HUMAN MUTATION 26(3), 199–204, 2005

RESEARCH ARTICLE

Worldwide Haplotype Diversity and Coding Sequence Variation at Human Bitter Taste Receptor Loci

Unkyung Kim[1†] Stephen Wooding[3†] Dante Ricci,[2] Lynn B. Jorde,[3] and Dennis Drayna[2*]

[1]Department of Biology, Kyungpook National University, Daegu, Republic of Korea; [2]National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Rockville, Maryland; [3]Department of Human Genetics, University of Utah, Salt Lake City, Utah

Communicated by Pui-Yan Kwok

Bitter taste perception in humans is mediated by receptors encoded by 25 genes that together comprise the *TAS2R* (or *T2R*) gene family. The ability to identify the ligand(s) for each of these receptors is dependent on understanding allelic variation in *TAS2R* genes, which may have a significant effect on ligand recognition. To investigate the extent of coding variation among *TAS2R* alleles, we performed a comprehensive evaluation of sequence and haplotype variation in the human bitter taste receptor gene repertoire. We found that these genes exhibit substantial coding sequence diversity. In a worldwide population sample of 55 individuals, we found an average of 4.2 variant amino acid positions per gene. In aggregate, the 24 genes analyzed here, along with the phenylthiocarbamide (PTC) receptor gene analyzed previously, specify 151 different protein coding haplotypes. Analyses of the ratio of synonymous and nonsynonymous nucleotide substitutions using the K_a/K_s ratio revealed an excess of amino acid substitutions relative to most other genes examined to date ($K_a/K_s = 0.94$). In addition, comparisons with more than 1,500 other genes revealed that levels of diversity in the *TAS2R* genes were significantly greater than expected ($\pi = 0.11\%$; $p < 0.01$), as were levels of differentiation among continental populations ($F_{ST} = 0.22$; $p < 0.05$). These diversity patterns indicate that unusually high levels of allelic variation are found within *TAS2R* loci and that human populations differ appreciably with respect to *TAS2R* allele frequencies. Diversity in the *TAS2R* genes may be accounted for by natural selection, which may have favored alleles responsive to toxic, bitter compounds found in plants. These findings are consistent with the view that different alleles of the *TAS2R* genes encode receptors that recognize different ligands, and suggest that the haplotypes we have identified will be important in studies of receptor-ligand recognition. Hum Mutat 26(3), 199–204, 2005. **© 2005 Wiley-Liss, Inc.**

KEY WORDS: bitter taste; *TAS2R*; cSNP; haplotypes; natural selection

INTRODUCTION

The sense of bitter taste is mediated by a group of bitter taste receptor proteins that reside on the surface of taste cells within the taste buds of the tongue. These proteins are seven transmembrane domain, G protein coupled receptors, encoded by members of the *TAS2R* (or *T2R*) gene family [Adler et al., 2000]. The human genome contains 25 apparently functional *TAS2R* genes that reside in three locations. Fifteen genes reside in a cluster on chromosome 12p, nine genes reside in a cluster on chromosome 7q, and a single family member resides on chromosome 5p [Shi et al., 2003].

A major goal of chemosensory science is to identify the ligands for each of these receptors. However, this task has proven difficult, and the ligand that binds to each receptor and initiates bitter taste perception is known in only a few cases. In humans, in vitro cell-based assays have shown that *TAS2R16* (taste receptor, type 2, member 16, MIM♯ 604867) responds to salicin and other beta-glucopyranosides, and *TAS2R10* (taste receptor, type 2, member 10, MIM♯ 604791) displays activity upon exposure to strychnine [Bufe et al., 2002]. An alternative human genetic approach has revealed that *TAS2R38* (*PTC*) (taste receptor, type 2, member 38, MIM♯ 607751) encodes the receptor for phenylthiocarbamide, a classic variant trait in humans [Kim et al., 2003].

The human *TAS2R38* gene has been shown to exist in seven different allelic forms, although only two of these, designated the major taster form and the major non-taster form, exist at high frequency outside of sub-Saharan Africa [Wooding et al., 2004]. These two forms have been maintained by balancing natural selection, and it has been suggested that the non-taster form

The Supplementary Material referred to in this article can be accessed at http://www.interscience.wiley.com/jpages/1059-7794/suppmat.

