Title

Worldwide diversity, association potential, and natural selection in the superimposed taste genes, CD36 and GNAT3

Authors

Ramirez, Vicente A

Wooding, Stephen P

Publication Date

2022

DOI

10.1093/chemse/bjab052

Peer reviewed

Chemical Senses, 2022, **47**, 1–12
https://doi.org/10.1093/chemse/bjab052
Advance access publication 31 December 2021
Original Article

Worldwide diversity, association potential, and natural selection in the superimposed taste genes, *CD36* and *GNAT3*

Vicente A. Ramirez BA MS[1] and Stephen P. Wooding BA MS PhD[2]

[1]Department of Public Health, University of California, Merced, Merced, CA, USA and
[2]Department of Anthropology, University of California, Merced, Merced, CA, USA
Corresponding author: Stephen P. Wooding, Department of Anthropology, University of California, Merced, 5200 North Lake Road, Merced, CA 95343, USA.
e-mail: swooding@ucmerced.edu

Abstract

CD36 and GNAT3 mediate taste responses, with CD36 acting as a lipid detector and GNAT3 acting as the α subunit of gustducin, a G protein governing sweet, savory, and bitter transduction. Strikingly, the genes encoding CD36 and GNAT3 are genomically superimposed, with *CD36* completely encompassing *GNAT3*. To characterize genetic variation across the *CD36-GNAT3* region, its implications for phenotypic diversity, and its recent evolution, we analyzed from ~2,500 worldwide subjects sequenced by the 1000 Genomes Project (1000GP). *CD36-GNAT3* harbored extensive diversity including 8,688 single-nucleotide polymorphisms (SNPs), 414 indels, and other complex variants. Sliding window analyses revealed that nucleotide diversity and population differentiation across *CD36-GNAT3* were consistent with genome-wide trends in the 1000GP ($\pi = 0.10\%$, $P = 0.64$; $F_{ST} = 9.0\%$, $P = 0.57$). In addition, functional predictions using SIFT and PolyPhen-2 identified 60 variants likely to alter protein function, and they were in weak linkage disequilibrium ($r^2 < 0.17$), suggesting their effects are largely independent. However, the frequencies of predicted functional variants were low ($\overline{P} = 0.0013$), indicating their contributions to phenotypic variance on population scales are limited. Tests using Tajima's *D* statistic revealed that pressures from natural selection have been relaxed across most of *CD36-GNAT3* during its recent history ($0.39 < P < 0.67$). However, *CD36* exons showed signs of local adaptation consistent with prior reports ($P < 0.035$). Thus, *CD36* and *GNAT3* harbor numerous variants predicted to affect taste sensitivity, but most are rare and phenotypic variance on a population level is likely mediated by a small number of sites.

Key words: taste, genetics, evolution, fat

Introduction

Taste perception is a fundamental mechanism of diet selection and control. By allowing animals to evaluate the nutritional properties and safety of foods before they are consumed, taste provides a powerful means of enhancing health and evolutionary fitness (Lindemann 2001; Reed and Knaapila 2010; Roper and Chaudhari 2017). For instance, bitter sensations, which are triggered by plant toxins, signal the presence of noxious components, allowing avoidance. Sweet sensations, which are triggered by sugars, signal carbohydrate richness. Salty, sour, and umami/savory sensations signal the presence of electrolytes, acidity indicative of ripeness, and protein content. Together these modalities provide a nutrient profile that can be used to guide intake, a major foraging advantage. The significance of this role is evident in the diversity of taste receptors found throughout vertebrates (Fischer et al. 2005; Wooding et al. 2006; Zhao et al. 2010, 2015; Wooding 2011; Jiang et al. 2012; Baldwin et al. 2014; Feng et al. 2014; Li and Zhang 2014; Antinucci and Risso 2017; Behrens et al. 2020).

A key feature of taste perception in humans is that it varies due to polymorphism in genes encoding receptors and other signaling components (Kim et al. 2004; Bachmanov et al. 2014). For example, *TAS2R38*, which encodes a bitter receptor, harbors alleles associated with taste responses to goitrin, a thyroid toxin synthesized by plants in the *Brassicaceae* family (Wooding et al. 2010). Similar associations are found between variants in TAS1R3 (an umami receptor subunit) and monosodium glutamate and between CA6 variants (a salivary carbonic anhydrase) and sodium salt (Chen et al. 2009; Feeney and Hayes 2014). Polymorphism in taste pathways also associates with preferences and consumption of foods such as alcoholic beverages and cruciferous vegetables, and health measures such as body mass index, susceptibility to colorectal cancers, and kidney disease (Greene 1974; Basson et al. 2005; Wooding et al. 2012; Behrens et al. 2013; Allen et al. 2014; Hayes et al. 2015; Choi et al. 2016; Barontini et al. 2017). These affect evolutionary fitness, and taste genes in humans harbor signatures of natural selection including evidence of local adaptation, balancing pressures, and purifying effects (Wooding et al. 2004; Drayna 2005; Kim et al. 2005, 2006; Campbell et al. 2012, 2014; Risso et al. 2016, 2018). Thus, modern patterns of variation in taste sensitivity, nutrition, and health reflect ancient evolutionary influences on taste.

Mounting evidence suggests that human taste abilities extend to the detection of fats, particularly long-chain fatty acids (LCFAs), and that fat taste sensitivity varies from person to person as the result of genetic polymorphism. In psychophysical assays, subjects are capable of discriminating fat content in controlled preparations even when non-gustatory cues are masked, supporting a role for taste

Editorial Decision 22 November 2021.

(Chalé-Rush et al. 2007; Cartoni et al. 2010; Mattes 2011). In addition, like other taste signals, neural signals generated by oral fat exposure originate in taste receptor cells and travel via the chorda tympani and glossopharyngeal nerves in mice (Gaillard et al. 2008). Oral fat exposure also activates the brain's insular cortex, which is activated during sweet perception (De Araujo and Rolls 2004). Several lines of evidence indicate that CD36, a fatty acid translocase, is the receptor accounting for these effects. It localizes to taste receptor cells, natively responds to fatty acids in vitro, and knockout of *CD36* in rats and mice alters their preferences for fat-containing solutions and foods (Laugerette et al. 2005). *CD36* also harbors alleles associated with both orosensory detection of fats and preferences for them (Keller et al. 2012; Pepino et al. 2012). In addition, CD36 plays known roles beyond taste, contributing to immune system function, lipid metabolism, and cell adhesion (Pepino et al. 2014). These findings raise questions about the extent of genetic polymorphism at *CD36* and its effects on fat perception and other phenotypes.

The potential contributions of CD36 to fat taste also raise evolutionary questions. The high nutritional value of lipids, which are calorically rich but environmentally scarce, suggests that CD36's role as a taste sensor placed it under selective pressures in the course of human evolution. In particular, humans' population expansion and migration out of Africa 50–60 thousand years ago introduced them to new physical and nutritional environments that likely altered the advantages of fat perception and metabolism. For instance, they may have been shaped by factors such as the accessibility of fats when hunting and foraging, or climate, which poses thermoregulatory challenges. They could also have arisen from CD36's non-gustatory roles in processes such as cell adhesion, which makes it vulnerable to exploitation by pathogens (Silverstein and Febbraio 2009). Such pressures leave signatures in genetic diversity including effects on allele frequencies and population differentiation (Bamshad and Wooding 2003). Thus, modern patterns of diversity in *CD36* may provide clues to the evolutionary factors driving responses to fats.

Strikingly, *CD36* genomically encompasses a second gene participating in taste perception, *GNAT3* (Fig. 1). *GNAT3* encodes a G-protein subunit mediating detection of sweet, savory, and bitter substances and, like *CD36*, harbors variants associated with taste sensitivity (Fushan et al. 2010; Farook et al. 2012; Pepino et al. 2012). GNAT3 plays non-gustatory roles as well, such as the detection of foreign compounds in the gut and airways (Egan and Margolskee 2008; Deshpande et al. 2010). The nested arrangement of the two genes suggests that patterns of diversity in them may be correlated due to their genetic linkage and shared evolutionary histories. If so, GNAT3-mediated taste responses (bitter, sweet, and umami) and CD36-mediated responses (fat) may be correlated. However, the *CD36-GNAT3* region is sufficiently large (~305 kb) that linkage disequilibrium (LD) may not be high across its entirety. LD can also be shaped by natural selection, which can affect its range and magnitude. Establishing the structure of genetic variation across *CD36-GNAT3* has the potential to reveal the extent of such effects.

