Analysis of Positive Selection at Single Nucleotide Polymorphisms Associated with Body Mass Index Does Not Support the “Thrifty Gene” Hypothesis

Highlights

- Signatures of positive selection examined at 115 SNPs linked to obesity
- Positive selection not observed more frequently than in randomly selected SNPs
- Positive selection at nine BMI linked SNPs, but five of these favored leanness
- These data do not support the “thrifty gene” hypothesis

In Brief

The “thrifty gene hypothesis” posits an evolutionary advantage for fat storage in humans to survive famines. Wang and Speakman do a comprehensive search for signatures of positive selection at loci previously associated with human obesity and do not find evidence for such selection, calling into question the “thrifty gene hypothesis.”

Authors

Guanlin Wang, John R. Speakman

Correspondence

j.speakman@genetics.ac.cn

Wang & Speakman, 2016, Cell Metabolism 24, 531–541

October 11, 2016 © 2016 Elsevier Inc.

http://dx.doi.org/10.1016/j.cmet.2016.08.014

Cell Press

Cell Metabolism
Analysis of Positive Selection at Single Nucleotide Polymorphisms Associated with Body Mass Index Does Not Support the “Thrifty Gene” Hypothesis

Guanlin Wang1,2 and John R. Speakman1,3,4,*

1State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, No. 1 West Beichen Road, Chaoyang District, Beijing 100101, PRC
2University of Chinese Academy of Sciences, 19 A Yuquan Road, Shijingshan District, Beijing 100049, PRC
3Institute of Biological and Environmental Sciences, University of Aberdeen, Zoology Building, Tillydrone Avenue, Aberdeen AB24 2TZ, UK
4Lead Contact
*Correspondence: j.speakman@genetics.ac.cn
http://dx.doi.org/10.1016/j.cmet.2016.08.014

SUMMARY

The “thrifty gene hypothesis” suggests genetic susceptibility to obesity arises because of positive selection for alleles that favored fat deposition and survival during famines. We used public domain data to locate signatures of positive selection based on derived allele frequency, genetic diversity, long haplotypes, and differences between populations at SNPs identified in genome-wide association studies (GWASs) for BMI. We used SNPs near the lactase (LCT), SLC24A5, and SLC45A2 genes as positive controls and 120 randomly selected SNPs as negative controls. We found evidence for positive selection (p < 0.05) at nine out of 115 BMI SNPs. However, five of these involved positive selection for the protective allele (i.e., for leanness). The widespread absence of signatures of positive selection, combined with selection favoring leanness at some alleles, does not support the suggestion that obesity provided a selective advantage to survive famines, or any other selective advantage.

INTRODUCTION

Obesity has become a worldwide health issue (Flegal et al., 2012; James et al., 2001), with more than 600 million obese individuals and almost 1.9 billion overweight adults from a world population of 7.3 billion in 2014 (Mendis et al., 2015). Obesity is a predisposing factor for a range of serious non-communicable diseases (NCDs), including type 2 diabetes mellitus (T2D) (Kahn et al., 2006), hypertension, cardiovascular disorders (Van Gaal et al., 2006), and cancer (Calle et al., 2003). Twin and family studies suggest that about 65% of the variance in the prevalence of obesity is genetic (Allison et al., 1996), and it is generally acknowledged that obesity is a consequence of a gene-environment interaction (Hill and Peters, 1998; Speakman et al., 2011). The high genetic contribution to obesity creates a conundrum concerning the selection pressures that may have favored evolution of a trait that has such manifest negative consequences. Three potential solutions to this conundrum have been advanced (Speakman, 2013). First, it is argued that during our evolutionary history, we were routinely exposed to periods of famine. Individuals with genes favoring fat storage might then have a selective advantage, as such fat would enable better survival during the periods when food was absent. In the modern world, such genes prepare us for a famine that never comes: the primary result of which is obesity (Chakravarthy and Booth, 2004; Eknoyan, 2006; Neel, 1962; Prentice, 2001; Watnick, 2006; Wells, 2006). This idea, generally called the “thrifty gene” hypothesis, has several problems, principal among which is the fact that any genetic polymorphism providing a significant advantage for fat storage and famine survival would move rapidly to fixation (Ayub et al., 2014; Speakman, 2008; Speakman et al., 2011): and consequently, we would all inherit these thrifty genes, and in modern society we would all become fat, which clearly we do not. A potential resolution to this problem (Prentice et al., 2008) is the suggestion that famines and hence strong selective pressure favoring fat storage has only happened during the last 12,000 years, since the dawn of agriculture (also called the “thrifty late” hypothesis; Ayub et al., 2014). Hence, most polymorphisms causing obesity have had insufficient time to move to fixation (Speakman, 2007a, 2008; but see Speakman et al., 2011). The second idea is that obesity risk is an unfortunate side effect of variable selection on some other trait such as thermogenic capacity (Sellayah et al., 2014; Speakman, 2013). Finally, a third idea is that obesity is a consequence of random mutation and drift in genes that regulate the upper limits of body weight control. It is argued that until around 2 million years ago, there would have been strong selective pressure from predation risk to avoid weight gain (Speakman, 2007b, 2008). However, the development of social behavior, fire, and weapons in early Homo species effectively reduced the predation risk to a level where the control system on body weight could mutate and drift: called the “drifty gene” hypothesis (Speakman, 2008).

