

Holocene Selection for Variants Associated With General Cognitive Ability: Comparing Ancient and Modern Genomes

Michael A. Woodley of Menie,^{1,2,†} Shameem Younuskuju,³ Bipin Balan,⁴ and Davide Piffer^{5,†}

¹Unz Foundation, Palo Alto, CA, USA

²Center Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel, Brussels, Belgium

³Weill Cornell Medicine, Cornell University, Dohar, Qatar

⁴Department of Agriculture and Forestry Science, University of Palermo, Palermo Italy

⁵Department of Psychology, Ben Gurion University of the Negev, Beer-Sheva, Israel

Human populations living during the Holocene underwent considerable microevolutionary change. It has been theorized that the transition of Holocene populations into agrarianism and urbanization brought about culture-gene co-evolution that favored via directional selection genetic variants associated with higher general cognitive ability (GCA). To examine whether GCA might have risen during the Holocene, we compare a sample of 99 ancient Eurasian genomes (ranging from 4.56 to 1.21 kyr BP) with a sample of 503 modern European genomes ($F_{st} = 0.013$), using three different cognitive polygenic scores (130 SNP, 9 SNP and 11 SNP). Significant differences favoring the modern genomes were found for all three polygenic scores (odds ratios = 0.92, $p = 0.001$; .81, $p = 0.037$; and .81, $p = .02$ respectively). These polygenic scores also outperformed the majority of scores assembled from random SNPs generated via a Monte Carlo model (between 76.4% and 84.6%). Furthermore, an indication of increasing positive allele count over 3.25 kyr was found using a subsample of 66 ancient genomes ($r = 0.22$, $p_{one-tailed} = .04$). These observations are consistent with the expectation that GCA rose during the Holocene.

■ **Keywords:** 1000 Genomes dataset, ancient genomes, Bronze Age, directional selection, polygenic scores

The Holocene (11.7 kyr BP to the present) was a time of considerable microevolutionary change among human populations (Hawks et al., 2007). Based on various converging lines of molecular evidence, it has been estimated that the rate of adaptive evolution among these populations may have been 100 times greater than during the preceding Pleistocene (Cochran & Harpending, 2009; Frost, 2011; Hawks et al., 2007). Novel adaptations that are known to have arisen and spread during this period include lactase persistence (McCracken, 1971) and alterations in hemoglobin, permitting enhanced tolerance to diminished oxygen levels among populations living at high altitude (Bigham et al., 2010).

Other significant adaptations that may have been favored by selection during this period are the learning and problem-solving mechanisms that give rise to general cognitive ability (GCA). At the level of individual differences in cognitive performance, GCA is characterized by positive correlations among different measures of cognitive abilities (Jensen, 1998). GCA is highly heritable, with behavior

genetic twin studies indicating that between 22% and 88% of its variance, depending on age, may be due to the action of additive genetic effects (Bouchard, 2004). Consistent with this, studies employing genome-wide complex trait analysis genomic-relatedness-based restricted maximum-likelihood (GCTA GREML) have found that a significant amount of the variance in GCA can be directly attributed to large numbers of variants with individually small, but cumulatively large (additive) effects on the phenotype (Davies et al., 2011).

RECEIVED 20 April 2017; ACCEPTED 29 May 2017

ADDRESS FOR CORRESPONDENCE: Michael Woodley of Menie, Center Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel, Brussels, Belgium. E-mail: Michael.Woodley@vub.ac.be; Davide Piffer, Department of Psychology, Ben Gurion University of the Negev, Beer-Sheva, Israel. E-mail: piffer@post.bgu.ac.il

† Joint first author.

GCA is not just present in human populations, but is present (at lower levels) also in a variety of animal taxa, including chimpanzees and other primates, raccoons, mice, rats, and ravens (Galsworthy et al., 2014). The existence of these species' differences furthermore indicates that GCA has been under directional selection, particularly in the primate clade. This seems to be especially true in the case of abilities associated with tool use, which in primates are among the most strongly associated with GCA, in addition to being the most heritable. They are also associated with the strongest signals of recent directional selection (Fernandes et al., 2014; Woodley of Menie et al., 2015). This is consistent with the theoretical expectation (Geary, 2005) that GCA is an adaptation to coping with *evolutionarily novel* challenges—these being challenges that occur infrequently throughout the phylogeny of a lineage—thus cannot be solved with recourse to specialized cognitive systems adapted to coping with recurrent problems (such as those associated with cheater detection or language acquisition). Instead, such problems require generalized and open-ended problem-solving systems, such as learning and working memory, in order to tailor solutions to them. The ability to *innovate* a solution to such a problem (via the development of a tool) is a key manifestation of the action of these generalized problem-solving systems (Geary, 2005).

