Title: Diversity and Functional Prediction of Non-Synonymous SNPs in Human Chemosensory Genes

Committee Members: Stephen Wooding (advisor), Ricardo Cisneros, Andrea Joyce

Approved on September 2018 for partial fulfillment of the Master's of Public Health

Abstract:

The discovery and understanding of genetic polymorphisms leading to traits and diseases is a major goal of modern genetics. The utilization of *in silico* prediction techniques has aided these discoveries by providing a cost-effective approach for seeking and prioritizing genetic candidates. There has been much interest in the genomic underpinnings that may contribute to phenotypic variability seen in taste sensitivity and perception in humans. We utilized whole genome and exome sequences from 2504 individuals across 26 global populations that are provided by the 1000 genomes project to explore global genetic variation in 56 genes that are known to contribute to taste and oral chemosensory pathways. We used annotation databases and tools provided by Ensembl's Variant Effect Predictor to perform this analysis. We utilized 38,000 sites of polymorphism to describe genetic variability and differentiation of global populations in taste genes. We report high levels of genetic variability between populations and between genes. We explored 2410 sites of exonic mutation, 1437 of which were non-synonymous mutations, to predict mutation that has functional effect on taste proteins. Together we report 463 non-synonymous mutations that are predicted to have deleterious effect on protein structure and function, most of which are rare variants MAF>0.1%. We report several prediction of deleteriousness in loci that show high levels of differentiation and are kept at high frequencies levels in populations. This warrants further investigation. Overall, we demonstrate that the use of in silico techniques has informed us of variants that may influence variability in taste phenotypes.

Introduction:

Understanding the relationship between genotype and phenotype is important. A common method now employed by clinical researchers and bioinformaticians for GWAS and whole genome sequencing experiments is to utilize *in silico* techniques to predict the functional significance of genetic variants in order to prioritize them(1–3). Prioritizing these markers to traits or diseases can be difficult when there are a large set of markers that may only contribute small effects due variability in the trait(4). Such methods have proven useful in studying complex diseases, and have given clinical researchers a cost-effective approach for improving these studies and assessing genotype-phenotypes associations(5).

Taste is an evolutionary gatekeeper of chemical substances from entering the body. It may have helped our evolutionary ancestors distinguish between toxins and poisonous foods, while simultaneously acting as a guiding mechanism to seek nutrient and energy rich foods. This had a great capacity to shape the evolutionary fitness of our ancestors, and has the capacity to shape modern human health(6–8). Naturally, there is much interest in how genetic variants shape difference in taste perception. There has been much that has been uncovered about the genetics of taste(9), and one of the best-characterized taste genes is TAS2R38. Over time our understanding the understanding of genetic variation and its influences on taste and health has increased, but there is still much to be discovered. As such, investigators remain intrigued at the molecular underpinnings that control the variability in taste phenotypes.

Humans are able to perceive at least five basic tastes: salty, bitter, sour, sweet, and umami(6); although support for oleogustus, the taste of fat, has accumulated(10–12). Sweet, umami, and bitter tastes are mediated by cell surface receptors belonging to the *TAS1R* and *TAS2R* gene families(13, 14). The human TAS2R gene family 25 functional GPCRs and 11 pseudogenes

along chromosomes 5, 7, and 12 that function as taste receptors for a wide variety of bitter compounds(7, 8). TAS2R38, for instance, is the most widely studied bitter receptor and is associated with taste sensitivity to phenylthiocarbamide(15, 16). Three single-nucleotide polymorphisms in the coding region of TAS2R38 give rise to five distinct haplotypes that account for 55-85% of variance in PTC sensitivity(17). Binding of bitter ligands to the TAS2R GPCRs activates G-proteins, including GNAT3, and signal transduction is mediated through a phospholipase c and IP3 mediated pathway(18). TRPM5, an ion-channel activated by the release of Ca2+, is associated along these taste pathways as a downstream target of the signaling cascade of the taste transduction pathway(19). Members of the RTP and REEP gene families seem to promote expression of GPCRs, and are likely associated with bitter taste sensitivity(20). These genes are all important members of the ligand binding and transduction mechanism that governs taste perception.

Salt is a vital component for life; our body utilizes ionic compounds to maintain homeostasis and for our cells to perform everyday tasks necessary for survival. When there is true sodium need animals will switch food preferences to sodium-rich foods over other foods(21). Salt taste transduction is hypothesized to be mediated by epithelial sodium channels (ENaCs). ENaCs themselves are composed of three subunits—α, β, and γ— encoded by SCNN1A, SCNN1B, and SCNN1G(22). The receptors underlying sour taste transduction mechanisms have been elusive. Over the years several candidates for receptor proteins have been proposed for sour taste. Among these are PKD2L1, and its related protein PKD1L3(23). These proteins have been supported to play a role in sour taste, as they are proposed for acid-sensitive taste receptor cells(19, 24). This has shown conflicting results, and evidence has pointed to they not being responsible for sour

taste, or at least not being the only mechanism for sour taste(25). Thus the mechanisms that underlie sour taste are not yet fully understood.

Taste of fatty acids, or oleogustus, has been more recently established as a true taste. The ability to taste fats is hypothesized to have been an evolutionary feature developed to help seek the most energy dense foods, and may have played a vital role in the evolution of the human brain(6). The most associated genetic components with this phenomenon are *CD36*, *GPR120*, and *GPR40*(12, 26, 27). Among these CD36 has been the most extensively researched and proposed to be a taste receptor for the detection of fatty acids(28, 29). This has mostly been established in mice, however it is unlikely that this mechanism differs much within humans and has been supported through some association studies in human populations(30, 31).

There are other mechanisms for oral chemosensation that humans and other mammals experience. The transient receptor protein family is of much interest for detection of environmental stimulus. TRPV1 is a widely expressed nocireceptor responsible for detecting a wide amount of stimuli, most notably the temperature in the ambient environment(32). Many of the TRP gene family proteins act as thermo-sensors including members of the TRPV family, TRPM family, and TRPA1. Some of these receptors react to cold like TRPM5, some are activated by warmth like TRPV3 and TRPM5, and finally some are activated by extreme or noxious heat like TRPV1(33, 34). These genes also encode surface receptors on the tongue and oral epithelia that are activated by chemical stimulus. Capsaicin is an irritant that is responsible for the sensation of spiciness and is detected by TRPV1. TRPM8 is activated by menthol. TRPA1 binds to molecules responsible for detecting certain pungent and noxious compounds like allyl isothiocyanate, or the taste of wasabi(35).

The pathways that govern taste are complex and their mechanisms are not completely established, especially in humans. Further the variability in emerging phenotypes between individuals resulting from genetic variation is not completely understood. As such we employed an in silico predictive analysis in order to explore the basis of genetic variation in human chemosensory pathways. In our investigation we have examined 56 gene regions associated with oral chemosensory signaling pathways. We have examined predicted consequence of non-synonymous variation upon protein function within these taste genes, and examined the population structure for these genetic variations. To do this we have utilized the 1000 genomes data repository to examine variation and global diversity within these taste genes pathways.

Methodology:

Data Selection:

A list of 56 gene regions that are implicated in taste were chosen based on review of the current literature on taste and gustatory perception. Genomic coordinates are in alignment to GRCH37/hg19 of the human genome annotation. Ensembl gene id's were obtained using the Ensembl BioMart tool(36). From this we found chose the most canonical transcript for analysis, and obtained the genomic coordinates. For each genomic transcript, a 1000 base pair flanking sequence was added to the 3' and 5' ends of the gene.

*Figure 1

Variation data were collected from the phase III repository of the 1000 Genomes project. The 1000 genomes project, completed in 2013, is a global catalog of human genomic variation encompassing whole genome and exome data from over 26 worldwide populations.(37, 38) For our analysis, we acquired phased haplotypes from 2504 non-related and self-reported healthy

individuals from across these 26 worldwide populations, which are described in Figure 2 and Table 1. Data were acquired from the 1000 genome phase III 2013 release.

*Figure 2

*Table 1

Data Cleaning:

The populations were aggregated into 5 super populations covering Africa, the Americas, South Asia, Eastern Asia, and Europe. Variant call format (VCF) files obtained from the 1000 genomes repository were indexed and sequences of data were extracted using tabix(39), Data were then processed to include exclude multiallelic sites and complex variants such as indels and structural variants. This was accomplished using bcftools, a command line tool for processing variant call format and binary variant call format files(40). The resulting files consisted of biallelic single nucleotide polymorphism. After processing our data we incorporated 38665 biallelic segregating sites into our population analysis.

Population Analysis:

VCF files were formatted to run in PopGenome, an R package for population genetic analysis(41). Bgzip was used to created bgzf compressed files, and tabix was used to create tabix index files for the biallelic VCFs. Formatting the VCF files in this way was necessary to read the data into PopGenome. In PopGenome, we collected the number of segregating site, nucleotide diversity and F_{ST}. Nucleotide diversity was measured as the mean pairwise differences between sequences per nucleotide site. F_{ST} values were calculated using Hudson's method as described in PopGenome documentation(41). These measures were calculated for the global sample, and for each continental population.

