bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A Genome Scan for Genes Underlying Adult Body Size Differences between Central African Pygmies and their Non-Pygmy Neighbors

Trevor J. Pemberton[1,#,*], Paul Verdu[2,#,*], Noémie S. Becker[3], Cristen J. Willer[4], Barry S. Hewlett[5], Sylvie Le Bomin[2], Alain Froment[6], Noah A. Rosenberg[7,*], and Evelyne Heyer[2,*]

[1]Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada; [2]CNRS-MNHN, Université Paris Diderot, Sorbonne Paris Cité, UMR 7206 Eco-Anthropologie et Ethnobiologie, Paris, France; [3]Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany; [4]Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, Michigan, USA; [5]Department of Anthropology, Washington State University, Vancouver, Washington, USA; [6]IRD–MNHN, UMR 208 Patrimoines locaux, Paris, France; [7]Department of Biology, Stanford University, Stanford, California, USA.

[#]These authors contributed equally to this work.

[*]Authors to whom correspondence should be addressed.

E-mail addresses:

TJP: Trevor.Pemberton@umanitoba.ca
PV: verdu@mnhn.fr
NSB: nbecker@bio.lmu.de
CJW: cristen@umich.edu
BSH: hewlett@vancouver.wsu.edu
SLB: sylvie.le-bomin@mnhn.fr
AF: froment@mnhn.fr
NAR: noahr@stanford.edu
EH: heyer@mnhn.fr

Running title: Genetics of Body Size in Central African Pygmies

Keywords: Adaptation; Admixture; Genetic Association Studies; Population Genetics; Single-Nucleotide Polymorphism

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Background: Central African hunter-gatherer Pygmy populations have reduced body size compared with their often much larger agricultural non-Pygmy neighbors, potentially reflecting adaptation to the anatomical and physiological constraints of their lifestyle in tropical rainforests. Earlier studies investigating the genetics of the pygmy phenotype have focused on standing height, one aspect of this complex phenotype that is itself a composite of skeletal components with different growth patterns. Here, we extend the investigations of standing height to the variability and genetic architecture of sitting height and subischial leg length as well as body mass index (BMI) in a sample of 406 unrelated West Central African Pygmies and non-Pygmies.

Results: In addition to their significantly reduced standing height compared with non-Pygmies, we find Pygmies to have significantly shorter sitting heights and subischial leg lengths as well as higher sitting/standing height ratios than non-Pygmies. However, while male Pygmies had significantly lower BMI compared with male non-Pygmies, the BMI of females were instead similar. Consistent with prior observations with standing height, sitting height and subischial leg length were strongly correlated with inferred levels of non-Pygmy genetic admixture while BMI was instead weakly correlated, likely reflecting the greater contribution of non-genetic factors to the determination of body weight compared with height. Using 196,725 SNPs on the Illumina Cardio-MetaboChip with genotypes on 358 Pygmy and 169 non-Pygmy individuals together with single- and multi-marker association approaches, we identified a single genomic region and seven genes associated with Pygmy/non-Pygmy categorization as well as 9, 10, 9, and 10 genes associated with standing and sitting height, sitting/standing height ratio, and subischial leg length, respectively. Many of the genes identified have putative functions consistent with a role in determining their associated trait as well as the complex Central African pygmy phenotype.

Conclusions: These findings highlight the potential of modestly sized datasets of Pygmies and non-Pygmies to detect biologically meaningful associations with traits contributing to the Central African pygmy phenotype. Moreover, they provide new insights into the phenotypic and genetic bases of the complex pygmy phenotype and offer new opportunities to facilitate our understanding of its complex evolutionary origins.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Background

Central African hunter-gatherer populations have historically been called "Pygmies" in reference to their proportionally reduced body size compared with their agricultural non-Pygmy neighbors [1]. Among Pygmy males, mean adult standing height ranges from 142 cm among the Efe from eastern Democratic Republic of Congo (DRC; 142 cm) to 161 cm among the Twa from western DRC (161 cm), averaging 150.6 cm across 23 Central African Pygmy populations (standard deviation [SD]=6.7 cm) [2,3]. This average is notably shorter than the average adult standing height of 167.3 cm among males from 252 Sub-Saharan non-Pygmy populations (SD=5.7 cm) [3]. Nevertheless, no discontinuity in adult standing height exists among Pygmies and non-Pygmies, and contemporary populations are categorized as Pygmy based on both cultural criteria and adult standing height [1,3-7].

It has been hypothesized that the diminutive body size of Central African Pygmies is the outcome of adaptive processes in response to hunter–gatherer lifestyles in tropical rainforests with high levels of pathogen exposure [8]. Hypotheses for the basis of adaptation have considered morphological adaptation to thermoregulation in a hot and humid environment [9], metabolic costs of hunting and gathering in an environment in which food is scarce [10-12], and an early onset of reproduction in a context of high mortality rates [13]. Physiological studies in both West- and East-Central African Pygmies have found altered glucose homeostasis, insulin secretion, and free fatty-acid profiles in the presence of normal human growth hormone (HGH) levels [14,15]. In addition, reduced growth rates during the first two years of life [16] as well as adolescence [17,18] compared with non-Pygmies have been reported, possibly reflecting perturbation of the insulin-like growth factor 1 (IGF1) [19,20] and HGH [21] receptor signaling pathways [22]. Limitations on physiological data, however, together with a paucity of demographic, epidemiological, and paleoanthropological data for Central Africa have contributed to the importance of genetics for assessing hypotheses about contemporary pygmy body size and its evolution [23].

Population-genetic studies have inferred a probable common ancestral origin for West Central African Pygmy populations ~3,000 years ago and for West- and East-Central African Pygmy populations ~25,000 years ago, following their divergence from ancestral Central African non-Pygmy populations ~50,000-70,000 years ago [24-28]. Substantial levels of non-Pygmy genetic admixture have been observed across Central African Pygmy populations [24-27,29-31], correlating positively with adult standing height [32-34]. The general genetic difference between Pygmies and non-Pygmies together with the correlation of genetic admixture and standing height suggests that adult body size differences among Central African Pygmies and neighboring non-Pygmies are attributable in large part to genetic factors, arguing against a view that diminutive Central African pygmy body size is the consequence solely of phenotypic plasticity in a challenging nutritional and parasitic environment [8].

Over the past ten years, numerous genome-wide association studies (GWAS) have sought to identify genetic factors determining adult standing height [35-51], a high-heritability trait (≥69% [52-54]) despite being strongly influenced by environmental factors [55,56]. The largest study to date identified 697 common single-nucleotide polymorphisms (SNPs) significantly associated with adult standing height in a cohort of ~250,000 individuals of recent European ancestry that together explained 16% of the variance and 20% of the heritability of adult standing height in their cohort [51]. Recently, two GWAS on separate admixed Afro-American cohorts identified novel loci significantly associated with adult standing height [57,58], illustrating the potential of populations with a component of recent African ancestry to contribute new information to the study of adult standing height. In this context, Central African Pygmy and non-Pygmy populations—with marked differences in adult standing height that are strongly correlated with levels of genetic admixture—provide a potentially powerful framework within which to understand the genetic basis of adult human height.

A number of recent studies explored the genetic basis of adult standing height variation in Central Africans, comparing genome-wide SNP data on ≤3 Pygmy populations and their non-Pygmy neighbors [33,34]. These studies uncovered signatures of differential polygenic adaptation between Pygmies and non-Pygmies primarily in genomic regions containing genes associated with immunity and metabolism, as well as evidence of differential signatures in Western and Eastern Pygmies suggesting a partially

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

convergent origin of the pygmy phenotype in these groups. Perry *et al*. [34] further identified genetic associations with standing height that included four regions previously implicated in adult standing height variation in Europeans. However, no association was detected with the ~100 European standing-height-associated SNPs present in these modestly sized datasets of 67–169 Pygmies and 58–61 non-Pygmies, strongly suggesting that the differing stature of Pygmies and non-Pygmies is attributable at least in part to genetic factors that have arisen since the ancestral split between Central Africans and Europeans.