It has been suggested that a strong test between these ideas would be the detection of signatures of positive selection for the “at risk” alleles at loci linked to obesity, which would strongly support the thrifty gene hypothesis and refute the other ideas (Prentice et al., 2008). An early study (Southam et al., 2009) indicated scant evidence for selection signatures based on
integrated haplotype scores and the differentiation fixation index at validated T2D and some obesity loci. This was confirmed more recently by a much more comprehensive survey of 65 loci associated with susceptibility to T2D (Ayub et al., 2014). Moreover, studies of selection at T2D SNPs noted that, contrary to the thrifty gene hypothesis, in many cases, it was the protective rather than the at risk allele that had been under positive selection (Ayub et al., 2014; Ségurel et al., 2013). However, refuting the hypothesis using loci linked to T2D only refutes the idea that genes linked to T2D were selected because they favored fat storage via effects on insulin resistance. This does not refute the wider idea that genes predisposing to fat storage, independent of T2D, may have been positively selected because of their advantageous effects with respect to famine survival. For example, a study focused exclusively on the fat mass and obesity associated gene (FTO) indicated there was a suggestion of balancing selection at rs17818902 in the third intron (Liu et al., 2015). However, a search for differences in selection between East-Asian and European populations concluded no evidence for strong positive selection differentiating these populations at BMI loci (Koh et al., 2014).

We provide here a comprehensive search for signatures of positive selection at 115 BMI-associated loci that have been previously validated across several genome-wide association studies (GWASs) (Locke et al., 2015; Okada et al., 2012; Sterio et al., 1010; Willer et al., 2009; Wen et al., 2012) to test the wider fat storage related context of the thrifty gene hypothesis. We used linkage disequilibrium (LD) data from the latest HapMap phase III release data (Gibbs et al., 2003) and various indices of positive selection using the 1000 Genomes Project data, specifically Tajima’s D, difference of derived allele frequency (DDAF), fixation index (Wright’s FST), integrated haplotype score (iHS), cross-population extended haplotype homozygosity (XPEHH), and cross-population composite likelihood ratio (XPCLR) locus-by-locus using dbPshp (Li et al., 2014). We used SNPs adjacent to, and in, the lactase gene (LCT) (Bersaglieri et al., 2004), SLC24A5 (Lamason et al., 2005), and SLC45A2 (Wilde et al., 2014), which have experienced positive selection in European populations as positive controls. We also used 120 randomly selected SNPs matched for similar recombination rate and allele frequencies as negative controls.

RESULTS

We conducted four kinds of tests which enable the detection of positive selection over different time frames throughout our evolutionary history (Sabeti et al., 2006). These were (1) reduction in genetic diversity (Tajima’s D and LD) (<250,000 years); (2) high-frequency derived alleles (<80,000 years [DDAF]); (3) differences between populations (<50,000 to 75,000 years) (FST and XPCLR); and (4) long haplotypes (<30,000 years) (iHS and XPEHH). Combining these tests, we can capture multiple indices of positive selection over the recent history of these genes. The positive selection signals thresholds were defined as Tajima’s D < 0, [DDAF] > 0.2, FST > 0.05, XPCLR > 5, XPEHH > 1, and iHS > 2 as recommended in a previous publication (Li et al., 2014).

