

SCIENTIFIC REPORTS

OPEN

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

Davide S. Rizzo^{1,2}, Massimo Mezzavilla^{3,4}, Luca Pagani^{2,5}, Antonietta Robino³, Gabriella Morini⁶, Sergio Tofanelli⁷, Maura Carrai⁷, Daniele Campa⁷, Roberto Barale⁷, Fabio Caradonna⁸, Paolo Gasparini³, Donata Luiselli², Stephen Wooding⁹ & Dennis Drayna¹

Received: 31 December 2015

Accepted: 18 April 2016

Published: 03 May 2016

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

More than 70 years ago A.L. Fox reported that phenylthiocarbamide (PTC) tastes extremely bitter to some people (defined as “tasters”) but not bitter at all to others (“non-tasters”)^{1,2}. Since then, numerous family, twin and population studies have shown that the inability to taste PTC is inherited in a nearly Mendelian recessive manner^{3–5}. In more recent studies, the use of the chemically similar 6-n-propylthiouracil (PROP) has often been substituted for PTC because of its ability to generate similar taste responses in humans and its better known toxicity profile⁶. In 2003, a locus that explained approximately 75% of the variation in PTC sensitivity was identified on chromosome 7⁴. At this locus, variation in the *TAS2R38* bitter receptor gene was subsequently found to underlie all of the bimodal distribution of this phenotype and to explain >70% of the total phenotypic variance⁷. *TAS2R38* is a member of the *TAS2R* bitter taste receptor gene family, which in humans consists of 25 functional genes and 11 pseudogenes, many of which show signatures of natural selection^{8–11}. Three single nucleotide polymorphisms (*rs714598*, *rs1726866*, *rs10246939*) at positions encoding amino acids 49, 262 and 296 represent the most common variant alleles of *TAS2R38*, and comprise the “taster” PAV (Proline, Alanine, Valine) and “non-taster” AVI (Alanine, Valine, Isoleucine) haplotypes. In addition, two rare (frequency <5%) (AAV and AAI) and two extremely rare (frequency <1%) (PAI and PVI) haplotypes have been identified. The other two possible haplotypes (AVV and PVV) have been individually reported in two studies^{12,13} but not otherwise observed.

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

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

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

A longstanding question has been the reason for the presence of two high-frequency haplotypes in worldwide populations. Because the perception of bitter taste is thought to protect us from the ingestion of toxic substances, how could the presumably non-functional AVI haplotype come to high frequency in populations worldwide? Analyses of the frequency distribution of *TAS2R38* haplotypes in different populations showed significant positive values for Tajima's D statistic, low F_{ST} values (0.001–0.05) and a deep coalescent time (TMRCA ~1 million years old) for this locus^{9,14}, providing evidence that balancing selection maintained both the taster and non-taster alleles at high frequency. However, a study of inter-specific variations of bitter taste receptor genes showed that *TAS2R* genes, and in particular *TAS2R38*, have undergone relaxation of selection in humans when compared with many other mammals¹⁵.

Balancing selection hypotheses have suggested the possibility that the AVI non-taster allele encodes a fully functional receptor for another hypothetical bitter substance⁸. It has also been suggested that pathogens may have been the real targets of natural selection, since bitter receptors are expressed in the respiratory and enteric system^{16,17} and one study suggested that common polymorphisms in the *TAS2R38* gene were linked to significant differences in the ability of the upper respiratory cells to clear and kill bacteria¹⁸. In addition, a recent study showed that *TAS2R38* genotypes regulate innate immune responses to oral bacteria¹⁹.