Received 12 November 2004; accepted revised manuscript 17 February 2005

*Correspondence to: Dennis Drayna, NIDCD/NIH, 5 Research Court, Rockville, MD 20850. E-mail: drayna@nidcd.nih.gov

Grant sponsor: NIH; Grant numbers: DC Z01-000046-04; HL070048; and ES12125; Grant sponsor: NSF; Grant number: BCS0218370.

†Unkyung Kim and Stephen Wooding contributed equally to this work.

DOI 10.1002/humu.20203
Published online in Wiley InterScience (www.interscience.wiley.com).

serves as a functional receptor for some other bitter substance not yet identified. These *TAS2R38* studies illustrate the possibility that different alleles of each gene may encode receptors that recognize different ligands. Thus, the problem of identifying specific ligands for each bitter taste receptor may be an allele-specific rather than a locus-specific task. Such variation could help explain both the presence of phenotypic variation in bitter-taste perception [Kim et al., 2004] and the mixed success of efforts to identify specific ligands to date.

To investigate patterns of haplotypic variation in the *TAS2R* genes, which will be an important resource in the identification of alleles with possibly different ligand specificities, we resequenced all the functional receptor loci to build a catalog of the variants at each locus, and we identified the protein-coding haplotypes that are specified by these variants. In addition, to identify those loci that show high levels of diversity consistent with the presence of functionally divergent alleles, we analyzed patterns of variation within and between three continental human populations: Africans, Asians, and Europeans. Finally, we investigated the possible role of natural selection in shaping diversity patterns in these genes using statistical tests of evolutionary neutrality.

MATERIALS AND METHODS

Population Samples

Human genomic DNA was obtained from 55 unrelated individuals in five different geographic populations including 21 Cameroonians, 10 Amerindians, 10 Japanese, nine Hungarians, and five Mbuti Pygmies from the Ituri Forest in Northeast Democratic Republic of Congo. All DNA samples except Cameroonian were obtained from Coriell Cell Repositories (http://locus.umdnj.edu/nigms/cells/humdiv.html).

PCR and DNA Sequencing

Human *TAS2R* genes consist of a single coding exon approximately 1 kb in length. We sequenced the open reading frame (ORF) of 21 out of 25 human *TAS2R* genes and combined this information with data from the *TAS2R38* (PTC) gene published previously [Kim et al., 2003]. Primers for PCR amplification and for sequencing based on the genomic sequence flanking the *TAS2R* genes in the human genome were designed (using software at the Primer3 Web site) to amplify the entire ORF of each *TAS2R* gene in humans, the sequences of which are available at: www.nidcd.nih.gov/research/scientists/draynad.asp.

PCR was performed in a total volume of 25 μl, containing 0.2 μM of each deoxynucleotide (Invitrogen), 15 pmol of each forward and reverse primers, 1.0–1.5 mM of $MgCl_2$, 10 mM Tris-HCl (pH8.3), 50 mM KCl, 0.75 U of *Taq* DNA polymerase (PE Biosystems), and 100 ng of genomic DNA. PCR conditions (PE9700, PE Biosystems) were as follows: 35 cycles of denaturation at 94°C for 30 sec; annealing at 55°C or 57°C, depending on the primers for 30 sec; and extension at 72°C for 1 min. The first step of denaturation and the last step of extension were 95°C for 2 min and 72°C for 10 min, respectively. PCR products (5 μl) were separated and visualized in a 2% agarose gel, and 15 μl of this PCR product was then treated with 0.3 U of shrimp alkaline phosphatase (USB) and 3 U of exonuclease I (USB) at 37°C for 1 hr, followed by incubation at 80°C for 15 min. This was diluted with an equal volume of dH_2O, and 6 μl was used for the final sequencing reaction. Sequencing reactions were performed in both directions on the PCR products in reactions containing 5 pmol of primer, 1 μl of Big Dye Terminator Ready Reaction Mix (PE Biosystems), and 1 μl of 5 × dilution buffer (400 mM Tris-HCl, pH 9, and 10 mM $MgCl_2$). Cycling conditions were 95°C for 2 min and 35 cycles of 94°C for 20 sec, 55°C for 20 sec, and 60°C for 4 min. Sequencing reaction products were ethanol precipitated, and the pellets were resuspended in 10 μl of formamide loading dye. An ABI 3730 DNA sequencer was used to resolve the products, and data was analyzed by using ABI Sequencing Analysis (v. 5.0) and LASERGENE-SeqMan software.