Reduction in Genetic Diversity

We downloaded the Tajima’s D values from the dbPshp for all the SNPs of 14 populations (CEU, GBR, IBS, FIN, TSI, MXL, CLM, PUR, ASW, LWK, YRI, CHB, CHS, and JPT) of the 1000 Genomes from locus-by-locus in a 10 kb window and 115 positive selection signals were found in 1,610 situations (locus × population) for BMI-associated SNPs, while 162 positive signatures were found in 1,680 situations for the negative controls (Figure S1). Overall signatures of positive selection were significantly less frequent in the BMI-associated SNPs compared to the control SNPs (Chi² = 5.632, p = 0.018 < 0.05). No obvious hitchhiking was found in the BMI-associated SNPs using LD data. LD data were retrieved from the International HapMap Project phase III for all 11 populations (ASW, CEU, CHB, CHS, GIH, JPT, LWK, MEX, MKK, TSI, and YRI). We calculated the running average LOD score using a window of ten SNPs in a 500 kbp range up and downstream around the target. Scatterplots of running average LOD scores as a function of the position were made. Obvious selective sweeps and hitchhiking to fixation were validated in the SNPs of the positive controls (LCT and SLC24A5 and SLC45A2) in the CEU population. The BMI-indexed SNPs showed different patterns, with no evidence supporting selective sweeps and hitchhiking (for example, see Figure 1, which illustrates three representative SNPs: a positive control rs4988235 [in MCM6: influence on LCT gene], a BMI linked SNP rs9939609 [in intron 1 of FTO], and a negative control random SNP rs2287652 [intron in ADCK1]). Plots for the all other SNPs we studied are available for request.

High-Frequency Derived Alleles

A high DAF may indicate positive selection and/or population expansion (Sabeti et al., 2006). We compared the DAF of all 115 BMI-associated SNPs with both positive and negative controls (Table S1). [DDAF] scores were also downloaded from dbPshp to detect positive selection for all the populations of the 1000 Genomes Project. In the positive control (related with LCT and SLC24A5 and SLC45A2) SNPs, we found strong positive selection signals as the [DDAF] was greater than 0.2 in all five
of the SNPs in the CEU population. In other European populations including GBR, IBS, FIN, and TSI, we obtained similar results, but not as strong as for the CEU population (Figure 2). We found 221 positive signals out of 1,610 tests for BMI-associated SNPs across all 14 populations and 275 out of 1,680 tests for the randomly selected SNPs (Figure 2). This frequency difference was not significant ($\chi^2 = 3.355, p = 0.067$) and suggested no difference in positive selection at alleles linked to BMI when compared to random SNPs in the genome. Across all the BMI-associated and control SNPs, we found more positive selection signals in East-Asian populations (CHB, CHS, and JPT) and African populations than the European populations (CEU, GBR, FIN, IBS, and TSI).

**Differences between Populations**

We also evaluated the population differentiation using $F_{ST}$ and XPCLR. We found significant difference in the frequency of positive tests for Wright’s $F_{ST}$ between the BMI-associated SNPs (82/1,610 of positive tests) and negative control SNPs (115/1,680 of positive tests) ($\chi^2 = 3.978$, $p = 0.046 < 0.05$) (Table 82/1,610 of positive tests). We also evaluated the population differentiation using $F_{ST}$ and linkage disequilibrium measures, we found 678 out of 1,610 tests indicated possible positive selection signals in East-Asian populations (CHB, CHS, and JPT) and African populations than the European populations (CEU, GBR, FIN, IBS, and TSI).

**Long Haplotypes**

If a mutation undergoes positive selection, the haplotype diversity will be reduced. A higher $i$HS score means haplotypes related to the ancestral alleles are longer than the derived alleles, and this can be used to measure the haplotype homozygosity extension (Voight et al., 2006). For BMI-associated SNPs, in 80 out of 1,610 tests, positive selection signals were observed using data from the *dbPshp* database. However, this did not differ significantly from the negative controls where 110 out of 1,680 tests were positive (Figure S2) ($\chi^2 = 3.355, p = 0.067$). We also compared the XP-EHH for the BMI-associated SNPs for each population. This index is complimentary to $i$HS and can detect alleles close to fixation under positive selection (Nielsen, 2005). In this case, 89 out of 1,610 tests indicated positive selection in the BMI-associated SNPs, which did not differ significantly from the 93 selection signals detected in 1,680 tests in the negative controls ($\chi^2 = 0, p = 0.993$) (Figure 4). All the positive control SNPs (five out of five) had values greater than 1.0, which indicated positive selection in the CEU population (Figure S3).

