Submitted: 22nd of April 2014
Published: 12th of May 2014
DOI: 10.26775/OBG.2014.05.12
ISSN: 2446-3876

Opposite selection pressures on stature and intelligence across human populations

Davide Piffer*

Open Behavioral Genetics

Abstract

Principal component analysis was used to detect signals of recent polygenic selection for higher stature. This selection was stronger among subSaharan Africans and weaker among East Asians. The principal component significantly correlates with mean male height in 9 populations. ANOVA shows that the frequencies of 46 heightincreasing alleles differ significantly across three continental races (Europeans, subSaharan Africans, East Asians). We can therefore reject the null hypothesis that random drift accounts for these differences. GWAS hits from different population samples have similar loadings on the principal component, implying that the same alleles have similar phenotypic effects across human populations. These heightincreasing alleles are distributed among the three geographic groups in a way that inversely mirrors what we see for intelligenceenhancing alleles. This pattern might be explained by Allen's rule and the cognitive demands imposed by cold climates.

Keywords: intelligence; polygenic selection; evolution; natural selection; random drift

1 Introduction

Piffer (2014a) provides evidence that recent selection has acted on cognitive abilities and has thus produced significant differences among human populations in the frequencies of alleles that tend to increase educational attainment and cognitive ability. In a subsequent paper, a suggestive signal of directional selection on human stature was also found by principal component analysis of allele frequencies, although ANOVA did not find significant differences between human populations in the frequencies of height-increasing alleles (Piffer, 2014b).

The present study has three aims:

1. increase the power to detect signals of recent selection on human height by including genome-wide association study (GWAS) hits from multiple studies, instead of focusing on a single study from the GIANT consortium (Allen et al., 2010), as in Piffer (2014b);
2. control for demographic biases by including GWAS hits from different population samples, i.e., if a GWAS is carried out on a specific population, it could have greater power to identify height-associated variants whose frequencies are closer to 0.5 in that population;
3. investigate how geographic variation in stature compares with geographic variation in cognitive ability. No specific a priori hypotheses on the evolutionary link between height and intelligence were formulated.

2 Materials and Methods

Height SNPs were selected from the NHGRI GWAS catalog (www.genome.gov/gwastudies).

The SNPs had to meet two criteria:

a) study published since 2010; and

b) p value $<10^{7}$.

If a study had more than 10 SNPs that met both criteria, only the first 10 (with the lowest p value) were chosen.

* E-mail: pifferdavide@gmail.com

To investigate the possibility of bias due to ethnicity, three GWAS were included from each of the following three racial groups: East Asians; SubSaharan Africans; and Europeans.

Frequencies of height-increasing alleles were obtained from 1000 Genomes (http://www.ensembl.org/index.html). For each study, a polygenic score (average frequency of height-increasing alleles across all SNPs) was calculated.

A total of 52 SNPs were chosen from 6 studies: Berndt et al. (2013); Hao et al. (2013); Carty et al. (2012); N'Diaye et al. (2011); Allen et al. (2010); Okada et al. (2010).

Another search was carried out on the Allele Frequency Database (ALFRED: alfred.med.yale.edu). The SNPs that could not be found were replaced with SNPs in strong linkage disequilibrium (r2 0.8 (LD)). LD was calculated with SNAP (SNP Annotation and Proxy Search, https://www.broadinstitute.org/mpg/snap/), using the 1000 Genomes pilot 1 dataset, CEU as population panel and a distance limit of 500 kB.

Mean population height was retrieved from Wikipedia (Human Height). All of the chosen studies provided measured height, were published after 2000, and were performed on young subjects (1739 years). This information is shown in Table 2.

Table 1: Component matrix (Polygenic scores). 1000 Genomes.

Publication	PC1
Berndt et al. (European)	0.850
Hao et al. (Chinese)	0.856
Carty et al. (African American)	0.804
N'Diaye et al. (African American)	0.958
Okada et al. (Japanese)	0.927
Allen et al. (European)	0.903

Table 2: Principal Component scores (Height-increasing alleles). 1000 Genomes. Mean male height (Wikipedia).

