Title

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

Journal

Chemical Senses, 41(8)

ISSN

0379-864X

Authors

Roudnitzky, Natacha
Risso, Davide
Drayna, Dennis
et al.

Publication Date

2016-10-01

DOI

10.1093/chemse/bjw067

Peer reviewed

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

Original Article

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

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

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

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

Abstract

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

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

Introduction

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

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

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

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

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

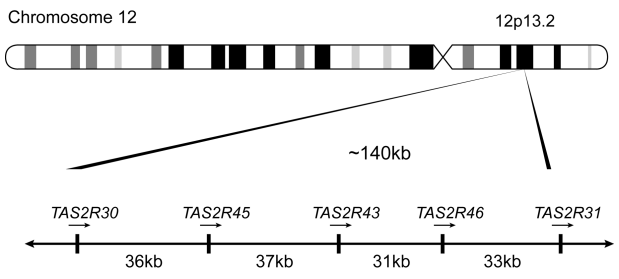

Figure 1. Gene locations. The *TAS2R30-31* cluster is located at 12p13.2. In the longest contig spanning the region (NW_001838055.2), their order in forward orientation is *TAS2R30, -45, -43, -46, -31*, with a mean separation of 34 kb.

Materials and methods

Subjects

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

Structural mapping

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

Copy-number estimates

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

Table 1. H952 population sample

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

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

Population genetic analysis

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

Results

Genomic structure

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

In addition to spanning the *TAS2R43* and *-45* coding regions, the deleted regions in NW_003571047.1 and NT_009714.17 extended far up- and downstream. While *-43* and *-45* are 930 and 900 bp in length, respectively, the deletions encompassing them were 37 787 and 32 247 bp. They were also asymmetrical, with the *TAS2R43* deletion extending from 11.2 kb upstream to 25.7 kb downstream of the coding region and the *TAS2R45* deletion extending from 0.7 kb upstream to 30.1 kb downstream. Further, the genomic region deleted in NW_003571047.1 overlapped with that deleted in NT_009714.17, producing a 5.2 kb intergenic segment absent in both relative to NW_003571050.1.

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

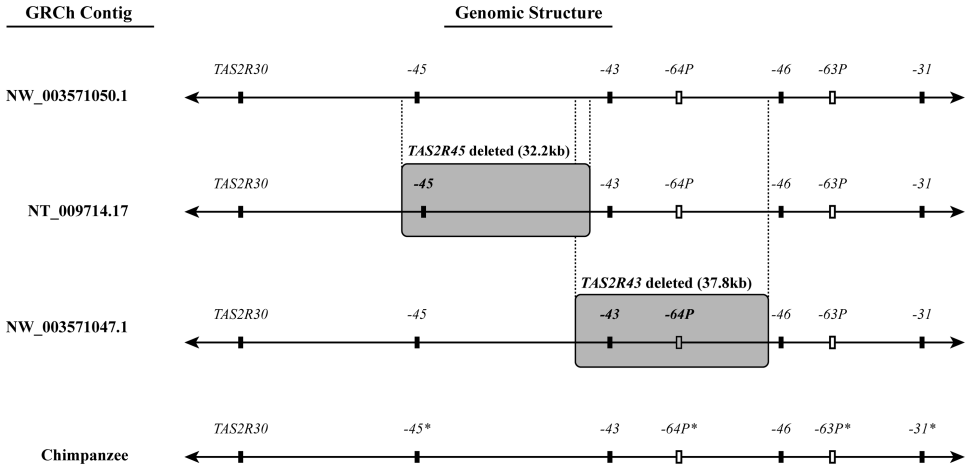

Figure 2. Genomic alignments. Alignments of three representative GRCh contigs are shown. Alignment gaps (highlighted), indicate deletion with respect to the reference contig, NW_003571050.1. The chimpanzee genome (bottom) was most similar to the reference contig, harboring homologs of all genes and pseudogenes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

