Check for updates

RESEARCH NOTE

REVISED Evidence of polygenic selection on human stature
inferred from spatial distribution of allele frequencies
[version 2; peer review: 1 approved with reservations]

Davide Piffer

Ulster Institute for Social Research, London, UK

v2 First published: 16 Jan 2015, 4:15
<https://doi.org/10.12688/f1000research.6002.1>
 Second version: 23 Dec 2015, 4:15
<https://doi.org/10.12688/f1000research.6002.2>
 Latest published: 25 Jan 2016, 4:15
<https://doi.org/10.12688/f1000research.6002.3>

Abstract

Spatial patterns of allele frequencies reveal a clear signal of natural (or sexual) selection on human height. The average frequency of all hits (N=693) and the top significant 66 common genetic variants (pruned for linkage disequilibrium) for 26 populations belonging to 5 sub-continental human groups was significantly correlated to average phenotypic population height. The method of correlated vectors provided additional evidence for a signal of natural selection in SNPs with higher significance. Factor analysis of the five top genome-wide association study (GWAS) hits revealed a clear factor indicating selection pressures on human height, peaking among northern Europeans and some African groups (Esan Nigeria) whilst reaching a nadir among South-East Asians.

Keywords

Height; Evolution; Polygenic Selection; Height

Open Peer Review**Approval Status** ? X

	1	2
version 3 (revision) 25 Jan 2016		view
version 2 (revision) 23 Dec 2015		
version 1 16 Jan 2015	view	

1. **Ben Busby**, NCBI/NLM/NIH, Bethesda, USA
2. **Andrew D Kern**, Rutgers University, Piscataway, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Davide Piffer (pifferdavide@gmail.com)

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2015 Piffer D. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Piffer D. **Evidence of polygenic selection on human stature inferred from spatial distribution of allele frequencies [version 2; peer review: 1 approved with reservations]** F1000Research 2015, 4:15
<https://doi.org/10.12688/f1000research.6002.2>

First published: 16 Jan 2015, 4:15 <https://doi.org/10.12688/f1000research.6002.1>

REVISED Amendments from Version 1

The frequencies of all the SNPs meeting GWAS significance were downloaded from 1000 Genomes (N=693).

The polygenic score computed from all the hits was correlated to average population height. Moreover, the method of correlated vector was applied to this bigger sample of SNPs and a weak signal was found, such that more significant hits had a stronger correlation to population height.

A new column has been added to [Table 1](#) and [Table 2](#). The values represent average frequencies of all the GWAS hits (N=693) for the 26 populations and continental groups, respectively.

See referee reports

Introduction

A recent GWAS ([Wood *et al.*, 2014](#)) based on a very large sample (N=250K) identified common variants responsible for normal variation in human height within populations.

Over the last few years, researchers have started moving away from the study of genetic evolution using a single-gene, Mendelian approach towards models that examine many genes together (polygenic). The more genes are involved in a given phenotype, the more the signal of natural selection will be “diluted” across different genomic regions (because each gene accounts for a tiny effect) making it difficult to detect it using approaches focused on a single gene ([Piffer, 2014a](#); [Pritchard *et al.*, 2010](#)). A first attempt at empirically identifying polygenic selection was made by [Turchin *et al.*, \(2012\)](#) on two populations (Northern and Southern Europeans) and evidence for higher frequency of height increasing alleles (obtained from GWAS studies) among Northern Europeans was provided. A drawback of that study was the reliance on populations from a single continent and that crude pairwise comparisons (e.g. French vs. Italian) were used without correlating frequency differences to average population height. Moreover, the strength of selection was not determined.

Two different approaches to identify selection based on the correlation of allele frequencies across different populations have been recently developed by [Piffer \(2013\)](#) and [Berg & Coop \(2014\)](#).

Piffer’s method uses factor analysis of trait increasing alleles (found by GWA studies) as a tool for finding a factor that represent the strength of selection on a phenotype and the underlying genetic variation ([Piffer, 2014a](#)). An additional methodology consists of computing the correlation between genetic frequencies and the average phenotypes of different populations; then, the resulting correlation coefficients are correlated with the corresponding alleles’ genome-wide significance (p value). If the alleles contain selection signals, a positive correlation will be found, as alleles with high p value (more likely to be false positives) have a weaker correlation to average population phenotype ([Piffer, 2014a](#)).