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

Results

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

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

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

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

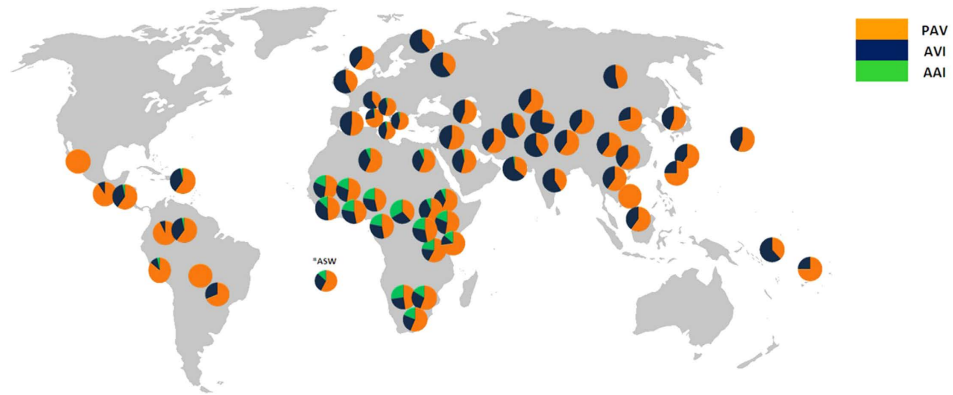


Figure 1. Worldwide distribution of *TAS2R38* PAV, AVI and AAI haplotypes in the studied populations. This map has been modified from its original version (<https://commons.wikimedia.org/wiki/File:BlankMap-World-noborders.png>).

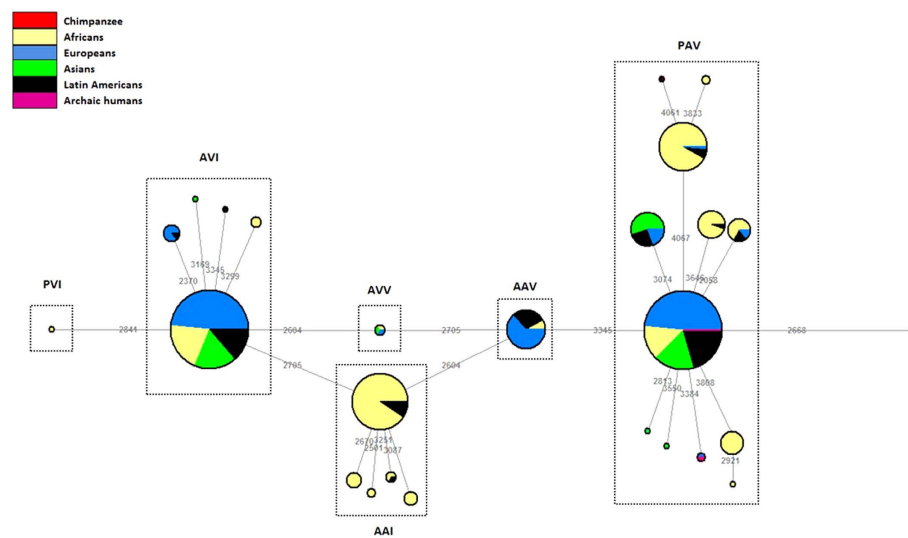


Figure 2. Neighbor-joining haplotype network illustrating the genealogical relationships between *TAS2R38* haplotypes in archaic hominids and modern African, European, Asian and Latin American populations. The position of the chimpanzee (outgroup) is highlighted in red.

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

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

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

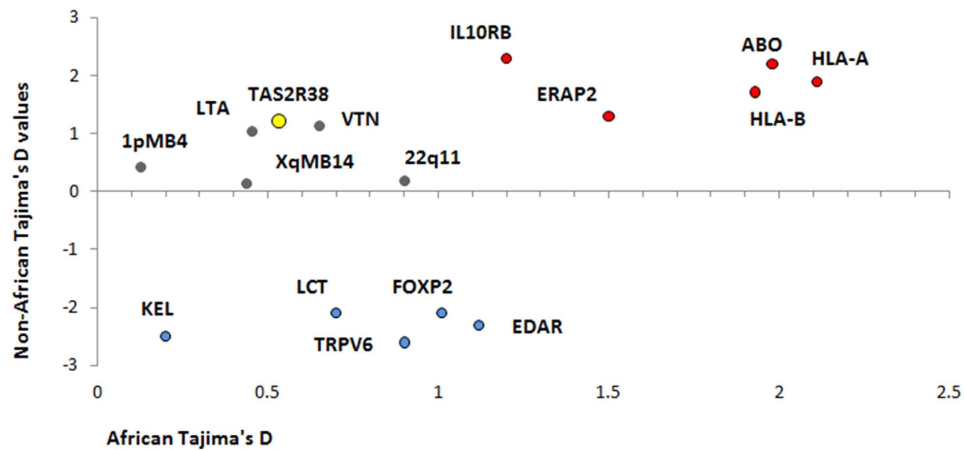


Figure 3. Comparison between Tajima's D values calculated in African and non-Africans for genes under different selective pressures. Genes under positive selection are represented with blue circles, genes under balancing selection with red circles and neutral regions with grey circles. The position of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

