Intelligence 53 (2015) 43–50

Contents lists available at ScienceDirect

Intelligence

A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation

Davide Piffer

Ulster Institute for Social Research, London, UK

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form 13 August 2015
Accepted 22 August 2015
Available online xxxx

Keywords:
Intelligence
GWAS
SNP

ABSTRACT

Published Genome Wide Association Studies (GWAS), reporting the presence of alleles exhibiting significant and replicable associations with IQ, are reviewed. The average between-population frequency (polygenic score) of nine alleles positively and significantly associated with intelligence is strongly correlated to country-level IQ ($r = .91$). Factor analysis of allele frequencies furthermore identified a metagene with a similar correlation to country IQ ($r = .86$). The majority of the alleles (seven out of nine) loaded positively on this metagene. Allele frequencies varied by continent in a way that corresponds with observed population differences in average phenotypic intelligence. Average allele frequencies for intelligence GWAS hits exhibited higher inter-population variability than random SNPs matched to the GWAS hits or GWAS hits for height. This indicates stronger directional polygenic selection for intelligence relative to height. Random sets of SNPs and Fst distances were employed to deal with the issue of autocorrelation due to population structure. GWAS hits were much stronger predictors of IQ than random SNPs. Regressing IQ on Fst distances did not significantly alter the results nonetheless it demonstrated that, whilst population structure due to genetic drift and migrations is indeed related to IQ differences between populations, the GWAS hit frequencies are independent predictors of aggregate IQ differences.

1. Introduction

Over the last few years, population geneticists have moved away from the study of genetic evolution using the single-gene, Mendelian approach, towards models that examine many genes together (i.e. polygenic models). The origin of this trend can be traced back to Fisher (1918). The more genes that are involved in a given phenotype, the more the signal of natural selection will be "dispersed" across different genomic regions (this is because each gene accounts for only a small proportion of the overall phenotypic variance) making individual genes difficult to detect using single gene-based approaches (i.e. Pritchard et al., 2010; Piffer, 2014). The first attempt at empirically identifying polygenic selection was made by Turchin et al. (2012) utilizing height alleles (obtained from Genome Wide Association Studies, henceforth GWAS) and two large populations (Northern and Southern Europeans). It was found that the Northern European population exhibited higher frequencies of height-increasing alleles (obtained from GWAS studies) — suggesting greater polygenic selection for height-enhancing alleles in this population relative to the Southern Europeans. The study had two principal drawbacks; i) the study relied on populations sourced from a single continent and, ii) crude country-level 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 and Coop (2014).

Piffer's method utilizes factor analysis of trait-increasing alleles (found by GWAS) as a tool for quantifying the strength of selection on a phenotype and the underlying genetic variation (Piffer, 2013, see also: Minkov & Bond, 2015). An additional methodology consists of computing the correlations between genetic frequencies and the average phenotypes of different populations; then, utilizing the method of correlated vectors (Jensen & Weng, 1994) to correlate the resulting vector of correlation coefficients with the corresponding vector of the alleles' factor loadings. If the alleles are associated with signals of genetic selection, a positive correlation will result, as alleles exhibiting high factor loadings can be expected to exhibit a stronger correlation with the phenotype of interest (Piffer, 2014).

To date, GWA studies have identified a handful of alleles with replicated association with intelligence and proxy phenotypes, such as educational attainment. Rietveld et al.'s (2013) meta-analysis identified 10 Single Nucleotide Polymorphisms (henceforth SNPs) that increased educational attainment, encompassing three with nominal genome-wide significance and seven with suggestive significance. A recent study (Ward et al., 2014) replicated the positive effect of these top

E-mail address: pifferdavide@gmail.com.

http://dx.doi.org/10.1016/j.intell.2015.08.008
0160-2896/

three SNPs (rs9320913, rs11584700 and rs4851266) on mathematics and reading performance in an independent sample of school children. These SNPs were also found to be associated with *g* (general intelligence) in a sub-sample of Rietveld et al.'s original study.

Another SNP (rs236330), located within gene *FNBP1L*, was found to be significantly associated with general intelligence — as was reported in two separate studies (Benyamin et al., 2013; Davies et al., 2011). This gene exhibits high levels of expression in neurons, including hippocampal neurons and developing brains, where it regulates neuronal morphology (Davies et al., 2011).

Rietveld et al. (2014), utilizing the proxy-phenotype method, found three additional SNPs significantly associated with cognitive performance in a sample overlapping that used in their previous study (Rietveld et al., 2013).

More recently, a GWAS focusing on fluid ability found 13 genetic variants exhibiting genome-wide significance (Davies et al., 2015), however, only three SNPs exhibited independent signals (i.e. were not in linkage disequilibrium [LD] with each other).

A significant GWAS association with executive function and processing speed was reported for the SNP rs17518584 ($p = 3.28 \times 10^{-9}$ after adjustment for age, gender and education, $N = 5429 - 32070$). This genetic variant is located in an intron of the gene Cell ADhesion Molecule 2 (*CADM2*) (Ibrahim-Verbaas et al., 2015).

The aim of the present paper is to analyze the population allele frequency patterns of all of the SNPs found to date (July 2015) to have genome-wide significant associations with general intelligence or related cognitive phenotypes (executive functioning, processing speed, educational attainment) in order to test the hypothesis that they predict country-level differences in cognitive ability above and beyond that predicted on the basis of neutral population genetic mechanisms (i.e. population structure due to migration and drift) which is a potential source of autocorrelation. The latter results from the non-independence of variables sampled in spatial or temporal proximity to one another (e.g. Mace & Pagal, 1994; for detailed criticisms of the concept as commonly, and naively applied in cross-cultural research see: Thornhill & Fincher, 2013). Autocorrelation is considered problematic because non-independence among supposedly spatially discrete variables violates the assumption of independently and identically distributed errors, hence leads to false positive (Type I) errors. Demonstrating that the alleles predict country-level differences in cognitive ability above and beyond that predicted on the basis of migration, drift etc, can be taken to evidence the theory that these differences have been shaped by diffuse polygenic selection operating on these alleles.