While these recent studies have focused on the differing standing height of Central African Pygmies and non-Pygmies, other morphological differences also exist between these two groups. For example, Pygmy body weight appears proportionally more reduced relative to standing height compared with non-Pygmies [59,60], a difference that cannot be wholly explained by differential nutrition levels in Pygmies and non-Pygmies [61]. In the context of the differential diets of hunter-gatherer Pygmies and agriculturalist non-Pygmies, and the greater seasonal variation in food availability experienced by hunter-gatherers compared with agriculturalists, such a difference might reflect the coevolution of differences in energy usage and storage in response to their different lifestyles. In addition, craniofacial and skeletal dissimilarities are also evident, including skull morphology [59,62] as well as leg and forearm length [59,63] where Pygmies generally have shorter legs and forearms relative to trunk length compared with non-Pygmies, potentially reflecting evolutionary changes in shared endoskeletal development pathways in response to the anatomical constraints of Pygmy hunting and gathering activities in the tropical rainforest. The reported dissimilarities in trunk and leg length patterns accord with their differential growth patterns in response to nutritional [64,65] and health [64] factors that vary between hunter-gatherer Pygmies and agriculturalist non-Pygmies [66-68], and are consistent with the view that both common and distinct pathways contribute to the determination of upper and lower body size [44]. In this context, while genes contributing to both trunk and leg length determination may be detected by genetic studies investigating standing height variation in Central Africans, those distinct to each trait may not because of the confounding effects of their variable contributions to standing height across groups. Thus, a joint analysis of variability in trunk and leg lengths in Pygmies and non-Pygmies might disentangle their relative contributions to Pygmy short stature and provide a more complete picture of the genetic architecture and anatomical constraints underlying the diminutive body size of contemporary Central African Pygmies.

Here, using body size and weight measurements available for individuals from seven Pygmy and three non-Pygmy populations from West Central Africa (**Figure 1**; **Table 1**) together with a genome-wide SNP data, we investigate the genetic basis of standing and sitting height, and subischial leg length in Central Africans. Our sample set includes 358 Pygmy and 169 non-Pygmy individuals genotyped on the Illumina Cardio-MetaboChip [69], a SNP genotyping microarray that interrogates a set of 68,126 SNPs previously identified in well-powered GWAS conducted by the Body Fat Percentage [70], CARDIoGRAM (coronary artery disease and myocardial infarction) [71] DIAGRAM (type 2 diabetes) [72], GIANT (anthropometric traits) [45,73,74], Global Lipids Genetics (lipids) [75], HaemGen (hematological measures) [76], ICBP (blood pressure) [77], MAGIC (glucose and insulin) [78-80], and QTIGC (QT interval) [81,82] consortia in addition to 146,453 SNPs chosen to facilitate fine-mapping of the genomic regions surrounding these 68,126 SNPs. The Cardio-MetaboChip therefore allows us to both investigate potential associations between the 1,050 SNPs associated with adult standing height variation [45] in Europeans with adult standing height variation in Central Africans. Moreover, it will allow us to explore Pygmy/non-Pygmy genetic differences in genes previously implicated in the determination of anthropometric, metabolic, and physiometric traits that we would expect to have been subject to historical natural selection under the hypotheses proposed for the evolution of the diminutive body size of Central African Pygmies [8-13].

Using a mixed-effect variance components model to correct for population stratification together with a multi-marker gene-based association method, we investigate the genetic determination of adult standing and sitting height, subischial leg length, and the ratio of sitting to standing height to identify genes and variants specific to their variation in Pygmies and in non-Pygmies. We identify three genomic regions containing SNPs with significant allele-frequency differences between Pygmies and non-Pygmies, and numerous genes significantly associated with each considered trait of which many have

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

known functions compatible with a deterministic role in Central African body size. Our results shed new light on the complex genetic architecture of the Central African pygmy phenotype, and highlight novel pathways contributing to its height and metabolic aspects that provide new opportunities for more comprehensive comparisons of Pygmies and non-Pygmies.

Results

Our objective was to perform a GWAS to advance our understating of the genetic basis of phenotypic differences between Central African Pygmies and their non-Pygmy neighbors, categorized as such based solely on cultural criteria and *not* on adult standing height (see **Materials and Methods**). Unlike normal GWAS where homogeneity among samples is desirable, we were instead conducting a GWAS in a more complex situation where significant phenotypic and genetic differences that provide potentially greater power to detect biologically meaningful associations are to be expected. We therefore first had to establish the levels phenotypic and genetic diversity that existed in our sample set to facilitate the design of the association analyses that we will perform.

Differences in measured traits between Pygmies and non-Pygmies

We first investigated whether the height and weight traits with measurements available for our sample set exhibited significant differences between Pygmies and non-Pygmies. Of the 527 individual in our sample set, we had trait measurements available for 115 Pygmy and 44 non-Pygmy individuals representing seven Pygmy and three non-Pygmy populations from West Central Africa (Additional File 1: **Table S1**).

Among the 115 Pygmies in our sample set, average male standing height was 155.67 cm (SD=6.55) while average female standing height was 148.95 cm (SD=5.82), in contrast to the 44 non-Pygmies whose average male and female standing heights were 167.21 cm (SD=4.61) and 154.45 cm (SD=6.23), respectively (**Figure 2A**). These values are consistent with average standing heights reported by prior studies of Central African Pygmy (males=154.85 cm with SD=3.08 and females=146.30 cm with SD=2.87) [3,9,83-88] and non-Pygmy (males=164.98 cm with SD=2.36 and females=155.96 cm with SD=0.85) [9,86,87,89] populations [59] and highlight the significant difference in stature that exists between Pygmies and non-Pygmies ($P_{males}=6.13\times10^{-8}$ and $P_{females}=2.65\times10^{-4}$; Wilcoxon rank-sum test).

Consistent with the high correlations expected between standing height and other height-related traits (Additional File 2: **Figure S1**), subischial leg lengths of male (mean=74.97 cm, SD=4.91) and female (mean=70.81 cm, SD=3.84) Pygmies were significantly shorter than those of male (mean=82.65 cm, SD=5.85) and female (mean=75.44 cm, SD=3.97) non-Pygmies ($P_{males}=0.007$ and $P_{females}=9.38\times10^{-5}$; **Figure 2B**). However, while the sitting heights of male Pygmies (mean=80.62 cm, SD=2.92) were significantly shorter than those of male non-Pygmies (mean=86.13 cm, SD=1.28; $P=0.002$), sitting heights of female Pygmies (mean=77.74 cm, SD=2.89) were instead similar to female non-Pygmies (mean=78.96 cm, SD=3.51; $P=0.073$; **Figure 2C**). These findings are consistent with leg length contributing more to the differential stature of Pygmies and non-Pygmies than upper body (trunk and head) length [59]. If we instead compare the ratio of sitting to standing height between Pygmies and non-Pygmies (**Figure 2D**), we observe significantly higher ratios among Pygmies than among non-Pygmies for females ($P=0.003$) but not for males ($P=0.203$) despite male and female Pygmies exhibiting a similar shift toward higher ratios compared with non-Pygmies (**Figure 2D**). The lack of significance for males likely reflects our decreased power to detect such a difference due to only four male non-Pygmies having both sitting and standing height data available compared with 67 male Pygmies.

Focusing on BMI instead of body weight since it accounts for the differential stature of Pygmies and non-Pygmies (**Figure 2A**), values were similar among male (mean=20.16, SD=1.95) and female (mean=20.06, SD=2.56) Pygmies ($P=0.587$), while those of male non-Pygmies (mean=22.31, SD=2.63) were significantly higher than those of female non-Pygmies (mean=20.16, SD=3.07; $P=0.031$; **Figure 2E**). Consequently, Pygmies had significantly lower BMI than non-Pygmies when considering males ($P=2.74\times10^{-4}$) but not females ($P=0.186$). While our BMI are consistent with those reported by prior studies for male and female Pygmies in 26 populations (mean=20.04 with SD=0.80 and mean=20.33 with

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SD=0.90, respectively) and male non-Pygmies in 10 populations (mean=21.48, SD=0.90), our female non-Pygmy BMI are notably lower than those of earlier studies (mean=21.65, SD=1.02) [59].

