker, J., Hersey, J., Homsi, G., Busey, A., Allen, J., and Vallone, D. (2013). Initiation with menthol cigarettes and youth smoking uptake. Addiction 108, 171–178. doi: 10.1111/j.1360-0443.2012.04045.x

Oliveira-Maia AJ, Stapleton-Kotloski JR, Lyall V, et al. Nicotine activates TRPM5-dependent and independent taste pathways. Proc Natl Acad Sci U S A. 2009;106(5):1596–1601

Oncken C, Feinn R, Covault J, Duffy V, Dornelas E, Kranzler HR et al. Genetic Vulnerability to Menthol Cigarette Preference in Women. Nicotine Tob Res 2015; 17(12):1416-20

Ortega FJ, AguÈera Z, Sabater M, Moreno-Navarrete JM, Alonso-Ledesma I, Xifra G et al. Genetic variations of the bitter taste receptor TAS2R38 are associated with obesity and impact on single immune traits. Mol Nutr Food Res 2016 Jul; 60(7):1673-83

Pearson JL, Abrams DB, Niaura RS, Richardson A, Vallone DM. A Ban on Menthol Cigarettes: Impact on Public Opinion and Smokers' Intention to Quit. Am J Public Health 2012. 102(11): e107–e114

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, Pritchard JK. Signals of recent positive selection in a worldwide sample of human populations. Genome Res 2009 May;19(5):826-37

Piper ME, Bolt DM, Kim SY, Japuntich SJ, Smith SS, Niederdeppe J, Cannon DS, Baker TB. Refining the tobacco dependence phenotype using the Wisconsin Inventory of Smoking Dependence Motives (WISDM). J Abnorm Psychol 2008 Nov;117(4):747-61

Pomerleau OF and Rosecrans J. Neuroregolatory effects of nicotine. Psychoneuroendocrinology. 1989;14(6):407-23

R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing. 2011. ISBN: 3-900051-07-0

Reading. The Tobacco Report (Chapter 4). 1998. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.9108&rep=rep1&type=pdf

Risso D, Tofanelli S, Morini G, Luiselli D, Drayna D. Genetic variation in taste receptor pseudogenes provides evidence for a dynamic role in human evolution. BMC Evol Biol. 2014 Sep 13;14:198

Risso DS, Mezzavilla M, Pagani L, Robino A, Morini G, Tofanelli S, Carrai M, Campa D, Barale R, Caradonna F, Gasparini P, Luiselli D, Wooding S, Drayna D. Global diversity in the *TAS2R38* bitter taste receptor: revisiting a classic evolutionary PROPosal. Sci Rep. 2016 May 3;6:25506

Rozas, J., J. C. Sancez–DelBarrio, X. Messeguer, and R. Rosaz. 2003. DnaSp, DNA polymorphism analyses by the coalescent and others methods. Bioinformatics 19: 2496–2497

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977; 74(12): 5463-5467

Shaik FA, Singh N, Arakawa M, Duan K, Bhullar RP, Chelikani P. 2016. Bitter taste receptors: extraoral roles in pathophysiology. The International Journal of Biochemistry & Cell Biology doi:10.1016/j.biocel.2016.03.011

Smith D, Margolskee R. 2001. Making sense of taste. Sci. Am. 284:32–39

Stähler F, Riedel K, Demgensky S, Neumann K, Dunkel A, Täubert A, Raab B, Behrens M, Raguse J, Hofmann T, Meyerhof W. A Role of the Epithelial Sodium Channel in Human Salt Taste Transduction? Chemosensory Perception. 2008 Mar,1 78-90

Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001; 68(4):978-89

Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. (November 18, 2011). The NSDUH Report: Recent Trends in Menthol Cigarette Use. Rockville, MD

Sutton CD, Robinson RG. The marketing of menthol cigarettes in the United States: populations, messages, and channels. Nicotine Tob Res. 2004 Feb;6 Suppl 1:S83-91