We addressed these issues in a population genetic analysis of *CD36-GNAT3* in >2,500 subjects from the 1000 Genomes Project (1000GP; Li 2011; The 1000 Genomes Project Consortium 2015). To establish the extent of diversity at *CD36-GNAT3* and its implications for genotype–phenotype associations, we comprehensively identified variable sites in the region, their allele frequencies in worldwide populations, and their linkage structure. We then used computational prediction to detect sites likely to have functional effects, and evolutionary analyses to determine the role of natural selection in shaping this variability. Our results shed light on the architecture of diversity in *CD36* and *GNAT3*, its potential contributions to taste and metabolism, and its evolutionary origins.

Methods

We examined genetic variation across *CD36-GNAT3* in 2,504 subjects included in Phase 3 of the 1000GP (The 1000 Genomes Project Consortium 2015). The 1000GP subjects

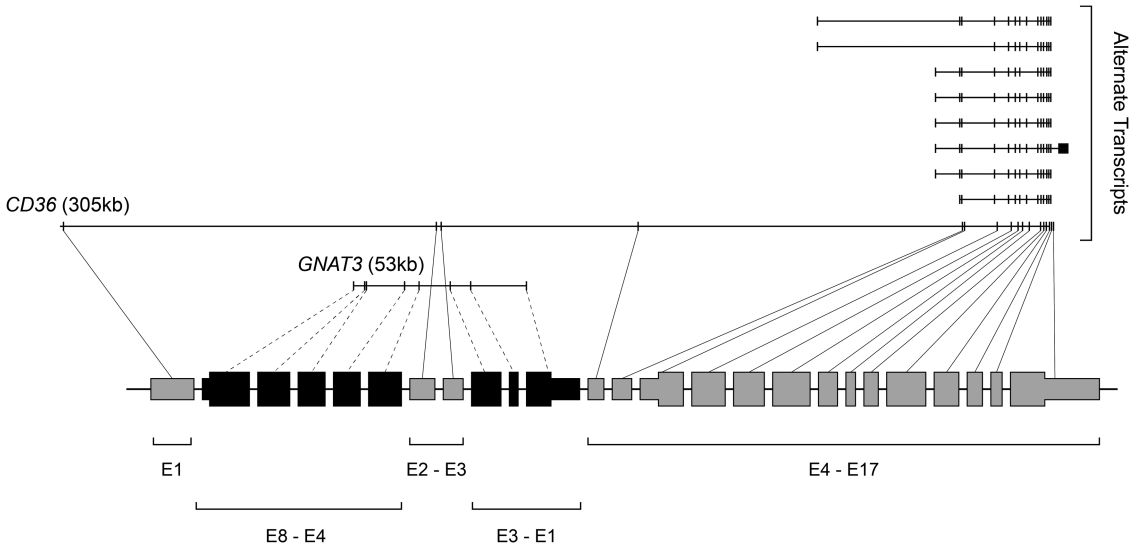

Fig. 1. Genomic organization of *CD36* and *GNAT3*. The *CD36-GNAT3* region is ~305 kb in length. *GNAT3* is nested within *CD36*, with exons 1–3 located in *CD36* intron 3 and exons 4–8 located in *CD36* exon 1.

comprise a random, demographically representative sample of 26 worldwide populations in five superpopulations, providing a diverse hierarchical perspective on human genetic variation (Table 1).

The genomic structure of the *CD36-GNAT3* region was determined from the Ensembl GRch37 human genome assembly, the reference for the 1000GP. These placed *CD36* (Ensembl ENSG00000135218, ENST00000435819) at position 7:79998891–7:80308593 (~305 kb) and *GNAT3* (ENSG00000214415, ENST00000398291) at position 7:80087987–7:80141336 (53 kb), with *GNAT3* located in introns 1 and 2 of *CD36*. Data for the region were extracted from 1000GP databases in variant call format (VCF) using the *Tabix* software package (Li 2011).

Genetic variation was assessed with respect to three factors, allelic polymorphism, population substructure, and LD. Nucleotide diversity (π), the mean pairwise nucleotide difference among sequences normalized to sequence length, was calculated across *CD36-GNAT3* as well as separately for *CD36* exons, *CD36* introns, *GNAT3* exons, and *GNAT3* introns (Tajima 1983). Population substructure was measured within and across 1000GP superpopulations using Weir and Hill's weighted F_{ST} (Weir and Hill 2002). These calculations were performed using *VCFtools* and the R packages PopGenome, pegas, hierfstat, and adgenet (Goudet 2004; Paradis 2010; Danecek et al. 2011; Jombart and Ahmed 2011; Pfeifer et al. 2014). LD was assessed for variants with frequencies >0.05 using two measures, D' and r^2 (Mueller 2004; Slatkin 2008). D', a measure of correlation between sites relative to the maximum possible given their allele frequencies, was used to determine the extent to which recombination has shaped diversity across the region. A raw measure of correlation among genetic markers, r^2, was used to determine the extent to which sites are expected to exhibit similar genotype–phenotype associations. These measures were calculated and visualized using the VariantAnnotation and Ldheatmap packages in the R statistical analysis environment and Bioconductor library (Ihaka and Gentleman 1996; Gentleman et al. 2004; Graham et al. 2006; Obenchain et al. 2014).

Table 1. Population sample.

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

Two algorithms were used to predict the potential functional impact of exon variants, PolyPhen-2 and SIFT (Kumar et al. 2009; Adzhubei et al. 2010). PolyPhen-2 predicts the impact of amino acid changes on protein function on the basis of the location of the changed site within the protein structure, level of conservation relative to homologous genes, and the biochemical characteristics of the substituted amino acids. It denotes the impact of substitutions on a scale from 0.0 (benign) to 1.0 (damaging). SIFT predicts whether amino acid substitutions affect protein function on the basis of probabilities estimated from gene homologs. It denotes impact on a scale from 0.0 to 1.0, categorizing scores below 0.05 as deleterious and higher scores as tolerated. Both scores were obtained using the Variant Effect Predictor (VEP) software package (McLaren et al. 2016). Regulatory variants were identified by using VEP to query the Ensembl Regulatory Build.

Tests for natural selection were performed using Tajima's *D* statistic (Tajima 1989). *D* compares the number of variable sites and mean nucleotide difference between alleles in a sample, which are differentially affected by selective processes. It is designed to test for selective effects nonrecombining genomic regions but is applicable with elevated conservativeness to regions with recombination. These analyses were performed using the PopGenome R package. As with pi estimates, these were calculated both overall and with respect to *CD36* and *GNAT3* exons and introns.

Because the vast majority of the human genome (>98%) is noncoding it is can be assumed to be evolving neutrally or nearly so with respect to natural selection. Therefore, to provide a neutral baseline for evaluating diversity measures in the *CD36-GNAT3*, we generated empirical distributions for three measures (π, F_{ST}, and *D*) using sliding window analyses. These were obtained by iteratively calculating each measure in ~270,000 adjacent 10 kb windows spanning the length of the 1000GP genomes, excluding known unstable and repetitive regions such as telomeres. The probability of the observed values given genome-wide trends was then determined by comparing the values observed in *CD36-GNAT3* with their genome-wide distributions. We denoted *P*-values from these empirical tests P_E to distinguish them from *P*-values obtained in parametric tests.

Results

The *CD36-GNAT3* region harbored extensive variation. A total of 9,111 polymorphic sites were identified. The majority (95.3%) were single-nucleotide polymorphisms

(SNPs; 8,653 biallelic and 32 multiallelic). A smaller number (4.5%) were insertion/deletion (indel) polymorphisms (390 biallelic and 21 multiallelic). The remainder (<0.2%) were rare complex variants, including three sites with both SNP and indel alleles, 11 copy number variants (CNVs), and one Alu insertion. *CD36* exons, which totaled 2,365 bp in length, contained 112 SNPs (all biallelic) and 8 indels, which occurred in 7 exons (Fig. 2). Four of the eight were frameshift deletions, two were 1 bp insertions in untranslated regions, and two were in-frame deletions. *GNAT3* exons, which totaled 1,159 bp in length, contained 28 SNPs (all biallelic), with one indel in the 5′-untranslated region of exon 1. The number of variants also differed among exons. No polymorphism was found in *GNAT3* exon 2, 17 SNPs were present in *CD36* exon 17, and the mean across exons was 6. The number of SNPs per nucleotide ranged from 0.0 (*GNAT3* exon 2) to 0.1 (*CD36* exon 12), with an average of 0.037.