Functional Effect Prediction:

Functional effects of variants were predicted using the SIFT and PolyPhen2 algorithms. SIFT uses sequence based homology approach and the physiochemical similarity of between alternate amino acids to predict the effect of non-synonymous substitutions on protein function(42, 43). PolyPhen2 uses homology-based methods, structural features, and conservation profiles in order to predict the effects of non-synonymous mutations. It predicts the functional significance of an amino acid substitution using a Naïve Bayes classifier trained using supervised machine learning(44).

SIFT scores range from 0-1 and are classified as deleterious at values <0.05. PolyPhen2 ranges from 0-1 and measures values >0.446 as "tolerated", 0.446-0.908 as "possibly damaging", and values greater than 0.908 as "probably damaging".

Due to the large number of SNPs used in this study, we report only on sites in which these two prediction algorithms are in agreement. This analysis was conducted using Ensembl's Variant Effect Predictor (release 92), an open source command-line based tool written in the Perl language which specializes in genomic annotation(45).

Statistical Analysis and Reporting:

Data were taken into R for statistical analysis. VEP results were filtered in R and various descriptive statistics were drawn. In order to measure concordance between the two predictive algorithms we employed a Spearman rank correlation test. We computed specificity and sensitivity of the prediction tools using the ClinVar annotations in the Ensembl database to classify benign and damaging variants. We defined our true positive (TP) as predictions of damaging variants that a tool gave for variants that have been verified in the literature. A false

positive (FP) was assigned when the prediction tool gave a positive assessment for a clinically benign variant. A true negative (TN) was assigned as the prediction tools assigning a benign classification to variant that is classified as benign in the dataset. A false negative was given when the predictor assigned a benign classification to a variant that is has been described in the literature as clinically significant.

Results:

Functional Annotation:

The functional consequence of non-synonymous variants was predicted for 47 genes. Because pseudogenes do not code for functional proteins, they were excluded from this analysis. The same holds true for PKDL13, which is not correctly annotated in the GRCh37 human genome annotation.

We loaded a total of 2410 variants rested across the coding regions of these genes. 1437 SNPs, were non-synonymous SNPs. 6 of these were start site loss variants. SIFT and PolyPhen scores were only available for the 1437 SNPs that result in amino acid substitutions 761 SNPs were synonymous. For the rest of the variants we did not assess their potential impact, but there was 17 frame shift deletions, 7 in-frame deletions, 59 premature stop gain variants, 4 stop loss variants, and 125 splice region variants catalogued in our sample **Supplementary 1**.

A total of 767 of these sites were predicted to have functional effect by either SIFT or PolyPhen. SIFT calculated 624 deleterious variant sites and PolyPhen calculated 606 damaging sites. Together both algorithms predicted 463 sites that are damaging towards protein function when in agreement. Due to the high likelihood of false negatives and positives, and the vast majority of sites to cover, we will discuss variants that are in agreement between the algorithms. For a

complete list of predictions from this analysis see Supplementary Figure 1. Of these 463 non-synonymous variants, 84 of them have MAF>0.1%.

Bitter Taste:

We predicted the functional consequence of 24 bitter taste receptors. Pseudogenes by definition do not encode functional proteins TAS2R2P, TAS2R12P, TAS2R15P, TAS2R18P, TAS2R62P, TAS2R63P, TAS2R64P, and TAS2R67P were excluded from this analysis. 772 variants were assigned by VEP to have low to high effect on bitter receptor genes. 247 of these were non-synonymous SNPs that are predicted by either SIFT or PolyPhen to have a damaging effect on protein function. Both tools to have a damaging effect on protein function predicted 133 of these. The number of predicted damaging variants ranged between 2-10 per TAS2R GPCR with and average of ~6 damaging variant per gene. 65 of these variants are singletons, and by this definition are only present on a single chromosome in a single individual. Of the remaining 68 non-synonymous SNPs predicted to have damaging effects, 38 of them are found at an allele frequency greater than or equal to 0.1%.

Table 3 reports the 38 non-synonymous SNPs predicted to be damaging with MAF>0.1%. Non-synonymous SNPs meeting these criteria were found in all TAS2Rs except TAS2R3, TAS2R4, TAS2R20, TAS2R40, TAS2R50, and TAS2R60. Supplemental Table 1 gives all predicted variants showing damaging affect in all bitter receptor genes included in this study. Two TAS2R variants predicted by both tools to have functional consequence on the their taste receptor protein were found in all study populations for TAS2R7 and TAS2R42. On TAS2R7, rs77050900 was found at a global allele frequency ~1.9% and ranged from 0.2-0.75% across our 5 super-populations. In our study populations, rs1669412, encoded in TAS2R42, was the common

predicted damaging variant in TAS2R GPCRs with frequencies ranging from 19-25% across the 5 super populations. 4 variants were found in all populations except Eastern Asian populations.

In TAS2R43 there are 6 copies of a start site loss variation. This mutation was in 6 heterozygous individuals all of African descent. Half of these were found in Nigerian populations with 2 of these individuals found in the Yoruban sample and 1 in the Esan sample. 2 individuals are of African descent living in the United States, and the final individual comes from the Barbados sample.

*Table 2

Sweet and Umami Taste Receptors:

The TAS1R taste receptor family is associated with sweet and umami sensitivity and perception. Three taste receptors located on chromosome 1 in the human genome make up this gene family; TAS1R1, TAS1R2, and TAS1R3. Analysis using VEP returned a total of 418 variants were present in the coding region of these genes. Across the 3 TAS1R genes there were a total of 253 non-synonymous variants. A total 134 variants were predicted to have damaging effect by either at least one tool. 96 non-synonymous variants were predicted both tools predicted to have damaging effect on receptor function across these three genes. These genes ranged from 27-40 potentially damaging non-synonymous variants with an average of 32 variants per gene. Most of these genes are rare variants with MAF<0.1% or singletons present on only one chromosome in the entire sample. There were 54 singletons, ~56% of the variants, present in our predictions.

Among the three TAS1R genes our prediction of variants potentially affect protein function resulted in 17 non-synonymous variants with an MAF>0.1%. 9 of these variants reside in TAS1R1, 2 in TAS1R2, and 8 reside in TAS1R3 exons. Rs35118458 in TAS1R1 was present in

all five populations at least once and rs147600530 in TAS1R1 had the highest minor allele frequency and its 76/79 the total copies of this variant were within the African population and 3 in the Admixed-American population.

*Table 3

Salty and Sour Taste Receptors

The proposed receptors for the taste of salt in humans is assembled as a heterotrimer composed of three subunits; α, ß, γ. 219 variants were reported in the coding sites of these three genes.. Among these variants 127 of them were non-synonymous variants with SCNN1A, SCNN1B, and SCNN1G carrying 52, 41, and 34 non-synonymous variants respectively. Across the three genes 70 variants were predicted to be damaging by either tool. In 44 of these sites the two prediction tools in agreement one another. For SCNN1A, B, and G the predicted damaging variants present in each gene were 23, 15, and 6 possibly damaging variants in each gene respectively. None of these had a MAF>0.1% among the entire sample in SCNN1G while SCNN1A possessed 3 predicted variants with MAF>0.1% and SCNN1B having two variants above this threshold.

Our analysis included only one sour receptor, PKD2L1. This was because PKD1L3 was not annotated in Ensembl when aligned to GRCh37 human genome release. The gene sequence contained 116 coding variants with low to high effect predicted by VEP. PKD2L1 possessed 65 non-synonymous variants. 47 of these variants were predicted by one of the tools to be damaging. There were 27 of these variable sites predicted by both tools.

* Table 4

Fat Taste Receptors:

Across fatty acid taste receptors VEP reported 144 variants, 35 in GPR120 and 109 in CD36, that were present in coding region CD36. Upon filtering this list of variable sites to only include non-synonymous SNPs, we report 76 non-synonymous variant sites between both genes. Across GPR120 there were 8 non-synonymous sites reporting a predicted damaging effect on normal protein function and across CD36 there were 47 non-synonymous sites that were predicted to have damaging effect by either tool. When variant predictions were filtered for agreement between both prediction algorithms about 44 variants were reported to have possible damaging effect on protein function. 5 of these variants are in GPR120 and 39 are in CD36.

The majority of these sites are very rare with MAF<0.1%. After we applied our filter, the sites in GPR120 all resulted in an MAF<0.1% and only 6 non-synonymous sites in CD36. Rs41478146 was present as 15 copies in the 1000 genomes project. These copies were almost exclusive to the African population with 14/15 copies of this variant being isolated to that continent and 1 copy of this variant is present in the American population.