–	PC Polygenic	PC1 6 repl. hits	Height
ASW	1.341	0.654	178
LWK	1.568	0.868	–
YRI	1.885	0.690	–
CLM	-0.281	-0.138	170.6
MXL	-0.398	0.077	170.6
PUR	-0.021	0.002	–
CHB	-1.352	-1.694	171.15
CHS	-1.221	-1.693	–
JPT	-1.183	-1.829	170.7
CEU	-0.107	0.929	179
FIN	-0.336	0.592	179
GBR	0.019	0.681	178
IBS	0.271	0.126	–
TSI	-0.185	0.735	177

3 Results

3.1 *Principal Component Analysis*

Six polygenic scores were calculated from the 52 SNPs with p< 10^7 described in the six published studies. As 6 of the SNPs were found to have a significant effect in more than 1 of the studies, these repetitions were deleted and the total was thus reduced to 46 SNPs. The findings are reported in Table 6 1. Principal component analysis was used to test the hypothesis that one factor largely determines the genetic variance, as a result of recent polygenic selection. Two components explain 78.23 % and 17.73 % of the variance, respectively. Kaiser-Meyer-Olkin is satisfactory (0.720). Chi-squared (Bartlett's test of sphericity)=115.24, df= 15, p=.000. The component matrix (Table 1 2) is consistent with the first component being a single selection pressure for higher stature. The second component is not clearly interpretable.

An analysis with rotation was performed but the solution was not clearly interpretable. Thus, the unrotated solution was preferred and is reported.

A PCA was carried out using the 6 replicated hits. Two components explain 69.68 and 18.06 % of the variance, respectively. Kaiser-Meyer-Olkin is satisfactory (0.702). Chi-squared (Bartlett's test of sphericity)= 65.07, df=15, p=0.000. The component matrix (Table 3) is consistent with the first component being a selection pressure for higher stature, whereas the second PC is not clearly interpretable. Component scores are reported in Table 2. When the first component is compared with total variance, the correlation is moderately high and significant (r= 0.745, p= 0.002). Correlations with mean population height are positive (r=.612) and marginally significant (p=.080) for PC1, and positive (r=.784) and significant (p=.012) for the PC extracted from the 6 replicated hits.

The first component for the height-increasing alleles was compared with the principal components for alleles affecting educational attainment and IQ (Piffer, 2014a; Piffer & Kirkegaard, 2014). The correlations are negative (r= .945 and 0.944, p<0.001). The first component for the 6 replicated hits was compared with the total variance for the same alleles that affect educational attainment and IQ. The correlations are likewise significant and negative (r=0.735 and 0.715, p<0.001).

Table 3: Component matrix (6 replicated hits). 1000 Genomes

–	PC1	PC2
Rs1351394T	0.933	0.112
Rs42235T	0.975	-0.136
Rs42235T	0.842	0.340
Rs7689420C	0.833	0.172
Rs11658329C	0.656	-0.734
Rs1787200A	0.080	0.986

Table 4: Component matrix. Top 4 height-increasing alleles. Proxies in parentheses. Frequencies from ALFRED.

SNPs	PC loadings
rs1991431-A (rs9846369 T)	0.842
rs42235 T (rs2282978 C)	0.828
rs11658329 C (rs9901507 A)	0.878
rs3791679 A	0.854

The method of correlated vectors was used to assess the predictive power of the factor scores and the presence of noise in the dataset, e.g., false positives (Piffer, 2014b). If the SNPs were true indicators of positive natural selection and if some of them were false hits, there should be a negative correlation between p value and PC loadings. A negative correlation was found but it did not reach statistical significance (N=46; r=0.203). Since the method of correlated vectors indicated the presence of some noise (false positives), the SNPs with the highest factor loadings (0.9) were searched on ALFRED to better assess the worldwide pattern of allele frequencies. These are reported in Table 7 5, for 50 populations. A principal component analysis was carried out. A single component explains 72.37 % of the variance. Kaiser-Meyer-Olkin is satisfactory (0.766). Chi-squared (Bartlett's test of sphericity)= 97.17, df=6, p=0.00. The component matrix (Table 4) shows that the component is clearly interpretable as a single selection pressure for higher stature. Component scores are reported in Table 7 (last column).