Data Analysis

The assigned SNP numbers in the *TAS2R* genes correspond to the cDNA sequence of each gene, with the A of the ATG translation initiation codon being taken as the +1 (detailed in Supplementary Table S1) (available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat). The protein nomenclature employed is consistent with the convention of numbering the initiation methionine as the +1 amino acid.

Haplotypes in the *TAS2R* genes were determined either explicitly from homozygous individuals or inferred from genotype data using the PHASE 2.0.2 computer program [Stephens et al., 2001; Stephens and Donnelly, 2003] (detailed in Supplementary Table S2).

Linkage disequilibrium among loci was estimated using the GOLD software package to measure D' [Abecasis and Cookson, 2000]. These estimates were obtained in analyses of all 55 sampled individuals. Fisher's exact test was used to determine statistical significance.

Patterns of synonymous and nonsynonymous nucleotide substitution were measured using K_a/K_s, the ratio of the number of nonsynonymous nucleotide substitutions per nonsynonymous site, and the number of synonymous nucleotide substitutions per synonymous site [Li et al., 1985]. This statistic can reveal biases in the rate at which synonymous and nonsynonymous substitutions occur. Observed K_a/K_s ratios were compared with values observed in 153 genes by Nekrutenko et al. [2002].

Levels of genetic diversity were measured using two methods. First, nucleotide diversity was measured for each locus using π, the mean pairwise difference between sequences, per nucleotide [Hartl and Clark, 1997]. Second, the fraction of genetic diversity due to population substructuring was measured for each locus using F_{ST}, an estimate of the genetic variance that is accounted for by differences among populations. F_{ST} values were calculated using the method of Slatkin and Voelm [1991], with the continental regions (Africa, Asia, and Europe) defining populations.

To determine whether trends were present in our data, we also tested whether the mean π and F_{ST} values in the *TAS2R* genes were greater than expected by chance. These tests were performed by comparing the observed π and F_{ST} values with 1,598 π values and 1,627 F_{ST} values reported by Schneider et al. [2003]. Comparisons were performed as follows. First, 21 random genes were chosen from the data of Schneider et al. [2003] to produce a randomized dataset of similar size to the *TAS2R* gene dataset. Next, the mean F_{ST} and π values calculated for the randomized dataset were compared with the mean observed in *TAS2R* genes. This procedure was repeated 10,000 times. The fraction of comparisons in which the mean F_{ST} and π values calculated from our data exceeded the mean calculated from the Schneider et al. [2003] dataset was treated as a p-value.

RESULTS AND DISCUSSION

We sequenced 24 human *TAS2R* (*T2R*) genes to find coding region single nucleotide polymorphisms (cSNPs) in 55 unrelated humans of African, Asian, European, and North American Native

ancestry. These genes displayed a high degree of nucleotide variation, ranging from one cSNP in *TAS2R13* to 12 cSNPs in *TAS2R48*. Combined with previous results from *TAS2R38* (*PTC*) [Wooding et al., 2004], we identified a total of 144 cSNPs in this set of *TAS2R* genes, with an average of six cSNPs per gene (detailed in Supplementary Table S1). Analysis using the GOLD software package indicated that linkage disequilibrium among the *TAS2R* genes was not statistically significant, except in the case of *TAS2R7* and *TAS2R8* ($p < 0.05$). For this reason, LD was treated as negligible in all further analyses.

For 32% of cSNPs, the minor allele was observed only once. The Cameroonian population displayed the greatest number of alleles, consistent with the view that African populations harbor higher levels of diversity than do other populations [Tishkoff and Verrelli, 2003]. The remainder of these cSNPs were approximately evenly divided into two classes: 36.2% had a minor allele frequency between 1% and 20%, and the remaining 31.5% had minor allele frequencies between 20% and 50%. Figure 1 shows the distribution of sharing of cSNPs across these populations. Excluding Pygmies, four populations shared 25 cSNPs although only six cSNPs exist in all the populations studied. In addition, the distribution of allele frequencies in Pygmies is different from other populations, with Pygmies showing reduced polymorphism for the majority of cSNPs.