By combining all the positive selection tests (except the DAF and linkage disequilibrium measures), we found 678 out of 9,660 tests indicated possible positive selection signals in the BMI-associated SNPs (7.0%). This was significantly lower than the occurrence of positive selection signals (8.7%) among the randomly selected SNPs (negative controls) ($\chi^2 = 15.523, p < 0.001$). Although there was no overall suggestion of widespread positive selection at BMI linked SNPs when compared to the rest of the genome, this overall analysis might mask strong selection at individual SNPs that is obscured when the data for all SNPs are combined. We therefore built an expected distribution of positive tests for each SNP based on the frequency spectrum of positive tests in the randomly selected SNPs (see Experimental Procedures), under the assumption that these randomly selected SNPs were not under selection. This modeling allowed us to define a false discovery rate (FDR) and establish a cutoff that might indicate positive selection in individual SNPs. The one-tailed upper 5% confidence limit of the cumulative binomial distribution (Figure S4) was $\geq 14$ to 17 positive tests across all tests and populations for each SNP. Using the more liberal of these criteria ($\geq 14$), four out of five of the positive control SNPs were confirmed as under significant positive selection. Additionally, we found nine BMI-associated SNPs for which there was also an indication of positive selection ($\geq 14$ positive tests across all statistics and populations). These SNPs and the nearest genes were rs12597579 [CDG2, glycoprotein 2], rs2206734 [CDK5 regulatory subunit associated protein 1-like 1], rs3810291 [TMEM160, transmembrane protein 160], rs12401738 [FUBP1, far upstream element (FUSE) binding protein 1], rs2121279 [LEP1B, leptin 1B], rs7903146 [TCF7L2, transcription factor 7-like 2], rs2365389 [FHT, fragile histidine triad], rs9925964 [KAT8, K(lysine) acetyltransferase 8]), and rs11583200 [ELAVL4, ELAV like neuron-specific RNA binding protein 4] (Table 1; Table S2). There was no indication that these positive tests were predominantly located in metrics that detect recent (<30,000 years) rather than more ancient (<200,000 years) selection (Figures 5 and 6). Importantly, in five of these nine SNPs, the positive selection was for the protective rather than the at risk allele (Table 1; Table S2) (Ahmad et al., 2015). That is, the data suggested that selection favored leanness rather than fat deposition. If we used the more stringent criterion of $>17$ positive tests, then three out of five of the positive controls remained significant, but none of the BMI selected SNPs reached significance.

**DISCUSSION**

Our analysis provides little evidence that there has been positive selection at SNPs linked to BMI. Of the nine where we did find evidence that they had been selected, four included selection for the at risk allele, which promotes fat deposition. However, for one of these SNPs (rs12407038), there was also an indication that the at risk allele slows down weight regain after loss and hence it could be considered protective (Papandonatos et al., 2015). In the other five, selection favored the protective allele, in other words, at 5/9 SNPs, selection favored leanness rather than obesity (Ahmad et al., 2015). For many of these genes, the functional nature of their linkage to obesity is uncertain and it is possible that positive selection has occurred because of some other functional outcomes related to the SNP in question. For example, rs11583200 (near to the gene ELAVL4) has been linked to differences in physical activity, which could clearly be a target of selection independent of any association to obesity (Ahmad et al., 2015). Finally, only three of the nine positive signatures of selection were detected in metrics that cover the last 30,000 years. This does not therefore support the thrifty late idea (Ayub et al., 2014) that most selection for thrifty genes
Figure 2. Results of the DDAF Test for 14 Populations of 1000 Genomes Project

In the left panel, the x axis is a list of BMI-associated SNPs, while in the right panel, the x axis refers to the negative control SNPs. In both images, the y axis is the absolute value of DDAF. The bold x axis is the threshold for the positive selection. The identities of the BMI linked SNPs that were above the threshold are indicated. In total, there were 221 signals above the threshold among the BMI SNPs and 275 above the threshold in the negative controls. The SNPs showing positive selection signals were identified except for the AMR and EAS populations owing to space limitation (refer to Table S4 for identities). Positive controls can be found in Figure S3.
Figure 3. XPCLR of BMI-Associated SNPs in the Left Panel and Negative Control SNPs in the Right Panel in 14 Populations from the 1000 Genomes Project