Piffer’s method ([Piffer, 2013](#); [Piffer, 2014a](#)) to identify signals of polygenic selection was used in this study and applied to the top five GWAS hits (ranked according to p value). [Piffer \(2014b\)](#) carried out a study on height SNPs but it was based on a smaller

GWAS sample and an older version (phase 1) of the 1000 Genomes data, containing data for only 14 populations. This paper uses the phase 3 1000 Genomes data and the GWAS meta-analysis was carried out on a much larger sample size, which produces more hits with better significance. The aim of this paper is to test the hypothesis that stature has undergone natural or sexual selection in populations after humans dispersed in different continents giving rise to distinct genetic clusters.

Methods

Frequencies of alleles with a positive effect (height increasing) were obtained from 1000 Genomes (phase 3): <http://browser.1000genomes.org/index.html> comprising 26 populations belonging to five racial groups.

Average population height was obtained from the references listed at: http://en.wikipedia.org/wiki/Human_height, considering only statistics published after 2000 and young age groups (18–40). Only 11 populations met these criteria (see references in [Table 1](#)).

First, the entire sample of SNPs meeting genome-wide significance (N=697) was downloaded and the population frequencies for the alleles with a positive Beta in the 250k GWAS Meta-analysis were calculated. 4 SNPs could not be included because they were absent from 1000 Genomes.

For each chromosome, the three alleles with the highest p values were selected, and these were all unlinked (>500Kb apart from each other). Only unlinked alleles were used to avoid the confounding influence of linkage on cross-population allele frequency. Selection was restricted only to the alleles with the highest significance because these are less likely to be false positives. The same number of SNPs (3) from each chromosome was used to get a representative sample of the entire genome, to avoid bias due to chromosome location. The conventional nominal p-value $<5 \times 10^{-8}$ was used as significance threshold ([Barsh *et al.*, 2012](#)).

A polygenic score was calculated as the mean frequency of height increasing alleles (defined as those with a positive Beta coefficient in the meta-analysis).

Analyses were carried out using R.

Results

Dataset 1. All Hits

<http://dx.doi.org/10.5256/f1000research.6002.d109935>

This dataset reports the 693 SNPs having genome-wide significance. Data derived from 1000 Genomes, phase 3 data.

Dataset 2. Hits 1+2+3

<http://dx.doi.org/10.5256/f1000research.6002.d41833>

This dataset reports the frequencies of 66 height increasing alleles, 3 from each autosomal chromosome. Data derived from 1000 Genomes, phase 3 data.

Table 1. Polygenic score and height.

Population	Polygenic Score (All Hits), %	Polygenic score (66 SNPs), %	Height (cm)	Reference (Height)
Afr.Car.Barbados	47.32	48.94		
US Blacks	47.65	48.71	178.00	McDowell <i>et al.</i> , 2008
Esan Nigeria	47.09	49.50		
Gambian	47.19	48.97		
Luhya Kenya	46.85	48.42		
Mende Sierra Leo	47.21	49.03		
Yoruba	46.98	48.52		
Colombian	48.32	46.05	170.60	Meisel & Vega, 2004
Mexican LA	47.88	46.95	170.6	McDowell <i>et al.</i> , 2008
Peruvian	47.40	46.48		
Puerto Rican	48.34	46.79		
Chinese Dai	46.75	44.88		
HanChineseBeijing	46.95	44.76	170.2	Yang <i>et al.</i> , 2005
HanChineseSouth	47.12	45.70	170.2	Yang <i>et al.</i> , 2005
Japanese	46.75	44.85	172.00	Ministry of Ed., 2004
Vietnam	46.98	44.76	165.70	Hung & Park, 2008
UtahWhites	50.22	47.62	178.9	McDowell <i>et al.</i> , 2008
Finns	49.61	48.09	180.70	National Institute for Health and Welfare, 2011
British	49.89	46.80	177.80	Moody (2013). Health Survey for England
Spanish	49.10	46.77		
TuscanItaly	49.17	47.11	177.00	Cacciari <i>et al.</i> , 2002*
Bengali Banglade	47.50	46.09		
Gujarati Ind. Tx	47.60	47.12		
Indian Telegu UK	47.76	47.62		
Punjabi Pakistan	48.08	47.21		
SriLankanUK	47.70	46.98		

*Not on Wikipedia. Region of a country, more specific statistics found elsewhere.

Dataset 3. Method of correlated vectors (MCV) for the top 66 SNPs

<http://dx.doi.org/10.5256/f1000research.6002.d109933>

This dataset reports SNP names, p value and the correlation between p value with polygenic score (col.B) and average height (col.D), for the top 66 SNPs. Data derived from 1000 Genomes, phase 3 data.