2. Methods

2.1. Variables

Allele frequencies for the SNPs were downloaded from the 1000 Genomes database (The 1000 Genomes Project Consortium, 1000 Genomes Project Consortium, 2012), using the final release of phase three data: http://browser.1000genomes.org/index.html

A common factor was extracted from among the SNPs, utilizing unweighted least squares factor analysis, yielding a metagene — this being a term utilized in genetics to describe patterns of covariance among genes.

Two alternative metagenes were constructed, one utilizing only the four SNPs (rs10457441, rs11584700, rs4851266, rs236330) exhibiting replicated GWAS associations with *g* and related phenotypes (henceforth "four SNPs metagene"; Piffer, 2015a), and a second utilizing all nine alleles discovered to date (henceforth "nine SNP metagene"). In addition to these metagene scores, a polygenic score was also computed by averaging across the allele frequencies.

Country-level average IQs were obtained from Lynn and Vanhanen (2012). The Finnish and Vietnamese IQs were both adjusted upwards to 101 (from 97 in Lynn & Vanhanen, 2012), to account for recent, more accurate estimates of the Finish population (Armstrong, Woodley, & Lynn, 2014) and a recent reanalysis of the Vietnamese data (Rindermann, Hoang, & Baumeister, 2013). IQ for Tuscany was calculated as the average between the IQs estimated from PISA Creative Problem Solving (Piffer & Lynn, 2014) and from PISA Math, Science, Reading. There were three populations (Chinese Dai, Gujarati Indian and Indian Telegu) for which no estimates of average IQ were available. These populations were therefore excluded from the analyses.

2.2. Autocorrelation control

The method for controlling for possible autocorrelation stemming from population structure is based on taking the correlation between Fst distances for the entire genome (or a random part of it) and distances (that is, the absolute number of the difference between any two populations) on the factor for all of the populations. Fst is a measure of population differentiation due to genetic structure, which is based on the partitioning of genetic diversity within and between populations. The vast majority of random SNPs all over the genome are believed not to be associated with specific phenotypes and therefore not to have been subject to selection. They rather represent population structure resulting from random genetic drift, migration, admixture, and similar processes.

Fst is calculated as the ratio of genetic variance in allele frequencies between populations to the sum of variance between gametes within individuals and within populations (Holsinger & Weir, 2009; Piffer & Dall'Olio, 2015; Weir & Cockerham, 1984).

Two matrices representing genetic distances with *N* unique pairwise comparisons are generated, where $N = n^*(n - 1)/2$. Another matrix representing phenotypic distances (computed for average population IQ) is then created.

Two steps are employed in testing the hypothesis that the factor does not merely represent population structure:

1. The correlation between the two matrices representing genetic distances is calculated. The lower the value, the more likely that the result is due to selection rather than population structure, as selection will skew distances away from background neutral variation due to random drift.
2. A regression of phenotypic distances on factor distances coupled with genome-wide Fst distances is carried out. If factor distances have an independent positive effect on the dependent variable (phenotypic distances), then the result is more likely to be indicative of polygenic selection.

All analyses were carried out in R (v. 3.2.1).

3. Results

The correlation between the most recent GWAS hit (allele T of rs17518584) and the four SNPs metagene was $r = .96$ ($N = 26$ populations).

The two hits (rs17522122.G and rs10119.G) from Davies et al. (2015) that were found not to be in LD with any of the four replicated SNPs (thus excluding rs10554471, which was found to be in LD with rs9320913 from Rietveld et al., 2013), were correlated to the four SNPs metagene. The Pearson's *r* values were − .514 and .698 (both $p < .05$) respectively.

One hit (rs1487441) located on chromosome six (position 98660615) from Rietveld et al. (2014) was in LD with one hit from Davies et al. (2015; rs10554471) and from Rietveld et al. (2013; rs9320913). The other two hits were however independent signals (rs7923609 and rs2721173). Their correlations with the four SNPs metagene were .322 (*ns*) and − .697 ($p < .05$) respectively.

An unweighted least squares factor analysis was carried out on all nine SNPs (one from Ibrahim-Verbaas et al., 2015, two from Davies

Table 1
Structure matrix for nine cognitive ability increasing alleles, showing the loading of the nine SNP metagene on each of its constituent SNPs.

Chromosome number	Location	SNP (ancestral vs derived)	Factor loading	Reference
6*	98572120	rs10457441.C (A)**	.70	Davies et al. (2015)
14	32372633	rs17522122.G (D)	−.52	Davies et al. (2015)
19	50098513	rs10119.G (D)	.78	Davies et al. (2015)
1	204576983	rs11584700.G (D)	.82	Rietveld et al. (2013)
2	100818479	rs4851266.T (D)	.93	Rietveld et al. (2013)
1	94059554	rs236330.C (D)	.91	Davies et al. (2011), Benyamin et al. (2013)
3	85604923	rs17518584.T (D)	.97	Ibrahim-Verbaas et al. (2015)
10	65133822	rs7923609.G (D)	.35	Rietveld et al. (2014)
8	145744429	rs2721173.C (A)	−.84	Rietveld et al. (2014)

* In linkage disequilibrium with two hits from two published GWA studies: rs9320913, location: 98,584,733 (Rietveld et al., 2013); rs1487441, location: 98,553,894 (Rietveld et al., 2014). Another recent study (Trampush et al., 2015) replicated the effect of this locus (specifically, rs1906252) on cognitive function.
** The alleles with a positive effect on IQ for the other SNPs linked to rs10457441 (minor allele) are derived for rs9320913 (allele A) and rs1487441 (allele A) and rs1906252 (allele A). This SNP — having an ancestral allele with a cognitive ability enhancing effect, seems to be an anomaly that contrasts with the other three GWAS hits in the same chromosomal region having derived alleles. This anomaly could be due to an error in the dbSNP database.

et al., 2015, two from Rietveld et al., 2014, in addition to the four SNPs analyzed in Piffer, 2015a).