XPCLR of BMI-associated SNPs (in the left panel) and negative control SNPs (in the right panel) in 14 populations in the 1000 Genomes Project. The bold x axis is the threshold for the positive selection. The identities of the BMI linked SNPs that were above the threshold are indicated. The SNPs showing positive selection signals were identified except for the EAS populations owing to space limitation (refer to Table S4 for identities). In total, there were 91 signals above the threshold among the BMI SNPs and 117 above the threshold in the negative controls.
Figure 4. XPEHH of BMI-Associated SNPs in the Left Panel and the Negative Control SNPs in the Right Panel in 14 populations from the 1000 Genomes Project

XPEHH of BMI-associated SNPs (in the left panel) and the negative control SNPs (in the right panel) in 14 populations of the 1000 Genomes Project. The bold x axis is the threshold for the positive selection. The identities of the BMI linked SNPs that were above the threshold are indicated. The SNPs showing positive selection signals were identified except for the EAS populations owing to space limitation (refer to Table S4 for identities). In total, there were 89 signals above the threshold among the BMI SNPs and 93 above the threshold in the negative controls. Almost all the positive control SNPs score were higher than 1, which indicated positive selection shown in Figure S3 in European populations (CEU, TSI, GBR, IBS, and FIN).
occurred in the last 10,000 to 12,000 years (Prentice et al., 2008). The widespread absence of signatures of strong positive selection at the BMI-associated SNPs, plus the fact that most of these signatures favored the protective allele, provide strong evidence refuting the widely held belief that the genetic contribution to obesity stems from the selective advantages of fat storage during our evolutionary history, particularly (but not exclusively due to) increased survival of periods of famine: the thrifty gene hypothesis (Neel, 1962). Moreover, the thrifty gene hypothesis was originally posited as an adaptive explanation for the high genetic contribution to diabetes (Neel, 1962) because insulin resistance might favor efficient fat storage (Neel, 1962). Recent work, however, has suggested that gene variants that predispose to insulin resistance are associated with reduced rather than elevated fat mass, notably in the limbs (Scott et al., 2014). Similar to the findings presented here, previous attempts to find signatures of positive selection at genetic loci linked to T2D have also proved largely negative (Ayub et al., 2014; Segurel et al., 2013).

A recent study of Inuit populations in Greenland suggested six SNPs had been under positive selection (Fumagalli et al., 2015). These included a cluster of SNPs linked to fatty acid desaturases that determine tissue levels of poly-unsaturated fatty acids and appear to be an adaptive response to a diet high in poly-unsaturates. These alleles were also linked to multiple metabolic phenotypes including height and weight. The association with weight seems likely to have come about because of indirect associations via the positive selection due to adaptation to the diet, rather than a direct result of selection on obesity, as posited by the thrifty gene hypothesis. We tested the three SNPs that were associated with body weight in the Greenland Inuit, but found no significant evidence of positive selection across the 14 populations studied here using the multiple metrics in the current study (rs12577256, n = 4 signals; rs174602, n = 12 signals; rs175470, n = 13 signals; and liberal significance cutoff ≥ 14 signals).

The main limitation in the current study was that BMI is an inaccurate measurement of body fatness (Romero-Corral et al., 2008). Hence, while these SNPs have been linked to BMI, for relatively few of them has further work been performed using more accurate methods for monitoring body fat to confirm that the genes are linked also to body fatness (Fox et al., 2012). There was no suggestion that those SNPs, where an association to body fatness has been proven, were more likely to be indicated as under positive selection. A second limitation was that we assumed that the main functional attribute of the SNP that affected whether it was positively selected or not was its impact on obesity (measured as BMI). Yet these SNPs (Table S2) mostly have unclear functions and may be functionally linked to other phenotypic traits that are under selection (such as the SNP rs11583200, which is linked to physical activity levels, mentioned above). A third problem was that we used the associated SNPs that emerged from GWASs to perform the tests. Experience with the FTO gene, where there has been extensive analysis of the roles of different adjacent SNPs (reviewed in Speakman, 2015) suggests that the index SNP may not be the causal variant for the associated variation in BMI. The extent of this problem will depend on the extent of linkage disequilibrium between the index and causal SNPs, which for all of these targets is unknown, but is presumably high or else the index SNPs would not emerge in the GWASs in the first place. A fourth problem was that together these SNPs explain less than 3% of the variance in body weight, from a total of around 65% that has been attributed to genetic factors. One might argue then that the loci that explain the remaining 60+% of the genetic variance are where the signals of strong selection are located. At present, we have no way to resolve this apart from the fact that the SNPs of largest effect are those that have emerged from the current GWASs and hence it would be anticipated that such SNPs would be more likely to be under selection than SNPs with much smaller impacts on fat storage levels (Speakman and Westerterp, 2013). The impact of other genetic variants such as copy number variants was not studied here and has been suggested to be potentially important in obesity susceptibility (cf. Amylase gene) (Falchi et al., 2014), but this role has been disputed (Usher et al., 2015). Finally, the statistics we used here are principally able to detect selection signatures associated with “hard” selective sweeps. More recently it has been suggested that selection may proceed on a background of standing variation or a “soft sweep”. This raises the scenario that perhaps both the thrifty gene and drifty gene ideas have been applicable at different