These findings reaffirm earlier observations of significant differences in standing height between Central African Pygmies and non-Pygmies, and highlight apparent differences in the contribution of trunk and leg length to these differences. However, our findings with BMI contrast with patterns in prior studies that indicate differences exist between Pygmies and non-Pygmies for both males (P=1.21×10^{-4}, Wilcoxon rank sum test of population means) and females (P=4.57×10^{-4}) and no male-female differences among Pygmies (P=0.858, Wilcoxon signed rank test) and non-Pygmies (P=1).

Genetic structure patterns

Now that we had established that significant differences existed between Pygmies and non-Pygmies for the height-related measurements available in our sample set, we next investigated the levels of genetic differentiation that are present among our Pygmy and non-Pygmy samples. Our genetic dataset consisted of 406 unrelated individuals from 20 populations with genotypes at 153,798 SNPs included on the Illumina Cardio-MetaboChip following the removal of lower quality SNPs and related individuals (see **Material and Methods**). In agreement with earlier studies [24-27,29,30,33], multidimensional scaling (MDS) analysis of pairwise allele-sharing dissimilarities (ASD) among these 406 unrelated individuals supported three main components to genetic differentiation among Central Africans (**Figure 3A**). First, differentiation between Pygmies and non-Pygmies─categorized as such based on cultural criteria and not on height─largely determined the first MDS dimension. Second, though non-Pygmies formed a single tight cluster in the first two MDS dimensions reflecting the recent history of these mostly Bantu-speaking populations [90-94], differentiation among Pygmy populations was evident in the second dimension, driven partly by differences between the West African Baka (CBK, EBK, SBK) and Aka (AKA, AKM) populations and the East African Nsua (NSU). Dispersion among Pygmies─measured as the variance among their ASD values─was significantly higher than among non-Pygmies (2.07×10^{-5} and 7.76×10^{-6}, respectively; P<10^{-16}, one-sided F-test), consistent with appreciable reproductive isolation following their ancestral divergences over the last ~25,000 years [24-26,92,95,96]. Third, the degree of genetic differentiation from non-Pygmies varied among Pygmy populations, consistent with what might be expected in the presence of variable levels of non-Pygmy gene flow into Pygmy populations. The Bezan, Kola and Bongo Pygmies clustered closer to the non-Pygmies (mean ASD=0.2047, SD=0.0027, across 20,413 pairs of individuals) than did the Baka and Aka (mean ASD=0.2051, SD=0.0025, 16,241 pairs) and Nsua (mean ASD=0.2072, SD=0.0036, 1,639 pairs) Pygmies; the first and third quantiles of the full distribution of ASD values were 0.1983 and 0.2053, respectively.

Genetic admixture patterns

To further explore the presence of potential signals of variable genetic admixture between Central African Pygmies and non-Pygmies suggested by our MDS analysis, we applied the Bayesian clustering algorithm implemented in *STRUCTURE* [97,98] to four non-overlapping sets of 10,106 low-linkage-disequilibrium (LD) genome-wide SNPs. At K=2 (**Figure 3B**), Pygmies were assigned greater membership in one cluster ("red") while non-Pygmies were assigned greater membership in the other cluster ("blue"). Per-individual membership proportions for the blue cluster varied among Pygmy populations from 0.173±0.103 in the Central/Eastern Baka (CBK) to 0.527±0.096 in the Eastern Bongo (EBG). In contrast, among non-Pygmies, per-individual membership proportions for the red cluster were on average 0.197±0.093.

Our *STRUCTURE* results accord with previous studies [24-27,29,31,34,99,100] and can be interpreted as signals of asymmetric admixture between Central African Pygmies and non-Pygmies. In this view, appreciable membership in the blue "non-Pygmy" cluster among Pygmies reflects substantial and variable levels of non-Pygmy admixture, whereas low levels of membership in the red "Pygmy" cluster among non-Pygmies suggest low levels of Pygmy introgression into non-Pygmies.

Correlation between non-Pygmy admixture and trait variation

Given the phenotypic and genetic differences observed between Pygmies and non-Pygmies in our sample

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

set, we sought to establish whether genetic factors might underlie observed phenotypic patterns. If genetic differences between Pygmies and non-Pygmies contribute to their phenotypic differences, we might expect measurements for these traits to be correlated with per-individual levels of genetic admixture. We therefore separately investigated each trait's correlation with per-individual membership proportions in the blue "non-Pygmy" *STRUCTURE* cluster at K=2 (**Figure 3B**; "non-Pygmy admixture" henceforth).

We observed a significant positive correlation between non-Pygmy admixture and adult standing height with the 76 males (Pearson r=0.585, P=2.90×10^{-8}; **Figure 4A**) and the 57 females (r=0.485, P=1.30×10^{-4}; **Figure 4B**) with standing height data available among the 406 unrelated individuals in our genotype dataset. The correlations remained significant when restricted to the 61 male and 31 female Pygmy individuals (r=0.311 with P=0.015 and r=0.441 with P=0.013, respectively). These findings are in agreement with previous studies on the relationship between levels of non-Pygmy admixture and adult standing height in Central African Pygmies [32-34] and provide further support for an appreciable genetic component to the determination of body size differences between Pygmies and non-Pygmies.

The correlations of adult sitting height and subischial leg length with non-Pygmy admixture were similar in males (Pearson r=0.431 with P=8.90×10^{-5} and r=0.475, P=1.41×10^{-5}, respectively; **Figures 4C** and **4E**). However, among females the correlation of non-Pygmy admixture with adult sitting height was markedly lower (r=0.226, P=0.040; **Figure 4C**) than with subischial leg length (r=0.574, P=6.53×10^{-7}; **Figure 4F**). Considered together, these findings are consistent with common and distinct genetic factors contributing to sitting height and subischial leg length variation among Central Africans, while male–female differences in the strength of the correlations with non-Pygmy admixture might perhaps reflect the differential contribution of genetic variants in estrogen-dependent growth pathways to the determination of trunk and leg length [101,102].

Finally, while BMI was significantly positively correlated with non-Pygmy admixture in females (r=0.262, P=0.015; **Figure 5B**), BMI was only marginally so in males (r=0.166, P=0.060; **Figure 5A**). These correlations were markedly weaker than those obtained with the height-related measures (**Figure 4**), in accordance with smaller Pygmy/non-Pygmy differences in BMI (**Figure 2C**) than in adult standing height (**Figure 2A**) and compatible with non-genetic factors contributing substantially to body weight differences between Central African Pygmies and non-Pygmies.

These findings are consistent with a significant genetic component in the determination of height-related differences between Central African Pygmies and non-Pygmies. However, our results do not provide similar support for an appreciable genetic component to BMI variation patterns in Pygmies and non-Pygmies. Consequently, we next sought to identify genetic factors correlated with variation in height-related traits in our sample set of Pygmies and non-Pygmies, but not with variation in BMI.

Association analyses of Pygmy/non-Pygmy categorization

We have found significant genetic and phenotypic differences between the Central African Pygmies and non-Pygmies in our sample set, with correlations between individual trait variation and inferred levels of non-Pygmy genetic admixture supporting an appreciable genetic component to the determination of their observed phenotypic differences. To investigate the underlying genetic component of their phenotypic differences, we first performed single- and multi-marker association tests to identify genomic regions harboring SNPs with significant allele frequency differences between the Pygmies and non-Pygmies in our genetic dataset of 406 unrelated individuals.

Single-marker association tests

To account for genetic structure (**Figure 3**) and cryptic genetic relatedness among individuals—which could inflate type-1 and type-2 errors [103-105]—in our single-marker association tests, we used *EMMAX*, which implements a linear mixed-effect regression model that corrects per-SNP association tests for structure and relatedness via a pairwise kinship matrix [105], and included village affiliation as a covariate (see **Materials and Methods**). We identified ten SNPs exhibiting a significant allele frequency differences between Pygmies and non-Pygmies after Bonferroni correction for multiple testing (α=5%), all lying within a single genomic region on chromosome 2 (**Figure 6C**). In addition, two further SNPs

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were marginally significant (α=10%; **Figure 6A** and **Table 2**), located on chromosomes 1 (Additional File 2: **Figure S2A**) and 7 (Additional File 2: **Figure S2B**). Our approach appeared to reasonably correct for structure and relatedness among individuals, as judged by the quantile-quantile plot (**Figure 6B**), while marginally inflated *P*-values were instead observed if village affiliation was not included as a covariate in the association tests (Additional File 2: **Figure S3**).