Of the cSNPs found, 106 (74%) are nonsynonymous and 35 (24%) are synonymous. Of nonsynonymous cSNPs, 57.5% substitute a non-conservative amino acid based on Grantham values [Grantham, 1974; Li et al., 1984], and amino acid changes were observed across the entire coding region of these genes. In addition to the 141 cSNPs that alter the amino acid sequence of a *TAS2R* protein, we observed three cSNPs that produce stop codons, resulting in segregating pseudogenes (SPGs) in the human sample. One SPG occurs in the *TAS2R46* gene, which has two nonsense alleles, c.749G>A and c862C>T. c.749G>A has a

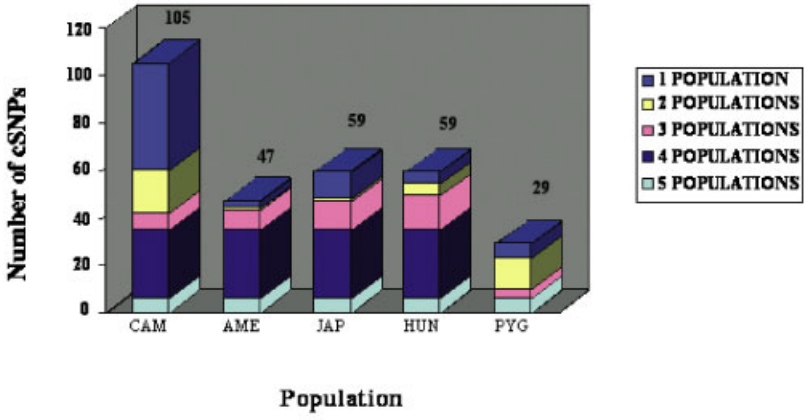

FIGURE 1. **The number and breadth of distribution of TAS2R SNPs in five populations. The number of cSNPs observed in each different population is indicated on the Y axis. The breadth of distribution of the SNPs is categorized by color, with different colors indicating the number of populations in which they are variable. Thus, the gray portion of each bar indicates the number of SNPs unique to that population, and the light blue portion indicates the number that are found in all populations tested. Population codes: Cameroonians (CAM); Amerindians (AME); Japanese (JAP); Hungarians (HUN); Pygmies (PYG). [The color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]**