*Table 5

Chemesthetic Receptors:

Transient receptor potential cation channels are activated by a wide variety of stimulus and found throughout the body playing many different biological and molecular roles. VEP returned 337 coding variants present these three genes; 106 variants, 122 variants, and 109 variants in TRPV1, TRPM8, and TRPA1 respectively. There were a total of 172 non-synonymous with 57, 59, and 56 variants in TRPV1, TRPM8, and TRPA1. 55 of these sites were predicted to have damaging effects by both tools.

The majority of these predicted variants were rare variants. For TRPV1, the 15/17 variants were singletons and none of them had an MAF>0.1. The same holds true for TRPA1 by which 16/19 variants are singletons and 2/19 sites have an MAF>0.01 and are isolated between populations.

Rs201141727 was present across all continental super-populations at frequencies between 1.8% among the European group and 8.3% among the American population and Asian populations.

*Table 6

Signal Transduction Molecules and Multi-pathway Molecules:

Various genes play mediating and moderating roles the signaling pathway following chemical stimulus to gustatory receptor. From these genes we derived 408 variant sites of polymorphism in the coding region. 207 of these are non-synonymous nucleotide polymorphism across 11 genes. 62 of these genes were predicted to have pathogenic effect on protein structure. When an MAF>0.1%, this resulted in 5 variants predicted to have damaging effects. GNB3 ss1388024686, PLCB3 rs12146487, CALHM1 rs535176093, CALHM1 rs145546138, RTP4 rs145014578, RTP4 rs1003995, REEP1 rs144874997. PLCB3 rs12146487 and RTP4 rs1003995 were the only two segregating sites that were shared between all populations.

Assessing the Predictors:

Although SIFT and PolyPhen-2 used separate approaches for predicting pathogenic variants, we observed concordance on their prediction choices. In order assess a correlation in SIFT and PolyPhen-2 scores, we employed a employed a spearman rank coefficient test. Our findings point to a significant correlation between the two predictors ($rho=-0.6996$, $p<0.01$). We employed a specificity and sensitivity analysis for the prediction tools used. A total of 25 ClinVar entries were included with 7 SNPS labeled as pathogenic or likely pathogenic. We

found sensitivity for PolyPhen-2 at 55.56% and specificity rate at 85.71%. The accuracy of the tool was then 64% accurate. For SIFT, specificity was measured at 50% and sensitivity was measured at 71.43%. Accuracy of SIFT predictions was 56% accurate. We employed both tools for prediction together, and returned positive prediction if either tool predicted damaging effect. For these parameters we found they returned a sensitivity at 85.71% and specificity at 44%. The accuracy prediction gave us 56% accurate predictions utilizing them both.

Population Analysis:

We analyzed a total of 38665 single nucleotide polymorphisms across the 56 genes. Nucleotide diversity, which measures the average pairwise differences between sequences, ranged between 0.01%-0.27% across the genes. These findings varied across populations with the African population demonstrating the highest average level of nucleotide diversity, and the Eastern Asian population demonstrating the lowest average level of nucleotide diversity. The TAS2R bitter receptors and represented the lowest and highest values on this spectrum.

F_{ST}, which is a measurement of genetic differentiation between populations, ranged from F_{ST} 0.007-0.34 across the global population in chemosensory genes. The African sample revealed the largest average F_{ST} across chemosensory genes and the Americas and Europe represented the lowest. The TAS2R pseudogenes represented both the low and high extremes of this spectrum.

Among genes TAS2R39 π=0.01% has the lowest average diversity per site and TAS2R20 had the highest π=0.27%. TAS2R15 demonstrated the lowest F_{ST} around F_{ST}=0.007 and TAS2R43 the highest at F_{ST}=0.291. TAS2R15 however is a pseudogene, and the lowest F_{ST} on a functional gene was TAS2R40 F_{ST}=0.028. Across all taste genes the average nucleotide diversity per site was π=0.12%. This value was π=0.10% when pseudogenes were excluded.

The analysis revealed that predicted sites classified as deleterious by one or both tools maintained high frequency and displayed high levels of differentiation. This occurred on rs35969491 on TAS2R42 (AF=0.54 F_{ST}~0.26). This occurred on rs68157013 TAS2R43 (AF=0.62 EAS-AF=94.25, F_{ST}~0.50). Similar levels of measured genetic differentiation occurred at F_{ST}>0.20 in PKD2L1, TAS1R1, TAS1R3, TAS2R1, TAS2R20, TAS2R3, TRPV1, and RTP4.

*Table 7

Discussion:

We sought to assess if we can predict variation in genes encoding taste perception pathways. To do this we put forth two predictive tools to elucidate candidate variants affecting protein function in taste genes. Using the Ensembl Variant Effect Predictor we annotated and predicted the functional impact of 1437 non-synonymous sites. This produced a list of 349 candidates predicted to be damaging to protein function and structure.

A concordance analysis revealed they were significantly correlated. We conducted a sensitivity analysis by which we compared our predictions with those established and published in ClinVar(46). Sensitivity did not increase when these two algorithms were used together. Instead this revealed that PolyPhen2 was most accurate at predicting functional effect within this particular set of genes.

We found that the genes that control taste pathways are diverse. Our analysis revealed major differences in nucleotide diversity, differentiation, and frequency of non-synonymous mutation. Interestingly, we found most non-synonymous mutations in these genes are at low frequency, with the exception of some intermediate sites. Among our genes, an excess of nonsynonymous variation was found in TAS2R13 which has 13 non-synonymous substitutions catalogued in the

1000 genomes repository and a single synonymous substitution. Most of the predicted damaging variants were present in less than 0.1% of the population and more commonly as singletons or isolated to a single geographic group. Additionally we found a number of putatively high impact variants in the taste gene repertoire. One such variation was the loss of the transcription start site on TAS2R43 that was present in multiple subpopulations of African descent. Such variants likely confer a loss of function for the gene.

We examined population genetic parameters to examine their effect on the detection of pathogenic variant. Variants of interest are those that were classified as deleterious and had a higher average F_{ST} for our gene sample. We particularly saw this occur in variants in TAS2R42 and TAS2R43. This differentiation between populations and high frequency of high impact variants may be indicative that some selective or demographic process has acted up these genes. These variants, classified as deleterious by our prediction tools, indicate that this method of variant prediction may describe outcomes that may offer a fitness or protective advantage in certain instances instead of being truly “deleterious”. This has been proposed previously in which deleteriously predicted variants, classified by SIFT and Polyphen2, frequently revealed signals of selection, which will drive up the frequency of mutations offering a fitness advantage and drive down the frequency of harmful mutations (47). Positive selection on TAS2R42 has been reported in the literature.(48). While selective processes and local adaptation may have contributed to the high differentiation levels of some of the putatively high impact SNPs in this study, demographic processes and genetic drift may also contribute to these phenomenon. Measuring and testing for signatures of natural selection was out of the scope of this study, but this detail requires further attention to better decipher these findings.

In all, we were able to elucidate meaningful information from the list of taste genes using variant annotation and prediction tools. The 1000 genomes data set is both a free and public data source for which to explore global genetic variation. Similarly, other biological data banks could further characterize functional variation in these genes. We integrated annotation sources catalogued in the Ensembl site through VEP. These tools, being open source and well documented, are a great resource for clinical and molecular researchers, and have demonstrated a resource efficient mechanism for elucidating and prioritizing causal candidates for traits and diseases.

In recent years there has been increased interest on the contribution that variation in chemosensory cues may have on health outcomes and to clinical research. These genes may play important roles in shaping behaviors and health outcomes. The bitter taste family has been previously implicated in preferences for food, tobacco, and alcohol(49, 50). This gives health researchers an opportunity to seek new fronts to solve issues like the rates obesity from over-nutrition and alcoholism. In the field of taste genetics there has been much documentation that genetic variability in genes involved in gustatory mechanism may also have effects in other physiological pathways, and have found receptors for taste in tissues outside of the oral epithelium(51, 52). For instance the expression of bitter GPCRs in the smooth airway, and their subsequent activation is of interest as a target for controlling inflammation(53, 54). Thus, these findings have implications for health science as a whole.

Acknowledgement:

I thank my committee members for providing feedback on the public health aspects of this manuscript. I thank Vania Hoang, who aided in creating and editing Figure 2.

References:

1. George Priya Doss,C. and Rajith,B. (2012) Computational Refinement of Functional Single Nucleotide Polymorphisms Associated with ATM Gene. *PLoS One*, **7**, e34573.

2. Akhoundi,F., Parvaneh,N. and Modjtaba,E.-B. (2016) In silico analysis of deleterious single nucleotide polymorphisms in human BUB1 mitotic checkpoint serine/threonine kinase B gene. *Meta Gene*, **9**, 142–150.