The Holocene is believed to have been an *environment of evolutionary relevance* for GCA as it was characterized by major transitions among human populations, especially in Eurasia, away from a hunter-gatherer mode of subsistence, towards a sedentary agriculture-based one and beyond that to urbanization (Cochran and Harpending, 2009; Hawks et al., 2007). This brought with it many evolutionarily novel problems, such as having to cope with increased population densities (Cochran and Harpending, 2009; Hawks et al., 2007). Innovations that played a major role in facilitating this transition would have included the domestication of cultivars and animals and the development of novel tools for raising the productivity of land (such as the plough; Cochran and Harpending, 2009). Cultural innovations such as monotheism, monarchy, aristocracy, feudalism, and currency-based economics arose in response to the need for coping with the hierarchical power distribution characteristic of large, static populations (Cochran and Harpending, 2009). Those populations that were successful in using innovations to solve novel problems would furthermore have had an advantage in warfare, being better able to innovate weapons and tactics, allowing them to replace less successful populations. Population growth would have increased the chances of rare GCA-enhancing mutations arising, which would have favored the aggregate fitness of those populations, permitting them to expand to the greatest extent (Cochran and Harpending, 2009; Darlington, 1970; Hawks et al., 2007). Thus, positive *culture-gene co-evolutionary feedback* occurring during

the Holocene might have rapidly increased GCA (Cochran and Harpending, 2009). Consistent with this, there is evidence that contemporary populations, which historically made the transition from a hunter-gatherer to an agrarian subsistence paradigm, have higher frequencies of working memory-enhancing genetic variants, compared to populations that never or only recently made this transition (Piffer, 2013).

High-resolution genome-wide association studies (GWAS) have identified a number of single nucleotide polymorphisms (SNPs) primarily associated with the development of the central nervous system that predict variance in both GCA and educational attainment—with which GCA shares approximately 60% of its linkage-pruned genetic variance (Davies et al., 2016; Domingue et al., 2015; Okbay et al., 2016; Rietveld et al., 2013, 2014; Selzam et al., 2017). These SNPs can be concatenated into *cognitive polygenic scores* (henceforth POLY_{COG}). The term 'cognitive' is being used here as a broader term than GCA, as it includes not only the cognitive ability measured by standardized tests but also other sources of variance that contribute to educational attainment. Polygenic scores are more or less normally distributed and can be used as a molecular index of GCA and educational attainment in standard regression-type analyses.

Recent advances in the recovery of ancient DNA have led to the accumulation of a wealth of genomic data on ancient human and other hominid populations living in the Holocene and prior epochs. These data permit comparisons to be made between ancient and modern human populations using POLY_{COG}, thus permitting the question of whether GCA rose during the Holocene to be tested directly. This analysis will here be conducted via the comparison of an ancient Eurasian genome sample from between 4.56 and 1.21 kyr BP with a modern, ancestrally matched European one.

Materials and Methods

Modern Genomes

Previous studies comparing ancient and modern genomes have utilized the Phase3 1000 Genomes dataset (1000 Genomes Project Consortium, 2015) as a modern reference set, on the basis that it is the most globally representative set of genomes currently available (Allentoft et al., 2015). The 2,504 samples in the Phase3 release are sourced from 26 populations, which can be categorized into five super-populations by continent. These are East Asian (EAS), South Asian (SAS), African (AFR), European (EUR), and American (AMR). These genomes were obtained in the form of VCF files, via the online 1000 Genomes FTP. The EUR sample is comprised of 503 individuals from five different countries (Finland, Great Britain, Italy, Spain, and the United States—specifically, north-western Europeans from Utah).

Ancient Genomes

Data on the ancient Holocene genomes were obtained from the European Nucleotide Archive (Leinonen et al., 2011) in the form of BAM files (accession number PRJEB9021). These genomes, which have been utilized in previous research to analyze the spread of variants associated with skin and eye color, among other traits (Allentoft et al., 2015), were aged between 4.56 and 1.21 kyr ($M = 3.44$ kyr, $SD = 0.62$ kyr), covering the early Bronze to early Iron Ages. Their genome-wide coverage is relatively low (0.01–7.4X average depth, overall average = 0.7X; Allentoft et al., 2015). There are 102 genomes in total; however, two samples (RISE507 and RISE508) were sourced from the same individual (Allentoft et al., 2015). The total allele count was therefore calculated employing the full sample, alternately using each version of the genome—the results of the two analyses were then averaged. The majority of the ancient genomes were sourced from regions that are presently considered to be parts of Europe, the remainder being sourced from Central Asia. Late Bronze Age European and Central Asian gene pools resemble present-day Eurasian genetic structure (Allentoft et al., 2015). Consistent with this, the F_{st} values range from 0 to 0.08, when present-day 1000 Genomes EUR are compared with the various populations comprising the Ancient samples, indicating little to modest genetic differentiation (little differentiation corresponds to an F_{st} range of 0 to 0.05, and modest to an F_{st} range of 0.05 to 0.15; Hartl & Clark, 1989). These values are lower than the distance between present-day Europeans and East Asians ($F_{st} = 0.11$; Allentoft et al., 2015). The two ancient genomes belonging to the Siberian Okunevo culture (RISE515 and RISE516) were somewhat exceptional however, exhibiting modest differentiation relative to the EUR sample when compared with the sample size and coverage weighted average F_{st} value computed for the additional 92 ancient genomes for which these data were available ($F_{st} = 0.074$ vs. 0.013 for the remainder of the sample). On this basis, they were removed from the final ancient sample in order to reduce the genetic differentiation between the ancient and modern samples. The final sample of ancient genomes was therefore comprised of 99 genomes, sourced from sites located in present-day Armenia (8%), Czech Republic (6%), Denmark (6%), Estonia (1%), Germany (10%), Hungary (10%), Italy (3%), Kazakhstan (1%), Lithuania (1%), Montenegro (2%), Poland (7%), Russia (36%), and Sweden (8%).