Multi-marker gene-based association tests

Because individual SNPs might fail to achieve the Bonferroni significance threshold owing to insufficient power in our modestly sized genetic dataset, we performed multi-marker gene-based association tests using *VEGAS* [106], a multivariate method that combines association signals across all SNPs located within a gene, correcting for non-independence between SNPs. We observed seven genes exhibiting a significant association with Pygmy/non-Pygmy categorization after Bonferroni correction (α=1%; **Table 3**); the genes identified did not differ appreciably if village affiliation was excluded as a covariate (Additional File 1: **Table S2**).

Relevance to the Central African pygmy phenotype

While none of the genes in the genomic regions encompassing the ten significant SNPs (**Figure 6C**) and the two marginally significant SNPs (Additional File 2: **Figure S2**) had putative functions consistent with previously reported phenotypic differences between Central African Pygmies and non-Pygmies, one of the seven genes identified in the multi-marker analysis had a putative function in limb development. *GPC6* is a growth factor receptor that is important for correct growth plate formation [107] whose abrogation has been observed to cause long-bone growth retardation [108], suggesting that GPC6 is important for their longitudinal growth. This association is compatible with the observation that Pygmies have significantly shorter legs compared with non-Pygmies (**Figure 2B**) that are proportionally more reduced relative to trunk height (**Figure 2D**).

Association analyses of height-related traits

Despite our modest sample size, we have identified genomic regions associated with an individual's categorization as Pygmy or non-Pygmy that include one gene with a putative function compatible with their observed differences in leg length. Building upon this success, we next sought to identify genomic regions containing SNPs significantly correlated with variation in the height-related traits that differed significantly between the Pygmies and non-Pygmies in our sample set (**Figure 2A-D**).

Single-marker association tests

We performed per-SNP association tests to identify genomic regions harboring genes that contribute to variation in adult standing and sitting height and their ratio as well as subischial leg length among the 132-159 individuals with these data available in our dataset (Additional File 1: **Table S1**) using *EMMAX* and including sex and Pygmy/non-Pygmy categorization as covariates. Again, *EMMAX* appeared to provide a reasonable correction for structure and relatedness in these analyses (Additional File 2: **Figures S4E-S4H**, respectively) while similar analyses controlling only for sex (Additional File 2: **Figures S4A-S4D**) or for sex and village affiliation (Additional File 2: **Figures S4I-S4L**) instead led to slightly inflated or deflated *P*-values, respectively.

We did not find any SNPs significantly associated with adult standing (Additional File 2: **Figure S5**) or sitting (Additional File 2: **Figure S6**) height or subischial leg length (Additional File 2: **Figure S7**) variation after Bonferroni corrections for multiple testing, likely reflecting insufficient power with our modest sample sizes. However, we did identify a SNP marginally associated with sitting/standing height ratio after Bonferroni correction for multiple testing (α=10%; **Figure 7A**).

Multi-marker gene-based association tests

Gene-based association tests using *VEGAS* identified 19 genes associated with adult height-based traits after Bonferroni correction for multiple testing (α=1%; **Table 4**): nine with standing height, ten with sitting height, nine with sitting/standing ratio, and ten with subischial leg length. The genes identified for each trait did not differ markedly if only sex, or sex and village affiliation, were instead considered as

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

covariates (Additional File 1: **Table S3**).

Relevance to the Central African pygmy phenotype

The single SNP marginally associated with sitting/standing height ratio, rs13097517, is located in an intron of *ERC2* (**Figure 7C**), a member of the Rab3-interacting molecule (RIM)-binding protein family that functions as a regulator of neurotransmitter release [109], which lies ~600 kb upstream of *WNT5A*, a signaling ligand whose loss-of-function causes craniofacial and skeletal abnormalities including shortened limbs [110,111]. Thus, this association might reflect the influence of an untyped variant in strong LD with rs13097517 that modifies the expression of *WNT5A*, and potentially contributing to the differential skull morphology [59,62] and leg and forearm lengths [59,63] of Pygmies compared with non-Pygmies.

Three of the 19 genes were associated with all four height-based traits and were also identified in our *VEGAS* analysis for Pygmy/non-Pygmy categorization (**Table 3**), of which one (*GPC6*) has been implicated in the longitudinal growth of long bones [108]. In addition, *ERC2*, which is associated with both standing/sitting height ratio and subischial leg length (**Table 4**), contains SNP rs13097517, which was marginally associated with standing/sitting height ratio in our SNP-based association tests (**Figure 7C**). Importantly, one gene significantly associated with standing and sitting height variation in our Central Africans data set (*MAGI2*) has also been associated with standing height variation in non-Africans [112], while another gene significantly associated with subischial leg length (*SHANK2*) lies within a genomic region previously associated with the pygmy phenotype in Eastern Batwa Pygmies [34].

Discussion

We have performed the first large-scale joint investigation of the variability and genetic architecture of the major components of adult standing height in Central African hunter-gatherer Pygmies and their agriculturalist non-Pygmy neighbors. Our findings accord with prior observations [59,63] that while Pygmy body size is generally proportionally reduced relative to non-Pygmies, their leg lengths are significantly shorter relative to their trunk length. Importantly, our results provide further support for an appreciable genetic component to the determination of body size differences between Pygmies and non-Pygmies, as implied by the correlations observed between the different measures and inferred levels of non-Pygmy admixture that replicate those reported previously for adult standing height [32-34].

Although our sample size is modest compared with traditional GWAS frequently conducted within populations of mainly European and Asian ancestry, our large-scale genetic association analyses using single- and multi-marker approaches identified three genomic regions as well as seven genes significantly associated with Pygmy/non-Pygmy categorization after a conservative Bonferroni correction for multiple testing. However, we were only able to identify a single genomic region associated with just one of the measured traits we considered (**Figure 7**). Given that for each trait a number of SNPs had notably lower *P*-values than the vast majority of those tested (Additional File 2: **Figures S5-S7**), our inability to identify additional genomic regions associated with variation in these traits likely reflects insufficient power with the much smaller sample sizes (132-159) than were available for our comparison of Pygmy/non-Pygmy categorization (406). Nevertheless, gene-based multi-marker test identified nine, ten, nine, and ten genes enriched for SNPs exhibiting an association with standing height, sitting height, sitting/standing height ratio, and subischial leg length, respectively. Importantly, quantile-quantile plots of the *P*-values created in each association analysis (Additional File 2: **Figures S3 and S4**) indicate that our approach appropriately corrected for structure and relatedness present among individuals in our dataset (**Figure 3**). Considered together with our identification of significantly associated genes and genomic regions, this would suggest that cross-population GWAS are a viable and powerful approach to identify the genetic basis of human phenotypes that naturally vary across populations.

A number of genes and genomic regions identified here potentially provide novel insight into the genetic basis of anatomical and physiological features of the Central African pygmy phenotype. Firstly, a number of genes associated with the height-related measures are reported to cause both skeletal and craniofacial abnormalities when perturbed: *MACROD2* [113,114], *GP1BB* [115,116], and *CACNA1C*

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[117,118]. It is therefore possible that craniofacial differences observed between Pygmies and non-Pygmies [59,62] may in part reflect the coevolution of these traits with Pygmy short stature; a hypothesis that remains to be formally tested in the absence of paleoanthropological data for Central Africa [3].

Secondly, the sole SNP marginally associated with sitting/standing height ratio (rs13097517) lies ~619 kb upstream of *WNT5A*, a signaling ligand whose abrogation causes craniofacial and skeletal abnormalities including shortened limbs [110,111]. Thus, in the context of the significantly higher sitting/standing height ratios we (**Figure 2D**) and others [59,63] have observed in Pygmies compared with non-Pygmies, this association would be compatible with a scenario in which Pygmy short stature is attributable in part to perturbation of *WNT5A* expression by one or more genetic variants in high LD with rs13097517. Intriguingly, WNT5A mediates the actions of growth factor receptor GPC6 [119], an important determinant of long bone growth [107,108] that was associated with Pygmy/non-Pygmy categorization (**Table 3**) and all four height-related traits (**Table 4**), suggesting that perturbation of the GPC6/WNT5A pathway may partly underlie Pygmy short stature.