A single factor was extracted, explaining 61% of the variance.

Structure matrix is shown in Table 1.

The factor scores for both the nine and four SNP metagenes, along with the polygenic score (the average of the allele frequencies) and population average IQs are shown in Table 2.

There was a positive and significant correlation between the nine SNPs metagene and IQ ($r = .863$, $N = 23$).

The method of correlated vectors was used to assess the predictive validity of factor analysis. The vector of the SNP's correlation with national IQs was correlated with the vector of their factor loadings.

There was a positive and significant correlation between the two vectors ($r = .986$).

3.1. Randomization

40 random SNPs matched to the nine GWAS hits were obtained using SNPsnap (Pers, Timshel, & Hirschhorn, 2015). These were used to test the hypothesis that the signal provided by the GWAS hits can be distinguished from background noise. That is to say, the factor extracted from the GWAS hits should be better predictive of national IQ than randomly matched SNPs.

Ten sets of four and four sets of nine SNPs were factor analyzed. The resulting factors were regressed against country IQ and the four and nine SNP metagene factors respectively. The polygenic score was also entered into the regression as a predictor along with polygenic scores obtained from the randomly selected SNPs. As was discussed in the methods, this polygenic score was calculated as the average frequency of the nine cognitive-ability increasing alleles. Its correlation with population-level IQ was $r = .91$ ($N = 23$ populations). The relationship is summarized in Fig. 1. Beta coefficients from the multiple-regression analysis are reported in Tables 3, 4 and 5.

The average Beta was calculated using the absolute value for the random SNPs and the real number for the GWAS hits. Whilst this inflated the values of the random SNP Betas, it is nonetheless based on the conservative assumption that the majority of the GWAS hits factor loadings are positive only by chance. GWAS hits produced higher Betas than the random SNPs (1.03 vs. .279).

In the analyses using the four and nine GWAS hit SNP metagenes, these were better predictors of IQ than the random SNPs factor in all

Table 2
Factor scores computed for both the nine and four SNP metagene, along with population-average IQ, polygenic score and population/continent of origin.

Continents*	Population	Polygenic score	Nine SNPs metagene scores	Four SNPs metagene score	Population average IQ
AFR	Afro Caribbean Barbados	.3726	−1.3174	−1.26112	83
AFR	African Americans	.3909	−1.2251	−1.21019	85
AFR	Esan Nigerian	.3607	−1.6754	−1.45081	71
AFR	Gambian	.3451	−1.5180	−1.44724	62
AFR	Luhya Kenyan	.334	−1.6720	−1.5391	74
AFR	Mende Sierra Leonean	.3516	−1.4886	−1.2412	64
AFR	Yoruban	.3391	−1.6197	−1.4649	71
HISP	Colombian	.4852	.1390	−.12223	83.5
HISP	Mexican (immigrants, Los Angeles)	.4871	.3748	.02157	88
HISP	Peruvian	.5006	.3496	−.30414	85
HISP	Puerto Rican	.4792	.1013	.00753	83.5
E.ASN	Chinese Dai	.5568	1.2361	1.18278	N/A
E.ASN	Han Chinese Bejing	.6182	1.2349	1.39839	105
E.ASN	Han Chinese South	.6	1.1606	1.30377	105
E.ASN	Japanese	.6076	.9399	1.2297	105
E.ASN	Vietnamese	.5914	1.2287	1.59826	99.4
EUR	American Whites (Utah)	.5298	.4879	.75587	99
EUR	Finnish	.54	.5797	.71432	101
EUR	British	.5427	.5357	.84863	100
EUR	Spanish	.5294	.3580	.59903	97
EUR	Italian (Tuscany)	.5229	.4469	.56805	99
SAS	Bengali Bangladeshi	.4858	.1920	−.25727	81
SAS	Gujarati Indian (immigrants, Texas)	.5126	.5075	.47096	N/A
SAS	Indian Telegu UK	.5066	.2838	−.60945	N/A
SAS	Punjabi Pakistani	.4976	.2230	.18886	84
SAS	Sri Lankan (immigrants, UK)	.4754	.1371	−.60945	79

* AFR = Sub-Saharan African; HISP = Hispanic/Latin American; E.ASN = East Asian; Eur = European; SAS = South Asian.