Reference Genome

Paired-end reads are aligned to the reference human genome obtained (in the form of a FASTA file) from the UCSC database (GRCh37/hg19).

Read Realignment and Base Recalibration

After removing duplicate reads, the reads were realigned around the known indels from the 1000 Genome sample us-

ing the GenomeAnalysisTKLite-2.3-9 (GATK) toolkit. The known indels set was obtained from the GATK resource page (McKenna et al., 2010). After performing realignment, the base recalibration step is performed (DePristo et al., 2011; Van der Auwera et al., 2013). Known variant position is taken into account to recalibrate the quality score.

Variant Calling and Computing POLY_{COG}

After performing realignment, the GenomeAnalysisTKLite-2.3-9 toolkit UnifiedGenotyper was used to identify the SNPs and short indels. The positive allele counts (positive here refers to the sign on the β value derived from GWAS regressions—indicating that they increase the trait of interest) and allele frequencies of the target SNPs were calculated from the Ancient Genome VCF files and also from the 1000 Genomes VCF files using a custom-made PERL script (available on request). Variant calling was used to construct the first POLY_{COG} score from a list of 162 SNPs that were found to have high genome wide significance ($p \leq 5 \times 10^{-8}$) as ‘hits’ for educational attainment in a recent large-scale, meta-analytic GWAS study (Okbay et al., 2016). After controlling for ascertainment bias by excluding those variants that were completely absent from the ancient genomes (due to the ancient samples having approximately 30% of the 1000 Genomes coverage) the resultant POLY_{COG} score consisted of 130 ‘hits’ that were present to some degree in both the ancient and modern genomes.

A second POLY_{COG} score was constructed using data from another GWAS study of educational attainment and related phenotypes (Davies et al., 2016). In this study, 1,115 SNPs reached GWAS significance, of which 15 were independent (linkage pruned) signals. Four SNPs were absent from the ancient genomes; thus, a POLY_{COG} was calculated using the remaining 11 SNPs. A third POLY_{COG} score was constructed utilizing a technique developed in Piffer (2015, 2017; cf. Woodley of Menie, Piffer et al., 2016), employing only the nine GWAS hits from Okbay et al. (2016) that were in close linkage disequilibrium (linkage cut-off $r \geq 0.8$, 500 kb linkage window) with ‘hits’ predicting educational attainment across two other large GWAS studies (Davies et al., 2016; Rietveld et al., 2013). Linkage was determined using the NIH LDLink program (Machiela & Chanock, 2015) with the 1000 Genomes Phase3 CEU population as a reference group. These are presented in Table 1.

By focusing on only the smaller numbers of linked replicates, the third POLY_{COG} score was developed as a means of negating the likely high numbers of false positive SNPs present in larger POLY_{COG} scores. This approach is based on maximizing the signal of directional selection by focusing on the GWAS hits with the highest significance (or replication rate across databases), as opposed to maximizing the amount of variance explained, which instead is the aim of the classical within-population GWAS design. While it is true that the effect of selection on a phenotype is

TABLE 1
Nine Linked Replicated SNPs With Linkage Disequilibrium (D') and R^2 Coefficients

Publication	Index SNP	Publication SNP	Linked SNP source	D'	R^2
Davies et al. (2016)	rs12042107_C	rs1008078_C	Okbay et al. (2016)	1	0.8
Rietveld et al. (2013)	rs11584700_G	rs11588857_A	Okbay et al. (2016)	1	0.94
Rietveld et al. (2013)	rs4851266_T	rs12987662_A	Okbay et al. (2016)	1	1
Davies et al. (2016)	rs13086611_T	rs148734725_A	Okbay et al. (2016)	1	1
Davies et al. (2016)	rs11130222_A	rs11712056_T	Okbay et al. (2016)	1	0.98
Davies et al. (2016)	rs55686445_C	rs62263923_G	Okbay et al. (2016)	1	0.98
Davies et al. (2016)	rs12553324_G	rs13294439_C	Okbay et al. (2016)	1	0.98
Davies et al. (2016)	rs4799950_G	rs12969294_G	Okbay et al. (2016)	1	0.92
Rietveld et al. (2013)	rs9320913_A	rs9320913_A	Okbay et al. (2016)	1	1

proportional to the additive genetic variance, the selection coefficient for a given SNP is proportional to the amount of variance explained by that SNP. Since the rate of allele frequency change is proportional to the selection coefficient, and because both GWAS data and evolutionary processes are notoriously very noisy, it is important to maximize signal by limiting the analysis to the SNPs with higher GWAS significance, which are also more likely to account for phenotypic variance. This is the same reason why detecting selection on phenotypes that are influenced by a few genes with large effects can be accomplished with traditional, less sensitive tools (e.g., F_{st} index) than detecting a signal of polygenic selection (Berg & Coop, 2014).