Another PCA using the 6 replicated hits did not produce a satisfactory solution (KMO= .450). Two components each explain only 38 % of the variance, and the component matrix is not clearly interpretable. For the 50 populations, however, the component scores are almost identical to those produced by the previous PCA, which used the 4 most significant SNPs (r= 0.92). Thus, only the latter component scores are reported.

This PC correlates with the first principal components for educational attainment (Piffer, 2014a) and IQ (Piffer & Kirkegaard, 2014) (reported in Table 5). The correlations are strongly negative and significant (r= 0.89; 0.92), p<0.01.

Table 5: IQ and Educational Attainment PC (Piffer & Kirkegaard, 2014).

–	IQ PC	Education PC
Afr.HG	-1.84	-2.1
Afr. Farmers	-1.32	-1.71
Middle Easterners	0.12	-0.19
Europeans	0.27	0
Central Asians	0.85	0.1
East Asians	0.83	0.97
Southeast Asians	1.15	0.32
Oceanians	-0.14	-0.685
Native Americans	0.07	-0.9

3.2 *ANOVA*

Because the height-increasing alleles showed a strong signal of recent selection, ANOVA was carried out to determine whether their frequencies differ significantly between human races. If there are no significant differences (the null hypothesis), we can infer that selection pressures on height have not differed appreciably among human populations in the course of evolution. If there are significant differences, we can infer that these selection pressures have indeed differed from one human population to another.

A oneway ANOVA was performed to test for different frequencies of height-increasing alleles across three human races (subSaharan Africans, East Asians, Europeans). Since 6 SNPs occurred in more than one study, these double occurrences were deleted and the total was reduced to 46 SNPs. Allele frequencies significantly differ between the three geographic groups, F (2, 137)= 3.196, p= 0.044.

Using a multiple comparison procedure with Tukey's post-hoc test, the three groups were compared to find out how their means differ from each other. It was found that these alleles have significantly higher frequencies among subSaharan Africans (M= 65.86 %) than among East Asians (M= 49.67 %), 95 % CI [0.62, 29.76], p= 0.039.

When Europeans (M= 54.43 %) were compared with the other two groups, the differences were not statistically significant at p<.05.

4 Discussion

Principal component analysis and ANOVA produced similar results. Height-increasing alleles are significantly more frequent among subSaharan Africans

than among East Asians, with Europeans falling in the middle. This pattern inversely mirrors the one found by (Piffer, 2014b) for alleles that enhance intellectual capacity, whose mean frequency is highest among East Asians, lower in Europeans, and lower still among subSaharan Africans.

The height-increasing alleles of this study had good predictive power, since the first component of their variance (obtained from the 6 replicated GWAS hits) significantly correlated with mean male height in each geographic group (Table 2).

No appreciable differences were found within the three geographic groups (e.g., Iberians vs. Finns or Chinese vs Japanese). Differences in height within the three continental races may thus be due to environmental factors, different kinds of genetic variation, or epigenetics not captured by the present analysis. Remarkably, although the SNPs came from genetically different groups, they all loaded highly (about 0.9) on the principal component. It thus seems that ascertainment bias does not dramatically skew the results towards frequently occurring alleles in the population under study. It also seems that the same alleles tend to have similar phenotypic effects across human populations.

There is one exception. The SNPs described by Hao et al. (2013) have negative loadings, probably because the sample size used for this GWAS (6,354 in the initial phase and 1,881 in the replication stage) was much smaller than in the other studies, whose much bigger samples (ranging from about 8k to 133k) gave them much higher statistical power. This is also highlighted by the higher p values of the Hao et al. (2013) SNPs. Moreover, whilst the other studies comprised 8 to 10 hits with $p<10^7$, the Hao et al. (2013) study had only 5 hits. Thus, the polygenic score overestimates their importance.

Since, the SNPs of Okada et al. (2010)'s Japanese sample loaded highly on the principal component, this conflicting result is not due to the East Asian ethnicity of Hao et al. (2013)'s sample.