3. Abdul Samad,F., Suliman,B.A., Basha,S.H., Manivasagam,T. and Essa,M.M. (2016) A Comprehensive In Silico Analysis on the Structural and Functional Impact of SNPs in the Congenital Heart Defects Associated with NKX2-5 Gene—A Molecular Dynamic Simulation Approach. *PLoS One*, **11**, e0153999.

4. Cantor,R.M., Lange,K. and Sinsheimer,J.S. (2010) Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application. *Am. J. Hum. Genet.*, **86**, 6–22.

5. Iczkowski,K.A. (2011) Cell adhesion molecule CD44: its functional roles in prostate cancer. *Am. J. Transl. Res.*, **3**, 1–7.

6. Breslin,P.A.S. (2013) An Evolutionary Perspective on Food Review and Human Taste. *Curr. Biol.*, **23**, R409–R418.

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

8. Risso,D., Tofanelli,S., Morini,G., Luiselli,D. and Drayna,D. (2014) Genetic variation in taste receptor pseudogenes provides evidence for a dynamic role in human evolution. *BMC Evol. Biol.*, **14**, 198.

9. Wooding,S. (2006) Phenylthiocarbamide: A 75-Year Adventure in Genetics and Natural

Selection. *Genetics*, **172**, 2015–2023.

10. Running,C.A., Craig,B.A. and Mattes,R.D. (2015) Oleogustus: The unique taste of fat. *Chem. Senses*, **40**, 507–516.

11. Gilbertson,T.A. (1998) Gustatory mechanisms for the detection of fat. *Curr. Opin. Neurobiol.*, **8**, 447–452.

12. Pepino,M.Y., Love-Gregory,L., Klein,S. and Abumrad,N.A. (2012) The fatty acid translocase gene CD36 and lingual lipase influence oral sensitivity to fat in obese subjects. *J. Lipid Res.* , **53**, 561–566.

13. Kim,U., Wooding,S., Riaz,N., Jorde,L.B. and Drayna,D. (2006) Variation in the human TAS1R taste receptor genes. *Chem. Senses*, **31**, 599–611.

14. He,W., Yasumatsu,K., Varadarajan,V., Yamada,A., Lem,J., Ninomiya,Y., Margolskee,R.F. and Damak,S. (2004) Umami Taste Responses Are Mediated by α-Transducin and α-Gustducin. *J. Neurosci.*, **24**, 7674 LP – 7680.

15. Bufe,B., Breslin,P.A.S., Kuhn,C., Reed,D.R., Tharp,C.D., Slack,J.P., Kim,U.K., Drayna,D. and Meyerhof,W. (2005) The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. *Curr. Biol.*, **15**, 322–327.

16. Risso,D.S., Mezzavilla,M., Pagani,L., Robino,A., Morini,G., Tofanelli,S., Carrai,M., Campa,D., Barale,R., Caradonna,F., *et al.* (2016) Global diversity in the TAS2R38 bitter taste receptor: revisiting a classic evolutionary PROPosal. *Sci. Rep.*, **6**, 25506.

17. Kim,U., Jorgenson,E., Coon,H., Leppert,M., Risch,N. and Drayna,D. (2003) Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science*, **299**, 1221–1225.

18. Caicedo,A., Pereira,E., Margolskee,R.F. and Roper,S.D. (2003) Role of the G-Protein Subunit α-Gustducin in Taste Cell Responses to Bitter Stimuli. *J. Neurosci.*, **23**, 9947 LP – 9952.

19. Ishimaru,Y., Abe,M., Asakura,T., Imai,H. and Abe,K. (2012) Expression Analysis of Taste Signal Transduction Molecules in the Fungiform and Circumvallate Papillae of the Rhesus

Macaque, Macaca mulatta. *PLoS One*, **7**, e45426.

20. Dunham,J.H. and Hall,R.A. (2009) Enhancement of G Protein-Coupled Receptor Surface Expression. *Trends Biotechnol.*, **27**, 541–545.

21. Geerling,J.C. and Loewy,A.D. (2008) Central regulation of sodium appetite. *Exp. Physiol.*, **93**, 177–209.

22. Butterworth,M.B. (2010) Regulation of the epithelial sodium channel (ENaC) by membrane trafficking. *Biochim. Biophys. Acta*, **1802**, 1166–1177.

23. Ishii,S., Misaka,T., Kishi,M., Kaga,T., Ishimaru,Y. and Abe,K. (2009) Acetic acid activates PKD1L3-PKD2L1 channel-A candidate sour taste receptor. *Biochem. Biophys. Res. Commun.*, **385**, 346–350.

24. Ishimaru,Y., Katano,Y., Yamamoto,K., Akiba,M., Misaka,T., Roberts,R.W., Asakura,T., Matsunami,H. and Abe,K. (2010) Interaction between PKD1L3 and PKD2L1 through their transmembrane domains is required for localization of PKD2L1 at taste pores in taste cells of circumvallate and foliate papillae. *FASEB J.*, **24**, 4058–4067.

25. Horio,N., Yoshida,R., Yasumatsu,K., Yanagawa,Y., Ishimaru,Y., Matsunami,H. and Ninomiya,Y. (2011) Sour Taste Responses in Mice Lacking PKD Channels. *PLoS One*, **6**, e20007.

26. Cartoni,C., Yasumatsu,K., Ohkuri,T., Shigemura,N., Yoshida,R., Godinot,N., le Coutre,J., Ninomiya,Y. and Damak,S. (2010) Taste Preference for Fatty Acids Is Mediated by GPR40 and GPR120. *J. Neurosci.*, **30**, 8376–8382.

27. Ozdener,M.H., Subramaniam,S., Sundaresan,S., Sery,O., Hashimoto,T., Asakawa,Y., Besnard,P., Abumrad,N.A. and Khan,N.A. (2014) CD36- and GPR120-mediated Ca(2+) Signaling in Human Taste Bud Cells Mediates Differential Responses to Fatty Acids and is Altered in Obese Mice. *Gastroenterology*, **146**, 995-1005.e5.

28. Silverstein,R.L. and Febbraio,M. (2009) CD36, a Scavenger Receptor Involved in Immunity, Metabolism, Angiogenesis, and Behavior. *Sci. Signal.*, **2**, re3–re3.

29. Laugerette,F., Passilly-Degrace,P., Patris,B., Niot,I., Febbraio,M., Montmayeur,J.-P. and

Besnard,P. (2005) CD36 involvement in orosensory detection of dietary lipids, spontaneous fat preference, and digestive secretions. *J. Clin. Invest.*, **115**, 3177–3184.

30. Sclafani,A., Ackroff,K. and Abumrad,N.A. (2007) CD36 gene deletion reduces fat preference and intake but not post-oral fat conditioning in mice. *Am. J. Physiol. - Regul. Integr. Comp. Physiol.*, **293**, R1823–R1832.

31. Sayed,A., Šerý,O., Plesnik,J., Daoudi,H., Rouabah,A., Rouabah,L. and Khan,N.A. (2015) CD36 AA genotype is associated with decreased lipid taste perception in young obese, but not lean, children. *Int. J. Obes.*, **39**, 920–924.

32. Immke,D.C. and Gavva,N.R. (2006) The TRPV1 receptor and nociception. *Semin. Cell Dev. Biol.*, **17**, 582–591.

33. Rosenzweig,M., Kang,K. and Garrity,P.A. (2008) Distinct TRP channels are required for warm and cool avoidance in Drosophila melanogaster. *Proc. Natl. Acad. Sci.*, **105**, 14668–14673.

34. Saito,S., Fukuta,N., Shingai,R. and Tominaga,M. (2011) Evolution of Vertebrate Transient Receptor Potential Vanilloid 3 Channels: Opposite Temperature Sensitivity between Mammals and Western Clawed Frogs. *PLOS Genet.*, **7**, e1002041.

35. Michlig,S., Merlini,J.M., Beaumont,M., Ledda,M., Tavenard,A., Mukherjee,R., Camacho,S. and Le Coutre,J. (2016) Effects of TRP channel agonist ingestion on metabolism and autonomic nervous system in a randomized clinical trial of healthy subjects. *Sci. Rep.*, **6**.

36. Zerbino,D.R., Achuthan,P., Akanni,W., Amode,M.R., Barrell,D., Bhai,J., Billis,K., Cummins,C., Gall,A., Girón,C.G., *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.

37. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.

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

39. Li,H. (2011) Tabix: Fast retrieval of sequence features from generic TAB-delimited files.

Bioinformatics, **27**, 718–719.

40. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T., *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

41. Pfeifer,B., Wittelsbürger,U., Ramos-Onsins,S.E. and Lercher,M.J. (2014) PopGenome: An efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.*, **31**, 1929–1936.

42. Ng,P.C. and Henikoff,S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, **31**, 3812–3814.

43. Kumar,P., Henikoff,S. and Ng,P.C. (2009) SIFT. *Nat. Protoc.*, **4**, 1073–1081.

44. Adzhubei,I., Jordan,D.M. and Sunyaev,S.R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.*, 10.1002/0471142905.hg0720s76.