### Table 1. Positive Selection Signals of Each SNP

<table>
<thead>
<tr>
<th>No.</th>
<th>SNP</th>
<th>Gene</th>
<th>Effect Allele</th>
<th>[DDAF]</th>
<th>TD</th>
<th>F_{ST}</th>
<th>[iHS]</th>
<th>XPCLR</th>
<th>XPEHH</th>
<th>Sum</th>
<th>Protective or Risky</th>
</tr>
</thead>
<tbody>
<tr>
<td>c2</td>
<td>rs12579757</td>
<td>GP2</td>
<td>T</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>14</td>
<td>protective</td>
</tr>
<tr>
<td>e3</td>
<td>rs2206734</td>
<td>CDKAL1</td>
<td>T</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>16</td>
<td>protective</td>
</tr>
<tr>
<td>e5</td>
<td>rs3810291</td>
<td>TEMEM160</td>
<td>A</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>15</td>
<td>risky</td>
</tr>
<tr>
<td>d9</td>
<td>rs12401738</td>
<td>FUBP1;USP33</td>
<td>T</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>risky/protective</td>
</tr>
<tr>
<td>b10</td>
<td>rs2121279</td>
<td>LEPIB</td>
<td>T</td>
<td>0</td>
<td>12</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>17</td>
<td>risky</td>
</tr>
<tr>
<td>f10</td>
<td>rs7903146</td>
<td>TCF7L2</td>
<td>C</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>16</td>
<td>risky</td>
</tr>
<tr>
<td>d11</td>
<td>rs2365389</td>
<td>FHIT</td>
<td>C</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>14</td>
<td>protective</td>
</tr>
<tr>
<td>e12</td>
<td>rs9925964</td>
<td>KAT8;ZNF646;VKORC1;ZNF668;STX1B;FBXL19</td>
<td>A</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>16</td>
<td>protective</td>
</tr>
<tr>
<td>e14</td>
<td>rs11583200</td>
<td>ELAVL4</td>
<td>C</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>14</td>
<td>protective</td>
</tr>
</tbody>
</table>

* Nine SNPs exceeded the criterion of >14 positive selection signals to be statistically significant. Four of them were risky SNPs (favoring obesity) and five were protective (favoring leanness). Each number refers to total positive selection signals of all 14 populations for each SNP. The sum is the total positive selection signal for each SNP across the 14 populations for all the six tests. The complete table including all the other BMI-associated SNPs that did not reach the significance criterion can be found in the Table S1.
periods in our evolutionary history and in different populations depending on the frequency and occurrence of famine events or changes in predation risk. Selection for famine survival may then have intermittently proceeded on a background of drifted allele frequencies. The available analytical techniques for genome analysis would not allow us to detect such effects. From our current results, there is little evidence to support the thrifty gene hypothesis that obesity is the consequence of selection under famines or indeed that obesity has been positively selected for any other reason.

**EXPERIMENTAL PROCEDURES**

**Identification of BMI Indexed SNPs**

There were 115 BMI-associated SNPs (Table S2) that were identified from previous GWAS publications (Locke et al., 2015; Okada et al., 2012; Speliotes et al., 2010; Willer et al., 2009; Wen et al., 2012) and included into the analysis. All SNPs reached genome-wide significance ($p < 10^{-8}$). The 115 SNPs were identified in adults from European and East-Asian populations. The majority of the SNPs (86 out of 115) were discovered in European populations, 24 were identified in East-Asian populations, and five were discovered in both European and East-Asian populations. All the 115 SNPs analyzed in our study are located on autosomes. We identified whether the risk allele is ancestral or derived allele through dbSNP (build 137: https://www.ncbi.nlm.nih.gov/snp/) and Ensembl database (http://www.ensembl.org/index.html).