Finally, a number of genes associated with Pygmy-non-Pygmy categorization have functions in immune rather than developmental processes. For example, *IKBKE* is an essential regulator of antiviral signaling pathways [120,121] and *CAPN2* is implicated in immune synapse formation in T-cell receptor signaling [122], while *FCGR1A* functions in monocyte phagocytosis [123] and has been implicated in dengue virus infection [124], a mosquito-borne disease present throughout Central Africa [125]. Our findings therefore add to the expanding body of evidence suggesting that differential adaptation in both immune- and auxologic-related processes have contributed to the evolution of the Central African pygmy phenotype [33,34,126].

Consistent with prior studies that investigated the contributions of >100 non-African standing-height-associated SNPs to standing height variation among 5 to 230 Central African Pygmies and non-Pygmies [33,34,126], none of the 949 SNPs listed in the Cardio-MetaboChip manifest [69] as associated with adult standing height were found to be individually associated with the height-related variation in our Central African sample set. Indeed, *P*-values for the 173 standing height SNPs reported by Wood *et al*. [51] and the 949 standing-height SNPs included on the Cardio-MetaboChip [69] based on the findings of Lango Allen *et al*. [45] did not depart from the uniform expectation to a greater extent than non-standing-height SNPs in our analyses of Pygmy/non-Pygmy categorization (**Figure 6B**) and subischial leg length (Additional File 2: **Figure S7B**). However, the Wood *et al*. SNPs did deviate to a greater extent in our standing height (Additional File 2: **Figure S5B**) and sitting/standing height ratio (**Figure 7B**) analyses while both the Wood *et al*. and Lango Allen *et al*. SNPs did deviated to a greater extent in our sitting height analysis (Additional File 2: **Figure S6B**). These observations indicate that previously reported non-African standing-height SNPs predominantly reflect perturbations of pathways contributing to trunk height but not to leg length. In contrast, the genes identified in our analysis of different height measures in Central Africans instead appear predominantly to reflect perturbation of pathways contributing to the determination of leg length. Our analyses therefore highlight the importance of conducting studies in Central African populations that can contribute novel information about the determinants of human standing height and body size proportions in general.

Despite our inability to detect SNP-wise associations with non-African standing height SNPs, gene-based association tests did identify one non-African standing height gene as being associated with standing and sitting height variation in Central Africans, *MAGI2* [112]. Thus, though the non-African standing height associated SNPs failed individually to reach significance in our association analyses, our identification of *MAGI2* using a multi-marker approach suggests that this is at least partly a consequence of insufficient power in our sample set. However, we cannot discount the possibility that for a subset of the non-African SNPs the absence of a significant association may be a result of discordance in LD patterns between Europeans and Central Africans disrupting the linkage between the genotyped SNP and the trait-determining variant [127].

Conclusions

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Despite the modest sample size in our study, we have identified genomic regions encompassing genes that are biologically meaningful in the context of both the traits themselves and phenotypic differences that exist between Central African Pygmies and non-Pygmies. Further, our findings provide additional support for an appreciable genetic component to the determination of phenotypic differences between Central African Pygmies and non-Pygmies. More generally, our results highlight the need for joint analyses considering different anthropometric and physiological measures in larger samples of Central African Pygmies and non-Pygmies to facilitate our understanding of the biological basis of the Central African pygmy phenotype. Future large-scale studies of worldwide short-stature populations have the potential to shed light on relative contributions of shared and distinct pathways in the development and maintenance of short stature. In particular, studying populations of different stature living in different ecologic environments will provide a greater understanding of adaptive processes underlying the appreciable variation in adult height observed across contemporary worldwide human populations.

Material and Methods

Samples

This study was conducted according to ethical principles of the Declaration of Helsinki. IRB approvals were obtained from the French Ministry of Higher Education and Research, University of Michigan, Stanford University, Washington State University and the University of Manitoba. Prior to sample collection, research authorizations were obtained from the Ministry of Public Health in Cameroon, the Ministry of Higher Education and Research in Gabon, the National Council for Science and Technology in Uganda, and the Ministry of Scientific Research in CAR, and informed consent was obtained from all research participants. In total, our study included 558 individuals from 20 Central African populations (**Figure 1**, **Table 1**). We conducted ethnographic interviews at each sampling site to categorize sampled populations *a priori* as "Pygmy" or "non-Pygmy" based on historical and cultural criteria that do not include adult height [4,6,7,23,128-131]. A community was categorized as "Pygmy" when it: (1) is recognized by outsiders as specialized in forest activities such as hunting-gathering and medical and magical knowledge of the rainforest; (2) shares complex socio-economic relationships with specific neighboring outsiders, such as exchanging forest products (e.g. game, wild honey) for iron tools (e.g. fishing hooks, iron blades, or spear heads); (3) is designated as "Pygmy" or its literal local translation or at least as "other than self" by neighboring populations; (4) distinguishes itself as a community with a different ethno-name from other neighboring communities, regardless of languages spoken; (5) has differing musical practices and instruments recognized as such by neighbors.

For each individual, DNA was extracted from either whole blood buffy coats with the DNeasy Blood & Tissue spin-column Kits (Qiagen, Valencia, CA), or saliva collected using Oragene kits (DNA Genotek Inc., Kanata, ON), following the manufacturer-recommended protocol. For a subset of the individuals from seven Pygmy─Baka (Center), Baka (East), Bezan (South), Bongo (Center), Bongo (East), Bongo (South), and Koya─and three non-Pygmy─Nzime, Tikar, and Bangando─populations, standing and sitting height were measured with a height gauge and weight with a standard weigh scale following standard anthropometric procedures [132]. Subischial leg length was calculated as standing minus sitting height, while BMI was calculated as body weight in kilograms divided by standing height in meters squared (kg m^{-2}). Accurate age data was unavailable for these individuals, as most of the communities do not keep track of birth dates. Though we cannot rule out the confounding effect of osteoporotic age-related shrinking in our study, we expect it to be minimal as our sample set includes only adults and elderly individuals were not considered.

Genotyping and quality control

DNA samples for 576 samples─558 individuals from 20 Central African Pygmy and non-Pygmy populations (**Table 1**) and 18 control samples─were genotyped at the University of Michigan Medical School DNA Sequencing Core (Ann Arbor, MI) using the Illumina MetaboChip that interrogates 196,725 genome-wide SNPs [69]. We focus on 196,091 SNPs whose genomic position had been independently

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

verified (Peter Chines, unpublished data). Genotype calling was performed using the GenomeStudio Genotyping Module (v.1.0; Illumina Inc., San Diego, CA). Following quality control procedures conducted at the genotype-calling level (Stage 1, **Figure S8** [Additional File 2]), the preliminary dataset contained 192,903 autosomal SNPs that were polymorphic in a sample of 543 Pygmy and non-Pygmy individuals. Next, we conducted population-level quality control procedures (Stage 2, **Figure S8** [Additional File 2]), creating an initial dataset of 154,106 autosomal SNPs polymorphic in a sample of 543 Pygmy and non-Pygmy individuals. Hardy-Weinberg equilibrium was evaluated separately in each population using Yates-corrected chi-squared tests [133], and we used the same exclusion criteria as in Pemberton *et al*. [134].

Pygmy and non-Pygmy neighbors interact socially and economically on a daily basis [6,7,130,135]. Thus, individuals from neighboring communities might be present at the time of sampling. On rare occasions, some non-Pygmies have fled their communities (to avoid taxes or military recruitment for instance) and taken refuge among neighboring Pygmy populations. Therefore, it is possible that some sampled individuals were wrongly categorized according to our Pygmy/non-Pygmy categorization criteria. To search for categorization errors, we performed MDS analysis of individual pairwise allele-sharing distances (see below) in our initial dataset. We identified 16 individuals─11 initially categorized as Pygmies and five as non-Pygmies─who did not cluster genetically with other individuals sharing the same categorization (data not shown) and who might have been miscategorized during sample collection. To be conservative, we removed these 16 putative misclassified individuals from the preliminary dataset, and repeated the population-level quality control procedures (Stage 3, **Figure S8** [Additional File 2]) to create a dataset containing 154,029 autosomal SNPs polymorphic in a sample of 527 Pygmy and non-Pygmy individuals ("527RELAT" henceforth; **Table 1**); all individuals possessed genotypes at ≥94.9% of SNPs.