Analysis of the ALFRED dataset provided a more thorough picture of worldwide genetic variation. Interestingly, Native Americans have higher PC values than do East Asians, despite their genetic relatedness. This finding inversely mirrors the finding by Piffer (2014a) that Native Americans have lower PC values for intelligence-enhancing alleles. It is noteworthy that two genetically similar populations (Native Americans and East Asians) have remarkably different allele frequencies for height and intelligence, and that these differences are consistent with observed phenotypic differences. These patterns cannot easily be attributed to random genetic drift or founder effects.

This between-group pattern contrasts with the within-group pattern already observed, i.e., IQ is slightly higher on average in taller people than in shorter people (Pearce et al., 2005; Humphreys et al., 1985). This correlation is in part due to common genetic factors (Marioni et al., 2014) and assortative mating (Beauchamp et al., 2011). Although some genes may have pleiotropic effects on stature and intelligence, we see a very different picture between human populations: selection for height has been strongest where selection for intelligence has been weakest. It could be that height correlates positively with intelligence within populations because both traits act as fitness indicators and are thus subject to a common selection pressure that pushes both of them in the same direction. Alternatively, mutational load has decreased both traits, hence creating the genetic correlation.

Why, then, do selection for height and selection for intelligence operate in opposite directions when we compare these three major geographic groups? It is possible that the same environments that selected for higher intelligence were also more advantageous for shorter people. Allen's rule posits that animals from colder climates usually have a lower ratio of surface area to volume, as seen in shorter limbs and a more rotund body shape, than do equivalent animals from warmer climates. This rule is based on the dynamics of heat loss in cold climates: the greater the exposed surface area, the greater the loss of heat and therefore energy. Animals in cold climates need to conserve as much energy as possible. A low surface area to volume ratio thus helps to conserve heat, as there is a smaller surface area for body heat to pass through. Human populations probably also follow Allen's rule (Katzmarzyk & Leonard, 1998).

The Arctic Mongoloid physical type seems to exhibit this kind of adaptation to cold climates (Steegman & Platner, 1968). This may be why East Asians have the shortest stature among the three major geographic groups. Parallel to this, colder climates select for higher intelligence as a means to cope with two evolutionarily novel problems: finding food and keeping warm (Templer & Arikawa, 2006). Thus, the negative link between intelligence and stature across populations could be the result of cold climates providing a fitness advantage for shorter stature (via Allen's rule) and higher intelligence (via "cold winters theory").

Because the component scores were high among African hunter-gatherers, including the Pygmies, it seems that the atypical height of the Pygmies is due only in part to selection on common variants. Their atypical growth pattern seems due to a few mutations that evolved independently among various Pygmy groups, possibly under the influence of natural selection for less growth and reduced thyroid function or sexual selection for smaller bodies (Migliano et al., 2013). Drastically reduced gene expression has been

proposed as another mechanism to account for the short stature of the Pygmies. In particular, the expression of Growth Hormone (GH) and Growth Hormone Receptor (GHR) genes has been reduced by factors of 1.8 and 8, respectively (Bozzola et al., 2009).

These explanations are tentative. For now, we can say with certainty only that stature and IQ have been subject to opposite selection pressures during recent human evolution.

References

Allen, H. L., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., ... Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*(7317), 832–838. doi: 10.1038/nature09410

Beauchamp, J. P., Cesarini, D., Johannesson, M., Lindqvist, E., & Apicella, C. (2011). On the sources of the height–intelligence correlation: New insights from a bivariate ace model with assortative mating. *Behavior Genetics*, *41*, 242–252. doi: 10.1007/s10519-010-9376-7

Berndt, S. I., Gustafsson, S., Mägi, R., Ganna, A., Wheeler, E., Feitosa, M. F., ... Ingelsson, E. (2013). Genomewide metaanalysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature Genetics*, *45*(5), 501–512. doi: 10.1038/ng.2606

Bozzola, M., Travaglino, P., Marziliano, N., Meazza, C., Pagani, S., Grasso, M., ... Arbustini, E. (2009, November). The shortness of pygmies is associated with severe under-expression of the growth hormone receptor. *Molecular Genetics and Metabolism*, *98*(3), 310–313. doi: 10.1016/j.ymgme.2009.05.009