45. McLaren,W., Gil,L., Hunt,S.E., Riat,H.S., Ritchie,G.R.S., Thormann,A., Flicek,P. and Cunningham,F. (2016) The Ensembl Variant Effect Predictor. *Genome Biol.*, **17**.

46. Landrum,M.J., Lee,J.M., Benson,M., Brown,G.R., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Jang,W., *et al.* (2018) ClinVar. *Nucleic Acids Res.*, **46**, D1062–D1067.

47. Ryu,H.J., Kim,Y.J., Park,Y.K., Kim,J.J., Park,M.Y., Seo,E.J., Yoo,H.W., Park,I.S., Oh,B. and Lee,J.K. (2008) Identification and Characterization of Human Genes Targeted by Natural Selection. *Genomics Inf.*, **6**, 173–180.

48. Kosiol,C., Vinař,T., da Fonseca,R.R., Hubisz,M.J., Bustamante,C.D., Nielsen,R. and Siepel,A. (2008) Patterns of Positive Selection in Six Mammalian Genomes. *PLOS Genet.*, **4**, e1000144.

49. Martinez-Ruiz,N.R., Lopez-Diaz,J.A., Wall-Medrano,A., Jimenez-Castro,J.A. and Angulo,O. (2014) Oral fat perception is related with body mass index, preference and consumption of high-fat foods. *Physiol. Behav.*, **129**, 36–42.

50. Dotson,C.D., Shaw,H.L., Mitchell,B.D., Munger,S.D. and Steinle,N.I. (2010) Variation in

the gene TAS2R38 is associated with the eating behavior disinhibition in Old Order Amish women. *Appetite*, **54**, 93–99.

51. Wölfle,U., Elsholz,F.A., Kersten,A., Haarhaus,B., Schumacher,U. and Schempp,C.M. (2016) Expression and functional activity of the human bitter taste receptor TAS2R38 in human placental tissues and JEG-3 cells. *Molecules*, **21**.

52. Gilca,M. and Dragos,D. (2017) Extraoral Taste Receptor Discovery: New Light on Ayurvedic Pharmacology. *Evid. Based. Complement. Alternat. Med.*, **2017**, 5435831.

53. Sharma,P., Yi,R., Nayak,A.P., Wang,N., Tang,F., Knight,M.J., Pan,S., Oliver,B. and Deshpande,D.A. (2017) Bitter Taste Receptor Agonists Mitigate Features of Allergic Asthma in Mice. *Sci. Rep.*, **7**, 46166.

54. An,S.S. and Liggett,S.B. (2018) Taste and smell GPCRs in the lung: Evidence for a previously unrecognized widespread chemosensory system. *Cell. Signal.*, **41**, 82–88.

Figure 1 Chromosome Locations of Human Taste Genes

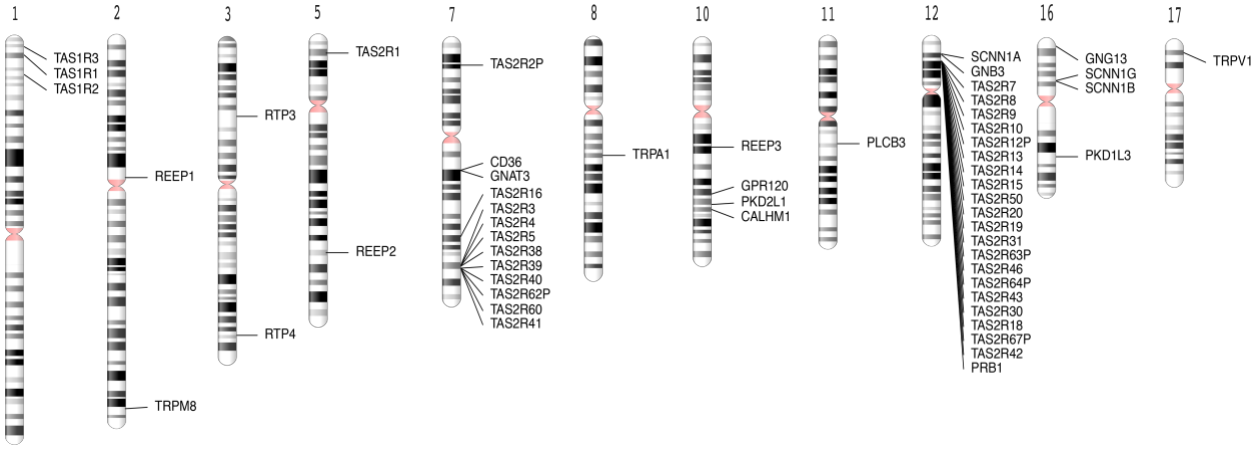

Figure 1 reveals the chromosomal locations of the human taste genes. The TAS2R as can be seen are spread through chromosomes 5, 7, and 12, forming clusters on chromosomes 7 and 12.

Figure 2 World Map of 1000 Genomes Project Populations

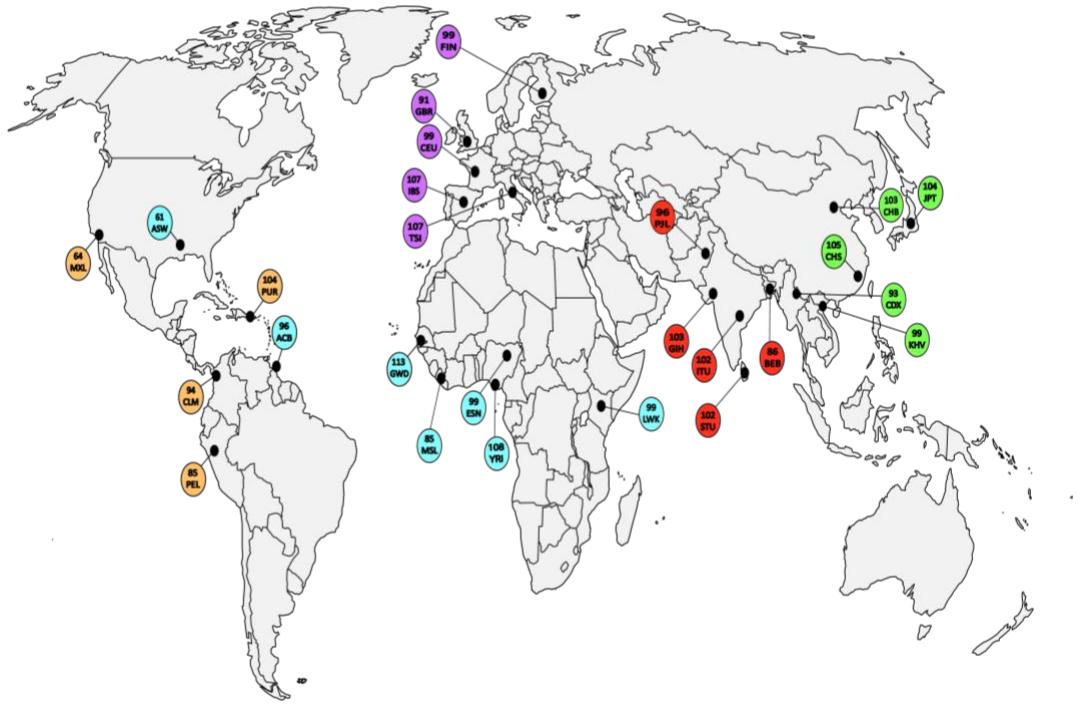

Figure 2 is a world map of the 1000 genomes project samples. These samples are aggregated into 5 super populations of Africa, America, Eastern Asia, Europe, and Southern Asia. These are African (blue), Admixed-American (orange), Eastern Asian(green), European (purple), and South Asian (red). These populations are broken down in Table 1

Table 1 1000 Genomes Project Populations

Population	N
Africa (N=661)	
Yoruba in Ibadan, Nigeria	108
Luhya in Webuye, Kenya	99
Gambian in Western Divisions in the Gambia	113
Mende in Sierra Leone	85
Esan in Nigeria	99
Americans of African Ancestry in Southwestern USA	61
African Caribbeans in Barbados	96
Americas (N=347)	
Mexican Ancestry from Los Angeles USA	64
Puerto Ricans from Puerto Rico	104
Colombians from Medellin, Colombia	94
Peruvians from Lima, Peru	85
East Asian (N=504)	
Han Chinese in Beijing, China	103
Japanese in Tokyo, Japan	104
Southern Han Chinese	105
Chinese Dai in Xishuangbanna, China	93
Kinh in Ho Chi Minh City, Vietnam	99
Europe (N=503)	
Utah Residents (CEPH) with European Ancestry	99
Toscani in Italia	107
Finnish in Finland	99
British in England and Scotland	91
Iberian Population in Spain	107
South Asian (N=489)	
Gujarati Indian from Houston, Texas	103
Punjabi from Lahore, Pakistan	96
Bengali from Bangladesh	86
Sri Lankan Tamil from the UK	102
Indian Telugu from the UK	102
Total	2504

Table 1 documents the population included in the 1000 genomes project. In the Phase III 1KG there are 26 populations, categorized into 5 super populations, that are representative of global variation. There are a total of 2504 unrelated individuals in this dataset.