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

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

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

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

Discussion

Although the presence of PTC or PROP has not been documented in nature, chemically similar compounds, known as glucosinolates, are found in common bitter foods, such as brussels sprouts, cabbage and broccoli²⁰. For this reason, it has been suggested that a correlation may exist between PROP taster status and dietary intake²¹, which could have important evolutionary consequences, a hypothesis supported by the results of a number of previous studies^{22–24}. Other studies have identified relationships between the intensity of PTC/PROP bitterness and the perception of several natural compounds^{25–28}. These correlations generally support the hypothesis that natural selection has been acting on *TAS2R38*, maintaining the AVI haplotype at roughly the same worldwide frequency as the PAV haplotype. In addition to this hypothesis, it has been suggested that the non-taster AVI haplotype may be a functional receptor for another bitter compound⁸ and that pathogens may have been the real targets of natural selection^{17,18}.

Using the largest sample available, we have confirmed the global predominance of the PAV and AVI haplotypes of *TAS2R38*. We also found that the AVI form is less common in Africa and that the AAI haplotype is primarily present in this continent, as previously reported⁹. The three common *TAS2R38* SNPs (i.e. *rs714598*, *rs1726866* and *rs10246939*) showed low F_{ST} values, indicating little genetic differentiation at these sites. This was previously interpreted⁹ as a footprint of balancing selection that maintained similar frequencies for the alternative alleles at these sites. However in contrast to previous studies^{8,9}, we failed to detect any recent departures from neutral expectations for the variation at the *TAS2R38* locus and surrounding regions. Although we found positive Tajima's D values in the *TAS2R38* coding region, these did not reach significance and were similar to those observed at evolutionary neutral loci. When correcting for population stratification and taking into account demographic events with Li's MFDM test, *TAS2R38* did not show significant P-values, suggesting that these variables may contribute to the observed positive Tajima's D values. Another study, although based on a much smaller number of individuals (N = 22), previously pointed out this possibility¹⁵. Other approaches that have previously successfully detected signs of balancing natural selection, including the HKA and Bayescan tests, failed to find any departures from neutrality in our data set. We also note that two other studies applying new methods to detect signatures of balancing selection failed to identify such signatures in *TAS2R38*^{29,30}.

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

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

The AAI haplotype seems to have a more complex evolutionary history. Its distribution is unique in Africa, aside from a few exceptions in Latin America presumably due to the African admixture in these populations³³ and in the Near East and southern Italy, regions with strong flow from Africa in historical times^{34,35}. From our simulations we hypothesize that the AAI haplotype has undergone weak but directional selection, in addition to the balancing selection that previously acted on PAV/AVI. This view is strengthened by a recent work studying a large collection of African populations, where the authors found that African *TAS2R38* haplotypes evolved under a more complex scenario that includes a combination of balancing and directional selective pressures⁹.

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

Materials and Methods

Sources of samples and genetic data. Archaic hominids studied in this project consisted of the Altai Neanderthal individual³⁶ and the Denisovan hominin from the same site³⁷. The modern human subjects consisted of both published databases and newly recruited samples. For the former, we included 1,361 individuals belonging to 17 worldwide populations from the 1000 Genomes Project Phase 3³⁸, 942 individuals from 53 different populations from the Human Genome Diversity Project (HGDP)³⁹ and 1,428 African individuals from 15 populations of the African Genome Variation Project (AGVP)^{40,41}. In addition, 419 individuals belonging to 20 communities living along the Silk Road in Caucasus and Central Asia^{42,43} and 1,439 adult healthy subjects from 9 different regions within Italy (this study) were included in the analysis. Individuals from different databases but belonging to the same populations were pooled together. A total of 11,178 chromosomes from 105 different worldwide populations were analyzed (Supplementary Table S3). All samples were anonymized, coded identifiers were assigned to them and each donor was given an information sheet. All experimental protocols were approved by the ethical committee of IRCCS-Burlo Garofolo Hospital, by the bioethics committee of Pisa University and by the National Institutes of Health (NIDCD protocol 01-DC-0230) combined Neurosciences/Blue Panel Institutional Review Board. Written informed consent was obtained from all subjects included in the study. All experiments and methods were performed in accordance with the approved guidelines and regulations.