Among the 140 SNPs in *CD36* and *GNAT3* exons, PolyPhen and SIFT detected 57 likely to alter function in the *CD36* and *GNAT3* proteins. Fifty-one had PolyPhen scores of Possibly or Probably Damaging and 51 had SIFT scores of Deleterious. The scores were largely in agreement between the two measures. Forty variants were scored as Possibly/Probably Damaging by PolyPhen and Deleterious by SIFT, and a further 15 were sites scored as Benign by PolyPhen and Tolerated by SIFT. Predictions disagreed for 12 sites, with 6 scored as tolerated by SIFT but Possibly/Probably Damaging by PolyPhen, and 6 scored as Deleterious by SIFT but tolerated by PolyPhen. Sites without SIFT and PolyPhen scores but likely to have functional impact were also found. Eighteen variants in *CD36* were indels causing frameshifts (4 sites) stop gains (10 sites), stop losses (2 sites), or in-frame deletions (2 sites). *GNAT3* harbored one stop gain (in exon 3) and one stop loss (in exon 8). For further analyses, the 60 sites with variants scored as possibly damaging by PolyPhen and Deleterious by SIFT, frameshifts, and those occurring in start or stop codons were denoted putatively high impact (PHI) sites, and their derived alleles denoted PHI alleles (Table 2).

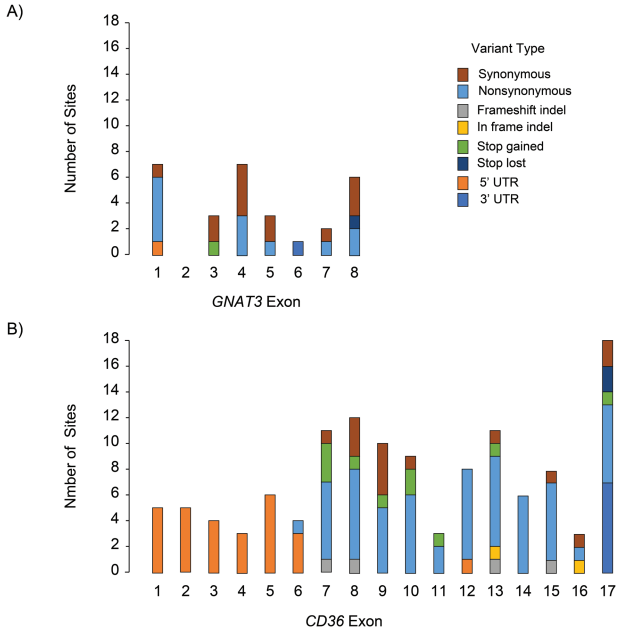

Fig. 2. Variant types and their frequencies across *CD36* and *GNAT3* exons. Both *GNAT3* and *CD36* harbored extensive variation, including numerous variants likely to affect function.

Potential regulatory variation was also found. VEP identified 48 regulatory regions and 10 features as transcription factor binding sites, which together harbored 685 variants annotated by Ensembl as expression modifiers. One hundred and thirty-one of these were most proximal to *GNAT3* exons, and 554 were most proximal to *CD36* exons. The majority (667; >95%) were SNPs, but other variant types were also present. Ten short insertions and 16 short deletions were found, along with six copy numbers and structural variants ranging from ~2 kb to ~130 kb in length, which spanned numerous regulatory and transcription factor binding sites.

As expected given that the majority of the *CD36-GNAT3* region is intronic, most variants (7,827) occurred in noncoding regions (Table 3). A smaller number, 686, occurred in regulatory regions. In addition, consistent with their relative lengths, *CD36* harbored more segregating sites than did *GNAT3* (112 vs. 28). Derived allele frequencies at noncoding sites ranged from 0.0002 (singletons) to 0.9998 with a mean of 0.033. Consistent with the low mean, the majority of alleles were rare, with 92% having frequencies below 0.05, and 83% having frequencies below 0.01 (Table 4). Alleles with intermediate frequencies accounted for a small proportion of sites, with 4% having frequencies between 0.25 and 0.75. These patterns extended to *CD36* and *GNAT3* exons and regulatory regions, with >80% of alleles having frequencies below 0.01 in all three cases.

Consistent with the abundance of low-frequency alleles, π across noncoding sites was low, 0.10% (Table 3). It had a P_E of 0.64 in the sliding window analysis, indicating that is consistent with expectations given the 1000GP sample. Values observed within superpopulations were similar to π values across the sample as a whole, ranging from 0.08 in Europe to 0.12 in Africa. Little difference in diversity was observed with respect to *CD36* and *GNAT3* exons within superpopulations, with π in *CD36* ranging from 0.04% (in Europe) to 0.05% (in Africa) and π in GNAT3 ranging from 0.03% (in East Asians) to 0.04% (in Africans). Differences in diversity between introns and exons at both *GNAT3* and *CD36* were also small, with *CD36* having π values of 0.08% and 0.05% in introns and exons, respectively, and *GNAT3* having values of 0.07% and 0.04%.

The frequencies of PHI alleles ranged from 0.0002 to 0.031 with a mean of 0.0013. Thus, while the minimum frequency of PHI alleles was identical to that across noncoding sites, the mean was substantially lower (0.0013 vs. 0.0333). Nucleotide diversity could not be calculated for PHI sites or regulatory variants because the denominator in π calculations, the length of the analyzed sequence, is indeterminate because not all sites in the genome have potential to harbor PHI or regulatory variants. However, heterozygosity estimates were consistent with the distributions of allele frequencies at both noncoding and PHI sites, with 80% of noncoding and 97% of PHI sites having heterozygosities below 0.025. As with π, this pattern extended to superpopulations. Alleles scored as modifiers in regulatory regions ranged in frequency from 0.0002 to 0.970 with a mean of 0.037, and 83% having frequencies below 1%.

The overall F_{ST} of *CD36-GNAT3* among superpopulations was 9.0% with a P_E of 0.57 (Table 3). Pairwise F_{ST} values

Table 2. Putative high impact (PHI) sites.