Relatedness among all pairs of the 527 individuals in the 527RELAT dataset was evaluated using identity-by-state allele sharing and the likelihood approach of *RELPAIR* (v.2.0.1) [136,137] following the methods of Pemberton *et al*. [134] restricted to sets of 9,999 SNPs. A total of 282 pairs of individuals were inferred by *RELPAIR* to be related at a level closer than first cousins: 230 intra-population and 52 inter-population pairs. All inter-population relative pairs involved geographically close populations (**Figure 1**); 47 involving the two Bezan populations (BZN and BZS), three involving two Bongo populations (CBG and EBG), and two with individuals from the nearby Fang (CFG) and Ngumba (NGB) populations. A dataset containing no first- or second-degree relatives was created by removing one individual from each of these 282 relative pairs. To minimize the number of individuals removed, we preferentially omitted individuals present in two or more relative pairs (either intra- or inter-population). In situations where either individual in a relative pair could be removed, we removed the individual with more missing data. After the exclusion of 121 related individuals─in addition to the 16 putative misclassified individuals─from the preliminary dataset of 543 individuals, many of whom were related to multiple individuals in the initial dataset, we repeated the population-level quality control procedures (Stage 4, **Figure S8** [Additional File 2]) to create an unrelated dataset with 153,798 autosomal SNPs polymorphic in a sample of 406 Pygmy and non-Pygmy individuals ("406UNRELAT" henceforth; **Table 1**); all individuals possessed genotypes at ≥94.9% of SNPs. Because of small sample sizes for the two Bezan populations (11 for BZN, 17 for BZS) in the 406UNRELAT dataset, we combined them into a single population ("Bezan" henceforth; BEZ, **Table 1**). Both the BZN and BZS samples belong to the same Bezan ethnic group located in Central Cameroon, which has a census size <400, divided between two nearby communities (~50 km apart) in frequent contact (P. Verdu and A. Froment, unpublished data).

Population genetic analyses

Multidimensional scaling

We performed MDS based on ASD matrices constructed for all pairs of individuals in the 406UNRELAT dataset using *asd* (v1.0; https://github.com/szpiech/asd). This program considers in the calculation for a given pair only those SNPs for which neither individual was missing genotypes. We applied classic metric MDS on the ASD matrix using the *cmdscale* function in R (v2.11.0) [138]. To estimate the

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proportion of variation in ASD explained by the first two dimensions of the MDS plot, we calculated the Spearman correlation coefficient ρ between the Euclidean distances for all pairs of individuals and their corresponding ASD values (using *cor.test* in R here and for other correlations). To evaluate dispersion levels among Pygmies and non-Pygmies in the MDS plot, we compared the variance among their ASD values with a one-sided *F*-test using *var.test* in R.

STRUCTURE

To investigate genetic structure in our 406UNRELAT dataset, we performed model-based Bayesian clustering analyses implemented in *STRUCTURE* (v2.3) [97,98], which probabilistically assigns proportions of each individual's genotypes to each of *K* genetic clusters, where *K* is set *a priori*, based on allele frequencies and irrespective of individual population categorizations. We used the *admixture* model with separate Dirichlet parameters in each cluster and correlated allele frequencies, and a burn-in period of 20,000 iterations followed by 10,000 iterations.

To minimize the number of linked loci owing to the gene-centric design of the Cardio-MetaboChip, in the *STRUCTURE* analyses we used 40,424 SNPs that have minimum spacing 18.75 kb. From these 40,424 SNPs, we created four non-overlapping panels of 10,106 SNPs, each with minimum marker spacing 75 kb. For SNP panel n ($1 \leq n \leq 4$), every fourth SNP along a vector of the considered 40,424 SNPs was selected, starting at position $n - 1$. In this vector, SNPs were numbered starting at 0 and ordered from chromosome 1 to 22 and by increasing distance along each chromosome (taking genomic positions from NCBI database build 37). For each panel, we computed 10 independent *STRUCTURE* runs for values of *K* between 2 and 4, producing 40 independent replicates for each *K*. We identified common modes among the 40 replicates using *CLUMPP* [139] with the *Greedy* algorithm and 1,000 random permutations. For each *K*, all pairs of runs with a symmetric similarity coefficient >0.9 were considered to belong to the same mode. For each mode, we computed individual membership proportions averaged across runs from that mode, visualizing the most frequent mode at each *K* using *DISTRUCT* [140].

To investigate the relationship between non-Pygmy admixture and adult height in Central Africans, we calculated the Pearson product-moment correlation coefficient *r* between individual membership proportions in the blue cluster at *K*=2 and measured adult height where available.

Association analyses

Single-marker tests

Per-SNP association tests were performed using *EMMAX* [105], which implements a mixed-effect regression model [141] to account for genetic structure in a sample set, incorporating variance components of random polygenic effects [142,143] and genetic relatedness among individuals using a pairwise Balding-Nichols kinship relatedness matrix [144] (equation 7 in [105]). We constructed the kinship matrix for all pairs of individuals in the 406UNRELAT dataset using *EMMAX*, including in the calculation only the 94,050 and 93,821 SNPs, respectively, inferred to be in linkage equilibrium (r^2<0.5). Association analyses for Pygmy/non-Pygmy categorization were performed by labeling Pygmies as "cases" and non-Pygmies as "controls." Association analyses were performed separately for each measured trait considering only those individuals with data available. To improve the normality of the body weight and BMI distributions and alleviate the impact of outliers, body weight and BMI values were rank-based inverse-normal transformed separately for each gender, and association tests performed using these transformed values. To control for population structure and sexual dimorphism in our sample set, in our analysis of Pygmy and non-Pygmy categorization we considered the residuals of the regression of Pygmy/non-Pygmy categorization on village affiliation, while for quantitative-traits we considered the residuals of a multiple regression for that trait on sex and Pygmy/non-Pygmy categories.

Multi-marker gene-based tests

As individual SNPs may fail to achieve the significance threshold due to insufficient power in our modest sample sizes, we performed multi-marker gene-based tests of association using *VEGAS* [106]. Separately for each gene, *VEGAS* computes a χ^2 test statistic from observed *EMMAX P*-values for all SNPs that lie within its boundaries (defined as ±50 kb from the ends of the transcribed region) and evaluates

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significance via simulations from a multivariate normal distribution with mean 0 and a covariance matrix of pairwise LD between SNPs. The transcribed region of each gene was defined using release hg19 of the UCSC gene database [145]; for genes with multiple isoforms, begin and end positions were defined as the outermost positions in the union of the transcribed regions of all isoforms. Pairwise LD estimates in our datasets considered in the calculation only unrelated individuals.

List of Abbreviations Used

AR: androgen receptor; ASD: allele-sharing dissimilarity; BMI: body mass index; DRC: Democratic Republic of Congo; GWAS: genome-wide association study; HGH: human growth hormone; IGF1: insulin-like growth factor 1; MDS: multidimensional scaling; SD: standard deviation; SNP: single nucleotide polymorphism.

Declarations

Ethics approval and consent to participate

This study was approved by institutional ethics review committees at the French Ministry of Higher Education and Research, University of Michigan, Stanford University, Washington State University and the University of Manitoba. Research authorizations were obtained from the Ministry of Public Health in Cameroon, the Ministry of Higher Education and Research in Gabon, the National Council for Science and Technology in Uganda, and the Ministry of Scientific Research in CAR.

Consent for publication

All authors read and approved the final manuscript.