Tobacco Products Scientific Advisory Committee. Menthol Cigarettes and Public Health: Review of the Scientific Evidence and Recommendations. 2011. Available at: http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/tobaccoproductsscientificadvisorycommittee/ucm269697.pdf

Uhl GR, Walther D, Behm FM, Rose JE. Menthol preference among smokers: association with TRPA1 variants. Nicotine Tob Res. 2011 Dec;13(12):1311-5

U.S. Department of Health and Human Services. The Health Consequences of Involuntary Exposure to Tobacco Smoke, Surgeon General Report. 2006. Available at: https://www.ncbi.nlm.nih.gov/books/NBK44324/

U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. 2014. Available at: http://www.surgeongeneral.gov/library/reports/50-years-of-progress/full-report.pdf

U.S. Food and Drug Administration. Preliminary scientific evaluation of the possible public health effects of menthol versus nonmenthol cigarettes. Silver Spring, MD: Center for Tobacco Products, Food and Drug Administration, 2013

Victor RG, Haley RW, Willett DL, Peshock RM, Vaeth PC, Leonard D et al. The Dallas Heart Study: a population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. Am J Cardiol. 2004; 93(12):1473-80

Wooding S et al. Natural Selection and Molecular Evolution in PTC, a Bitter-Taste Receptor Gene. Am J Hum Genet. 2004 Apr; 74(4): 637–646

Wright S. Evolution and the Genetics of Populations Volume 2: the Theory of Gene Frequencies. Univ. of Chicago Press, Chicago, 1969 pg 294–295

Yerger VB and McCandless PM. Menthol sensory qualities and smoking topography: a review of tobacco industry documents. Tob Control. 2011 May; 20(Suppl_2): ii37–ii43

Wang, X., Thomas, S. D. & Zhang, J. Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. 2004. Hum Mol Genet. 13, 2671–8

8. Acknowledgments

Premessa -originale: scrivo deliberatamente la parte dei ringraziamenti in Italiano, per rendere omaggio a questa fantastica lingua, la qual musicalità e poesia la rendono, secondo molti, più bella al mondo.
Premessa -traduzione: i miei genitori non parlano l'Inglese.

Potrei continuare a scrivere altre 60 pagine di soli ringraziamenti, citando tutte le persone senza le quali questo progetto non sarebbe stato possibile. Preferisco pero' salvare qualche albero quindi cerchero' di riassumere il tutto in poche righe. Le persone citate di seguito sono state indispensabili e di grande aiuto durante questi tre anni di Dottorato e sono grato a tutte, in egual misura.

In ordine assolutamente casuale:

- **Donata Luiselli & Dennis Drayna**. Tutor della mia tesi, mi hanno insegnato come funziona il mondo della scienza in Italia e America, rispettivamente. Grazie.

- **Eduardo Sainz**. Oracolo di Sapienza. Mi ha avvicinato alla Biologia Molecolare e Cellulare, e fatto pagare un affitto molto basso. Gracias.

- **Julia Kozlitina**. Per qualunque dubbio di origine statistico-matematico che ha risolto. Thanks.

-**Sergio Tofanelli**. Colui che mi ha iniziato alla ricerca scientifica e insegnato ad essere corretti, sempre. Merci.

- **Gabriella Morini.** Amica, ancor prima di essere una collega. Grande ricercatrice e persona, prima di tutto. Obrigado.

- **Luca Pagani.** Da anni oramai un amico e guida nel mondo della bio-informatica. Cheers.

- Wolfgang Meyerhof, Maik Behrens. Esperti mondiali di recettori del gusto, hanno condiviso parte di questa conoscenza nel loro laboratorio a Potsdam. Danke.

- Oliver & Gaby. Grazie per le fusa e le feste. Miao e bau.

- Marco Persiano, Chiara Della Santina, Riccardo Cerrato, Lorenzo Mazzotta, Marco Puccetti. Due parole: migliori amici.

- **Franca e Piero**. I migliori genitori che abbia mai avuto.