Analysis

All statistical analyses were conducted using R v. 3.3.2 (R Core Team, 2016). In order to test the hypothesis that POLY_{COG} had different proportions of the positive and negative alleles between the two populations against the null hypothesis that there would be no differences, we used Fisher's Exact and chi-squared tests, since the data are binary (positive vs. negative allele/ancient vs. modern populations). At the same time, the contingency table dealt with unequal representation of genomes for each SNP within the ancient group (i.e., some SNPs were called if they were present in only one individual) by incorporating this information (each SNP is assigned a weight proportional to the number of genomes in which the SNP was called); odds ratios (OR) computed utilizing Fisher's Exact Test were preferred over the chi-squared test because it allows exact estimation of the likelihood of the null hypothesis, rather than relying on an approximation. An alternative to Fisher's Exact test is the log-likelihood ratio G -test, which is employed here as an additional test of significance. This was implemented using the R package DescTools (Signorell, 2016). Additional analyses were also performed on the two smaller POLY_{COG} (11 and 9 SNPs) using two unconditional tests (considered more powerful alternatives); specifically, Barnard's (1945, 1947) and Boschloo's (1970) Exact tests, utilizing the R package Exact (Calhoun, 2016).

In constructing POLY_{COG} for conventional GWAS-type analyses, it is common practice to weight each SNP by its associated GWAS regression β value, as these polygenic

scores are typically the sum of very large numbers of SNPs with highly heterogeneous effect sizes as predictors of the phenotype of interest (e.g., Okbay et al., 2016). These regression weights were not found to significantly correlate with the frequency differences between modern and ancient populations for the largest of our POLY_{COG}, however (130 SNP; $r = -0.128$, $p = .07$), likely because selecting only the SNPs with the highest p values introduced range restriction among the regression weights. Therefore, given that the SNPs comprising our POLY_{COG} could be considered to be roughly 'equally good' as predictors of the trait of interest, the inclusion of regression β in the analysis as weights is therefore unlikely to add additional power, which in turn justifies the use of a more straightforward and also parsimonious binarization approach in comparing the ancient and modern genomes.

A second analysis was carried out on a subsample of 66 ancient genomes for which sample age was known (age range: 1.21–4.46 kyr). The correlation was computed between individual positive allele counts derived using the 130 SNP POLY_{COG} with genome age (ascertained by utilizing radiocarbon dating), in order to test for change over time. Since on average the ancient samples had approximately 30% of the 1000 Genomes coverage, carrying out this analysis using the smaller POLY_{COG} would add too much noise and produce unreliable estimates for individual genomes. Two additional genomes for which age data were present (RISE413 and RISE42) were identified as having significantly outlying values of positive allele count based on the outlier-labeling rule (Hoaglin et al., 1986) because there was only one target SNP present per genome. For the analysis only genomes whose coverage was limited to two or more of the 130 SNPs were retained. Removing these two outlying genomes also reduced the skew of the regression residual from 1.16 to 0.698, making the remaining dataset more amenable to parametric regression.

Random SNP Polygenic Scores

As an additional robustness test, a method developed by Piffer (2017) was employed, in which the Monte Carlo model was run using randomly drawn and non-overlapping sets of SNPs matched for minor allele frequency (MAF). The expectation is that the three POLY_{COG} should

TABLE 2

Two by Two Contingency Tables With Positive and Negative GWAS Effect Allele Counts for Ancient and Modern Genomes for Each of the Three POLY_{COG}, Along with Fisher's Exact Test OR and G-Test Log-Likelihood Ratios

POLY _{COG}	Genome grouping	Positive allele count	Negative allele count	Fisher's Exact Test OR; 95% CI; p-value	G-test; p value
Nine linked-hits POLY _{COG} . N = 9	Ancient genome	177	273	0.81 [0.665, 0.990]; p = .037	G = 4.445; p = .035
	Modern genome	4,017	5,037		
Pooled analysis POLY _{COG} (Okbay et al., 2016). N = 130	Ancient genome	3298.5	3997.5	0.92 [0.882, 0.969]; p = .001	G = 10.486; p = .001
	Modern genome	61,666	69,114		
Significant hits POLY _{COG} (Davies et al., 2016). N = 11	Ancient genome	268.5	282.5	0.81 [0.683, 0.969]; p = .02	G = 5.536; p = .019
	Modern genome	5,960	5,106		

Note: The non-integer allele counts result from averaging the genomes of two samples (RISE507 and RISE508) that came from the same individual. The samples had different allele counts likely resulting from different coverage.

outperform the majority of random polygenic scores in terms of revealing a frequency difference, favoring modern populations, consistent with theory. The random polygenic scores were computed for non-overlapping sets of SNPs for the modern and ancient populations and for each allele (A, B). A 2×2 matrix was computed for each comparison (Allele A/Ancient population, Allele A/Modern population, Allele B/Ancient population, Allele B/Modern population). As with the POLY_{COG}, OR for each matrix were then computed. Finally, the proportion of ORs with a weaker effect (i.e., higher OR value) than what was observed for the three POLY_{COG} was computed. Frequency matching was carried out using SNPSnap (Pers et al., 2015), by entering the nine linked replicate SNPs, which it must be recalled are likely the closest SNPs to the true causal variants, and setting LD at $r^2 < 0.1$ (for EUR). This resulted in a total of 7,369 SNPs being retrieved.