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

Data analysis. Three *TAS2R38* SNPs (*rs714598*, *rs1726866*, *rs10246939*) were analyzed in the HGDP, AGVP, Silk Road and Italian samples. The program PHASE⁴⁴ was used to infer *TAS2R38* haplotypes in the Italian and Silk Road subjects. Individuals from the 1000 Genomes Project Phase 1 were used as a reference to impute *rs714598* in both the Silk Road and HGDP populations and *rs10246939* in some individuals of the AGVP population (N = 1,200), since these SNPs were not present in the genotyping arrays used to type these

individuals. Only haplotypes with posterior probability of 0.9 or above were considered. VCF files containing sequence data at the *TAS2R38* locus and its 500 bp upstream and downstream regions (7:141972131-141974273, GRCh38) were retrieved from both the 1000 Genomes Phase 3 Database (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>) and from the 50 and 30-fold coverage Neanderthal and Denisovan genomes aligned to the hg19 human reference sequence, after obtaining the genomic coordinates as appropriate (<http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/VCF/> and http://cdna.eva.mpg.de/denisova/VCF/hg19_1000g/) using VCFtools⁴⁵.

The software Arlequin v.3.5⁴⁶ was used to calculate basic population genetics statistics, such as nucleotide diversity, number of polymorphic sites and estimated heterozygosity, and to perform the Analysis of Molecular Variance (AMOVA). Allele frequencies between populations were compared performing Chi-square tests, and the Bonferroni correction was used to correct the nominal P-values (i.e. adjusted $P = P \text{ value} \times \text{number of individual tests}$), with the software PLINK v 1.07⁴⁷. This software was also used to calculate pairwise LD measures and to estimate haplotype blocks which were also confirmed with Haploview⁴⁸. For this analysis, we expanded the analyzed region to 50,000 base pairs in both directions. The haplotype network was constructed with Network 4.5⁴⁹ using a median-joining algorithm.

Testing natural selection. Departures from neutrality were tested using several approaches: we calculated Tajima's D values using the DNASP package⁵⁰ in *TAS2R38* and surrounding regions. In addition, we compared these values to the ones calculated across the genome for coding loci with sizes similar to *TAS2R38* (e.g. 1,143 bases) and to those calculated in candidate genes selected from literature known to have undergone some kind of selective pressure. Since it has been shown that Tajima's test may not be powerful enough to detect departures from neutrality in regions shorter than 5,000 bases⁵¹, Tajima's D values were also calculated in a 10,000 bases region encompassing the selected loci. Pairwise F_{ST} values for all pairs of populations and within the same sub-population for *TAS2R38* and surrounding regions were calculated with the software Arlequin v.3.5⁴⁶ and the significance of these statistics was tested using a coalescent simulation adapted from Hudson⁵². In addition, we calculated F_{ST} values for the three *TAS2R38* common SNPs. These values were then compared to the ones calculated across the genome for SNPs of similar frequency (e.g. MAF ranging from 0.42 to 0.47) and to the genome-wide F_{ST} distribution. In order to further explore the signatures of natural selection at *TAS2R38*, we performed Li's MFDM test⁵³, which has been shown to be very robust in distinguishing selection from demography, even in presence of balancing selection. In addition, the Bayesian regression method implemented in Bayescan v2.1⁵⁴ and the HKA test⁵⁵ were applied to our dataset. Finally, simulations of the evolution *TAS2R38* haplotypes outside Africa were performed with the python library simuPOP v1.1.4⁵⁶. We simulated several scenarios for the six *TAS2R38* haplotypes with increasing level of expansions (i.e. 10 fold, 20 fold and 40 fold) and different natural selection pressures (i.e. $s = 0.05$, $s = 0.01$ and $s = 0.001$). Each scenario was replicated 1000 times to generate a prediction of the resulting haplotype frequencies. Finally, sliding-window analyses (100 kb) of heterozygosity were calculated on the entire chromosome 7 in the the 1000 Genome Project populations. More details on the analytical methods can be found in the Supplementary Information.

References

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

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

Acknowledgements

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

Author Contributions

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

Additional Information

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

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

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



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