Gene	rsid	Exon	Variant type	Reference codon	Alternate codon	Reference amino acid	Alternate amino acid
GNAT3	rs533524866	1	Ns	aCc	aAc	T	N
GNAT3	rs573082324	3	Sg	tTg	tAg	L	Stop
GNAT3	rs200010494	4	Ns	Gat	Aat	D	N
GNAT3	rs571120313	4	Ns	cTg	cCg	L	P
GNAT3	rs570030158	5	Ns	ttG	ttT	L	F
GNAT3	rs186877232	8	Sl	Taa	Caa	Stop	Q
GNAT3	rs534902139	8	Ns	Ttc	Gtc	F	V
CD36	rs559876270	6	Ns	gGg	gAg	G	E
CD36	rs75326924	7	Ns	Cct	Tct	P	S
CD36	rs139067066	7	Sg	tgG	tgA	W	Stop
CD36	rs150037612	7	Ns	aCg	aTg	T	M
CD36	rs534577878	7	Ns	tgG	tgC	W	C
CD36	rs545489204	7	Sg	Cag	Tag	Q	Stop
CD36	rs556181210	7	Ns	aTc	aAc	I	N
CD36	rs571975065	7	Fs	Aaa	aa	K	na
CD36	rs574416705	7	Sg	taC	taG	Y	Stop
CD36	rs70961715	8	Ns	cGt	cCt	R	P
CD36	rs201765331	8	Ns	tCa	tTa	S	L
CD36	rs548507859	8	Ns	Tca	Cca	S	P
CD36	rs556438655	8	Sg	Gaa	Taa	E	Stop
CD36	rs563097847	8	Ns	Ctc	Ttc	L	F
CD36	rs572295823	8	Fs	aAC	a	N	na
CD36	rs201759307	9	Ns	Tgg	Cgg	W	R
CD36	rs568503917	9	Ns	gGc	gTc	G	V
CD36	rs569959776	9	Sg	taT	taG	Y	Stop
CD36	rs35776095	10	Ns	gGa	gAa	G	E
CD36	rs373829578	10	Sg	Aaa	Taa	K	Stop
CD36	rs200067322	10	Ns	Gga	Aga	G	R
CD36	rs535150936	10	Ns	gGt	gTt	G	V
CD36	rs201245766	10	Sg	agg	agGTAAg	R	Stop
CD36	rs149178142	11	Ns	aCa	aTa	T	I
CD36	rs149985988	11	Sg	tgC	tgA	C	Stop
CD36	rs557732736	11	Ns	aTt	aAt	I	N
CD36	rs142186404	12	Ns	Ttt	Gtt	F	V
CD36	rs145908803	12	Ns	cCa	cTa	P	L
CD36	rs199681631	12	Ns	aGg	aCg	R	T
CD36	rs201155452	12	Ns	cCt	cAt	P	H
CD36	rs535549168	12	Ns	tTg	tCg	L	S
CD36	rs3211938	13	Sg	taT	taG	Y	Stop
CD36	rs200757788	13	Ns	aGa	aTa	R	I
CD36	rs554019170	13	Ns	gAc	gGc	D	G
CD36	rs558115067	13	Fs	ctG	ct	L	na
CD36	rs567491856	13	Ns	tGt	tTt	C	F
CD36	rs571553184	13	If	aAAGaa	aaa	KE	K
CD36	rs147903735	14	Ns	Cat	Tat	H	Y
CD36	rs370701210	14	Ns	Gca	Cca	A	P
CD36	rs371884082	14	Ns	cAt	cGt	H	R
CD36	rs376311045	14	Ns	Cca	Tca	P	S
CD36	rs564971571	14	Ns	Cct	Tct	P	S
CD36	rs148910227	15	Ns	Cgg	Tgg	R	W
CD36	rs200194486	15	Ns	aCt	aGt	T	S
CD36	rs200906462	15	Ns	Act	Cct	T	P

Table 2. Continued

Gene	rsid	Exon	Variant type	Reference codon	Alternate codon	Reference amino acid	Alternate amino acid
CD36	rs201355711	15	Ns	cAg	cTg	Q	L
CD36	rs551607784	15	Fs	Gca	ca	A	na
CD36	rs550565800	16	If	taTATTGTGCCTATt	tat	YIVPI	Y
CD36	rs201558608	17	Ns	Ggt	Agt	G	S
CD36	rs550163799	17	Sl	Taa	Gaa	Stop	E
CD36	rs559916528	17	Ns	gGt	gCt	G	A
CD36	rs563772337	17	Sg	Caa	Taa	Q	Stop
CD36	rs570171917	17	Sl	tAa	tCa	Stop	S

Table 3. Genetic diversity in populations and superpopulations. Numbers in parentheses indicate nonsynonymous variants.

Site category	Worldwide	Africa	Americas	East Asia	Europe	South Asia
S						
Noncoding	7,827	3,885	2,624	2,124	2,120	2,443
CD36 exons	112 (98)	43 (38)	22 (19)	30 (27)	21 (17)	24 (19)
GNAT3 exons	28 (15)	10 (6)	10 (6)	5 (2)	6 (2)	9 (5)
PHI sites	60	24	10	24	8	14
Regulatory sites	686	333	237	180	194	227
π (%)						
Noncoding	0.100	0.121	0.087	0.092	0.079	0.087
CD36 exons	0.045	0.054	0.038	0.042	0.041	0.042
GNAT3 exons	0.036	0.038	0.035	0.026	0.033	0.034
PHI sites	—	—	—	—	—	—
Regulatory sites	—	—	—	—	—	—
F_{ST}						
Noncoding	0.090	0.015	0.019	0.015	0.002	0.005
CD36 exons	0.045	0.079	0.000	0.011	0.004	0.000
GNAT3 exons	0.093	0.006	0.030	0.001	0.007	0.015
PHI sites	0.066	0.108	0.000	0.015	0.000	0.007
Regulatory sites	0.080	0.016	0.031	0.003	0.001	0.007

Table 4. Allele frequency quantiles.

	Frequency quantile					
Site category	<0.01	<0.02	<0.03	<0.04	<0.05	<0.05
Noncoding	0.83	0.04	0.02	0.02	0.01	0.09
CD36 exon	0.96	0.01	0.00	0.01	0.00	0.02
GNAT3 exon	0.96	0.00	0.00	0.00	0.00	0.04
PHI	0.97	0.02	0.00	0.02	0.00	0.00
Regulatory	0.83	0.05	0.02	0.01	0.00	0.09

ranged from a low of 1.1% (between Europe and the Americas) to a high of 13.2% (between Africa and Europe). F_{ST} values among populations within superpopulations were smaller, ranging from 0.003 in Europe to 0.018 in the Americas. The F_{ST} among superpopulations was lower for both PHI sites (6.6%) and regulatory sites (8.0%) than for noncoding sites (9.0%; Table 3). This pattern held for pairwise F_{STs} between superpopulations, in which F_{ST} for noncoding variants was always higher than F_{ST} for PHI and regulatory sites. However, F_{ST} within superpopulations did not follow this pattern (Table 3). F_{ST} for PHI sites was similar to or less than F_{ST} for noncoding sites in four populations (Americas, East Asia, Europe, and South Asia). In contrast, F_{ST} in Africa was substantially higher for *CD36* exons than for noncoding sites (7.9% vs. 1.5%) and higher still for PHI sites, 10.8%.

D´ values across *CD36-GNAT* were high (0.7–1.0) across localized regions separated by regions of lower LD (<0.5), pointing to the presence of 7 haplotype blocks ranging from ~10 kb to ~100 kb in length (Fig. 3). A ~70 kb *D´* block was centered on *GNAT3* and spanned its full length (53 kb). *CD36* exons were distributed across 5 of the 7 blocks, which contained exons 1, 2–3, 4, 5–8, and 9–17, respectively. In contrast to *D´*, r^2 was low across *CD36-GNAT3* with the exception of highly localized areas. *CD36* exons 10–17 were located in a block with r^2 near 1.0. LD calculated among all PHI sites was consistent with patterns