Results

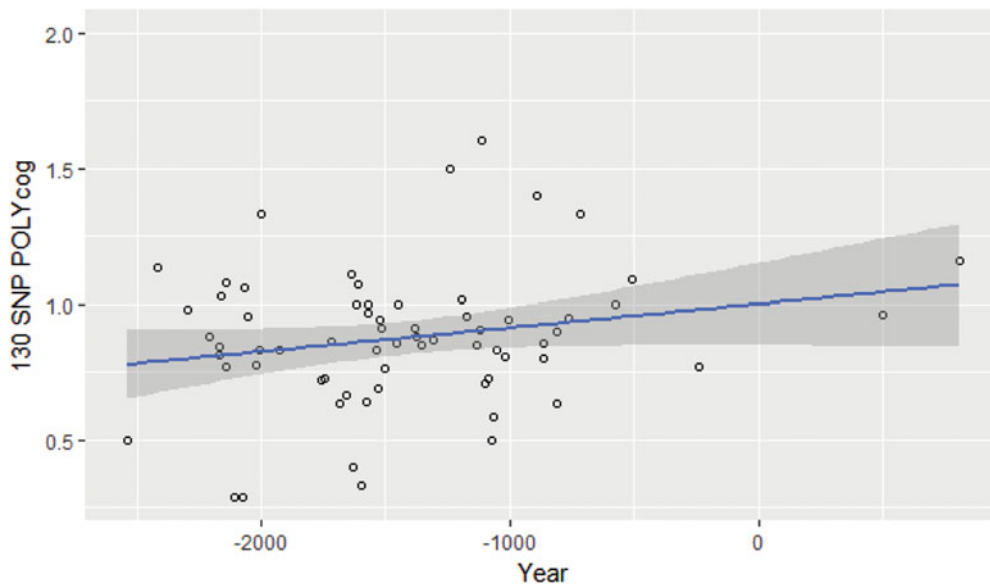
A 2×2 contingency table was created for each of three alternative POLY_{COG} assigning allele counts by GWAS effect status (positive or negative effect) to ancient and modern genomes. OR of the association between positive or negative alleles and ancient or modern populations were computed and are reported in Table 1. As Fisher's Exact and chi-squared tests revealed nearly identical results, only the former are reported. In addition, the G-test (log-likelihood ratio) is reported as an alternative test of the significance of the difference. As a further robustness test, the difference in POLY_{COG} between the two sets of genomes was examined using the unconditional Barnard's and Boschloo's Exact tests. These statistics are extremely computationally intensive when large numbers of pair-wise comparisons are employed; therefore, they could only be used for the two smaller POLY_{COG} (9 and 11 SNPs). The results were identical to those obtained via the conditional techniques.

The Monte Carlo model was run with 6,740 random SNPs (after removal of the ones absent from the ancient genome in order to control for ascertainment bias). Hence, 748, 613, and 52 random SNP polygenic scores (corresponding to the 9, 11, and 130 SNPs POLY_{COG}, respectively) were computed. The POLY_{COG} outperformed the random polygenic scores (producing lower OR values) in the majority of instances (9 SNPs = 76.4% [572 out of 749 draws]; 130 SNPs = 84.6% [44 out of 52 draws]; 11 SNPs = 77.8% [477 out of 613 draws]), which indicates a low probability that the results presented in Table 2 are due to chance.

These results indicate that modern European genomes have higher POLY_{COG} relative to ancient ones (sourced from Europe and Central Asia).

A second analysis was conducted to investigate the association between POLY_{COG} and sample year within the subset of 66 ancient genomes for which radiocarbon age estimates were available (from Allentoft et al., 2015). If selection is operating on these variants throughout the 3.25 kyr covered by this sample, then less ancient genomes should have higher positive allele counts relative to more ancient ones. The correlation of positive allele frequency with the sample year (scaled using the BCE/CE calendar era) was in the expected positive direction (samples from more recent years had higher positive allele counts); the result is furthermore statistically significant when a one-tailed test is used ($r = 0.217$, 95% CI [-0.026, 0.043], $p_{\text{one-tailed}} = .04$), which is theoretically justified on the basis that the direction of the effect was anticipated (Kimmel, 1957). The scatter plot is presented in Figure 1.

Coverage was not correlated with either genome age ($r = 0.022$) or positive allele count ($r = -0.054$). Additionally, a multiple linear regression was run, with positive allele count along with coverage and age as dependent and independent variables respectively. This did not alter the effect size compared to the sample without the outliers for models with and without coverage controlled ($\beta = 0.209$ vs.

**FIGURE 1**

(Colour online) Scatter plot of 130 SNP POLY_{COG} positive allele counts for 66 ancient genomes as a function of year (scaled in terms of the BCE/CE calendar eras). Note: $r = 0.217$, $p_{\text{one-tailed}} = .04$, $N = 66$. Gray area around trend line corresponds to the 95% CI.

0.211). The regression models were run with each duplicate genome (RISE508 and RISE507) separately, the resultant β values were then averaged.

Discussion

The Holocene appears to have been an environment of evolutionary relevance for GCA, as modern European genomes have higher POLY_{COG} relative to those sourced (predominantly) from Bronze Age Europe and Central Asia. A Monte Carlo model employing polygenic scores comprised of non-overlapping sets of random SNPs revealed that the three POLY_{COG} outperformed between 76.4% and 84.6% of random polygenic scores. It must be noted also that a major strength of our analysis is that it used real allele frequency data, as opposed to traditional Monte Carlo simulation-based approaches, which rely on computer-generated (virtual) data.