Table 2 Prediction of Pathogenic Variants In Bitter Receptors

Gene	rsID	AF	EAS_AF	EUR_AF	AFR_AF	AMR_AF	SAS_AF	MPWD	F_{ST}	Amino.Acids
TAS2R1	rs2234232	0.0076	0.0000	0.0000	0.0265	0.0043	0.0000	0.0151	0.0207	C/Y
	rs2234231	0.0010	0.0000	0.0000	0.0038	0.0000	0.0000	0.0020	0.0025	P/L
TAS2R5	rs2234013	0.0030	0.0000	0.0080	0.0000	0.0058	0.0031	0.0060	0.0033	G/S
	rs2234014	0.0080	0.0000	0.0010	0.0280	0.0029	0.0000	0.0159	0.0219	P/L
TAS2R7	rs139604652	0.0046	0.0000	0.0109	0.0000	0.0115	0.0041	0.0091	0.0058	V/E
	rs202246571	0.0034	0.0010	0.0000	0.0000	0.0000	0.0164	0.0068	0.0141	L/H
	rs77050900	0.0194	0.0754	0.0070	0.0023	0.0058	0.0072	0.0380	0.0513	I/T
TAS2R8	rs114977408	0.0020	0.0000	0.0000	0.0076	0.0000	0.0000	0.0040	0.0060	Y/H
	rs61737282	0.0026	0.0000	0.0000	0.0091	0.0014	0.0000	0.0052	0.0063	I/T
	rs142540719	0.0054	0.0000	0.0089	0.0008	0.0101	0.0102	0.0107	0.0041	R/I
TAS2R9	rs113883583	0.0122	0.0000	0.0030	0.0257	0.0043	0.0215	0.0241	0.0117	G/E
	rs148917754	0.0022	0.0099	0.0000	0.0000	0.0000	0.0010	0.0044	0.0076	S/L
TAS2R10	rs114006371	0.0038	0.0000	0.0000	0.0129	0.0029	0.0000	0.0076	0.0091	I/N
	rs142507813	0.0038	0.0179	0.0000	0.0000	0.0000	0.0010	0.0076	0.0156	M/I
	rs117936881	0.0082	0.0407	0.0000	0.0000	0.0000	0.0000	0.0162	0.0399	W/R
TAS2R13	rs34885344	0.0042	0.0000	0.0000	0.0129	0.0058	0.0000	0.0084	0.0083	H/R
TAS2R14	rs35804287	0.0060	0.0000	0.0249	0.0000	0.0043	0.0020	0.0119	0.0182	L/F
TAS2R16	rs28371575	0.0014	0.0000	0.0000	0.0000	0.0000	0.0072	0.0028	0.0063	L/P
TAS2R19	rs115193179	0.0038	0.0000	0.0119	0.0008	0.0058	0.0020	0.0076	0.0056	K/T
	rs192199862	0.0014	0.0000	0.0000	0.0000	0.0101	0.0000	0.0028	0.0100	G/D
TAS2R30	rs568940139	0.0010	0.0000	0.0000	0.0000	0.0000	0.0051	0.0020	0.0042	F/L
TAS2R31	rs139069360	0.0144	0.0000	0.0249	0.0159	0.0072	0.0215	0.0283	0.0062	W/C
	rs116926686	0.0102	0.0000	0.0249	0.0000	0.0072	0.0215	0.0202	0.0137	L/F
	rs202165116	0.0020	0.0000	0.0000	0.0076	0.0000	0.0000	0.0040	0.0060	L/S
	rs140958087	0.0034	0.0159	0.0000	0.0000	0.0000	0.0010	0.0068	0.0137	G/V
	rs143614038	0.0010	0.0000	0.0030	0.0000	0.0029	0.0000	0.0020	0.0015	D/H
TAS2R38	rs115966953	0.0020	0.0000	0.0000	0.0076	0.0000	0.0000	0.0040	0.0060	R/Q
TAS2R39	rs200380921	0.0014	0.0000	0.0000	0.0053	0.0000	0.0000	0.0028	0.0039	L/R
	rs184819681	0.0058	0.0000	0.0030	0.0151	0.0086	0.0000	0.0115	0.0076	S/N
TAS2R41	rs75955374	0.0040	0.0000	0.0000	0.0144	0.0014	0.0000	0.0080	0.0113	V/D
	rs189299466	0.0014	0.0010	0.0060	0.0000	0.0000	0.0000	0.0028	0.0039	S/Y
	rs146786143	0.0016	0.0000	0.0000	0.0061	0.0000	0.0000	0.0032	0.0046	I/S
	rs200985152	0.0012	0.0000	0.0060	0.0000	0.0000	0.0000	0.0024	0.0050	L/F
TAS2R42	rs139960283	0.0052	0.0000	0.0000	0.0197	0.0000	0.0000	0.0103	0.0173	L/W
	rs1669412	0.2200	0.1944	0.2167	0.2474	0.2305	0.2055	0.3433	0.0018	R/Q
TAS2R43	rs138563991	0.0050	0.0000	0.0199	0.0000	0.0058	0.0010	0.0099	0.0136	W/C
	rs139372865	0.0012	0.0000	0.0000	0.0045	0.0000	0.0000	0.0024	0.0032	M/I
TAS2R46	rs139412224	0.0012	0.0060	0.0000	0.0000	0.0000	0.0000	0.0024	0.0050	I/M

Predictions for the TAS2R family of bitter receptors. SIFT and PolyPhen-2 were in concordance for these predictions. MAF was cut off at 0.1%. All predictions can be found on Supplementary Table 1

Table 3 Prediction of Pathogenic Variants In Sweet and Umami Receptors

Gene	rsID	AF	EAS_AF	EUR_AF	AFR_AF	AMR_AF	SAS_AF	MPWD	F_{ST}	Amino.Acids
TAS1R1	rs140548974	**0.0022**	**0.0000**	**0.0000**	**0.0083**	**0.0000**	**0.0000**	**0.0044**	**0.0067**	**L/V**
	rs41307749	**0.0016**	**0.0000**	**0.0020**	**0.0008**	**0.0058**	**0.0010**	**0.0032**	**0.0015**	**S/C**
	rs114597256	**0.0082**	**0.0327**	**0.0000**	**0.0061**	**0.0000**	**0.0000**	**0.0162**	**0.0234**	**I/T**
	rs61740593	**0.0034**	**0.0000**	**0.0000**	**0.0121**	**0.0014**	**0.0000**	**0.0068**	**0.0092**	**R/G**
	rs35118458	**0.0120**	**0.0010**	**0.0209**	**0.0015**	**0.0144**	**0.0266**	**0.0237**	**0.0109**	**R/Q**
	rs150869784	**0.0010**	**0.0000**	**0.0000**	**0.0038**	**0.0000**	**0.0000**	**0.0020**	**0.0025**	**F/S**
	rs148892133	**0.0016**	**0.0000**	**0.0000**	**0.0045**	**0.0029**	**0.0000**	**0.0032**	**0.0022**	**V/A**
	rs41278022	**0.0020**	**0.0000**	**0.0070**	**0.0000**	**0.0000**	**0.0031**	**0.0040**	**0.0038**	**R/C**
	rs150612979	**0.0032**	**0.0060**	**0.0000**	**0.0000**	**0.0115**	**0.0020**	**0.0064**	**0.0054**	**P/L**
TAS1R2	rs200812417	**0.0022**	**0.0109**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0044**	**0.0100**	**R/H**
	rs74056647	**0.0048**	**0.0000**	**0.0000**	**0.0166**	**0.0029**	**0.0000**	**0.0095**	**0.0126**	**L/M**
TAS1R3	rs139632532	**0.0010**	**0.0000**	**0.0000**	**0.0038**	**0.0000**	**0.0000**	**0.0020**	**0.0025**	**C/R**
	rs140582284	**0.0010**	**0.0000**	**0.0000**	**0.0015**	**0.0043**	**0.0000**	**0.0020**	**0.0018**	**G/R**
	rs76584377	**0.0040**	**0.0000**	**0.0000**	**0.0144**	**0.0014**	**0.0000**	**0.0080**	**0.0113**	**K/N**
	rs186067621	**0.0026**	**0.0000**	**0.0000**	**0.0000**	**0.0173**	**0.0010**	**0.0052**	**0.0160**	**A/V**
	rs147600530	**0.0158**	**0.0000**	**0.0000**	**0.0575**	**0.0043**	**0.0000**	**0.0311**	**0.0494**	**G/C**
	rs139115619	**0.0012**	**0.0000**	**0.0000**	**0.0038**	**0.0014**	**0.0000**	**0.0024**	**0.0018**	**L/I**
	rs565867201	**0.0028**	**0.0000**	**0.0000**	**0.0106**	**0.0000**	**0.0000**	**0.0056**	**0.0089**	**W/L**
	rs149196451	**0.0012**	**0.0000**	**0.0020**	**0.0000**	**0.0058**	**0.0000**	**0.0024**	**0.0032**	**L/V**

Findings for the TAS1R family. These findings were filtered to include predictions by which SIFT and PolyPhen were in agreement and the minimal allele frequency for reporting was at 0.1% (MAF>0.1%). For a complete list of all variants found in this study refer to Supplemental Table 1.