**Identification of Control SNPs**

Positive Control

Lactose tolerance (Bersaglieri et al., 2004) has been shown to have undergone strong positive selection in individuals of European descent after the domestication of cattle. Lactose tolerance is linked to activity of the lactase gene ($LCT$) and $MCM6$ (upstream of $LCT$). Three $LCT$ related SNPs were chosen from dbSNP database (build 137: https://www.ncbi.nlm.nih.gov/snp/) as positive controls. In addition, alleles that favor lighter skin, hair, and eye pigmentation (Lamason et al., 2005; Wilde et al., 2014) in European populations have also experienced strong positive selection. We also chose two of these SNPs (rs1426654 [$SLC24A5$] and rs16891982 [$SLC45A2$]) as positive controls. We used selection on these five SNPs in the CEU population as a positive control for the efficiency of the various metrics we used to quantify positive selection.

Negative Control

120 randomly selected SNPs matched for their recombination and allele frequencies to the BMI-associated SNPs were selected using the Ensembl database. Two-sample $t$ tests were conducted on the ancestral allele frequency after arcsin transformation. No significant differences ($p > 0.05$) were found in European (CEU, TSI, GBR, IBS, and FIN), American (MXL, CLM, and PUR), East-Asian (CHB, CHS, and JPT), and African (ASW, LWK, and YRI) populations, after Bonferroni correction for multiple testing.

We calculated the total positive selection signals of Tajima’s D, $\mu$DAF, $F_{ST}$, $\mu$HS, XPEHH, and XPCLR. We also counted positive selection signals for each SNP. Because we had six tests and 14 populations, the maximum number of
The frequency spectrum of these occurrences was used to calculate the probability of each frequency between 0 and 30. A typical running average LOD score was calculated using a ten SNP window over 500 kb upstream and downstream around the target SNPs (BMI-associated SNPs).

We used a Bayesian conjugate beta-binomial analysis to model the distribution of positive signals in the control SNPs, using the WinBUGS program. First, we set the parameters of a beta distribution (prior), knowing that the mean probability of success across all tests and populations was 0.0802. We set parameter “a” of the beta distribution at values between 1.6 and 4.4 and then varied “b” to obtain the mean probability of 0.0802 for the prior distribution. We made 10,000 MCMC simulations for each posterior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution to obtain the probability of each frequency between 0 and 30. A typical comparison between a MCMC simulation (in this example with a = 1.75 and b = 19.9 for the beta distribution) and the actual distribution is shown in Figure S4. We used these posterior distributions based on various priors to define a one-tailed 95% upper confidence limit for positive selection across all the various values of parameters entered into the prior beta distribution, the minimum value of the 95% upper confidence limit was 14 positive selection signals that could occur for any given SNP was 6 x 14 = 84.

The actual number of positive signals in the negative control SNPs ranged from 0 to 26 (Figure S4). The frequency spectrum of these occurrences was used to generate a theoretical distribution that would account for false discoveries (FDR) in the main data set for BMI-associated SNPs. We set the parameters of a beta distribution at values between 1.6 and 4.4 and then varied “b” to obtain the mean probability of 0.0802 for the prior distribution, combining the beta prior with the likelihood binomial distribution at values between 1.6 and 4.4 and then varied “b” to obtain the mean probability of 0.0802 for the prior distribution. We then performed a conjugate beta-binomial analysis to model the distribution of positive signals in the control SNPs, using the WinBUGS program. First, we set the parameters of a beta distribution (prior), knowing that the mean probability of success across all tests and populations was 0.0802. We set parameter “a” of the beta distribution at values between 1.6 and 4.4 and then varied “b” to obtain the mean probability of 0.0802 for the prior distribution. We made 10,000 MCMC simulations for each posterior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution to obtain the probability of each frequency between 0 and 30. A typical comparison between a MCMC simulation (in this example with a = 1.75 and b = 19.9 for the beta distribution) and the actual distribution is shown in Figure S4. We used these posterior distributions based on various priors to define a one-tailed 95% upper confidence limit for positive selection across all the various values of parameters entered into the prior beta distribution, the minimum value of the 95% upper confidence limit was 14 positive selection signals that could occur for any given SNP was 6 x 14 = 84.