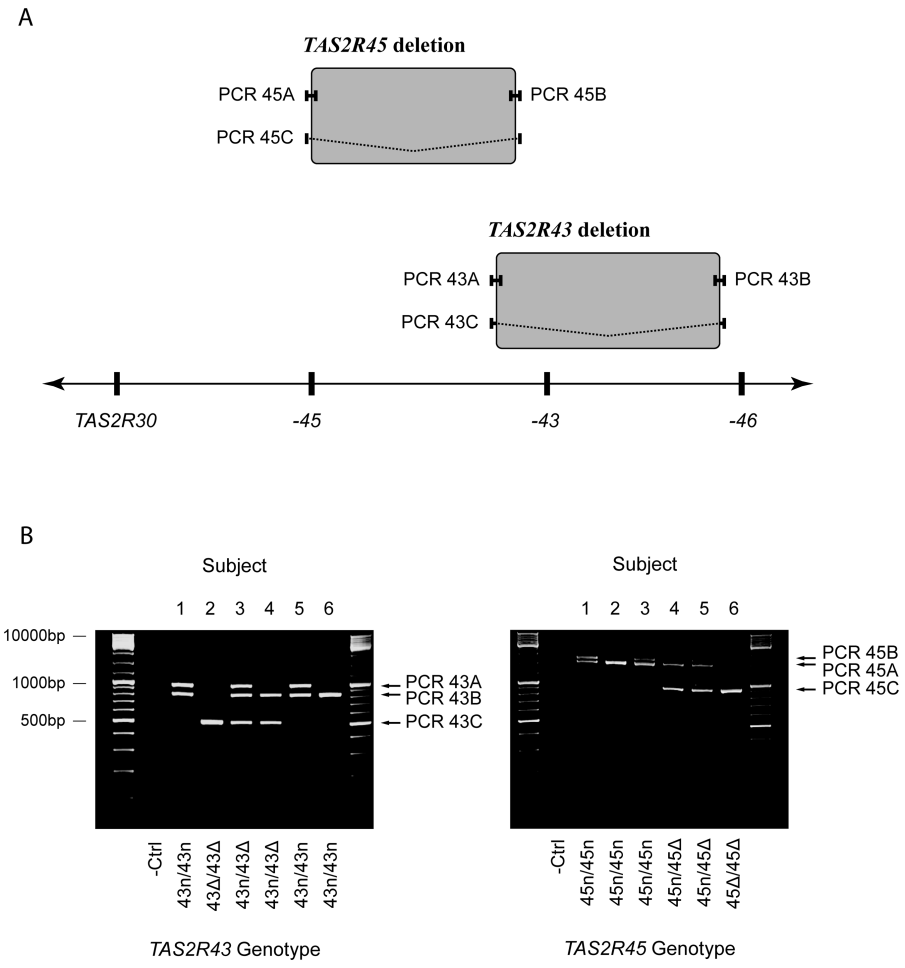

Figure 3. Amplification maps. (**A**) Iteratively amplifying targets in the *TAS2R43-45* region confirmed the boundaries of the deletions discovered in contig alignments. Three defining amplifications were established for each: one crossing the upstream limit of the deletion (43A and 45A), one crossing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