Carty, C. L., Johnson, N. A., Hutter, C. M., Reiner, A. P., Peters, U., Tang, H., & Kooperberg, C. (2012). Genome-wide association study of body height in african americans: the women's health initiative snp health association resource (share). *Human molecular genetics*, *21*(3), 711–720. doi: 10.1093/hmg/ddr489

Hao, Y., Liu, X., Lu, X., Yang, X., Wang, L., Chen, S., ... Gu, D. (2013). Genome-wide association study in han chinese identifies three novel loci for human height. *Human genetics*, *132*(6), 681–689. doi: 10.1007/s00439-013-1280-9

Humphreys, L. G., Davey, T. C., & Park, R. K. (1985). Longitudinal correlation analysis of standing height and intelligence. *Child Development*, *56*(6), 1465–1478. Retrieved from http://www.jstor.org/stable/1130466

Katzmarzyk, P., & Leonard, W. (1998). Climatic influences on human body size and proportions: Ecological adaptations and secular trends. *American Journal of Physical Anthropology*, *106*(4), 483–503. doi: 10.1002/(SICI)1096-8644(199808)106:4<483::AID-AJPA4>3.0.CO;2-K

Marioni, R., Batty, G., Hayward, C., Kerr, S., Campbell, A., Hocking, L., ... Deary, I. (2014). Common genetic variants explain the majority of the correlation between height and intelligence: the generation scotland study. *Behavior Genetics*, *44*, 91–96. doi: 10.1007/s10519-014-9644-z

Migliano, a., Romero, I., Metspalu, M., Leavesley, M., Pagani, L., Antao, T., ... Kivisild, T. (2013, 12). Evolution of the pygmy phenotype: Evidence of positive selection fro genome-wide scans in african, asian, and melanesian pygmies. *Human biology*, *85*, 251-84. doi: 10.3378/027.085.0313

N'Diaye, A., Chen, G. K., Palmer, C. D., Ge, B., Tayo, B., Mathias, R. A., ... Haiman, C. A. (2011, 10). Identification, replication, and fine-mapping of loci associated with adult height in individuals of african ancestry. *PLOS Genetics*, *7*(10), 1-11. doi: 10.1371/journal.pgen.1002298

Okada, Y., Kamatani, Y., & Takahashi, A. (2010). A genomewide association study in 19 633 japanese subjects identified lhx3qsox2 and igf1 as adult height loci. *Human Molecular Genetics*, *19*, 2302–2312.

Pearce, M. S., Deary, I. J., Young, A. H., & Parker, L. (2005, 03). Growth in early life and childhood IQ at age 11 years: the Newcastle Thousand Families Study. *International Journal of Epidemiology*, *34*(3), 673-677. doi: 10.1093/ije/dyi038

Piffer, D. (2014a, 01). 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. *The mankind quarterly*, *54*, 168-200. doi: 10.46469/mq.2013.54.2.3

Piffer, D. (2014b). Simple statistical tools to detect signals of recent polygenic selection. *Interdisciplinary Bio Central*. doi: 10.4051/ibc.2014.6.1.0001

Piffer, D., & Kirkegaard, E. O. W. (2014). The genetic correlation between educational attainment, intracranial volume and iq is due to recent polygenic selection on general cognitive ability. *Open Behavioral Genetics*. doi: 10.26775/OBG.2014.04.12

Steegman, A., & Platner, W. (1968). Experimental cold modification of cranio-facial morphology. *American Journal of Physical Anthropology*, *28*, 17–30.

Templer, D. I., & Arikawa, H. (2006). Simple statistical tools to detect signals of recent polygenic selection. *Intelligence*, *34*, 121–139.

Table 6: Polygenic scores from 6 Height GWAS. 1000 Genomes.