The actual number of positive signals in the negative control SNPs ranged from 0 to 26 (Figure S4). The frequency spectrum of these occurrences was used to generate a theoretical distribution that would account for false discoveries (FDR) in the main data set for BMI-associated SNPs. We set the parameters of a beta distribution at values between 1.6 and 4.4 and then varied “b” to obtain the mean probability of 0.0802 for the prior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution to obtain the probability of each frequency between 0 and 30. A typical comparison between a MCMC simulation (in this example with a = 1.75 and b = 19.9 for the beta distribution) and the actual distribution is shown in Figure S4. We used these posterior distributions based on various priors to define a one-tailed 95% upper confidence limit for positive selection across all the various values of parameters entered into the prior beta distribution, the minimum value of the 95% upper confidence limit was 14 positive selection signals that could occur for any given SNP was 6 x 14 = 84.

The actual number of positive signals in the negative control SNPs ranged from 0 to 26 (Figure S4). The frequency spectrum of these occurrences was used to generate a theoretical distribution that would account for false discoveries (FDR) in the main data set for BMI-associated SNPs. We set the parameters of a beta distribution at values between 1.6 and 4.4 and then varied “b” to obtain the mean probability of 0.0802 for the prior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution to obtain the probability of each frequency between 0 and 30. A typical comparison between a MCMC simulation (in this example with a = 1.75 and b = 19.9 for the beta distribution) and the actual distribution is shown in Figure S4. We used these posterior distributions based on various priors to define a one-tailed 95% upper confidence limit for positive selection across all the various values of parameters entered into the prior beta distribution, the minimum value of the 95% upper confidence limit was 14 positive selection signals that could occur for any given SNP was 6 x 14 = 84.

The actual number of positive signals in the negative control SNPs ranged from 0 to 26 (Figure S4). The frequency spectrum of these occurrences was used to generate a theoretical distribution that would account for false discoveries (FDR) in the main data set for BMI-associated SNPs. We set the parameters of a beta distribution at values between 1.6 and 4.4 and then varied “b” to obtain the mean probability of 0.0802 for the prior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution to obtain the probability of each frequency between 0 and 30. A typical comparison between a MCMC simulation (in this example with a = 1.75 and b = 19.9 for the beta distribution) and the actual distribution is shown in Figure S4. We used these posterior distributions based on various priors to define a one-tailed 95% upper confidence limit for positive selection across all the various values of parameters entered into the prior beta distribution, the minimum value of the 95% upper confidence limit was 14 positive selection signals that could occur for any given SNP was 6 x 14 = 84.

The actual number of positive signals in the negative control SNPs ranged from 0 to 26 (Figure S4). The frequency spectrum of these occurrences was used to generate a theoretical distribution that would account for false discoveries (FDR) in the main data set for BMI-associated SNPs. We set the parameters of a beta distribution at values between 1.6 and 4.4 and then varied “b” to obtain the mean probability of 0.0802 for the prior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution, combining the beta prior with the likelihood binomial distribution for r successes from n = 84 (tests x populations), with p drawn from the prior. We made 10,000 MCMC simulations for each posterior distribution to obtain the probability of each frequency between 0 and 30. A typical comparison between a MCMC simulation (in this example with a = 1.75 and b = 19.9 for the beta distribution) and the actual distribution is shown in Figure S4. We used these posterior distributions based on various priors to define a one-tailed 95% upper confidence limit for positive selection across all the various values of parameters entered into the prior beta distribution, the minimum value of the 95% upper confidence limit was 14 positive selection signals that could occur for any given SNP was 6 x 14 = 84.
AUTHOR CONTRIBUTIONS

J.R.S. conceived and designed the project. G.W. and J.R.S. did the analysis. J.R.S. and G.W. co-wrote and revised the paper.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (NSFC grant 91431102), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDB13030100), and the “1000 Talents Program.” We are grateful to all the members of the Molecular Energetics Group, especially Chaoqun Niu, for their support and discussion of the results and to Professor Wenfeng Qian for helpful discussion and constructive comments on the manuscript. We are grateful to three anonymous referees who made useful comments on the original submitted version, which greatly improved the paper.

Received: March 8, 2016
Revised: July 4, 2016
Accepted: August 23, 2016
Published: September 22, 2016

REFERENCES


Hayakawa, K. (1996). The heritability of body mass index among an internal