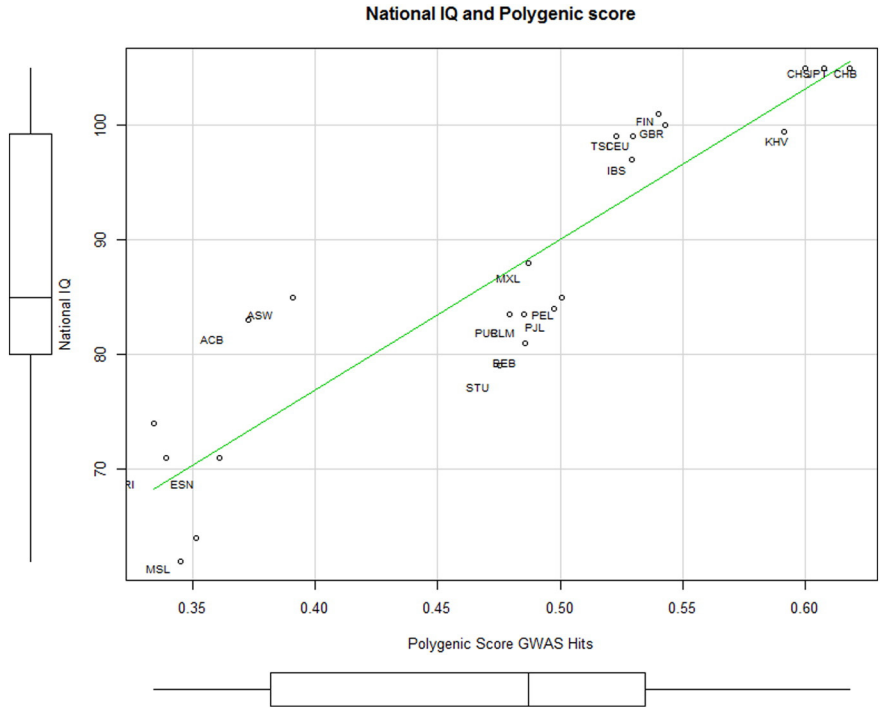

Fig. 1. Relationship between national IQ and polygenic score.

14 instances. For the analyses using the polygenic scores, the GWAS hits were a better predictor in all four instances.

The amount of multicollinearity was calculated with the variance inflation factor (VIF). A rule of thumb is that if VIF > 10, then multicollinearity is high (Kutner, Nachtsheim, & Neter, 2004). The regressions with the factors of four SNPs and with random polygenic scores (Tables 3 and 5, respectively) had VIFs well below 10. However, the regressions with the factors of nine SNPs exhibited high multicollinearity with a VIF index well in excess of 10 (Table 4).

Table 3
Beta coefficients resulting from the regression of population IQ on the four SNPs metagene and the Random SNP factors.

	Random SNP Factor	Four SNP metagene (GWAS hits)	VIF*
1	−.058	.867	3.623
2	.130	.993	1.521
3	−.29	.655	5.503
4	−.087	.866	1.540
5	−.092	.991	2.860
6	.081	.985	3.303
7	.038	.946	2.382
8	−.155	1.051	3.889
9	−.172	.759	6.626
10	.017	.930	2.347
Average Beta	.112	.9043	

* Variance inflation factor: command "vif", R package "car".

Table 4
Beta coefficients resulting from the regression of population IQ on the nine SNPs metagene and the Random SNP factors.

	Random SNP Factor	Nine SNP metagene (GWAS hits)	VIF
1	−.372	.569	2.682
2	.014	.849	28.84
3	.765	1.611	22.47
4	.869	1.695	12.02
Average Beta	.505	1.181	

MCV was run on the four random sets of SNPs. Contrary to expectations it yielded high magnitude correlations. These are reported in Table 6.

This result is likely due to the presence of *phylogenetic* autocorrelation. In this particular instance, the autocorrelation results from populations that are closer in space also being genetically more similar.

This could also account for the observation of high-magnitude correlations between the between factors extracted from the random sets of SNPs and population IQ, and the GWAS hits metagenes (see the correlation matrix reported in the supplementary files).

Whilst the regression analyses indicate that the GWAS hits factors are better predictors of IQ than the random factors in the majority of cases, the latter still show a strong correlation with country-level IQ. The average correlation with IQ (absolute correlation coefficient) is

Table 5
Beta coefficients resulting from the regression of population IQ on the nine GWAS hits polygenic score and a random SNP polygenic score.

	Random SNPs Polygenic Score	9 GWAS hits Polygenic Score	VIF
1	−.049	.875	2.028
2	−.180	.964	1.097
3	.487	1.344	4.826
4	−.169	.825	1.348
Average Beta	.221	1.002	

Table 6
Results of MCV correlating the vector of IQ–allele frequency correlations with the vector of the loadings of a nine random matched SNP factor on its constituent alleles. Four vector correlations were computed in total — one for each set of nine random SNPs.

First random SNP set	Second random SNP set	Third random SNP set	Fourth random SNP set
−.926*	.957**	−.978	−.974

* All $p < .05$*
** A positive correlation indicates that SNPs with higher factor loadings have a higher correlation to country IQ.

$r = .74$ ($p < .05$). The average correlation between the GWAS hits metagene and IQ is $r = .89$ ($p < .05$).

It is thus necessary to deal with autocorrelation. This can be achieved by controlling for genome-wide genetic distances utilizing the procedure employed by Piffer (2015b) and described in Section 2.2.

3.2. Controlling for population structure

The same procedure applied by Piffer (2015b) will be extended to the nine GWAS hits metagene. Fst distances (Weir & Cockerham, 1984) for Chromosome 21 and 1 (largest and smallest) will be utilized. These were calculated using *VCFtools* on the 1000 Genomes, phase 3 files. *VCFtools* is a program for working with Variant Call Format (VCF) files, like those generated by the 1000 Genomes Project (Danecek et al., 2011). The R code accompanying *VCFtools* is reported in the Appendix (A1).

In addition, distances calculated for the fourteen random SNP factors will be used in a simulation: they will be entered as independent variables in a regression of IQ on Fst distances, instead of the GWAS hits metagene. If they fail to predict IQ above and beyond Fst distances, whilst the metagene does, then this will indicate that the metagene has predictive power independent of population structure.

Inspection of the correlation matrix[1] reveals that population IQ distances exhibit a stronger correlation with the metagene and polygenic score distances (*r* x gdist = .78; gdist_metagene = .65; Polygenic = .76) than to the distances representing population structure ('randomnine') (*r* x Fst Chromosome 1 = .58; Fst Chromosome 21 = .58; randomnine 1 = .52; randomnine 2 = .64; randomnine 3 = .57; randomnine 4 = .52).