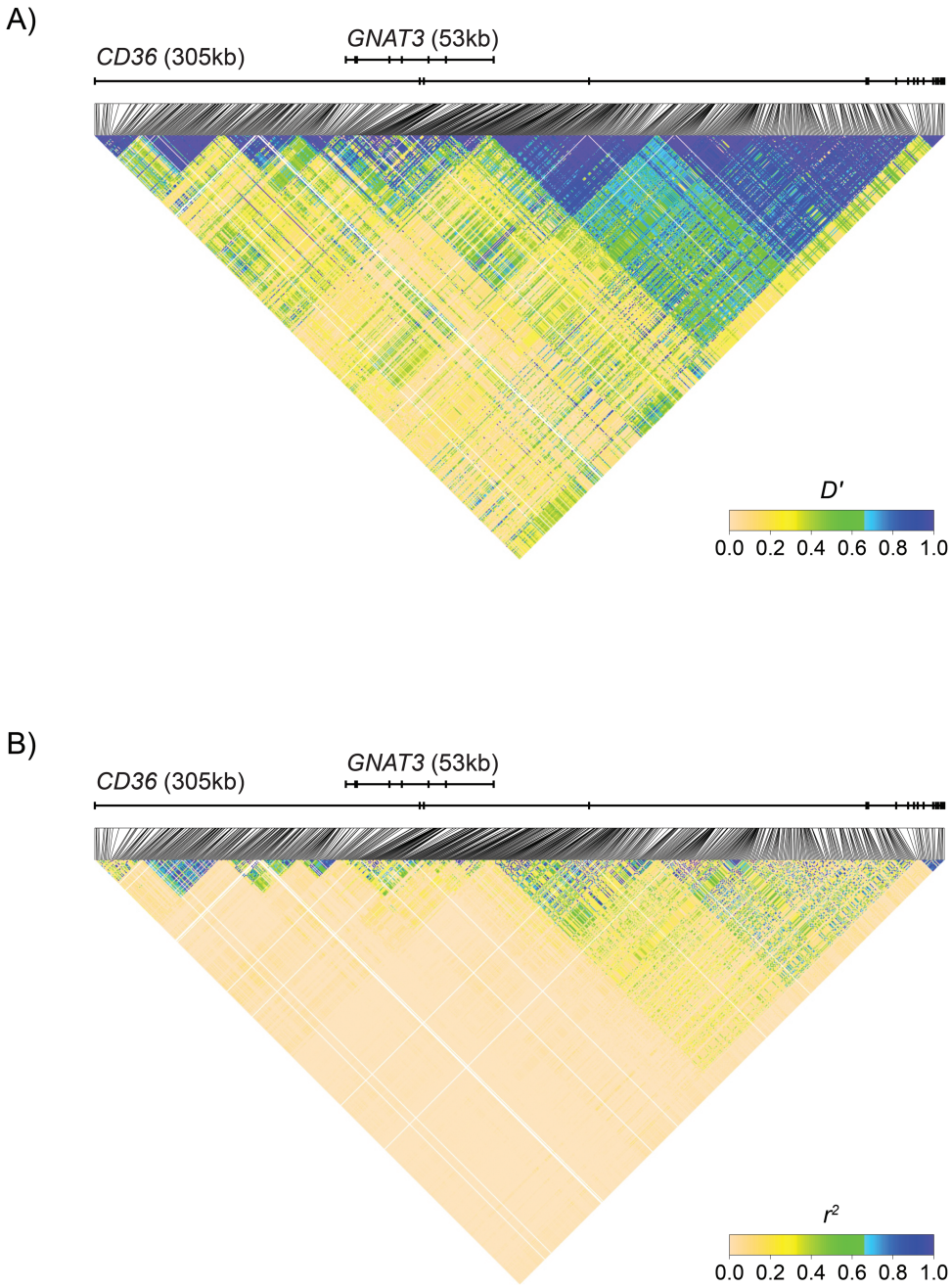

Fig. 3. Linkage disequilibrium across *CD36-GNAT3*. (A) Pairwise *D′*. (B) Pairwise r^2.

expected when allele frequencies are low. In general, pairwise LD measures between rare variants take on extreme values, with *D′* values near 1 and r^2 values near 0. This pattern held across all but one pair of PHI sites. No pair of PHI sites had an r^2 above 0.015 or *D′* below 0.995 with the exception of rs563097847 and rs558115067, which had an r^2 of 0.17.

Tajima's *D* values were strongly negative for the primary site categories (all sites, exons, and introns; Table 5). However, their statistical significance depended on the test used. While the standard *D* test rejected neutrality at a high level of significance ($P < 0.001$), comparisons with the sliding window distribution yielded different results. In the sliding window analyses, only the *D* value of *CD36* exons departed from expectations, and marginally so with $P_E = 0.035$. Across the other four categories, P_E ranged from 0.390 to 0.665, well within expectations (Table 5).

Discussion

Contemporary human populations are the product of a rapid expansion of small ancestral populations out of Africa 50–60 thousand years ago (Bergström et al. 2021). This process is evident in diversity patterns in human genes, which harbor signatures of ancient demography and natural selection (Bamshad and Wooding 2003; Marth et al. 2003). Most obvious on a genome-wide scale are a paucity of variation and downward skew in allele frequencies, patterns attributed to a combination of early population bottlenecks and pervasive pressure from purifying selection (Rogers 1995; Cvijović et al.

Table 5. Neutrality Tests. Tests rejecting the null hypothesis with $P < 0.001$ are indicated by ***. Tests rejecting the null hypothesis with $P < 0.05$ are indicated by *.

Region	Tajima's *D*	*P*	P_E
All sites	−1.87	<0.001***	0.608
CD36 exons	−2.41	<0.001***	0.035*
CD36 introns	−1.86	<0.001***	0.620
GNAT3 exons	−1.81	<0.001***	0.665
GNAT3 introns	−2.05	<0.001***	0.390

2018). However, while prevalent, these processes can act simultaneously with others, and patterns of diversity in individual genes can reveal the evolutionary underpinnings of specific traits (Hancock and Di Rienzo 2008). For instance, unexpectedly high LD and F_{ST} in the lactase (*LCT*) gene indicate that positive selection favored variants conferring lactase persistence in early herding peoples, where ability to digest milk was a fitness advantage (Bersaglieri et al. 2004). Conversely, high diversity and low LD in calpain-10 (*CAPN10*) signal the presence of long-term balancing selection at a locus implicated in energy use and storage, with the explanation being that more than one allele has been selectively maintained for an extended period and LD has decayed (Vander Molen et al. 2005). The overlapping structures of *CD36* and *GNAT3* together with their roles in taste and metabolism raise questions about the extent of diversity at these loci, its potential impact on phenotypes, and the roles of demography and natural selection in shaping it.

Worldwide diversity across *CD36-GNAT3*

We found that patterns of nucleotide diversity across *CD36-GNAT3* were consistent with prevailing genome-wide trends and inferences about human origins. Diversity in noncoding regions, π = 0.10%, was well within expectations given the empirical distribution (P_E = 0.64) and similar to numerous prior estimates, which vary depending on the populations sampled but are typically 0.075%–0.125% (Table 3; The 1000 Genomes Project Consortium 2010). Such values fall far below theoretical expectations given humans' current population size. They are more consistent with paleoanthropological and genetic evidence that ancient populations sizes were small and attained their current size relatively recently (Henn et al. 2012; Osada 2015; Bergström et al. 2021). Nucleotide diversity within continents was similarly consistent with contemporary findings on human evolution. In particular, it was somewhat higher in Africa (π = 0.12%) than in other superpopulations (π = 0.09%). This pattern is widely observed and usually ascribed to the antiquity of African populations, which have diverged over a longer period than populations founded during humans' recent expansion (Yu et al. 2002; Campbell and Tishkoff 2008; Henn et al. 2012; Bergström et al. 2021).

Patterns of population differentiation with respect to noncoding variation in our sample were also consistent with genome-wide trends and human origins. F_{ST} among superpopulations with respect to noncoding sites, 9%, was typical for values in the 1000GP (P_E = 0.57) and similar to previous estimates among continents, which are typically near 10% (Table 3; Holsinger and Weir 2009). For instance, it is only slightly higher than the mean observed in HapMap phase 3 data (8.0%), which have a similar population composition (Elhaik 2012). Pairwise F_{ST} values between superpopulations were also consistent with previous estimates, varying widely depending on the continents being compared (1.1%–13.2%; Elhaik 2012). The mean within-superpopulation F_{ST} in our sample, 1.2%, was similarly consistent with previous genome-wide estimates, such as the 1% reported in HapMap phase 3 populations by Elhaik (2012). As with π, these values reflect expectations following humans' history of population growth and dispersal, with groups near each other having more recent common ancestry and higher rates of intermigration than populations farther apart.

The presence of nonsynonymous SNPs and PHI sites in *CD36* and *GNAT3* exons suggests that functional polymorphism is present in both genes, which could affect fat taste and metabolism as well as non-gustatory phenotypes. However, patterns of diversity at PHI sites indicated that their contributions to phenotypic variance on a population scale are limited. Of the 113 coding changes we found, only 60 were scored as PHI variants. Further, the frequencies of PHI alleles were low, with only two of the 60 having frequencies above 1% and the most common having a frequency of 3% (Table 4). These patterns extended to diversity within superpopulations, with the number of PHI sites being lower than the number of nonsynonymous variants on all continents (Table 3). These patterns indicate that while polymorphism affecting phenotypes is almost certainly present, it is low in frequency.

F_{STs} within superpopulations revealed potentially important regional trends. In particular, F_{STs} were far higher in Africa than elsewhere for two site categories (Table 3). The F_{ST} of *CD36* in Africa was 7-fold greater than in any other superpopulation (0.079 vs. 0.011 in East Asia) and 5-fold greater than in noncoding sites in Africa (0.079 vs. 0.015). The F_{ST} of PHI sites in Africa was also 7-fold greater than in any other superpopulation (0.108 vs. 0.015 in East Asia) and 5-fold greater than in noncoding sites. These patterns suggest that populations may be differentiated with respect to functional variation and, if so, it is likely most pronounced with respect to *CD36* in Africa. However, it is important to note that while patterns in F_{ST} with respect to PHI sites point to the presence of phenotypic differences among populations they provide little information about the magnitude of the differences, which depend not just on allele frequencies, but on effect sizes as well.