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

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

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

Population genetics

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

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

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

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

A

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

B

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

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

A

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

B

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

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

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

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

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

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

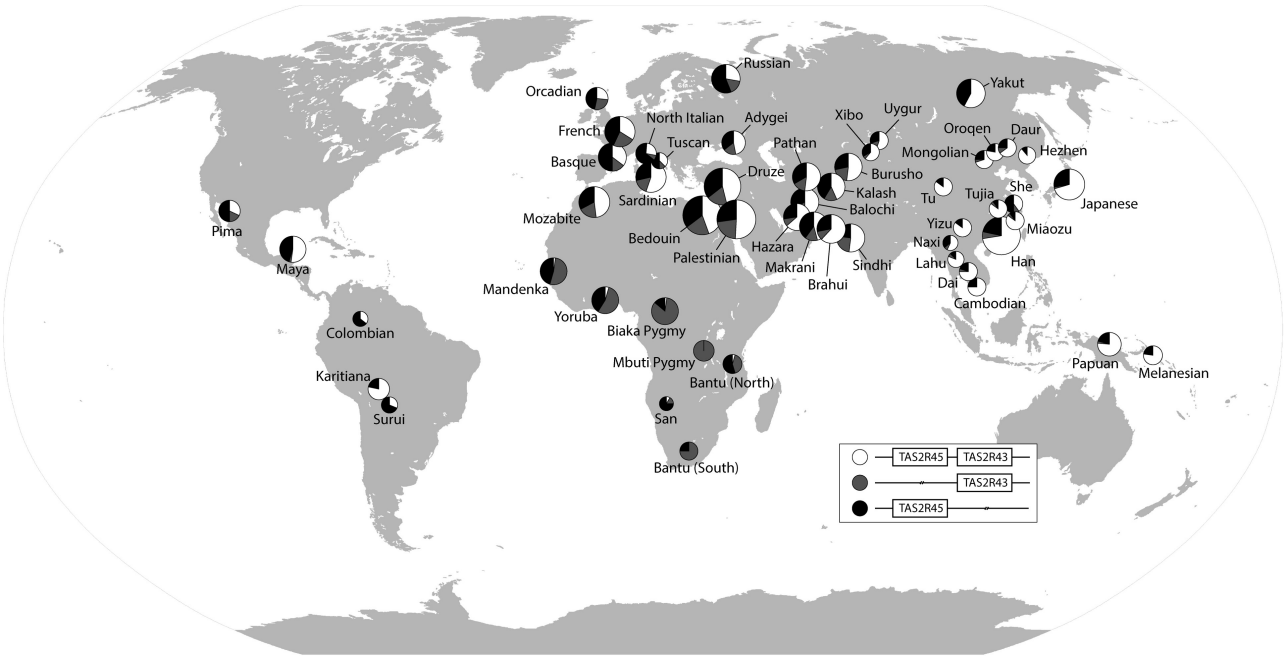

Figure 6. Global haplotype frequencies. Each point represents a sampled H952 population. Point size is proportional to sample size. Sections indicate the relative haplotype frequency within populations.

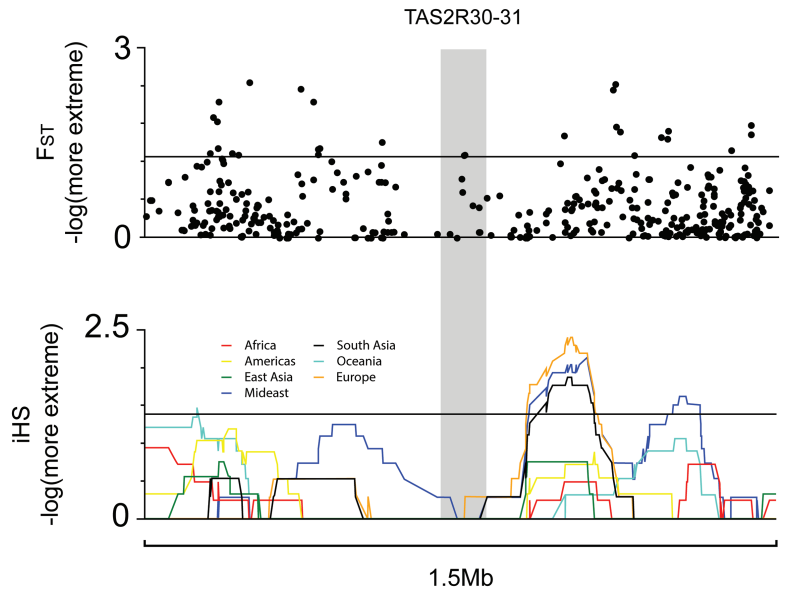

Figure 7. Tests for natural selection. Patterns across a 1.5 Mb region centered on the *TAS2R30-31* cluster in the HGDP are shown. (**A**) Plot of SNP F_{ST} values relative to the distribution across all SNPs in the HGDP. Y-axis indicates the negative log of the fraction of HGDP SNPs with more extreme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