The R function "CIrdif" (package "psychometric") was used to calculate the confidence intervals for the difference between correlation coefficients, comparing the correlations of each pair of variables with country IQ. Two of the three correlation differences are significantly greater than zero. The 95% Confidence Intervals for the comparisons are *r* x Fst Chromosome 1 vs *r* x four SNPs metagene: .068 to .32; *r* x Fst Chromosome 1 vs *r* x nine SNPs metagene: −.06 to .21; *r* x random nine SNPs[2] vs Polygenic score (nine Hits): .068 to .33.

3.3. Regression analysis

Regression analysis was employed to estimate the relative strength of each predictor.

A total of 325 pairwise comparisons were obtained for the 26 populations from the 1000 Genomes database utilized in the present study. Distances were calculated as the absolute value of the difference between population pairs on the selected variables (country IQ, GWAS hit metagene and polygenic score). The country IQ variable had missing values, so a total of 253 distances was calculated. Fst distances were calculated for Chromosome 1, using the methodology employed by Piffer (2015b). Chromosome 1 and 21 Fst distances were almost identical ($r = .992$, $p < .05$), hence only the bigger chromosome (1) was used. Beta coefficients are reported in Tables 7a and 7b.

The ratio of average Betas (Average Beta random factor/Average Beta Fst) is .3/.366 = .82. The ratio of Betas of GWAS hits to Fst Betas is .885/.167 = 5.3.

There was a moderate amount of multicollinearity but this was not high according to the commonly accepted threshold (Tables 7a and 7a).

Table 7a
GWAS hits metagene distances and Fst: Standardized Beta coefficients. Dependent variable: IQ differences.

	Fst distances		VIF
	−.093	Four SNPs metagene distances: .850	2.704
	−.033	Nine SNPs metagene distances: .685	5.155
	−.409	Polygenic score distances: 1.227	4.553
Average Beta	−0.178	0.92	

Table 7b
Random SNPs factor distances and Fst: Standardized Beta coefficients. Dependent variable: IQ differences.

	Fst distances		VIF
	.431	Random factor 1 distances: .317	1.295
	.036	Random factor 2 distances: .609	5.033
	.372	Random factor 3 distances: .227	6.454
	.626	Random factor 4 distances: −049	5.652
Average Beta	0.366	0.276	

3.4. ANOVA

The average frequencies of the nine SNP hits for the five 1000 Genomes continental groups were calculated. These are represented in a boxplot (Fig. 2) and reported in Table 8.

An analysis of variance (ANOVA) was conducted in order to analyze the difference between group means. $F(1,235)$, $Pr(>F) = .311$.

Tukey's post-hoc test was used to compare means. Confidence intervals are reported in Table 9.

3.5. Estimating inter-population variability from average allele frequencies

Inter-population variability is usually employed as a means of detecting signals of selection at specific loci. The Fst index is measured at a single locus, as it compares inter-population variability to within-population variability. Deviation from normality (the average genome-wide Fst value between two populations) suggests the presence of selection at that locus. Another approach could be applied to polygenic traits based on analyzing many loci together. Once the average allele frequency of trait-increasing alleles is calculated, it is possible to obtain simple measures of inter-population variability, such as the standard deviation (SD). The SD of the average allele frequency of the nine GWAS hits was .088. This was higher in magnitude than the SD estimated for the average frequency of the four sets of nine random SNPs: .043, .032, .078 and .031. For comparison, seven sets of nine SNPs were also constructed from the 66 height GWAS hits reported in Piffer

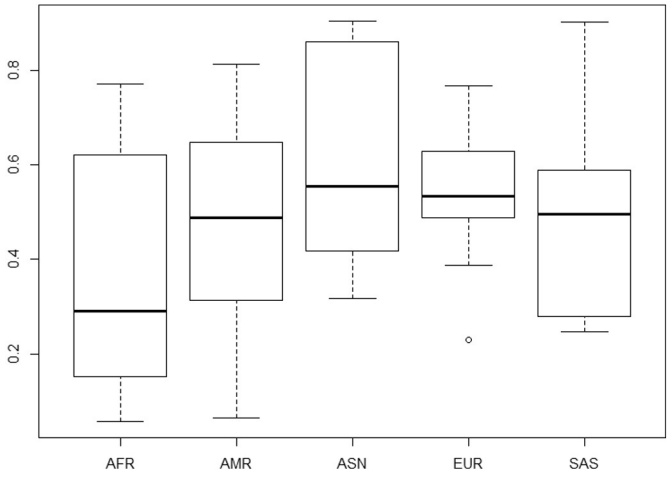

Fig. 2. Average frequency of cognitive ability increasing alleles by continental group.

[1] This is available for viewing at the following URL: https://docs.google.com/spreadsheets/d/1EanD-Jj15a3OBJpu5Tr5e86iLTLZPpa9vUEPTwaVsHU/edit?usp=sharing.

[2] Average correlation of the four correlation coefficients for the four random polygenic scores comprising nine SNPs picked at random.

Table 8
Allele frequencies by continental group.

Populations*	rs10457441 C	rs17522122 G	rs10119 G	rs11584700 G	rs4851266 T	rs236330C	rs17518584 T	rs7923609 G	rs2721173 C
AFR	.195	.696	.621	.057	.104	.34	.151	.29	.772
AMR	.352	.648	.793	.094	.064	.813	.591	.313	.488
ASN	.417	.622	.904	.317	.554	.862	.86	.337	.487
EUR	.533	.511	.713	.229	.387	.767	.628	.487	.539
SAS	.279	.588	.786	.255	.247	.903	.495	.525	.412

* AFR = Sub-Saharan African; AMR = Hispanic/Latin American; ASN = East Asian; EUR = European; SAS = South Asian.