Table 4 Predictions for Salt and Sour Genes

Genes	rsID	AF	EAS_AF	EUR_AF	AFR_AF	AMR_AF	SAS_AF	MPWD	F_{ST}	Amino.Acids
PKD2L1	rs147426900	0.0152	0.0000	0.0139	0.0280	0.0072	0.0204	0.0298	0.0078	R/C
	rs143064336	0.0016	0.0000	0.0000	0.0061	0.0000	0.0000	0.0032	0.0046	S/T
SCNN1A	rs148749888	0.0014	0.0000	0.0040	0.0000	0.0000	0.0031	0.0028	0.0018	S/L
	rs5742912	0.0134	0.0000	0.0249	0.0030	0.0259	0.0204	0.0264	0.0107	W/R
	rs55797039	0.0072	0.0000	0.0209	0.0000	0.0072	0.0102	0.0143	0.0102	R/W
SCNN1B	rs72654356	0.0028	0.0000	0.0000	0.0106	0.0000	0.0000	0.0056	0.0089	L/Q

Findings for the salt and sour taste genes. These findings were filtered to include predictions by which SIFT and PolyPhen were in agreement and the minimal allele frequency for reporting was at 0.1% (MAF>0.1%). SCNN1G did not contain sites with MAF>0.1, but its predictions can be found in Supplementary table 1.

Table 5 Prediction for Fat Taste Gene

Gene	rsID	AF	EAS_AF	EUR_AF	AFR_AF	AMR_AF	SAS_AF	MPWD	F_{ST}	Amino.Acids
CD36	rs75326924	0.00199681	0.0099	0	0	0	0	0.00398643	0.00901247	P/S
	rs70961715	0.00119808	0.006	0	0	0	0	0.00239377	0.00500353	R/P
	rs201765331	0.000998403	0.005	0	0	0	0	0.00199521	0.00400126	S/L
	rs41478146	0.00299521	0	0	0.0106	0.0014	0	0.00597367	0.00783704	Y/F
	rs148910227	0.00139776	0.003	0	0.0023	0.0014	0	0.00279218	0.000328296	R/W
	rs559916528	0.00159744	0	0	0	0	0.0082	0.00319042	0.00731539	G/A

Findings for the fatty acid taste gene. These findings were filtered to include predictions by which SIFT and PolyPhen were in agreement and the minimal allele frequency for reporting was at 0.1% (MAF>0.1%). GPR120 did contain relevant functional sites, but all at MAF<0.1%. Refer to Supplemental Table 1 for information of GPR120

Table 6 Prediction for TRP Channels and Multipath Way Genes

Genes	rsID	AF	EAS_AF	EUR_AF	AFR_AF	AMR_AF	SAS_AF	MPWD	F_{ST}	Amino.Acids
TRPM8	rs17868387	0.0481	0.0833	0.0368	0.0182	0.0836	0.0389	0.0916	0.0183	Y/C
	rs28902173	0.0060	0.0000	0.0000	0.0227	0.0000	0.0000	0.0119	0.0201	M/T
	rs201940567	0.0038	0.0000	0.0000	0.0008	0.0000	0.0184	0.0076	0.0163	R/C
TRPA1	rs61758121	0.0010	0.0000	0.0000	0.0038	0.0000	0.0000	0.0020	0.0025	R/C
	rs144498143	0.0010	0.0000	0.0000	0.0023	0.0029	0.0000	0.0020	0.0005	R/H
GNB3	ss1388024686	0.0196	0.0000	0.0666	0.0008	0.0317	0.0082	0.0384	0.0421	G/S
PLCB3	rs12146487	0.0769	0.0565	0.1551	0.0272	0.0836	0.0798	0.1420	0.0338	R/H
CALHM1	rs535176093	0.0022	0.0000	0.0000	0.0000	0.0000	0.0112	0.0044	0.0104	R/C
	rs145546138	0.0080	0.0387	0.0000	0.0000	0.0000	0.0010	0.0159	0.0365	R/H
RTP4	rs145014578	0.0078	0.0000	0.0109	0.0045	0.0014	0.0215	0.0155	0.0086	D/H
	rs1003995	0.2879	0.4504	0.3419	0.0348	0.4524	0.2904	0.4101	0.1573	A/D
REEP1	rs144874997	0.0012	0.0000	0.0000	0.0045	0.0000	0.0000	0.0024	0.0032	R/W

Findings for the TAS1R family. These findings were filtered to include predictions by which SIFT and PolyPhen were in agreement and the minimal allele frequency for reporting was at 0.1% (MAF>0.1%). RTP4 interestingly has a high frequency variant predicted to be pathogenic with high levels of differentiation.