Availability of data and material

The Illumina Cardio-MetaboChip SNP genotype dataset for the 406 unrelated individuals analyzed in this study is available from the European Genome-Phenome archive (**<accession no. pending>**), and will be accessible for population and evolutionary history studies in accord with the informed consent documents used for this study.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the France–Stanford Center for Interdisciplinary Studies (N.A.R.); the French Assistance Publique Hôpitaux de Paris, ATM-MNHN 'Les relations Sociétés-Natures dans le long terme' 2009–2012 and the Fondation pour la Recherche Médicale and the ANR-Blanc program 'GrowingAP' (E.H.); a University of Michigan Center for Genetics in Health and Medicine postdoctoral fellowship (T.J.P.); and a Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2015-04739; T.J.P.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

EH, NAR, PV, TJP, NSB, and CJW conceived and designed the study. AF, BSH, PV, and SLB collected the samples. NSB prepared DNA samples for genotyping. PV performed the genotype calling and quality control with assistance from TJP and NSB. TJP performed data preparation and population-level quality control. PV performed population-genetic analyses with the assistance of TJP. TJP and PV designed and performed the association analyses. TJP, PV and NAR wrote the manuscript with the assistance of all other authors. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the volunteers from Cameroon, Central African Republic, Gabon, and Uganda who participated in this study. The authors thank Peter Chines for providing verified Cardio-MetaboChip SNP positions, Sen Li for providing ancestral and derived allelic states for the MetaboChip SNPs, Hyun Min Kang for assistance with *EMMAX*, and Frédéric Austerlitz, Erkan Buzbas, Zachary Szpiech, and

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lawrence Uricchio for useful comments and discussions.

Description of Additional Data Files

Additional File 1: Tables S1-S3
Additional File 2: Figures S1-S8

References

1. Hewlett BS: **Introduction**. In: *Hunter-gatherers of the Congo Basin: cultures, histrories, and biology of African Pygmies.* New Brunswick, New Jersey: Transaction Publishers; 2014: vxii-xxix.
2. Hiernaux J: **The people of Africa**. London: Weidenfeld and Nicolson; 1974.
3. Froment A: **Adaptation biologique et variation dans l'espèce humaine : le cas des Pygmées d'Afrique.** *Bulletins et Mémoires de la Société d'Anthropologie de Paris* 1993, **5**:417-448.
4. Bahuchet S: **L'invention des Pygmées**. *Cahiers d'Etudes africaines* 1993, **33**(1):153-181.
5. Bahuchet S: **La Rencontre des Agriculteurs. Les Pygmées parmi les Peuples d'Afrique Centrale.** Paris; 1993.
6. Hewlett B: **Cultural diversity among African Pygmies**. In: *Cultural diversity among twentieth-century foragers: An African perspective*. Edited by Kent S, vol. 1. Cambridge: Cambridge University Press; 1996: 215-244.
7. Bahuchet S: **Changing language, remaining Pygmy**. *Hum Biol* 2012, **84**(1):11-43.
8. Perry GH, Dominy NJ: **Evolution of the human Pygmy phenotype**. *Trends Ecol Evol* 2009, **24**(4):218-225.
9. Cavalli-Sforza LL: **African Pygmies**. Orlando, FL: Academic Press; 1986.
10. Bailey RC, Head G, Tenike M, Owen B, Rechtman R, Zechenter E: **Hunting and gathering in tropical rainforest: is it possible?** *Amer Anthropol* 1989, **91**:59-82.
11. Bailey RC: **The comparative growth of Efe Pygmies and African farmers from birth to age 5 years**. *Ann Hum Biol* 1991, **18**(2):113-120.
12. Diamond JM: **Anthropology. Why are Pygmies small?** *Nature* 1991, **354**(6349):111-112.
13. Migliano AB, Vinicius L, Lahr MM: **Life history trade-offs explain the evolution of human Pygmies**. *Proc Natl Acad Sci U S A* 2007, **104**(51):20216-20219.
14. Rimoin DL, Merimee TJ, Rabinowitz D, Cavalli-Sforza LL, McKusick VA: **Peripheral subresponsiveness to human growth hormone in the African Pygmies**. *N Engl J Med* 1969, **281**(25):1383-1388.
15. Merimee TJ, Rimoin DL, Cavalli-Sforza LL: **Metabolic studies in the African Pygmy**. *J Clin Invest* 1972, **51**(2):395-401.
16. Rozzi FV, Koudou Y, Froment A, Le Bouc Y, Botton J: **Growth pattern from birth to adulthood in African pygmies of known age**. *Nat Commun* 2015, **6**:7672.
17. Merimee TJ, Zapf J, Hewlett B, Cavalli-Sforza LL: **Insulin-like growth factors in Pygmies. The role of puberty in determining final stature**. *N Engl J Med* 1987, **316**(15):906-911.
18. Hagino I, Hayashi K, Kawamura K, Sato H, Yamauchi T: **Adolescent growth spurt and growth pattern factors related to the short stature of Pygmy hunter-gatherers of Southeast Cameroon**. *Ann Hum Biol* 2013, **40**(1):9-14.
19. Geffner ME, Bersch N, Bailey RC, Golde DW: **Insulin-like growth factor I resistance in immortalized T cell lines from African Efe Pygmies**. *J Clin Endocrinol Metab* 1995, **80**(12):3732-3738.
20. Hattori Y, Vera JC, Rivas CI, Bersch N, Bailey RC, Geffner ME, Golde DW: **Decreased insulin-like growth factor I receptor expression and function in immortalized African Pygmy T cells**. *J Clin Endocrinol Metab* 1996, **81**(6):2257-2263.
21. Bozzola M, Travaglino P, Marziliano N, Meazza C, Pagani S, Grasso M, Tauber M, Diegoli M, Pilotto A, Disabella E, Tarantino P, Brega A, Arbustini E: **The shortness of Pygmies is associated**

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with severe under-expression of the growth hormone receptor. *Mol Genet Metab* 2009, **98**(3):310-313.

22. Becker NS, Verdu P, Georges M, Duquesnoy P, Froment A, Amselem S, Le Bouc Y, Heyer E: **The role of *GHR* and *IGF1* genes in the genetic determination of African Pygmies' short stature**. *Eur J Hum Genet* 2013, **21**(6):653-658.
23. Hewlett BS: **Hunter-Gatherers of the Congo Basin: Cultures, Histories, and Biology of African Pygmies**. New Brunswick, New Jersey: Transaction Publishers; 2014.
24. Patin E, Laval G, Barreiro LB, Salas A, Semino O, Santachiara-Benerecetti S, Kidd KK, Kidd JR, Van der Veen L, Hombert JM, Gessain A, Froment A, Bahuchet S, Heyer E, Quintana-Murci L: **Inferring the demographic history of African farmers and Pygmy hunter-gatherers using a multilocus resequencing data set**. *PLoS Genet* 2009, **5**(4):e1000448.
25. Verdu P, Austerlitz F, Estoup A, Vitalis R, Georges M, Thery S, Froment A, Le Bomin S, Gessain A, Hombert JM, Van der Veen L, Quintana-Murci L, Bahuchet S, Heyer E: **Origins and genetic diversity of Pygmy hunter-gatherers from Western Central Africa**. *Curr Biol* 2009, **19**(4):312-318.
26. Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D: **Insights into the demographic history of African Pygmies from complete mitochondrial genomes**. *Mol Biol Evol* 2011, **28**(2):1099-1110.
27. Verdu P, Becker NS, Froment A, Georges M, Grugni V, Quintana-Murci L, Hombert JM, Van der Veen L, Le Bomin S, Bahuchet S, Heyer E, Austerlitz F: **Sociocultural behavior, sex-biased admixture, and effective population sizes in Central African Pygmies and non-Pygmies**. *Mol Biol Evol* 2013, **30**(4):918-937.
28. Hsieh P, Veeramah KR, Lachance J, Tishkoff SA, Wall JD, Hammer MF, Gutenkunst RN: **Whole-genome sequence analyses of Western Central African Pygmy hunter-gatherers reveal a complex demographic history and identify candidate genes under positive natural selection**. *Genome Res* 2016, **26**(3):279-290.
29. Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglia A, Tofanelli S, Spedini G, Capelli C: **Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors**. *Mol Biol Evol* 2004, **21**(9):1673-1682.
30. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM: **The genetic structure and history of Africans and African Americans**. *Science* 2009, **324**(5930):1035-1044.
31. Patin E, Siddle KJ, Laval G, Quach H, Harmant C, Becker N, Froment A, Regnault B, Lemee L, Gravel S, Hombert JM, Van der Veen L, Dominy NJ, Perry GH, Barreiro LB, Verdu P, Heyer E, Quintana-Murci L: **The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists**. *Nat Commun* 2014, **5**:3163.
32. Becker NS, Verdu P, Froment A, Le Bomin S, Pagezy H, Bahuchet S, Heyer E: **Indirect evidence for the genetic determination of short stature in African Pygmies**. *Am J Phys Anthropol* 2011, **145**(3):390-401.
33. Jarvis JP, Scheinfeldt LB, Soi S, Lambert C, Omberg L, Ferwerda B, Froment A, Bodo JM, Beggs W, Hoffman G, Mezey J, Tishkoff SA: **Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African Pygmies**. *PLoS Genet* 2012, **8**(4):e1002641.
34. Perry GH, Foll M, Grenier JC, Patin E, Nedelec Y, Pacis A, Barakatt M, Gravel S, Zhou X, Nsobya SL, Excoffier L, Quintana-Murci L, Dominy NJ, Barreiro LB: **Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers**. *Proc Natl Acad Sci U S A* 2014, **111**(35):E3596-3603.
35. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S, Helgadottir A, Ingason A, Steinthorsdottir V, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T,