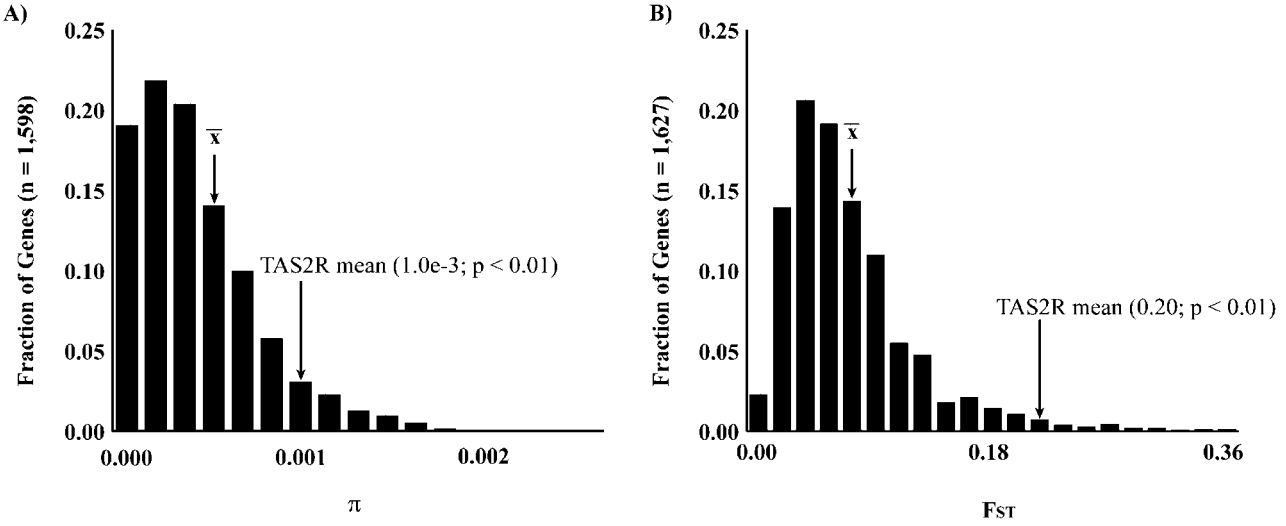

FIGURE 2. **Comparisons of π and F_{ST} in *TAS2R* genes and in ~1,600 genes reported by Schneider et al. [2003]. A: Distribution of π values in 1,598 genes. X-bar indicates the distribution mean. B: Distribution of FST values in 1,627 genes. X-bar indicates the distribution mean.**

null allele frequency of ~24% in all populations except Pygmies, while c.862C>T was observed only once, in Cameroonians. The other SPG was observed in *TAS2R7*, and although it was observed only in Amerindians, the null allele displayed a frequency of 30% in this population. Although only three segregating null alleles seem to exist in the *TAS2R* gene repertoire, our result is consistent with observations in the olfactory receptor gene family and suggests that *TAS2R* SPGs exist at different frequencies in different populations [Gilad and Lancet, 2003; Menashe et al., 2002].

Analyses of the K_a/K_s ratio confirmed that an appreciable excess of nonsynonymous nucleotide substitutions was found in the *TAS2R* genes. Under theoretical expectation, the ratio of nonsynonymous substitutions per nonsynonymous nucleotide position (K_a) to synonymous substitutions per synonymous position (K_s) in a given coding region is approximately 1 [Li et al., 1985]. In our sample, K_a/K_s was near the theoretical expectation, 0.94 ($P>0.20$); however, the observed value was much higher than expected given observations in other genes. For example, the mean observed in our sample was more than five standard deviations higher than that reported by Nekrutenko et al. [2002] in an analysis of 152 genes, which had an average K_a/K_s of 0.11 and a standard deviation of 0.164. The low K_a/K_s ratios observed in most human genes are thought to arise from evolutionary constraints, likely imposed by natural selection, which reduce the K_a/K_s ratio to around 0.10 [Li et al., 1985]. Thus, the K_a/K_s ratio observed in our sample suggests that the rate at which *TAS2R* genes have experienced nonsynonymous nucleotide substitutions is much higher than in most other genes examined to date. The preponderance of amino acid substitutions among alleles at *TAS2R* loci suggests that these alleles likely differ in their physiologic properties.

To further explore the possible functional importance of amino acid substitutions in the *TAS2R* genes, we determined identified haplotypes differing by one or more amino acid substitutions using two methods: 1) direct ascertainment from individuals homozygous at all or all but one nucleotide positions at a locus, and 2) Bayesian methods as implemented in the PHASE software package [Stephens et al., 2001; Stephens and Donnelly, 2003]. These analyses revealed that the number of cSNP haplotypes ranged from two to 11 for each gene. In the 24 *TAS2R* genes together we identified a total of 144 different haplotypes (detailed in Supplementary Table S2). The PTC gene, analyzed separately, specifies seven cSNP haplotypes to yield a total of 151. In general, one or two haplotypes of each *TAS2R* gene were common worldwide, and were observed in all five different populations. Given the size of our sample (110 chromosomes), we estimate that we identified more than 75% of haplotypes with frequencies from 0.01 to 0.03 and more than 99% of haplotypes with frequencies from 0.03 to 1.0.

Population genetic analyses revealed high levels of nucleotide diversity. The mean pairwise difference per nucleotide between sequences (π) at *TAS2R* loci ranged from 5.2e-5 (in *TAS2R1*) to 3.9e-3 (in *TAS2R49*), with a mean of 1.14e-3 (Table 1, Fig. 2A). This value was more than 33% greater than that found in a genome-wide analysis reported by The International SNP Map Working Group [Sachidanandam et al., 2001], 7.5e-4, and more than 75% greater than the mean of 1,598 values reported by Schneider et al. [2003], 5.6e-4.