(2015c). The average SD was .06 (.073, .057, .0658, .0751, .0286, .0607 and .0587).

4. Discussion

The average frequency (polygenic score) of nine alleles positively associated with IQ and proxy phenotypes at the individual differences level in published GWAS is strongly and significantly correlated to population, or country IQ ($r = .91$). Factor analysis of allele frequencies yielded a metagene factor with a similar correlation to IQ (.86). The majority of alleles (seven out of nine) loaded positively on this factor. 40 unrelated SNPs were drawn at random and their frequencies factor analyzed for use as a control. The pattern of very high-magnitude positive or negative correlations suggests that spatial autocorrelation might be inflating the relationships between variables. That is to say, factors extracted utilizing random SNPs exhibited very high correlations to the GWAS hits factors ($r = .6$ to .98) and similarly high correlations with country IQ distances. Unexpectedly, the method of correlated vectors produced very high values also when run using the random SNPs, rendering the extremely high magnitude and significant correlation (.99, $p < .05$) found for the GWAS hits somewhat less impressive. However, the correlation of IQ with the GWAS hits metagene (.89) was somewhat higher than the IQ correlation with the random SNP factors (.74). When entered in a regression, the GWAS hits metagene and the polygenic score were both higher-magnitude predictors of country IQ than the random SNPs factor in all instances (Average Betas: 1.03 and .28, respectively). The variance inflation factor showed the presence a moderate amount of multicollinearity but in most instances this was not high (<10, see Tables 3 and 5). However, in some regressions this was very high (e.g. VIF > 20, see Table 4) and likely accounted for the inflated Betas (>1).

Additionally, Fst distances were calculated in order to provide an estimate of genetic distances between populations, reflecting neutral evolutionary processes (e.g. migration and drift) but not directional selection pressure. Distances between populations were calculated as the absolute number of the difference between pairs (325 unique pairs). First, the four and nine GWAS hits metagenes and the polygenic score emerged as better predictors of IQ than Fst distances (Beta = .68 to 1.12 vs 0 to .41). Paradoxically, Fst distances had a negative relationship with IQ (specifically, IQ differences between populations) when regressed with the polygenic score (i.e. differences between populations). The latter had a very high predictive power (Beta = 1.12, see Table 7a). A set of regressions were run with Fst utilizing random factors instead of the GWAS hits metagenes as predictors. This revealed that random factors did not predict IQ better than Fst distances. Specifically, GWAS hits exhibited Beta coefficients that were about five times higher in magnitude than those associated with the random SNPs. Conversely, the latter had Betas equal to or lower than those associated with Fst distances (ratio = .82). Although the variance inflation factor was within acceptable limits, it showed the presence of moderate multicollinearity that might explain the somewhat inflated Betas (Tables 7a and 7a).

Comparison of allele frequency means for the five continental groups from the 1000 Genomes database revealed frequency differences that closely correspond to observed continent-level aggregate IQ, yielding the following pattern: East Asian > European > South Asian > American (Hispanic) > African. However, ANOVA did not yield p values that meet the conventional significance threshold ($p < .05$), furthermore Tukey's test produced confidence intervals that bisected zero. The lack of statistical significance is clearly due to the very small sample size ($N = 9$). Increasing numbers of GWA studies will undoubtedly provide more hits in the future permitting the generation of an increasingly accurate picture of cognitive-related genetic variation, both within and between populations.

When comparing the average frequencies of height and intelligence GWAS hits, a striking difference is the much higher inter-population variability of the latter, ranging from 33.4% (Luhya, Kenyan) to 61.82% (Han Chinese, Bejing). Conversely, the average frequencies of height increasing alleles range from 44.76% (Vietnamese) to 49.5% (Esan Nigerian) (Piffer, 2015c). This might indicate a stronger signal of selection for intelligence than for height, thus creating larger variability across isolated populations.

Another possibility is that environmental or social (e.g. cultural, sexual) factors that affect intelligence-related fitness differentials vary more across geographic regions and human populations compared to factors affecting height-related fitness differential. Relevant to this is the distinction between directional and diversifying selection (Holsinger & Weir, 2009). For example, if there has been geographically homogeneous sexual selection for taller males (due to intrasexual or intersexual competition) this would have produced an increase in the frequency of alleles with a positive effect on height, but the frequency differentials would have remained relatively small. Thus, this would result in strong directional selection with relatively little diversifying selection. On the other hand, if factors affecting intelligence-related fitness differentials, such as cultural differences in economic conditions and reproductive habits such as marriage customs and use of contraception, have varied more across populations, this would have resulted in high diversifying selection even in the presence of relatively weak directional selection. Evolutionary mechanisms such as gene–culture co-evolution (e.g. Feldman & Laland, 1996) are good candidates for the explanation of the apparent high diversifying selection on intelligence (Woodley, 2011). For example, higher "genotypic intelligence" could cause the development of more complex societies, and life in these complex societies might increasingly select for high-IQ genes. Culture–gene co-evolutionary mechanisms of this sort have been proposed by Clark (2006) to account for the apparent rise in human capital salient traits among among pre-industrial populations in England.

Table 9
Tukey's test with 95% confidence intervals for difference between continental group means.