A second analysis revealed that within a subset of the ancient genome sample, there was a (one-tailed) significant trend toward increased positive allele counts over time. As was mentioned in the section *Introduction*, these SNPs are assumed to exert their effects on educational attainment mainly through GCA, but also through other traits relevant for educational attainment. While the increase in these variants over time is certainly consistent with the expectation of rising GCA, the possibility that their increase indicates a simultaneous rise in other factors that make unique contributions to educational attainment (such as 'slow' life history or 'high-*K*' social cognitive characteristics; Giosan, 2006) cannot be ruled out.

This increase in POLY_{COG} can realistically be accounted for in two different ways: (1) microevolutionary selection pressures on standing genetic variation, resulting in soft sweeps (Hermisson & Pennings, 2005; Orr & Betancourt, 2001), and (2) population expansion, replacement, and admixture due to migration.

The idea that selection takes many millennia to produce noticeable changes in a population is at odds with data indicating that within human populations, changes in the means of heritable traits such as height, weight, cholesterol levels, blood pressure, age at first birth, and age at menopause can occur due to selection over a relatively short period of time (Stearns et al., 2010). Selection operating over the course of millennia (as in the present case) would be expected to produce quite considerable microevolutionary change. As was discussed in the section *Introduction*, Holocene populations appear to have undergone accelerated adaptive microevolution relative to those living in the Pleistocene (Cochran & Harpending, 2009; Hawks et al., 2007). Increasing cultural complexity and technological sophistication among these populations may therefore have arisen in part from selection favoring GCA. Cultural and technological change can in turn create, via culture-gene co-evolutionary feedback, conditions favoring higher GCA (Cochran & Harpending, 2009; Piffer, 2013). This process likely continued until the Late Modern Era, where it has been noted that among Western populations living between the 15th and early 19th centuries, those with higher social status (which shares genetic variance with, and is therefore a proxy for GCA; Trzaskowski et al., 2014) typically produced the most surviving offspring. These in turn tended

toward downward social mobility due to intense competition, replacing the reproductively unsuccessful low-status stratum and effectively 'bootstrapping' those populations via the application of high levels of skill to solving problems associated with production and industry, eventually leading to the Industrial Revolution in Europe (Clark, 2007, 2014). The millennia-long microevolutionary trend favoring higher GCA not only ceased, but likely went into reverse among European-derived populations living in the 19th century (Lynn, 1996; Lynn & Van Court, 2004), largely in response to factors such as the asymmetric use of birth control and prolonged exposure to education among those with high GCA (Lynn, 1996). Consistent with this, it has been found that various POLY_{COG} negatively predict reproductive success in contemporary Western populations (Beauchamp, 2016; Conley et al., 2016; Kong et al., 2017; Woodley of Menie, Schwartz et al., 2016). It is important to note that this recent microevolutionary trend (working in the opposite direction) has likely attenuated the difference in POLY_{COG} between the modern and ancient genomes noted in the present study.

Changes in allele frequencies can also occur via population expansion and replacement. The end of the Last Glacial Maximum initiated major cultural changes, leading to the development of agriculture, population growth, and the subsequent Neolithization process (8–5 kyr BP), which involved the spread of genes as well as culture from the Middle East (Sokal et al., 1991). Another big cultural change associated with major population movements was the Bronze Age, whose culture started replacing the Neolithic farming cultures in temperate Eastern Europe 5 kyr BP (Allentoft et al., 2015). This Pontic-Caspian Steppe genetic component among contemporary Europeans is associated with the Corded Ware and Yamnaya peoples, and possibly accounts for the spread of Indo-European languages (Allentoft et al., 2015). Being potentially highly fitness enhancing, admixture among populations may have furthermore led to the spread of the causal variants that are tapped by POLY_{COG}. However, recent evidence suggests that the European and Central Asian gene pools toward the end of the Bronze Age closely mirror present-day West Eurasian genetic structure (Allentoft et al., 2015). Hence, population movements and replacements likely played a larger role in the evolution of GCA prior to the period covered by the present sample. Sicily and Sardinia, and the southern fringes of Europe in general are notable exceptions however, as they tend to cluster genetically with Neolithic farmers (Allentoft et al., 2015).

The first theoretical scenario (microevolutionary change involving standing genetic variation) most likely therefore played the largest role in the POLY_{COG} increase during the time period covered by our samples.

It has been noted that concomitant in time with the apparent increase in POLY_{COG} observed presently is an apparent parallel selection trend favoring smaller brains (Hawks,

2011). These trends may at first appear to be contradictory, as GCA is positively associated with brain volume. It must be noted, however, that meta-analysis reveals that the phenotypic association is weak ($\rho = .24$; Pietschnig et al., 2015) with the genetic correlation being slightly stronger ($r_g \sim 0.30$; Okbay et al., 2016), indicating that the majority of the phenotypic and genotypic variance in brain volume is unrelated to variance in GCA. Furthermore, studies utilizing POLY_{COG} have found that it does not predict variation in brain volume, despite both POLY_{COG} and brain volume making independent contributions to GCA (Deary et al., 2016). These findings suggest that declining brain volume during the Holocene may have been a consequence of enhanced brain efficiency stemming from increased corticalization and neuronal connectivity, with more bioenergetically optimized brains simply requiring less mass to achieve greater processing power. It may therefore be this process that the increase in POLY_{COG} is tracking in our samples. A prediction stemming from Hawks (2011) is that variants that predict brain volume and not GCA should show the opposite trend in time to POLY_{COG} when examined in the context of the present samples.