Signatures of natural selection

The high-calorie content of fats makes them highly valuable nutritionally, so the mechanisms underlying their detection must be under strong selective pressure. This is a familiar issue in human evolutionary biology, most famously as part of Neel's "thrifty gene" hypothesis, which posits that ability to store fat provides fitness advantages (Neel 1962; Reales et al. 2017). Under the thrifty gene hypothesis, the advantages of perceiving and metabolizing fats are evident. However, its implications specifically for *CD36* and *GNAT3* are not. On the one hand, the roles of CD36 and GNAT3 in sensory signaling, lipid metabolism, and other processes imply that they are intolerant of novel mutations and under pressure from purifying natural selection. However, humans' diffusion out of Africa placed *CD36* and *GNAT3* into novel environments, which could have produced pressures to adapt. Moreover, a single genomic region can be under more than one selective pressure

Table 6. PHI and regulatory sites with reported genotype–phenotype associations.

rsid	Effect	Allele frequency						Reports
		Worldwide	Africa	Americas	East Asia	Europe	South Asia	
rs1527479	Expression modifier	0.35	0.22	0.46	0.31	0.54	0.30	Bokor et al. 2010 Jayewardene et al. 2016 Lecompte et al. 2011
rs1534314	Expression modifier	0.26	0.26	0.19	0.44	0.07	0.30	Ghosh et al. 2011
rs3211883	Expression modifier	0.64	0.34	0.76	0.62	0.90	0.69	Bokor et al. 2010 Ghosh et al. 2011 Heni et al. 2011
rs3211938	Stop gained (*CD36* Exon 13)	0.03	0.12	0.00	0.00	0.00	0.00	Fry et al. 2009 Love-Gregory et al. 2011 Love-Gregory et al. 2008
rs6960369	Expression modifier	0.23	0.33	0.13	0.34	0.07	0.28	Ghosh et al. 2011
rs12706912	Expression modifier	0.56	0.60	0.45	0.59	0.44	0.68	Ghosh et al. 2011

at a time. For instance, selective pressures on the *LCT* gene appear to have been strong in some parts of Europe but not others (Bersaglieri et al. 2004). Similarly, selective pressures on *TAS2R38* have been stronger on some haplotype backgrounds than others (Risso et al. 2016).

The results of our Tajima's *D* tests across *CD36-GNAT3* as a whole were most consistent with an absence of selective pressure during its recent history. As expected given the low diversity and downward skew in allele frequencies, observed *D* values were strongly negative (Table 5). This can result from two common selective pressures: purifying selection and positive selection. However, it can also be caused by rapid population growth, during which genetic drift slows and new variants accumulate (Wooding 2003; Adams and Hudson 2004). In our study, the observed D was below expectations at a high level of statistical significance ($P < 0.001$). However, this did not reveal whether the cause of the shift was selection, demography, or both. *D* tests using the empirical distribution in the 1000GP clarified the probable cause. Because the vast majority of the human genome (>98%) is noncoding it is can be assumed to be evolving neutrally or nearly so with respect to selection, but it is still shaped by demography. Thus, the empirical distribution of *D* we obtained from the 1000GP represents expectations adjusted for demography while the standard distribution does not. We found that when compared with the empirical distribution, *D* values were well within expectations for four of the five site categories we analyzed ($0.390 < P_E < 0.665$; Table 5). The exception was in *CD36* exons, where the observed *D* was significantly lower than expected, albeit marginally so ($D = -2.41$, $P_E = 0.035$). This is consistent with localized purifying or weak positive selection. In our view, the empirical tests are the more convincing and conclude that on a worldwide scale selective pressures have been largely absent across *CD36-GNAT3* during its recent history with exception of exonic regions in *CD36*.

Evidence for selection in *CD36* exons in the 1000GP is consistent with prior findings by Fry et al. (2009). In a study of *CD36* in Africa, Fry et al. detected signatures of positive selection favoring a premature stop allele at rs3211938, which is located in *CD36* exon 13. Fry et al.'s investigation centered on resistance to malaria infection, which is hypothesized to be affected by variation in *CD36* because the receptor is an antigenic target for *Plasmodium falciparum*. The study found signatures of selection surrounding rs3211938 but rejected the malaria resistance hypothesis because association studies failed to detect correlations with malaria severity. However, the selective signatures were robust; suggesting that factors other than malaria resistance must be responsible. Evidence that lipid perception and metabolism are mediated by *CD36* offers a possible explanation. In clinical studies, variation in *CD36* associates with lipid perception and metabolic phenotypes, and rs3211938 is specifically implicated as the source of the associations (Love-Gregory et al. 2008, 2011). This supports the hypothesis that *CD36* harbors genotypes affecting responses to fats, which could exert selective pressures on genes mediating fat taste, metabolism, or both. Thus, a speculative explanation for the high frequency of the rs3211938 stop allele in localized areas of Africa is that it was favored by the nutritional environment, and contemporary associations are a holdover from those pressures.

Our findings on population differentiation are also consistent with Fry et al.'s (2009) proposal that selection has promoted differentiation among African populations. In our data, F_{ST} with respect to noncoding sites in Africa (F_{ST} = 1.5%) was similar to or slightly greater than in other superpopulations (F_{ST} = 0.2%–1.8%), indicating they are differentiated to roughly the same extent with respect to neutral variation (Table 3). However, F_{ST} with respect to *CD36* exons was far higher in Africa than in other superpopulations (7.9% vs. <1.5%), suggesting that some factor has driven differentiation specifically at that locus. Moreover, the trend was amplified in PHI sites, which is expected under local adaptation because selection most affects sites with functional effects. Further, differentiation in Africa was greatest specifically with respect to the premature stop in *CD36* (rs3211938), a major mutation particularly likely to experience selective

effects. These patterns are consistent with the effects of local adaptation and that evolutionary pressures on fat responses had important effects on diversity in the continent.

Implications for association studies

Our finding that *CD36-GNAT3* harbors numerous sites with putative functional effects supports predictions that the region bears variants affecting lipid perception and metabolism. To date, more than 80 variants in *CD36* and *GNAT3* have been reported to associate with both gustatory and non-gustatory phenotypes, and the potential for connections of *CD36* with obesity and related diseases draws ongoing interest (MacArthur et al. 2017). Our diversity estimates and functional predictions support these findings. They specifically buttress evidence for associations at eight sites identified in six previous studies (Table 6). Cross tabulating our list of PHI and regulatory sites against the list of previously reported associations revealed six with putative effects: rs1527479, rs1534314, rs3211883, rs3211938, rs12706912, and rs6960369. And of these, three (rs1527479, rs3211883, and rs3211938) showed evidence of associations in more than one study. One (rs3211938, the premature stop in *CD36*) showed evidence of being under pressure from positive natural selection, which can only act on functional sites. Thus, these are particularly strong candidates for future investigation.

Statistical power is a critical consideration in association studies aimed at dissecting phenotypic effects, and our findings predict that it will be weak in *CD36-GNAT3*. Although numerous sites predicted to affect protein function and gene expression were present in the region, their frequencies were low. For instance, 83% of regulatory alleles had frequencies below 1%, and 97% of PHI alleles had frequencies below 1%. This is important because even when such alleles have functional impact they cannot contribute much phenotypic variance overall. Conversely, it suggests that if *CD36-GNAT3* does harbor alleles contributing substantial variance, they must be limited to a small number of sites where allele frequencies are elevated. This pattern is evident in the cross-tabulation of PHI and regulatory sites against reported associations, in which all alleles exhibiting associations were found at high frequencies (>10%) in at least one superpopulation (Table 6).

Patterns of LD in our sample also have implications for association studies. When high, LD can produce false associations where noncausal variants cosegregate with causal ones, making their effects difficult to discriminate. However, it can enhance efforts to dissect associations by reducing the density of genotyping needed to localize the sources of effects, which can subsequently be pinpointed using fine mapping strategies. In contrast, when LD is low, sites' independent effects on phenotypes can be pinpointed directly through dense genotyping; however, this approach is vulnerable to missing effects if marker density is insufficient. In the case of *CD36* and *GNAT3*, the structure of LD is a particularly important consideration because the two genes underlie similar chemosensory traits. Thus, if present, high LD could cause GNAT3- and CD36-mediated phenotypes to spuriously associate with variants in both genes, making their functional contributions difficult to distinguish. We saw low potential for such effects in *CD36-GNAT3*. While the high D' values we observed are consistent with a risk of confounds, r^2 was low or 0 between almost all site pairs, including PHI and regulatory sites. This makes confounds unlikely.