To determine whether the mean π-value in the *TAS2R* genes was significantly greater than expected, we compared the mean π in *TAS2R* genes with means generated by resampling the 1,598 values reported by Schneider et al. [2003]. This analysis revealed that the mean value of π in the *TAS2R* genes was significantly greater than expected ($p<0.01$). Thus, alleles at *TAS2R* loci are, on average, more divergent from one another than are alleles at other loci. Taken together, the finding that π-values in the *TAS2R* genes are relatively high, along with the finding that *TAS2R* genes harbor an unusually high number of amino acid substitutions, suggests that multiple, functionally divergent alleles are likely found at *TAS2R* loci.

TABLE 1. **Summary Statistics***

Gene	S	mpd	π	F_{ST}
TAS2R01	**3**	**0.05**	**5.22E-05**	**0.23**
TAS2R03	**3**	**0.55**	**5.73E-04**	**0.10**
TAS2R04	**8**	**1.7**	**1.89E-03**	**0.07**
TAS2R05	**7**	**0.97**	**1.08E-03**	**0.07**
TAS2R07	**6**	**0.35**	**3.64E-04**	**0.13**
TAS2R08	**6**	**0.91**	**9.74E-04**	**0.28**
TAS2R09	**7**	**0.64**	**6.84E-04**	**0.14**
TAS2R10	**6**	**0.56**	**6.15E-04**	**0.14**
TAS2R13	**1**	**0.44**	**4.80E-04**	**0.34**
TAS2R14	**4**	**0.54**	**5.70E-04**	**0.47**
TAS2R16	**8**	**0.15**	**1.74E-04**	**0.15**
TAS2R38	**5**	**1.47**	**1.50E-03**	**0.06**
TAS2R39	**2**	**0.11**	**1.05E-04**	**0.02**
TAS2R40	**3**	**0.37**	**3.82E-04**	**0.15**
TAS2R41	**4**	**0.75**	**8.06E-04**	**0.23**
TAS2R43	**6**	**1.1**	**1.19E-03**	**0.80**
TAS2R44	**11**	**2.5**	**2.69E-03**	**0.24**
TAS2R45	**6**	**2.35**	**2.35E-03**	**0.14**
TAS2R46	**6**	**0.98**	**1.05E-03**	**0.12**
TAS2R47	**5**	**0.92**	**9.55E-04**	**0.16**
TAS2R48	**12**	**1.85**	**2.05E-03**	**0.26**
TAS2R49	**11**	**3.59**	**3.86E-03**	**0.33**
TAS2R50	**7**	**1.27**	**1.41E-03**	**0.25**
TAS2R55	**5**	**2.12**	**2.25E-03**	**0.39**
TAS2R60	**2**	**0.52**	**5.38E-04**	**0.24**
Mean	**5.8**	**1.07**	**1.14E-3**	**0.22**

***S, number of polymorphic nucleotide positions; mpd, mean pairwise difference between sequences; π, mean pairwise difference per nucleotide; F_{ST} = (HT − HS)/HT, where HT is the heterozygosity of the total sample and HS is the average heterozygosity of the individual populations.**

To determine the extent to which alleles vary in frequency among sampled populations we analyzed the F_{ST} statistic, which takes values near zero when populations have similar allele frequencies and values near unity when populations have dramatically different frequencies [Slatkin and Voelm, 1991]. F_{ST} values for the individual genes in our sample ranged from 0.02 (for *TAS2R39*) to 0.80 (for *TAS2R43*), with a mean of 0.22 (Table 1, Fig. 2). Published F_{ST} values based on DNA sequence variation in humans are usually lower, often falling around 0.15 [Tishkoff and Verrelli, 2003]. To determine whether the mean F_{ST} in the *TAS2R* genes was significantly greater than expected, we compared the mean F_{ST} in *TAS2R* genes with means generated by resampling the 1,627 values reported by Schneider et al. [2003]. This analysis revealed that the mean value of F_{ST} in the *TAS2R* genes was significantly greater ($p<0.01$) than expected given the empirical distribution. Thus, Africa, Asia, and Europe differ more with respect to variation in *TAS2R* genes than they do with respect to most other genes.