Populations	Difference	95% Confidence interval
AMR − Afr	.103	−.217, .424
ASN − AFR	.237	−.083, .557
EUR − AFR	.174	−.146, .494
SAS − AFR	.140	−.180, .461
ASN − AMR	.133	−.186, .454
EUR − AMR	.070	−.249, .391
SAS − AMR	.037	−.283, .357
EUR − ASN	−.062	−.383, .257
SAS − ASN	−.097	−.417, .224
SAS − EUR	−.034	−.354, .287

Similar mechanisms have been proposed to account for apparent dysgenic trends with respect to *g* among Western countries in the period 1800 to the present (Woodley & Figueredo, 2013).

The higher number of SNPs comprising the height polygenic score ($N = 66$) also likely account to some extent for the smaller differences, however subsets of nine height increasing SNPs had a lower average SD relative to the cognitive ability GWAS hits (.06 vs .088). Moreover, the polygenic scores comprising nine random SNPs exhibit smaller average inter-population variability (SD = .046 vs .088).

Encouraging results for the study of the evolution of height and intelligence come from a recent paper presenting evidence of directional dominance on height and intelligence (but not on other traits such as blood pressure and low density lipoprotein cholesterol). This study provides evidence that increased stature and cognitive capacity have been recently positively selected in human evolution (Joshi et al., 2015) and provide additional context for the results of the present paper.

It should be noted that all of the nine alleles are present at significant frequencies (>5%) among all the five major races (Sub-Saharan African, South Asian, European, East Asian, American) (see Table 8). Thus, the intelligence polymorphisms do not appear to be race-specific but were already present in *Homo sapiens* prior to the African exodus circa 60–100 Kya. This is even more remarkable, given that the GWAS samples consisted mostly of individuals of European descent and that none of the GWAS hits appears to be European-specific polymorphisms (Table 8). It is thus likely that the vast majority of mutations affecting intelligence were already present in the ancestral African population and as humans settled in different parts of the worlds, these polymorphisms were subject to directional selection pressure, which produced an overall increase in human intelligence at different rates in different geographical areas. For the same reason, if non-European intelligence increasing polymorphisms exist, these are likely to represent a minority of the additive genetic variation contributing to differences in intelligence.

Population structure due to migrations and genetic drift constitute noise that this study attempted to isolate from the selection signals. However, in addition there may be pleiotropic effects on traits other than IQ so that an "IQ gene" could be subject to other kinds of selection pressure (e.g. via selection on blood pressure, body mass index, etc.) This complicates the picture and makes it more difficult to isolate a selection signal from genetic variants each with a small effect on the phenotype. Analyzing greater numbers of genes is thus very important, as the confounding effects of pleiotropic genes will cancel each other out over a large number of alleles as selection pressures will operate in different directions for different phenotypes. However, pleiotropy can "deviate" the frequency of a single allele. Reliance on a relatively limited number of populations (N = 26) also limits the significance of the results and does not enable us to completely rule out the null hypothesis that migrations or genetic drift account for the reported country-level effects.

Molecular population genetics is a rapidly expanding field and in the near future more cognitive ability-related genes will undoubtedly be identified. These will permit the hypothesis presented in this paper to be more comprehensively tested. Also the rapid progress in the sequencing of ancient genomes from fossils and also human remains from historical periods will enable us to calculate their "intelligence polygenic score" and to examine historical change in intelligence. We will be able employ these data in addressing paradoxes. For example, why did the apparently large brained Neanderthal lack the ecological flexibility necessary to outcompete *H.sapiens*? Alternatively, on a smaller time-scale, recent dysgenic trends in *g* in industrial societies can be quantified at the molecular level, corroborating possible phenotypic evidences of this (Woodley of Menie, 2015; Woodley of Menie, Fernandes, Figueredo, & Meisenberg, 2015).

Finally, this method can be "reverse-engineered" to aid in the detection of new GWAS hits by selecting polymorphisms whose frequencies correlate with the polygenic score or selection factor. These genes (or "polygenes") will have a higher probability of being intelligence-related genes, thus reducing the need for extremely large samples and the reliance upon 'chance capitalization' typical of current intelligence GWA studies (Piffer & Gilfoyle, 2014).

Acknowledgments

I would like to thank two anonymous reviewers for helpful comments on an earlier draft of this manuscript, and also Dr. Michael A. Woodley of Menie for editorial guidance in revising this manuscript.

Appendix A1

*Vcftools code:*cd c:/folder/… #set to directory containing 1000 Genomes vcf file

c:/Users/…/vcftools/bin/vcftools --vcfALL.chr1.phase3_shapeit2_mvncall_integrated_v5.20130502.genotypes.vcf --weir-fst-pop POP1.txt --weir-fst-pop POP2.txt --out fst.POP1.POP2

R code and datasets can be downloaded from: osf.io/jt73x

Appendix B. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.intell.2015.08.008.

References

Armstrong, E. L., Woodley, M. A., & Lynn, R. (2014). Cognitive abilities amongst the Sàmi population. *Intelligence*, *46*, 35–39. http://dx.doi.org/10.1016/j.intell.2014.03.009.

Benyamin, B., Pourcain, B., St. Davis, O. S., Davies, G., Hansell, N. K., Brion, M. -J. A., et al. (2013). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular Psychiatry*, *19*, 253–258. http://dx.doi.org/10.1038/mp.2012.184.

Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics*, *10*, e1004412.

Clark, G. (2006). Survival of the richest: The Malthusian mechanism in pre-Industrial England. *The Journal of Economic History*, *66*, 1–30.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCF tools. *Bioinformatics*, *27*, 2156–2158.

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chourake, V., Giddaluru, S., et al. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N = 53949). *Molecular Psychiatry*, *20*, 183–192. http://dx.doi.org/10.1038/mp.2014.188.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., et al. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*, 996–1005. http://dx.doi.org/10.1038/mp.2011.85.

Feldman, M. W., & Laland, K. N. (1996). Gene-culture coevolutionary theory. *Trends in Evolution and Ecology*, *11*, 453–457.