Conflict of Interest

None declared.

References

Adams AM, Hudson RR. 2004. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. *Genetics*. 168(3):1699–1712.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods*. 7(4):248–249.

Allen AL, McGeary JE, Hayes JE. 2014. Polymorphisms in *TRPV1* and *TAS2Rs* associate with sensations from sampled ethanol. *Alcohol Clin Exp Res*. 38(10):2550–2560.

Antinucci M, Risso D. 2017. A matter of taste: lineage-specific loss of function of taste receptor genes in vertebrates. *Front Mol Biosci*. 4:81. doi:10.3389/fmolb.2017.00081.

Bachmanov AA, Bosak NP, Lin C, Matsumoto I, Ohmoto M, Reed DR, Nelson TM. 2014. Genetics of taste receptors. *Curr Pharm Des*. 20(16):2669–2683.

Baldwin MW, Toda Y, Nakagita T, O'Connell MJ, Klasing KC, Misaka T, Edwards SV, Liberles SD. 2014. Sensory biology. Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science*. 345(6199):929–933.

Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. *Nat Rev Genet*. 4(2):99–111.

Barontini J, Antinucci M, Tofanelli S, Cammalleri M, Dal Monte M, Gemignani F, Vodicka P, Marangoni R, Vodickova L, Kupcinskas J, et al. 2017. Association between polymorphisms of *TAS2R16* and susceptibility to colorectal cancer. *BMC Gastroenterol*. 17(1):104.

Basson MD, Bartoshuk LM, Dichello SZ, Panzini L, Weiffenbach JM, Duffy VB. 2005. Association between 6-*n*-propylthiouracil (PROP) bitterness and colonic neoplasms. *Dig Dis Sci*. 50(3):483–489.

Behrens M, Di Pizio A, Redel U, Meyerhof W, Korsching SI. 2020. At the root of *T2R* gene evolution: recognition profiles of coelacanth and zebrafish bitter receptors. *Genome Biol Evol*. 13(1). doi:10.1093/gbe/evaa264.

Behrens M, Gunn HC, Ramos PC, Meyerhof W, Wooding SP. 2013. Genetic, functional, and phenotypic diversity in TAS2R38-mediated bitter taste perception. *Chem Senses*. 38(6):475–484.

Bergström A, Stringer C, Hajdinjak M, Scerri EML, Skoglund P. 2021. Origins of modern human ancestry. *Nature*. 590(7845):229–237.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet*. 74(6):1111–1120.

Bokor S, Legry V, Meirhaeghe A, Ruiz JR, Mauro B, Widhalm K, Manios Y, Amouyel P, Moreno LA, Molnar D, et al. 2010. Single-nucleotide polymorphism of *CD36* locus and obesity in European adolescents. *Obesity (Silver Spring)*. 18:1398–1403.

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(4):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(2):288–302.

Campbell MC, Tishkoff SA. 2008. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*. 9:403–433. doi:10.1146/annurev.genom.9.081307.164258.

Cartoni C, Yasumatsu K, Ohkuri T, Shigemura N, Yoshida R, Godinot N, le Coutre J, Ninomiya Y, Damak S. 2010. Taste preference for fatty acids is mediated by GPR40 and GPR120. *J Neurosci*. 30(25):8376–8382.

Chalé-Rush A, Burgess JR, Mattes RD. 2007. Evidence for human orosensory (taste?) sensitivity to free fatty acids. *Chem Senses*. 32(5):423–431.

Chen QY, Alarcon S, Tharp A, Ahmed OM, Estrella NL, Greene TA, Rucker J, Breslin PA. 2009. Perceptual variation in umami taste and polymorphisms in *TAS1R* taste receptor genes. *Am J Clin Nutr*. 90(3):770S–779S.

Choi JH, Lee J, Choi IJ, Kim YW, Ryu KW, Kim J. 2016. Genetic variation in the *TAS2R38* bitter taste receptor and gastric cancer risk in Koreans. *Sci Rep*. 6:26904. doi:10.1038/srep26904.

Cvijović I, Good BH, Desai MM. 2018. The effect of strong purifying selection on genetic diversity. *Genetics*. 209(4):1235–1278.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al.; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics*. 27(15):2156–2158.

De Araujo IE, Rolls ET. 2004. Representation in the human brain of food texture and oral fat. *J Neurosci*. 24(12):3086–3093.

Deshpande DA, Wang WC, McIlmoyle EL, Robinett KS, Schillinger RM, An SS, Sham JS, Liggett SB. 2010. Bitter taste receptors on airway smooth muscle bronchodilate by localized calcium signaling and reverse obstruction. *Nat Med*. 16(11):1299–1304.

Drayna D. 2005. Human taste genetics. *Annu Rev Genomics Hum Genet*. 6:217–235. doi:10.1146/annurev.genom.6.080604.162340.

Egan JM, Margolskee RF. 2008. Taste cells of the gut and gastrointestinal chemosensation. *Mol Interv*. 8(2):78–81.

Elhaik E. 2012. Empirical distributions of F_{ST} from large-scale human polymorphism data. *PLoS One*. 7(11):e49837.

Farook VS, Puppala S, Schneider J, Fowler SP, Chittoor G, Dyer TD, Allayee H, Cole SA, Arya R, Black MH, et al. 2012. Metabolic syndrome is linked to chromosome 7q21 and associated with genetic variants in *CD36* and *GNAT3* in Mexican Americans. *Obesity (Silver Spring)*. 20(10):2083–2092.

Feeney EL, Hayes JE. 2014. Exploring associations between taste perception, oral anatomy and polymorphisms in the carbonic anhydrase (gustin) gene *CA6*. *Physiol Behav*. 128:148–154. doi:10.1016/j.physbeh.2014.02.013.

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.

Fry AE, Ghansa A, Small KS, Palma A, Auburn S, Diakite M, Green A, Campino S, Teo YY, Clark TG, et al. 2009. Positive selection of a *CD36* nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes. *Hum Mol Genet*. 18(14):2683–2692.

Fushan AA, Simons CT, Slack JP, Drayna D. 2010. Association between common variation in genes encoding sweet taste signaling components and human sucrose perception. *Chem Senses*. 35(7):579–592.

Gaillard D, Laugerette F, Darcel N, El-Yassimi A, Passilly-Degrace P, Hichami A, Khan NA, Montmayeur JP, Besnard P. 2008. The gustatory pathway is involved in CD36-mediated orosensory perception of long-chain fatty acids in the mouse. *FASEB J*. 22(5):1458–1468.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. *Bioconductor*: open software development for computational biology and bioinformatics. *Genome Biol*. 5(10):R80.

Ghosh A, Murugesan G, Chen K, Zhang L, Wang Q, Febbraio M, Anselmo RM, Marchant K, Barnard J, Silverstein RL. 2011. Platelet CD36 surface expression levels affect functional responses to oxidized LDL and are associated with inheritance of specific genetic polymorphisms. *Blood*. 117: 6355–6366.

Goudet J. 2004. *hierfstat*, a package for R to compute and test hierarchical *F*-statistics. *Mol Ecol Notes*. 5:184–186. doi:10.18637/jss.v016.c0.

Graham J, McNeney B, Blay S, Shin J-H. 2006. *LDheatmap*: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw*. 16:c03. doi:10.18637/jss.v016.c03.

Greene LS. 1974. Physical growth and development, neurological maturation, and behavioral functioning in two Ecuadorian Andean communities in which goiter is endemic. II. PTC taste sensitivity and neurological maturation. *Am J Phys Anthropol*. 41(1):139–151.

Hancock AM, Di Rienzo A. 2008. Detecting the genetic signature of natural selection in human populations: models, methods, and data. *Annu Rev Anthropol*. 37:197–217. doi:10.1146/annurev.anthro.37.081407.085141.

Hayes JE, Feeney EL, Nolden AA, McGeary JE. 2015. Quinine bitterness and grapefruit liking associate with allelic variants in *TAS2R31*. *Chem Senses*. 40(6):437–443.