Table 7 Genetic Diversity Across 56 Taste Genes

Gene	Taste	Chr	Segregating Sites						Nucleotide Diversity						F_{ST}					
			Global	AFR	AMR	EAS	EUR	SAS	Global	AFR	AMR	EAS	EUR	SAS	Global	AFR	AMR	EAS	EUR	SAS
TAS2R1	Bitter	5	93	43	27	24	17	26	0.04%	0.04%	0.04%	0.05%	0.03%	0.03%	0.125	0.103	0.07	0.22	0.11	0.1
TAS2R3	Bitter	7	79	37	24	27	19	25	0.08%	0.09%	0.07%	0.06%	0.07%	0.08%	0.085	0.125	0.05	0.12	0.06	0.06
TAS2R4	Bitter	7	99	48	37	29	29	33	0.12%	0.10%	0.12%	0.10%	0.11%	0.11%	0.103	0.18	0.06	0.14	0.06	0.06
TAS2R5	Bitter	7	99	47	34	26	26	30	0.10%	0.10%	0.09%	0.09%	0.09%	0.10%	0.112	0.184	0.07	0.15	0.07	0.07
TAS2R7	Bitter	12	100	53	28	24	24	24	0.03%	0.03%	0.02%	0.05%	0.01%	0.02%	0.039	0.043	0.03	0.05	0.03	0.04
TAS2R8	Bitter	12	95	36	30	25	25	38	0.06%	0.07%	0.04%	0.05%	0.04%	0.04%	0.138	0.245	0.08	0.11	0.11	0.09
TAS2R9	Bitter	12	97	40	32	24	21	32	0.05%	0.04%	0.05%	0.05%	0.04%	0.05%	0.131	0.266	0.07	0.12	0.1	0.08
TAS2R10	Bitter	12	85	37	23	26	15	21	0.05%	0.06%	0.04%	0.05%	0.03%	0.04%	0.093	0.173	0.06	0.07	0.08	0.06
TAS2R13	Bitter	12	96	40	23	30	22	25	0.06%	0.05%	0.06%	0.05%	0.06%	0.06%	0.178	0.351	0.1	0.17	0.11	0.12
TAS2R14	Bitter	12	100	46	26	25	29	32	0.07%	0.08%	0.07%	0.05%	0.06%	0.06%	0.068	0.134	0.04	0.06	0.04	0.04
TAS2R16	Bitter	7	94	38	26	27	20	32	0.08%	0.11%	0.06%	0.06%	0.06%	0.06%	0.109	0.184	0.07	0.11	0.07	0.08
TAS2R19	Bitter	12	82	46	30	19	21	26	0.12%	0.11%	0.12%	0.07%	0.12%	0.10%	0.112	0.143	0.08	0.15	0.1	0.09
TAS2R20	Bitter	12	109	57	44	40	38	38	0.27%	0.13%	0.25%	0.19%	0.26%	0.24%	0.23	0.386	0.13	0.29	0.13	0.19
TAS2R30	Bitter	12	114	59	39	41	40	38	0.23%	0.23%	0.24%	0.16%	0.24%	0.20%	0.084	0.086	0.06	0.12	0.08	0.07
TAS2R31	Bitter	12	114	56	35	34	44	45	0.17%	0.18%	0.18%	0.09%	0.20%	0.14%	0.08	0.082	0.06	0.11	0.08	0.06
TAS2R38	Bitter	7	83	43	24	24	13	23	0.06%	0.07%	0.05%	0.05%	0.05%	0.05%	0.076	0.05	0.08	0.08	0.06	0.11
TAS2R39	Bitter	7	83	31	19	17	18	18	0.01%	0.01%	0.01%	0.00%	0.02%	0.01%	0.061	0.052	0.03	0.08	0.1	0.04
TAS2R40	Bitter	7	77	34	23	11	14	17	0.01%	0.03%	0.01%	0.00%	0.00%	0.01%	0.028	0.036	0.02	0.04	0.02	0.02
TAS2R41	Bitter	7	85	39	22	27	24	30	0.11%	0.05%	0.12%	0.12%	0.12%	0.12%	0.056	0.121	0.03	0.05	0.04	0.04
TAS2R42	Bitter	12	78	38	19	33	25	30	0.17%	0.17%	0.16%	0.11%	0.15%	0.14%	0.156	0.29	0.09	0.15	0.1	0.1
TAS2R43	Bitter	12	105	66	42	22	41	50	0.12%	0.10%	0.10%	0.02%	0.12%	0.10%	0.291	0.478	0.17	0.35	0.17	0.19
TAS2R46	Bitter	12	94	44	32	24	25	26	0.07%	0.08%	0.07%	0.05%	0.07%	0.05%	0.078	0.114	0.06	0.09	0.07	0.06
TAS2R50	Bitter	12	77	32	26	25	21	26	0.11%	0.10%	0.10%	0.07%	0.09%	0.09%	0.158	0.268	0.09	0.17	0.1	0.13
TAS2R60	Bitter	7	82	45	16	22	15	19	0.06%	0.05%	0.05%	0.04%	0.06%	0.06%	0.081	0.124	0.05	0.09	0.06	0.08
TAS2R2P	Pseudogene	7	108	31	24	20	19	20	0.16%	0.06%	0.08%	0.06%	0.09%	0.08%	0.156	0.323	0.08	0.15	0.09	0.1
TAS2R12P	Pseudogene	12	67	48	38	30	36	39	0.08%	0.19%	0.15%	0.12%	0.16%	0.13%	0.073	0.115	0.06	0.07	0.06	0.06
TAS2R15	Pseudogene	12	84	46	32	30	25	28	0.19%	0.22%	0.19%	0.17%	0.17%	0.16%	0.007	0.011	0	0.01	0	0.01
TAS2R18	Pseudogene	12	78	38	23	28	22	23	0.16%	0.15%	0.14%	0.08%	0.14%	0.13%	0.18	0.276	0.11	0.23	0.13	0.13
TAS2R62P	Pseudogene	7	77	35	26	20	19	18	0.07%	0.07%	0.07%	0.05%	0.07%	0.07%	0.065	0.09	0.04	0.07	0.06	0.07
TAS2R63P	Pseudogene	12	84	40	36	17	25	29	0.12%	0.15%	0.12%	0.07%	0.12%	0.09%	0.081	0.088	0.07	0.1	0.08	0.07
TAS2R64P	Pseudogene	12	70	33	25	16	22	27	0.07%	0.07%	0.05%	0.01%	0.05%	0.04%	0.34	0.509	0.21	0.37	0.22	0.25
TAS2R67P	Pseudogene	12	63	33	29	21	21	28	0.17%	0.19%	0.16%	0.10%	0.15%	0.14%	0.175	0.308	0.1	0.19	0.11	0.12
TAS1R1	Umami	1	899	491	269	282	243	264	0.08%	0.13%	0.05%	0.07%	0.05%	0.05%	0.096	0.153	0.08	0.07	0.08	0.07
TAS1R2	Sweet	1	816	436	304	277	239	325	0.17%	0.19%	0.15%	0.14%	0.16%	0.16%	0.059	0.059	0.05	0.09	0.05	0.04
TAS1R3	Umami and sweet	1	339	141	83	80	97	102	0.05%	0.09%	0.03%	0.04%	0.03%	0.04%	0.066	0.103	0.04	0.05	0.06	0.05
PKD1L3	Sour	16	2523	1215	864	773	763	820	0.15%	0.16%	0.14%	0.12%	0.15%	0.13%	0.083	0.109	0.07	0.09	0.06	0.08
PKD2L1	Sour	10	1273	640	420	352	392	400	0.09%	0.12%	0.07%	0.07%	0.06%	0.08%	0.084	0.122	0.06	0.07	0.07	0.08
SCNN1A	Salt	12	1036	532	347	304	294	380	0.11%	0.12%	0.10%	0.10%	0.10%	0.10%	0.093	0.15	0.07	0.08	0.09	0.07
SCNN1B	Salt	16	3066	1553	1074	798	906	945	0.09%	0.12%	0.07%	0.07%	0.06%	0.07%	0.113	0.176	0.09	0.1	0.1	0.08

SCNN1G	Salt	16	1020	486	337	338	315	409	0.11%	0.11%	0.12%	0.07%	0.13%	0.09%	0.069	0.096	0.06	0.09	0.05	0.05
GPR120	Fat	10	746	353	278	238	246	242	0.09%	0.12%	0.06%	0.05%	0.07%	0.08%	0.086	0.101	0.06	0.13	0.06	0.07
CD36	Fat	7	8483	4234	2874	2327	2323	2684	0.10%	0.12%	0.08%	0.09%	0.08%	0.08%	0.081	0.121	0.06	0.07	0.07	0.06
PRB1	Multiple	12	271	125	95	98	72	85	0.15%	0.17%	0.14%	0.14%	0.13%	0.14%	0.064	0.087	0.04	0.07	0.07	0.05
CALHM1	Multiple	10	165	68	42	56	35	46	0.03%	0.03%	0.03%	0.04%	0.04%	0.03%	0.054	0.092	0.04	0.04	0.05	0.05
RTP3	Multiple	3	165	79	57	52	41	41	0.11%	0.09%	0.11%	0.11%	0.10%	0.11%	0.094	0.16	0.06	0.11	0.08	0.06
RTP4	Multiple	3	186	103	75	55	73	67	0.14%	0.14%	0.13%	0.10%	0.14%	0.13%	0.093	0.123	0.07	0.16	0.06	0.06
REEP1	Multiple	2	3444	1762	1219	1022	1042	1167	0.10%	0.10%	0.08%	0.08%	0.08%	0.09%	0.129	0.229	0.1	0.11	0.09	0.08
REEP2	Multiple	5	263	108	65	85	73	78	0.08%	0.07%	0.07%	0.07%	0.07%	0.08%	0.077	0.075	0.06	0.12	0.05	0.08
REEP3	Multiple	10	2425	1157	702	676	686	731	0.07%	0.07%	0.06%	0.05%	0.06%	0.06%	0.083	0.124	0.05	0.11	0.06	0.06
TRPV1	Hot/capsaicin	17	1225	589	403	408	405	463	0.13%	0.11%	0.11%	0.09%	0.10%	0.10%	0.188	0.314	0.12	0.21	0.13	0.13
TRPM8	Cool/menthol	2	3188	1590	1016	981	1010	1037	0.11%	0.11%	0.09%	0.11%	0.07%	0.10%	0.112	0.161	0.09	0.09	0.14	0.07
TRPA1	Pungent/	8	1630	798	585	528	538	534	0.14%	0.14%	0.12%	0.11%	0.14%	0.14%	0.058	0.055	0.05	0.09	0.06	0.04
GNAT3	Transduction	7	1467	692	476	403	399	503	0.08%	0.11%	0.05%	0.06%	0.05%	0.05%	0.134	0.204	0.1	0.13	0.1	0.09
GNB3	Transduction	12	306	148	79	77	74	91	0.07%	0.08%	0.05%	0.07%	0.05%	0.04%	0.101	0.133	0.07	0.11	0.1	0.09
GNG13	Transduction	16	300	160	105	93	88	105	0.16%	0.20%	0.12%	0.13%	0.07%	0.23%	0.073	0.074	0.06	0.05	0.09	0.09
PLCB3	Transduction	11	578	261	148	153	129	162	0.05%	0.04%	0.04%	0.04%	0.04%	0.05%	0.103	0.139	0.08	0.11	0.1	0.08
Mean			690	341	229	201	201	225	0.10%	0.11%	0.09%	0.08%	0.09%	0.09%	0.106	0.162	0.07	0.12	0.08	0.08

The number of segregating sites, nucleotide diversity and level of population differentiation (F_{ST}) are displayed for 56 taste genes. We see overall the highest levels of diversity occur in the African samples and the lowest occurring in the Eastern Asian samples. Levels of differentiation and diversity also vary across genes, with genes with TAS2Rs representing both extremes of high and low. TAS2R39 and TAS2R40 represent the lowest level of genetic diversity, and TAS2R20 the highest. TAS2R43 has the highest level of differentiation (F_{ST}). Kept into this table are the TAS2R pseudogenes, to use as a reference and comparison group for levels of nucleotide diversity and differentiation.