Fisher, R. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Philosophical Transactions of the Royal Society of Edinburgh*, *52*, 399–433.

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting Fst. *Nature Reviews Genetics*, *639–650*.

Ibrahim-Verbaas, C. A., Bressler, J., Debette, S., Schuur, M., Smith, A. V., Bis, J. C., et al. (2015). GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Molecular Psychiatry*http://dx.doi.org/10.1038/mp.2015.37.

Jensen, A. R., & Weng, L. J. (1994). What is a good *g*? *Intelligence*, *18*, 231–258.

Joshi, P. K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., et al. (2015). Directional dominance on stature and cognition in diverse human populations. *Nature*, *523*, 459–462.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression models* (4th ed.). McGraw-Hill Irwin.

Lynn, R., & Vanhanen, T. (2012). *Intelligence: A unifying construct for the social sciences*. London: Ulster Institute for Social Research.

Mace, R., & Pagal, M. (1994). The comparative method in anthropology. *Current Anthropology*, *35*, 549–564.

Minkov, M., & Bond, M. H. (2015). Genetic polymorphisms predict national differences in life history strategy and time orientation. *Personality and Individual Differences*, *76*, 204–215.

Pers, T. H., Timshel, P., & Hirschhorn, J. N. (2015). SNPsnap: A web-based tool for identification and annotation of matched SNPs. *Bioinformatics*, *31*, 418–420.

Piffer, D. (2013). 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*, *54*, 168–200.

Piffer, D. (2014). Simple statistical tools to detect signals of recent polygenic selection. *Interdisciplinary Bio Central*, *6*, 1–6. http://dx.doi.org/10.4051/ibc.2014.6.1.0001.

Piffer, D. (2015a). *Estimating the genotypic intelligence of populations and assessing the impact of socioeconomic factors and migrations*. The winnowerhttp://dx.doi.org/10.15200/winn.142299.93508 (e142299.93508).

Piffer, D. (2015b). *Intelligence GWAS hits: Selection signal or population structure? A test of the null hypothesis*. The winnowerhttp://dx.doi.org/10.15200/winn.142593.30710 (e142593.307).

Piffer, D. (2015c). *Evidence of polygenic selection on human stature inferred from spatial distribution of allele frequencies [v1; ref status: awaiting peer review, http://f1000r.es/4ye]. F1000Research*, *4*. (pp. 15), 15. http://dx.doi.org/10.12688/f1000research.6002.1.

Piffer, D., & Dall'Olio, M. (2015). *Using Vcftools to calculate Weir and Cockerham's variance components for 1000 genomes phase 3 data.* Figsharehttp://dx.doi.org/10.6084/m9.figshare.1450792.

Piffer, D., & Gilfoyle, B. (2014). *Detecting "polygenes" using signals of polygenic selection.* Tools for increasing the power of GWAS. Figsharehttp://dx.doi.org/10.6084/m9.figshare.1172301.

Piffer, D., & Lynn, R. (2014). New evidence for differences in fluid intelligence between north and south Italy and against school resources as an explanation for the north–south IQ differential. *Intelligence*, *46*, 246–249. http://dx.doi.org/10.1016/j.intell.2014.07.006.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current biology*, *20*, 208–215.

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., et al. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences, USA*, *111*, 13790–13794. http://dx.doi.org/10.1073/pnas.1404623111.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*, 1467–1471. http://dx.doi.org/10.1126/science.1235488.

Rindermann, H., Hoang, Q. S. N., & Baumeister, A. E. E. (2013). Cognitive ability, parenting and instruction in Vietnam and Germany. *Intelligence*, *41*, 366–377.

The 1000 Genomes Project Consortium (2012). An integrated map of genetic variationfrom 1,092 human genomes. *Nature*, *491*, 56–65.

Thornhill, R., & Fincher, C. L. (2013). The comparative method in cross-cultural and cross-species research. *Evolutionary Biology*, *40*, 480–493.

Trampush, J. W., Lencz, T., Knowles, E., Davies, G., Guha, S., Pe'er, I., et al. (2015). Independent evidence for an association between general cognitive ability and a genetic locus for educational attainment. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*, 363–373. http://dx.doi.org/10.1002/ajmg.b.32319.

Turchin, M. C., Chiang, C. W., Palmer, C. D., Sankararaman, S., Reich, D., Genetic Investigation of, A.T.C, et al. (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics*, *44*, 1015–1019.

Ward, M. E., McMahon, G., St Pourcain, B., Evans, D. M., Rietveld, C. A., Benjamin, D. J., et al. (2014). Genetic variation associated with differential educational attainment in adults has anticipated associations with school performance in children. *PLoS ONE*, *9*, e100248. http://dx.doi.org/10.1371/journal.pone.0100248.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Woodley, M. A. (2011). The cognitive differentiation-integration effort hypothesis: A synthesis between the fitness indicator and life history models of human intelligence. *Review of General Psychology*, *15*, 228–245.

Woodley of Menie, M. A. (2015). How fragile is our intellect? Estimating losses in general intelligence due to both selection and mutation accumulation. *Personality and Individual Differences*, *75*, 80–84.

Woodley of Menie, M. A., Fernandes, H. B. F., Figueredo, A. J., & Meisenberg, G. (2015). By their words ye shall know them: Evidence of genetic selection against general intelligence and concurrent environmental enrichment in vocabulary usage since the mid 19th century. *Frontiers in Psychology*, *6*, 361.

Woodley, M. A., & Figueredo, A. J. (2013). *Historical variability in heritable general intelligence: Its evolutionary origins and socio-cultural consequences.* Buckingham, UK: Buckingham University Press.