Heni M, Mussig K, Machicao F, Machann J, Schick F, Claussen CD, Stefan N, Fritsche A, Haring HU, Staiger H. 2011. Variants in the *CD36* gene locus determine whole-body adiposity, but have no independent effect on insulin sensitivity. *Obesity (Silver Spring)*. 19: 1004–1009.

Henn BM, Cavalli-Sforza LL, Feldman MW. 2012. The great human expansion. *Proc Natl Acad Sci USA*. 109(44):17758–17764.

Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nat Rev Genet*. 10(9):639–650.

Ihaka R, Gentleman R. 1996. R: a language for data analysis and graphics. *J Comput Graph Stat*. 5(3):299–314.

Jayewardene AF, Mavros Y, Gwinn T, Hancock DP, Rooney KB. 2016. Associations between CD36 gene polymorphisms and metabolic response to a short-term endurance-training program in a young-adult population. *Appl Physiol Nutr Metab*. 41: 157–167.

Jiang P, Josue J, Li X, Glaser D, Li W, Brand JG, Margolskee RF, Reed DR, Beauchamp GK. 2012. Major taste loss in carnivorous mammals. *Proc Natl Acad Sci USA*. 109(13):4956–4961.

Jombart T, Ahmed I. 2011. *adegenet* 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. 27(21):3070–3071.

Keller KL, Liang LC, Sakimura J, May D, van Belle C, Breen C, Driggin E, Tepper BJ, Lanzano PC, Deng L, et al. 2012. Common variants in the *CD36* gene are associated with oral fat perception, fat preferences, and obesity in African Americans. *Obesity (Silver Spring)*. 20(5):1066–1073.

Kim UK, Breslin PA, Reed D, Drayna D. 2004. Genetics of human taste perception. *J Dent Res*. 83(6):448–453.

Kim UK, Wooding S, Riaz N, Jorde LB, Drayna D. 2006. Variation in the human *TAS1R* taste receptor genes. *Chem Senses*. 31(7):599–611.

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.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 4(7):1073–1081.

Laugerette F, Passilly-Degrace P, Patris B, Niot I, Febbraio M, Montmayeur JP, Besnard P. 2005. CD36 involvement in orosensory detection of dietary lipids, spontaneous fat preference, and digestive secretions. *J Clin Invest*. 115(11):3177–3184.

Lecompte S, Szabo de Edelenyi F, Goumidi L, Maiani G, Moschonis G, Widhalm K, Molnar D, Kafatos A, Spinneker A, Breidenassel C, et al. 2011. Polymorphisms in the *CD36/FAT* gene are associated with plasma vitamin E concentrations in humans. *Am J Clin Nutr*. 93: 644–651.

Li H. 2011. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics*. 27(5):718–719.

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

Lindemann B. 2001. Receptors and transduction in taste. *Nature*. 413(6852):219–225.

Love-Gregory L, Sherva R, Schappe T, Qi JS, McCrea J, Klein S, Connelly MA, Abumrad NA. 2011. Common *CD36* SNPs reduce protein expression and may contribute to a protective atherogenic profile. *Hum Mol Genet*. 20(1):193–201.

Love-Gregory L, Sherva R, Sun L, Wasson J, Schappe T, Doria A, Rao DC, Hunt SC, Klein S, Neuman RJ, et al. 2008. Variants in the *CD36* gene associate with the metabolic syndrome and high-density lipoprotein cholesterol. *Hum Mol Genet*. 17(11):1695–1704.

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 45(D1):D896–D901.

Marth G, Schuler G, Yeh R, Davenport R, Agarwala R, Church D, Wheelan S, Baker J, Ward M, Kholodov M, et al. 2003. Sequence variations in the public human genome data reflect a bottlenecked population history. *Proc Natl Acad Sci USA*. 100(1):376–381.

Mattes RD. 2011. Accumulating evidence supports a taste component for free fatty acids in humans. *Physiol Behav*. 104(4):624–631.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl variant effect predictor. *Genome Biol*. 17(1):122.

Mueller JC. 2004. Linkage disequilibrium for different scales and applications. *Brief Bioinform*. 5(4):355–364.

Neel JV. 1962. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am J Hum Genet*. 14(4):353–362.

Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M. 2014. *VariantAnnotation*: a *Bioconductor* package for exploration and annotation of genetic variants. *Bioinformatics*. 30(14):2076–2078.

Osada N. 2015. Genetic diversity in humans and non-human primates and its evolutionary consequences. *Genes Genet Syst*. 90(3):133–145.

Paradis E. 2010. *pegas*: an R package for population genetics with an integrated-modular approach. *Bioinformatics*. 26(3):419–420.

Pepino MY, Kuda O, Samovski D, Abumrad NA. 2014. Structure-function of CD36 and importance of fatty acid signal transduction in fat metabolism. *Annu Rev Nutr*. 34:281–303. doi:10.1146/annurev-nutr-071812-161220.

Pepino MY, Love-Gregory L, Klein S, Abumrad NA. 2012. The fatty acid translocase gene *CD36* and lingual lipase influence oral sensitivity to fat in obese subjects. *J Lipid Res*. 53(3):561–566.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. *PopGenome*: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol*. 31(7):1929–1936.

Reales G, Rovaris DL, Jacovas VC, Hünemeier T, Sandoval JR, Salazar-Granara A, Demarchi DA, Tarazona-Santos E, Felkl AB, Serafini MA, et al. 2017. A tale of agriculturalists and hunter-gatherers: exploring the thrifty genotype hypothesis in native South Americans. *Am J Phys Anthropol*. 163(3):591–601.

Reed DR, Knaapila A. 2010. Genetics of taste and smell: poisons and pleasures. *Prog Mol Biol Transl Sci*. 94:213–240. doi:10.1016/B978-0-12-375003-7.00008-X.

Risso DS, Mezzavilla M, Pagani L, Robino A, Morini G, Tofanelli S, Carrai M, Campa D, Barale R, Caradonna F, et al. 2016. Global diversity in the *TAS2R38* bitter taste receptor: revisiting a classic evolutionary PROPosal. *Sci Rep*. 6(10):25506.

Risso D, Sainz E, Morini G, Tofanelli S, Drayna D. 2018. Taste perception of *Antidesma bunius* fruit and its relationships to bitter taste receptor gene haplotypes. *Chem Senses*. 43(7):463–468.

Rogers AR. 1995. Genetic evidence for a Pleistocene population explosion. *Evolution*. 49(4):608–615.

Roper SD, Chaudhari N. 2017. Taste buds: cells, signals and synapses. *Nat Rev Neurosci*. 18(8):485–497.

Silverstein RL, Febbraio M. 2009. CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior. *Sci Signal*. 2(72):re3.

Slatkin M. 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 9(6):477–485.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 105(2):437–460.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123(3):585–595.

The 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature*. 467(7319):1061–1073.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature*. 526(7571):68–74.

Vander Molen J, Frisse LM, Fullerton SM, Qian Y, Del Bosque-Plata L, Hudson RR, Di Rienzo A. 2005. Population genetics of *CAPN10* and *GPR35*: implications for the evolution of type 2 diabetes variants. *Am J Hum Genet*. 76(4):548–560.

Weir BS, Hill WG. 2002. Estimating *F*-statistics. *Annu Rev Genet*. 36:721–750. doi:10.1146/annurev.genet.36.050802.093940.

Wooding S. 2003. PopHist: inferring population history from the spectrum of allele frequencies. *Bioinformatics*. 19(4):539–540. doi:10.1186/1471-2350-13-96.

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

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.

Wooding S, Bufe B, Grassi C, Howard MT, Stone AC, Vazquez M, Dunn DM, Meyerhof W, Weiss RB, Bamshad MJ. 2006. Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature*. 440(7086):930–934.

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

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.

Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue SK, Li WH. 2002. Larger genetic differences within Africans than between Africans and Eurasians. *Genetics*. 161(1):269–274.

Zhao H, Li J, Zhang J. 2015. Molecular evidence for the loss of three basic tastes in penguins. *Curr Biol*. 25(4):R141–R142.

Zhao H, Zhou Y, Pinto CM, Charles-Dominique P, Galindo-González J, Zhang S, Zhang J. 2010. Evolution of the sweet taste receptor gene *Tas1r2* in bats. *Mol Biol Evol*. 27(11):2642–2650.