Dataset 4. Method of correlated vectors (MCV) for all SNPs

<http://dx.doi.org/10.5256/f1000research.6002.d109934>

This dataset reports SNP names, p value and the correlation between p value with polygenic score (col.B) and average height (col.D), for all SNPs. Data derived from 1000 Genomes, phase 3 data.

Polygenic score

Polygenic scores and average country height are reported in Table 1. The Pearson correlation between polygenic score including all hits (N=693) and average country height was $r=0.79$ (N=11, $p=0.004$).

The correlation with the 66 hits polygenic score was $r=0.83$ (N=11, $p=0.002$). Table 2 reports average frequencies by sub-continental populations.

Frequencies in descending order for the two (66 hits and all hits, respectively) polygenic scores are: 1) Africans (AFR); 2) Europeans (EUR); 3) South Asians (SAS); 4) Latin Americans/Hispanics (AMR); 5) East Asians (ASN). 1) Europeans; 2) Latin Americans/Hispanics; 3) South Asians; 4) Africans; 5) East Asians

Table 2. Frequencies of height increasing alleles for sub-continental populations.

Continent	66 Hits Polygenic score (%)	All hits Polygenic Score
AFR	47.69	47.10
AMR	45.92	48.07
ASN	45.52	46.98
EUR	46.65	49.63
SAS	46.549	47.79

Method of correlated vectors (MCV)

Spearman's rank order correlation between each allele's p value and its correlation with the polygenic score (66 hits) and with height were respectively -0.26 and -0.34 (N=66, p=0.037 and 0.0053). The "rcorr" and "cor" functions in R produced slightly different results due to differences in dealing with ties (equal values). "cor" produced slightly stronger coefficients (-0.28 and -0.37). The correlations using all hits were much smaller, but in the right direction (N=693; r= -0.059, p=0.121 and r= -0.111, p=0.003 for height and polygenic score, respectively).

This provides evidence for the hypothesis that more significant GWAS hits (alleles) are enriched with natural selection signal. A similar phenomenon was observed in a previous analysis of genes affecting human height (Piffer, 2014b).

Factor analysis of the top 5 hits

Factor analysis requires a satisfying cases to variable ratio, thus only a handful of SNPs could be used and these had necessarily to be those with the lowest p value, as they are more likely to be genuine hits (see previous section, MCV).

The top 5 alleles (i.e. those with the lowest p value) all correlated with the polygenic score and with average height in the expected direction (positively), as shown in Table 3 (see Dataset 3). The average correlations were 0.58 and 0.69, respectively, which is a significant improvement compared to the average of the correlations with polygenic score and height of all the 66 alleles (r=0.03 and 0.04, respectively; see Dataset 2, cells BP38-39).

A factor analysis using minimum residuals was carried out. A single factor was extracted that explained 42% of the variance. Factor loadings are displayed in Table 4. These are all positive (in the expected direction).

Factor scores were extracted with the Thurstone method (Thurstone, 1947), and are reported in Table 5.

The Pearson correlation between average country height and the factor score was strongly positive (r=0.88, N=11, p=0.001). This factor was also significantly correlated to the 66 hits polygenic score (r=0.78, N=26, p<0.001) and to the 693 hits polygenic score (r=0.545, N=26, p=0.004).

Table 3. Top five SNPs. (p value and r with polygenic (pol) score).

SNP	rs724016.G	rs1812175.G	rs42039.T	rs143384.G	rs8756.C
GWAS p value	3.2E-158	2.1E-86	3.8E-88	1.2E-121	4.5E-90
r with pol. score	0.78	0.26	0.22	0.84	0.78
r with average pop. height	0.62	0.75	0.75	0.49	0.88

Table 4. Top 5 SNPs. Standardized loadings (pattern matrix) based upon correlation matrix.

Gen.coordinate	SNP ID	Factor loading
142.588.260 (Chr.3)	rs724016.G	0.62
145.794.294 (Chr.4)	rs1812175.G	0.33
92.082.358 (Chr.7)	rs42039.T	0.62
33.489.170 (Chr.20)	rs143384.G	0.48
64.646.019 (Chr.12)	rs8756.C	1

Table 5. Factor scores.