Population*	Berndt et al.	Hao et al.	Carty et al.	NDyaie et al.	Okada et al.	Allen et al.
ASW	66	53	48.86	64.13	69.6	71.4
LWK	65.67	52	53.57	66.75	69.7	72.8
YRI	68.33	50.6	56	68.13	72.5	72.7
CLM	55.83	59.4	38.57	50.38	63.6	58.6
MXL	55.33	61	35.43	51.13	61.7	61.2
PUR	56	54.6	39.43	51.88	64.4	61.2
CHB	57	70.2	37.43	46.25	47.9	43
CHS	59	71.8	39.43	47.5	48.8	41.9
JPT	55.67	69.4	38.86	48	52.3	43.2
CEU	57.67	57.4	36.43	50.88	62.6	64.6
FIN	55.67	59.2	33.57	52.5	61	62.7
GBR	60	56.4	35.29	52.88	62.6	63.2
IBS	58.33	48.6	39.14	52.25	64.9	62.7
TSI	58.5	57.8	34.14	49.25	62.9	64.1

ASW: African ancestry in SW USA; LWK: Luhya, Kenya; YRI: Yoruba, Nigeria; CLM: Colombian; MXL: Mexican ancestry from LA, California; PUR: Puerto Ricans from Puerto Rico; CHB: Han Chinese in Bejing, China; CHS: Southern Han Chinese; JPT: Japanese in Tokyo, Japan; CEU: Utah Residents with Northern and Western European Ancestry; FIN: Finnish in Finland; GBR: British in England and Scotland; IBS: Iberian population in Spain; TSI: Toscani in Italy.

–	rs1991431A (rs9846369 T)	rs42235 T (rs2282978 C)	rs11658329 C (rs9901507 A)	rs3791679 A	PC
African H.G.					2.235
San	83	58	100	100	2.462
Biaka	94	53	90	100	2.419
Mbuti	77	50	70	93	1.825
African Farmers					2.025
Bantu	78	45	87	100	2.021
Yoruba	73	52	85	96	2.021
Mandenka	83	44	94	90	2.021
Middle Easterners					2.025
Mozabite	40	28	40	80	0.452
Bedouin	30	61	46	71	0.833
Druze	27	38	34	55	0.084
Palestinian	28	43	30	62	0.221
Europeans					0.425
Adygei	22	40	41	68	0.271
Basque	27	52	33	83	0.628
French	50	43	22	72	0.551
Italians	47	52	20	75	0.664
Orcadian	44	38	22	72	0.383
Russian	34	32	20	74	0.142
Sardinian	29	50	18	73	0.340
Central Asians					−0.150
Burusho	28	30	8	82	−0.021
Kalash	26	14	10	86	−0.249
Pashtun	28	28	22	72	−0.008

Balochi	34	38	21	64	0.141
Brahui	32	20	22	66	−0.154
Hazara	21	17	17	50	−0.595
Sindhi	20	28	14	76	−0.166
East Asians					−0.916
Dai	40	20	0	25	−0.752
Mongolia	25	5	10	35	−0.988
Daur	17	11	11	44	−0.887
Han	24	11	6	26	−1.048
Hezhe	39	11	0	28	−0.884
Japanese	21	5	5	29	−1.177
Koreans	15	19	4	17	−1.165
Lahu	25	25	0	20	−0.935
Miao	15	20	10	25	−0.989
Naxi	44	28	22	17	−0.403
Oroquen	25	30	5	30	−0.682
She	25	0	10	15	−1.297
Tu	20	10	5	35	−1.031
Tujia	20	10	5	40	−0.975
Uyghur	25	10	15	45	−0.735
Xibe	28	11	6	28	−0.970
Yi	40	10	0	25	−0.920
Yakut	27	20	6	245	−0.648
Southeast Asians					−0.537
Cambodians,Khmer	24	22	14	47	−0.537
Oceanians					−0.434
Papuan New Guinean	21	26	0	65	−0.467
Melanesian, Nasioi	45	0	0	81	−0.401
Native Americans					0.319
Pima, Mexico	28	6	12	69	−0.524
Maya, Yucatan	37	54	16	82	0.598
Amerindians	77	15	4	92	0.480
Karitiana	79	46	4	93	1.039
Surui	40	31	5	71	0.006

Table 7: Frequencies and PC for top 4 (GWAS significance) heightincreasing alleles (ALFRED). Proxies in parentheses.