A potential limitation of the present work is linkage disequilibrium (LD) decay. This takes place when a pair of SNPs on a chromosome in a population moves from linkage disequilibrium to linkage equilibrium over time, due to recombination events eventually occurring between every possible point on the chromosome (Bush & Moore, 2012). LD decay can be an issue when comparing polygenic scores computed using genetically different populations (i.e., either across time or space). Since most GWAS hits are actually tag SNPs, decay in LD implies that the causal SNPs will be less efficiently flagged by the tag SNPs, resulting in the tag SNPs coming to resemble a sample of random SNPs. Since the frequency of the average SNP allele is 50%, the GWAS hits will tend to converge towards an average frequency of 50%, with increasing LD decay. The implication of this for our analysis is that our estimates of the difference in POLY_{COG} between the reference population (modern European) and the ancient population are therefore conservative when the mean frequency is lower than 50%, while the opposite is true when the frequency is higher (the mean frequencies in the modern 1000 Genomes samples were the following: for the 130 SNP POLY_{COG} = 47%, 9 SNP POLY_{COG} = 44%, and the 11 SNP POLY_{COG} = 53%). Furthermore, because GCA likely involves around 10,000 causal variants (Hsu, 2014), and as there is relatively little genetic differentiation between the ancient and modern genomes, as indicated by the sample size and coverage weighted F_{st} value of 0.013, the results of simulations indicate that LD decay is ultimately expected to be minimal (Scutari et al., 2016) and should not be confounding the results.

The present effort constitutes a direct attempt to show the theoretically anticipated change in POLY_{COG} over

time utilizing relatively simple methodology, which makes few modeling assumptions. The difference in POLY_{COG} between the ancient and modern genomes is indifferent to the use of alternate versions of POLY_{COG} and to the use of various statistical techniques for quantifying the significance of the group difference. The results of the Monte Carlo model indicate that the majority of frequency-matched random SNP polygenic scores underperform relative to the POLY_{COG}. Among a subsample of the ancient genomes, indications of increasing positive allele counts as a function of time are also present. It is acknowledged that there are limitations in the data employed—larger samples of ancient genomes would certainly increase confidence in the present finding. Furthermore, it is acknowledged that any of the discrete lines of evidence that have been brought to bear on the question can, in isolation, of course be criticized, with some lines being considered weaker than others. It must be noted, however, that multiple operationalizations of the variable of interest and multiple methods have been employed here, which, even if each individual line of evidence is to be considered ‘weak’, should nevertheless greatly increase confidence in our conclusion when the evidence is considered collectively as part of a larger nomological network (Cronbach & Meehl, 1955).

References

- 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526, 68–74.
- Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., ... Willerslev, E. (2015). Population genomics of Bronze Age Eurasia. *Nature*, 522, 167–174.
- Barnard, G. A. (1945). A new test for 2×2 tables. *Nature*, 156, 783–784.
- Barnard, G. A. (1947). Significance tests for 2×2 tables. *Biometrika*, 34, 123–138.
- Beauchamp, J. P. (2016). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 7774–7779.
- Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics*, 10, e1004412.
- Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., ... Shriver, M. D. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genetics*, 6, e1001116.
- Boschloo, R. D. (1970). Raised conditional level of significance for the 2×2 -table when testing the equality of two probabilities. *Statistica Neerlandica*, 24, 1–35.
- Bouchard, T. J. Jr. (2004). Genetic influence on human psychological traits – A survey. *Current Directions in Psychological Science*, 13, 148–151.
- Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Computational Biology*, 8, e1002822.
- Calhoun, P. (2016). Exact: unconditional exact test. R Package Version 1.7. Retrieved from <https://CRAN.R-project.org/package=Exact>
- Clark, G. (2007). *Farewell to alms: A brief economic history of the world*. Princeton, NJ: Princeton University Press.
- Clark, G. (2014). *The son also rises: Surnames and the history of social mobility*. Princeton, NJ: Princeton University Press.
- Cochran, G., & Harpending, H. (2009). *The 10,000 year explosion: How civilization accelerated human evolution*. New York, NY: Basic Books.
- Conley, D., Laidley, T., Belsky, D. W., Fletcher, J. M., Boardman, J. D., & Domingue, B. J. (2016). Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 6647–6652.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Darlington, C. D. (1970). *The evolution of man and society*. New York, NY: Simon & Schuster.
- Davies, G., Marioni, R. E., Liewald, D. C., Hill, W. D., Hagenaars, S. P., ... Deary, I. J. (2016). Genome-wide association study of cognitive functions and educational attainment in UK Biobank ($N = 112151$). *Molecular Psychiatry*, 21, 758–767.
- Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., ... Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, 16, 996–1005.
- Deary, I. J., Cox, S. R., & Ritchie, S. J. (2016). Getting Spearman off the skyhook: One more in a century (since Thomson, 1916) of attempts to vanquish g. *Psychological Inquiry*, 27, 192–199.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491–498.
- Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M., & Boardman, J. D. (2015). Polygenic influence on educational attainment. *AERA Open*, 1, e13.
- Fernandes, H. B. F., Woodley, M. A., & te Nijenhuis, J. (2014). Differences in cognitive abilities among primates are concentrated on G: Phenotypic and phylogenetic comparisons with two meta-analytical databases. *Intelligence*, 46, 311–322.
- Frost, P. (2011). Human nature or human natures? *Futures*, 43, 740–748.
- Galsworthy, M. J., Arden, R., & Chabris, C. F. (2014). Animal models of general cognitive ability for genetic research into cognitive functioning. In D. Finkel & C. A. Reynolds (Eds.), *Behavior genetics of cognition across the lifespan* (pp. 257–278). New York, NY: Springer.
- Geary, D. C. (2005). *The origin of mind: Evolution of brain, cognition, and general intelligence*. Washington, DC: American Psychological Association.