An important feature of the dataset reported by Schneider et al. [2003] is that it is intended to represent populations in the United States, which are admixed and thus have a relatively low F_{ST} value, 0.06. For this reason, we replicated the test of F_{ST} using the data of Akey et al. [2002], who found an average FST of 0.123 in an analysis of 25,549 genome-wide SNPs, which is more typical of human populations [Tishkoff and Verrelli, 2003]. This comparison

confirmed that the mean F_{ST} in the *TAS2R* genes was significantly greater than expected given the empirical distribution ($p < 0.01$).

The preponderance of high F_{ST} values in human *TAS2R* genes indicates that human populations differ more with respect to variation in the bitter taste receptor genes than they do with respect to most other regions of the genome. This finding suggests that major human populations could differ appreciably in the frequency of functionally important variants and, consequently, in the frequency of bitter-taste related phenotypes. A key implication of this finding is that efforts to identify the various functionally divergent alleles at bitter-taste receptor loci may benefit from sampling diverse human populations.

Evidence that the *TAS2R* gene family harbors unusually high levels of genetic diversity, including a very high number of amino acid substitutions, suggests the hypothesis that the effects of natural selection might be relaxed on these genes. The removal of selective constraints would explain both the excess of amino acid substitutions and the presence of segregating pseudogenes at the TAS2R loci. However, while the relaxation of selective constraints would explain these patterns, it would fail to explain the high levels of population differentiation, reflected by high mean F_{ST} values. An alternative hypothesis that would explain the high mean F_{ST} value observed in the *TAS2R* genes, and the high mean values of π and K_a/K_s would be that local adaptation has occurred. Under local adaptation, loci under selection diverge as different adaptive mutations rise in frequency in different populations, causing increases in all three of the measures described here [Bamshad and Wooding, 2003]. In addition, patterns similar to those seen in the *TAS2R* family are found in other genetic systems in which local adaptation is thought to have occurred [Gilad et al., 2002; Hamblin et al., 2002; Hollox et al., 2001; Rana et al., 1999; Tishkoff et al., 2001]. We hypothesize that local adaptation in human bitter taste receptor genes is common and has been driven by the fitness advantages of avoiding toxins found in plants.

It has long been recognized that numerous toxins produced by plants as a means of defense against herbivores taste bitter [Drewnowski and Gomez-Carneros, 2000]. These toxins vary geographically in both composition and abundance [Berenbaum and Robinson, 2003; Coley and Barone, 1996; Tepper, 1998]. Further, several of the known ligands for bitter-taste receptors are plant toxins [Behrens et al., 2004; Bufe et al., 2002]. One possibility is that bitter-taste receptor genes in humans have adapted to recognize those toxins found in human environments.

The patterns of population genetic variation in our data have a variety of implications for studies of bitter taste perception in vitro and in vivo. In particular, evidence that numerous amino acid substitutions are present at *TAS2R* loci, that these loci are more diverse than are most loci in humans, and that natural selection may have fostered intragenic diversification at *TAS2R* loci, combine to suggest that functionally important variants exist within many *TAS2R* genes. For in vivo studies, our results suggest that homogeneous human populations may fail to identify many important variants found between, rather than within populations, and support the use of geographically dispersed populations in gene mapping and association studies.

The *TAS2R* gene SNPs we have found are consistent with those recently reported by Wang et al. [2004]. However, we have identified approximately 60 additional variants in these 24 genes, presumably because of the larger representation of sub-Saharan African individuals, which tend to be the most genetically diverse of all human groups, in our subject population. Our results add to previous reports regarding haplotypes of *TAS2R* genes and of human genes in general, demonstrating that only a fraction of all the possible haplotypes exist. For example, while there are seven SNPs within the *TAS2R38* (PTC) gene that could generate $2^7 = 128$ different haplotypes, only seven haplotypes have been observed worldwide. All *TAS2R* genes appear to follow this pattern. The nature, frequency, and geographic distribution of each of these *TAS2R* gene haplotypes will be important because it is the haplotype, not any particular SNP allele, that determines the receptor protein that is encoded. This is likely to be of prime importance in the task of assigning ligands for each of the *TAS2R* genes. If, like *TAS2R38*, different alleles of these genes differ significantly in their responsiveness to any one bitter ligand, this task will need to be re-conceptualized. We proposed it will no longer have the goal of identifying ligands for each of the 25 different *TAS2R* genes; instead it will need to focus on identifying ligands for each of the 151 different protein-coding *TAS2R* haplotypes specified in these genes.