Population	Height Top 5 SNP factor	Height (cm)
Afr.Car.Barbados	1.08	
US Blacks	0.24	178.00
Esan Nigeria	1.29	
Gambian	0.73	
Luhya Kenya	0.38	
Mende Sierra Leo	0.38	
Yoruba	1.08	
Colombian	0.08	170.60
Mexican LA	-0.27	172.00
Peruvian	0.15	
Puerto Rican	0.44	
Chinese Dai	-1.76	
HanChineseBeijing	-1.41	172.10
HanChineseSouth	-1.62	172.10
Japanese	-1.41	172.00
Vietnam	-1.69	165.70
UtahWhites	1.43	179.00
Finns	1.29	180.70
British	1.01	177.80
Spanish	0.58	
TuscanItaly	0.72	177.00
Bengali Banglade	-0.41	
Gujarati Ind. Tx	-0.41	
Indian Telegu UK	-0.69	
Punjabi Pakistan	-0.48	
SriLankanUK	-0.70	

Discussion

Polygenic scores, created by averaging frequencies from 26 populations of all SNPs having attained genome-wide significance ($N=693$; $p<5*10^{-8}$) in the largest and most recent human height GWAS, was positively correlated with the average height of 11 populations ($r=0.79$). Another polygenic score, obtained from 66 height increasing alleles pruned for linkage disequilibrium, was positively correlated with the average height of 11 populations ($r=0.83$). The method of correlated vectors revealed that alleles with lower p values had a higher correlation with phenotypic height and polygenic score, suggesting that they tend to be enriched with signal of natural selection. A factor analysis of the top five GWAS hits produced a factor (whose loadings are all in the expected direction) which is significantly and strongly correlated both to population average height and to the polygenic scores. This showed an improvement over the correlation of the five single alleles with population height (Table 3, last row) which averaged 0.66, which in turn improved over the average correlation of the 66 alleles, which was near zero.

The rankings of polygenic and factor scores match with the folk perception on the stature of various racial groups: Africans>Europeans>South/Central Asians>Hispanics>East Asians (Table 2). However, the ranking of Africans was lower in the polygenic score computed using all the GWAS hits, whereas the others were little altered with respect to each other.

South East Asians had the lowest scores, a result which matches with their anthropometric description.

Within Europe, northern Europeans (Finns and White Americans) had a higher genotypic stature than their southern counterparts (Italians and Spaniards), confirming the results from a previous study on GWAS loci which compared northern vs southern Europeans (Turchin *et al.*, 2012).

A limitation was the unavailability of sound statistics on the average height of many populations. Moreover, although human height is largely heritable, it is also heavily influenced by nutrition and living conditions. The importance of environment is suggested by the dramatic secular trend which took place in the 20th century in developed countries (e.g. Arcaleni, 2006; Webb *et al.*, 2008); an association with dietary intakes (i.e. milk consumption) and socio-economic status has also been observed (Mamidi *et al.*, 2011; Webb *et al.*, 2008). Most of the missing data were for developing countries which likely have not reached their full growth potential or ethnic groups living in Western societies (Indian Telegu or Gujarati) for which anthropometric statistics are not easily available. If the allele frequency factor represents a genuine signal of natural selection, then the difference between it and current phenotypic height could be used as an indicator of the quality of diet and living conditions in general.

Conclusion

Factor analysis of allele frequencies is a promising method for detecting signals of recent selection on polygenic traits.

Data availability

F1000Research: Dataset 1. All Hits., [10.5256/f1000research.6002.d109935](https://doi.org/10.5256/f1000research.6002.d109935) (Piffer, 2015a).

F1000Research: Dataset 2. Hits 1+2+3, [10.5256/f1000research.6002.d41833](https://doi.org/10.5256/f1000research.6002.d41833) (Piffer, 2014c).

F1000Research: Dataset 3. Method of correlated vectors (MCV) for the top 66 SNPs., [10.5256/f1000research.6002.d109933](https://doi.org/10.5256/f1000research.6002.d109933) (Piffer, 2015b).

F1000Research: Dataset 4. Method of correlated vectors (MCV) for all SNPs., [10.5256/f1000research.6002.d109934](https://doi.org/10.5256/f1000research.6002.d109934) (Piffer, 2015c).