- Giosan, C. (2006). High-K strategy scale: A measure of the high-K independent criterion of fitness. *Evolutionary Psychology*, 4, 394–405.
- Hartl, D. L., & Clark, A. G. (1989). *Principles of population genetics* (2nd ed.). Sunderland, MA: Sinauer Associates.
- Hawks, J. (2011). Selection for smaller brains in Holocene human evolution. *arXiv*: 1102.5604.
- Hawks, J., Wang, E. T., Cochran, G. M., Harpending, H. C., & Moyzis, R. K. (2007). Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 20753–20758.
- Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169, 2335–2352.
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81, 991–999.
- Hsu, S. D. H. (2014). On the genetic architecture of intelligence and other quantitative traits. *arXiv*: 1408.3421.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kimmel, H. D. (1957). Three criteria for the use of one-tailed tests. *Psychological Bulletin*, 54, 351–353.
- Kong, A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., ... Daly, M. J. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E727–E732.
- Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cedeno-Tárraga, A., ... Cochrane, G. (2011). The European Nucleotide Archive. *Nucleic Acids Research*, 39, D28–D31.
- Lynn, R. (1996). *Dysgenics: Genetic deterioration in modern populations*. Westport, CT: Praeger.
- Lynn, R., & Van Court, M. (2004). New evidence of dysgenic fertility for intelligence in the United States. *Intelligence*, 32, 193–201.
- Machiela, M. J., & Chanock, S. J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, 31, 3555–3557.
- McCracken, R. D. (1971). Lactase deficiency: An example of dietary evolution. *Current Anthropology*, 12, 479–517.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303.
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., ... Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533, 539–542.
- Orr, H. A., & Betancourt, A. J. (2001). Haldane's sieve and adaptation from the standing gene variation. *Genetics*, 157, 875–884.
- Pers, T. H., Timshel, P., & Hirschhorn, J. N. (2015). SNPsnap: A web-based tool for identification and annotation of SNPs. *Bioinformatics*, 31, 418–420.
- Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience & Biobehavioral Reviews*, 57, 411–432.
- Piffer, D. (2013). Correlation of the COMT Val158Met polymorphism with latitude and a hunter-gatherer lifestyle suggests culture-gene coevolution and selective pressure on cognition genes due to climate. *Anthropological Science*, 121, 161–171.
- Piffer, D. (2015). A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence*, 53, 43–50.
- Piffer, D. (2017). Evidence for recent polygenic selection on educational attainment and underlying cognitive abilities inferred from GWAS hits: A Monte Carlo simulation using random SNPs. *Preprints 2017*, 2017010127, doi:10.20944/preprints201701.0127.v3.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., ... Koellinger, P. D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 13790–13794.
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., ... Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 340, 1467–1471.
- Scutari, M., Mackay, I., & Balding, D. (2016). Using genetic distance to infer the accuracy of genomic prediction. *PLOS Genetics*, 12, e1006288.
- Selzam, S., Krapohl, E., von Stumm, S., O'Reilly, P. F., Rimfeld, K., ... Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry*, 22, 267–272.
- Signorell, A. (2016). DescTools: Tools for descriptive statistics. R package version 0.99.18. Retrieved from <https://cran.r-project.org/package=DescTools>
- Sokal, R. R., Oden, N. L., & Wilson, C. (1991). Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature*, 351, 143–145.
- Stearns, S. C., Byars, S. G., Govindaraju, D. R., & Ewbank, D. (2010). Measuring selection in contemporary human populations. *Nature Reviews Genetics*, 11, 611–622.
- Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., ... Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence*, 42, 83–88.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., ... DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43, 11.10.1–11.10.33.

Woodley of Menie, M. A., Fernandes, H. B. F., & Hopkins, W. D. (2015). The more g-loaded, the more heritable, evolvable, and phenotypically variable: Homology with humans in chimpanzee cognitive abilities. *Intelligence*, *50*, 159–163.

Woodley of Menie, M. A., Piffer, D., Peñaherrera, M. A., & Rindermann, H. (2016). Evidence of contemporary poly-

genic selection on the *Big G* of national cognitive ability: A cross-cultural sociogenetic analysis. *Personality & Individual Differences*, *102*, 90–97.

Woodley of Menie, M. A., Schwartz, J. A., & Beaver, K. M. (2016). How cognitive genetic factors influence fertility outcomes: A mediational SEM analysis. *Twin Research & Human Genetics*, *19*, 628–637.