Discussion

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

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

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

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

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

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

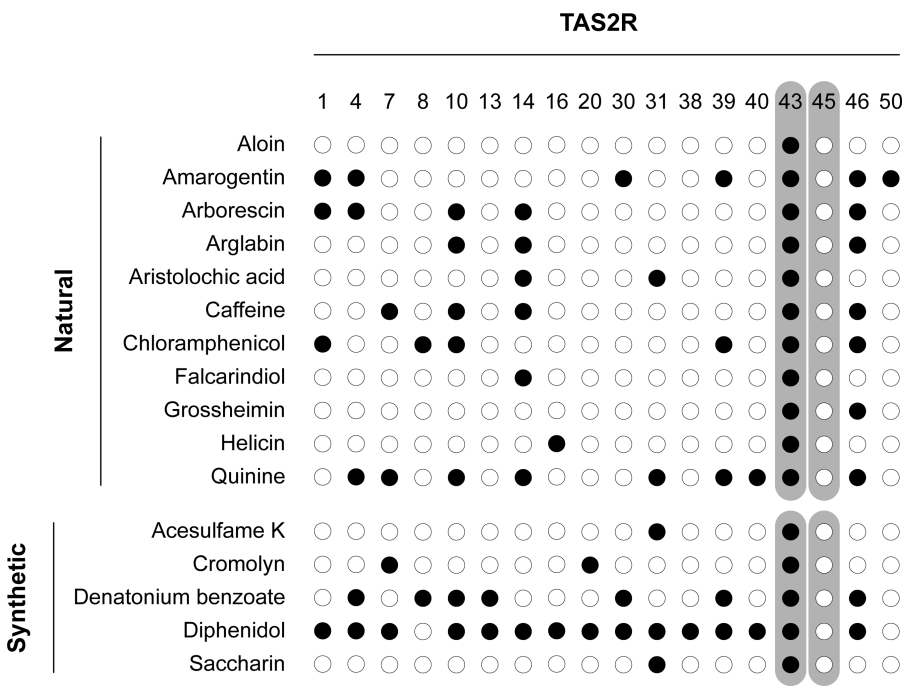

Figure 8. TAS2Rs and agonists. Responses of 18 TAS2Rs to each of the 16 known agonists of TAS2R43, reported by Meyerhof *et al.* (2010), are shown. TAS2R45 has no known agonists (Meyerhof *et al.* 2010; Thalmann *et al.* 2013). Filled circles indicate responsiveness. Of TAS2R43's 16 agonists, 15 are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

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

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

Supplementary material

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

Acknowledgments

The authors thank Stefanie Schultz for technical assistance.

Conflict of Interest

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

References

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

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

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

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

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

Campbell MC, Ranciaro A, Zinshteyn D, Rawlings-Goss R, Hirbo J, Thompson S, Woldemeskel D, Froment A, Rucker JB, Omar SA, *et al.* 2014. Origin and differential selection of allelic variation at *TAS2R16* associated with salicin bitter taste sensitivity in Africa. *Mol Biol Evol*. 31:288–302.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kim UK, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. 2003. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science*. 299(5610):1221–1225.

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

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

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

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

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

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

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

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

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

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

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

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

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

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, *et al*. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res*. 19:826–837.

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

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

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

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

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

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

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

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

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

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

Soranzo N, Bufe B, Sabeti PC, Wilson JF, Weale ME, Marguerie R, Meyerhof W, Goldstein DB. 2005. Positive selection on a high-sensitivity allele of the human bitter-taste receptor *TAS2R16*. *Curr Biol*. 15(14):1257–1265.

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

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

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

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

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

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

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

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

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

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

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

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

Wooding S, Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D. 2004. Natural selection and molecular evolution in *PTC*, a bitter-taste receptor gene. *Am J Hum Genet*. 74(4):637–646.

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

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

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