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

References

- Arcaleni E: **Secular trend and regional differences in the stature of Italians, 1854–1980.** *Econ Hum Biol.* 2006; **4**(1): 24–38.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Barsh GS, Copenhaver GP, Gibson G, *et al.*: **Guidelines for genome-wide association studies.** *PLoS Genet.* 2012; **8**(7): e1002812.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Berg JJ, Coop G: **A population genetic signal of polygenic adaptation.** *PLoS Genet.* 2014; **10**(8): e1004412.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cacciari E, Milani S, Balsamo A, *et al.*: **Italian cross-sectional growth charts for height, weight and BMI (6–20 y).** *Eur J Clin Nutr.* 2002; **56**(2): 171–80.
[PubMed Abstract](#)
- Hung MV, Pak S: **The impact of environment on morphological and physical indexes of Vietnamese and South Korean students.** *VNU Journal of Science, Natural Science and Technology.* 2008; **24**: 50–55.
[Reference Source](#)
- Mamidi RS, Kulkarni B, Singh A: **Secular trends in height in different states of India in relation to socioeconomic characteristics and dietary intakes.** *Food Nutr Bull.* 2011; **32**(1): 23–34.
[PubMed Abstract](#)
- McDowell MA, Fryar CD, Hirsch R, *et al.*: **“Anthropometric Reference Data for Children and Adults: United States, 2003–2006”.** *National Health Statistics Reports.* 2008; **10**.
[Reference Source](#)
- Meisel A, Vega M: **“A tropical success story: a century of improvements in the biological standard of living, Colombia 1910–2002”.** Paper prepared for The Fifth World Congress of Cliometrics, Venice International University, Venice, Italy, July 8–11, 2004.
[Reference Source](#)
- Ministry of Education, Culture, Sports, Science and Technology. Japan, 2004.
[Reference Source](#)
- Moody A: **10: Adult anthropometric measures, overweight and obesity.** In Craig, Rachel; Mindell, Jennifer. *Health Survey for England – 2012.* 2013.
[Reference Source](#)
- National Institute for Health and Welfare: **Lasten kasvun seurannan uudistaminen, Asiantuntijaryhmän raportti.** 2011.
[Reference Source](#)
- Piffer D: **Factor Analysis of Population Allele Frequencies as a Simple, Novel Method of Detecting Signals of Recent Polygenic Selection: The Example of Educational Attainment and IQ.** *Mankind Quarterly.* 2013; **54**(2): 168200.
[Reference Source](#)
- Piffer D: **Simple statistical tools to detect signals of recent polygenic selection.** *IBC.* 2014a; **6**(1): 1–6.
[Reference Source](#)
- Piffer D: **Opposite selection pressure on stature and intelligence across human populations.** *Open Behavioral Genetics.* 2014b.
[Reference Source](#)
- Piffer D: **Dataset 1 in: Evidence of polygenic selection on human stature inferred from spatial distribution of allele frequencies.** *F1000Research.* 2015a.
[Data Source](#)
- Piffer D: **Hits 1+2+3.** *F1000Research.* 2014c.
[Data Source](#)
- Piffer D: **Dataset 3 in: Evidence of polygenic selection on human stature inferred from spatial distribution of allele frequencies.** *F1000Research.* 2015b.
[Data Source](#)
- Piffer D: **Dataset 4 in: Evidence of polygenic selection on human stature inferred from spatial distribution of allele frequencies.** *F1000Research.* 2015c.
[Data Source](#)
- Pritchard JK, Pickrell JK, Coop G: **The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation.** *Curr Biol.* 2010; **20**(4): R208–215.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Thurstone LL: **Multiple Factor Analysis.** University of Chicago Press. 1947.
[Publisher Full Text](#)
- Turchin MC, Chiang CW, Palmer CD, *et al.*: **Evidence of widespread selection on standing variation in Europe at height-associated SNPs.** *Nat Genet.* 2012; **44**(9): 1015–1019.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Webb EA, Kuh D, Pajak A, *et al.*: **Estimation of secular trends in adult height, and childhood socioeconomic circumstances in three Eastern European populations.** *Econ Hum Biol.* 2008; **6**(2): 228–236.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wood AR, Esko T, Yang J, *et al.*: **Defining the role of common variation in the genomic and biological architecture of adult human height.** *Nat Genet.* 2014; **46**(11): 1173–86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Yang XG, Li YP, Ma GS, *et al.*: **[Study on weight and height of the Chinese people and the differences between 1992 and 2002].** *Zhonghua Liu Xing Bing Xue Za Zhi.* 2005; **26**(7): 489–93.
[PubMed Abstract](#)

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 30 October 2015

<https://doi.org/10.5256/f1000research.6422.r9015>

© 2015 Busby B. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ben Busby

The School of Biochemistry, NCBI/NLM/NIH, Bethesda, MD, USA

The only major thing that, in my opinion, stands between this manuscript is the availability of the R scripts used to produce the data tables (especially given the difference between corr and rcorr). I looked for any linking to or availability of these scripts in Piffer 2013 or 2014a-d, and could not find it. Without these scripts, I do not think this work can be considered reproducible.

Two additional (minor) revisions are as follows.

First, the phenotypic data in table one seems to rely solely on Wikipedia. It seems likely that the author could find additional height information beyond Wikipedia.

Second, the last four paragraphs of the article could be combined into one.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research