ACKNOWLEDGMENTS

We thank Drs. Susan Sullivan and Robert Morell for helpful comments on the manuscript. Manish Pungliya, Benjamin Salisbury, and J. Claiborne Stephens (Genaissance Pharmaceuticals) kindly shared data used in several comparisons. We gratefully acknowledge support from the NIH, DC Z01-000046-04 to D.D., HL070048 to L.B.J., and ES12125 to S.W., and from the NSF, BCS0218370 to L.B.J.

REFERENCES

Abecasis GR, Cookson WO. 2000. GOLD—graphical overview of linkage disequilibrium. Bioinformatics 16:182–183.

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

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

Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111.

Behrens M, Brockhoff A, Kuhn C, Bufe B, Winnig M, Meyerhof W. 2004. The human taste receptor hTAS2R14 responds to a variety of different bitter compounds. Biochem Biophys Res Commun 319:479–485.

Berenbaum MR, Robinson GE. 2003. Chemical communication in a post-genomic world. Proc Natl Acad Sci USA 100(Suppl 2): 14513.

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

Coley PD, Barone JA. 1996. Herbivory and plant defenses in tropical forests. Ann Rev Ecol Syst 27:305–335.

Drewnowski A, Gomez-Carneros C. 2000. Bitter taste, phytonutrients, and the consumer: a review. Am J Clin Nutr 72: 1424–1435.

Gilad Y, Rosenberg S, Przeworski M, Lancet D, Skorecki K. 2002. Evidence for positive selection and population structure at the human MAO-A gene. Proc Natl Acad Sci USA 99:862–867.

Gilad Y, Lancet D. 2003. Population differences in the human functional olfactory repertoire. Mol Biol Evol 20:307–314.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. Science 185:862–864.

Hamblin MT, Thompson EE, Di Rienzo A. 2002. Complex signatures of natural selection at the Duffy blood group locus. Am J Hum Genet 70:369–383.

Hartl DL, Clark AG. 1997. Principles of population genetics. Sunderland, MA: Sinauer Associates.

Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N, Kozlov AI, Swallow DM. 2001. Lactase haplotype diversity in the Old World. Am J Hum Genet 68:160–172.

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

Kim UK, Breslin PA, Reed D, Drayna D. 2004. Genetics of human taste perception. J Den Res 83:448–453.

Li WH, Wu CI, Luo CC. 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71.

Li WH, Wu CI, Luo CC. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174.

Menashe I, Man O, Lancet D, Gilad Y. 2002. Population differences in haplotype structure within a human olfactory receptor gene cluster. Hum Mol Genet 11:1381–1390.

Nekrutenko A, Makova KD, Li WH. 2002. The K(A)/K(S) ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. Genome Res 12: 198–202.

Rana BK, Hewett-Emmett D, Jin L, Chang BH, Sambuughin N, Lin M, Watkins S, Bamshad M, Jorde LB, Ramsay M, Jenkins T, Li WH. 1999. High polymorphism at the human melanocortin 1 receptor locus. Genetics 151:1547–1557.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggil PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D, International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933.

Schneider JA, Pungliya MS, Choi JY, Jiang R, Sun XJ, Salisbury BA, Stephens JC. 2003. DNA variability of human genes. Mech Ageing Dev 124:17–25.

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

Slatkin M, Voelm L. 1991. F_{ST} in a hierarchical island model. Genetics 127:627–629.

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

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–1169.

Tepper BJ. 1998. 6-n-Propylthiouracil: a genetic marker for taste, with implications for food preference and dietary habits. Am J Hum Genet 63:1271–1276.

Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, Piro A, Stoneking M, Tagarelli A, Tagarelli G, Toumah EH, Williams SM, Clark AG. 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293:455–462.

Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340.

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 13:2671–2678.

Wooding S, Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D. 2004. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. Am J Hum Genet 74:637–646.