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pedersen O, Aben KK, Witjes JA, Swinkels DW, den Heijer M, Franke B, Verbeek AL, Becker DM, Yanek LR, Becker LC *et al*: **Many sequence variants affecting diversity of adult human height**. *Nat Genet* 2008, **40**(5):609-615.

36. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C, Illig T, Hackett R, Heid IM, Jacobs KB, Lyssenko V, Uda M, Boehnke M, Chanock SJ, Groop LC, Hu FB, Isomaa B, Kraft P, Peltonen L, Salomaa V, Schlessinger D, Hunter DJ, Hayes RB, Abecasis GR, Wichmann HE, Mohlke KL *et al*: **Identification of ten loci associated with height highlights new biological pathways in human growth**. *Nat Genet* 2008, **40**(5):584-591.
37. Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, Bonnycastle LL, Shen H, Timpson N, Lettre G, Usala G, Chines PS, Stringham HM, Scott LJ, Dei M, Lai S, Albai G, Crisponi L, Naitza S, Doheny KF, Pugh EW, Ben-Shlomo Y, Ebrahim S, Lawlor DA, Bergman RN, Watanabe RM, Uda M, Tuomilehto J, Coresh J, Hirschhorn JN, Shuldiner AR *et al*: **Common variants in the *GDF5-UQCC* region are associated with variation in human height**. *Nat Genet* 2008, **40**(2):198-203.
38. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Samani NJ, Shields B, Prokopenko I, Farrall M, Dominiczak A, Johnson T, Bergmann S, Beckmann JS, Vollenweider P, Waterworth DM, Mooser V, Palmer CN, Morris AD, Ouwehand WH, Zhao JH, Li S, Loos RJ, Barroso I, Deloukas P, Sandhu MS *et al*: **Genome-wide association analysis identifies 20 loci that influence adult height**. *Nat Genet* 2008, **40**(5):575-583.
39. Estrada K, Krawczak M, Schreiber S, van Duijn K, Stolk L, van Meurs JB, Liu F, Penninx BW, Smit JH, Vogelzangs N, Hottenga JJ, Willemsen G, de Geus EJ, Lorentzon M, von Eller-Eberstein H, Lips P, Schoor N, Pop V, de Keijzer J, Hofman A, Aulchenko YS, Oostra BA, Ohlsson C, Boomsma DI, Uitterlinden AG, van Duijn CM, Rivadeneira F, Kayser M: **A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation**. *Hum Mol Genet* 2009, **18**(18):3516-3524.
40. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, Yoon D, Lee MH, Kim DJ, Park M, Cha SH, Kim JW, Han BG, Min H, Ahn Y, Park MS, Han HR, Jang HY, Cho EY, Lee JE, Cho NH, Shin C, Park T, Park JW, Lee JK, Cardon L, Clarke G, McCarthy MI, Lee JY, Lee JK *et al*: **A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits**. *Nat Genet* 2009, **41**(5):527-534.
41. Johansson A, Marroni F, Hayward C, Franklin CS, Kirichenko AV, Jonasson I, Hicks AA, Vitart V, Isaacs A, Axenovich T, Campbell S, Dunlop MG, Floyd J, Hastie N, Hofman A, Knott S, Kolcic I, Pichler I, Polasek O, Rivadeneira F, Tenesa A, Uitterlinden AG, Wild SH, Zorkoltseva IV, Meitinger T, Wilson JF, Rudan I, Campbell H, Pattaro C, Pramstaller P *et al*: **Common variants in the *JAZF1* gene associated with height identified by linkage and genome-wide association analysis**. *Hum Mol Genet* 2009, **18**(2):373-380.
42. Lei SF, Yang TL, Tan LJ, Chen XD, Guo Y, Guo YF, Zhang L, Liu XG, Yan H, Pan F, Zhang ZX, Peng YM, Zhou Q, He LN, Zhu XZ, Cheng J, Liu YZ, Papasian CJ, Deng HW: **Genome-wide association scan for stature in Chinese: evidence for ethnic specific loci**. *Hum Genet* 2009, **125**(1):1-9.
43. Lei SF, Tan LJ, Liu XG, Wang L, Yan H, Guo YF, Liu YZ, Xiong DH, Li J, Yang TL, Chen XD, Guo Y, Deng FY, Zhang YP, Zhu XZ, Levy S, Papasian CJ, Hamilton JJ, Recker RR, Deng HW: **Genome-wide association study identifies two novel loci containing *FLNB* and *SBF2* genes underlying stature variation**. *Hum Mol Genet* 2009, **18**(9):1661-1669.
44. Soranzo N, Rivadeneira F, Chinappen-Horsley U, Malkina I, Richards JB, Hammond N, Stolk L, Nica A, Inouye M, Hofman A, Stephens J, Wheeler E, Arp P, Gwilliam R, Jhamai PM, Potter S, Chaney A, Ghori MJ, Ravindrarajah R, Ermakov S, Estrada K, Pols HA, Williams FM, McArdle WL, van Meurs JB, Loos RJ, Dermitzakis ET, Ahmadi KR, Hart DJ, Ouwehand WH *et al*: **Meta-**

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet* 2009, **5**(4):e1000445.

45. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G *et al*: **Hundreds of variants clustered in genomic loci and biological pathways affect human height**. *Nature* 2010, **467**(7317):832-838.
46. Kim JJ, Lee HI, Park T, Kim K, Lee JE, Cho NH, Shin C, Cho YS, Lee JY, Han BG, Yoo HW, Lee JK: **Identification of 15 loci influencing height in a Korean population**. *J Hum Genet* 2010, **55**(1):27-31.
47. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, Daigo Y, Yamamoto K, Kubo M, Nakamura Y, Kamatani N: **A genome-wide association study in 19 633 Japanese subjects identified *LHX3-QSOX2* and *IGF1* as adult height loci**. *Hum Mol Genet* 2010, **19**(11):2303-2312.
48. Croteau-Chonka DC, Marvelle AF, Lange EM, Lee NR, Adair LS, Lange LA, Mohlke KL: **Genome-wide association study of anthropometric traits and evidence of interactions with age and study year in Filipino women**. *Obesity* 2011, **19**(5):1019-1027.
49. Lanktree MB, Guo Y, Murtaza M, Glessner JT, Bailey SD, Onland-Moret NC, Lettre G, Ongen H, Rajagopalan R, Johnson T, Shen H, Nelson CP, Klopp N, Baumert J, Padmanabhan S, Pankratz N, Pankow JS, Shah S, Taylor K, Barnard J, Peters BJ, Maloney CM, Lobmeyer MT, Stanton A, Zafarmand MH, Romaine SP, Mehta A, van Iperen EP, Gong Y, Price TS *et al*: **Meta-analysis of dense genecentric association studies reveals common and uncommon variants associated with height**. *Am J Hum Genet* 2011, **88**(1):6-18.
50. Berndt SI, Gustafsson S, Magi R, Ganna A, Wheeler E, Feitosa MF, Justice AE, Monda KL, Croteau-Chonka DC, Day FR, Esko T, Fall T, Ferreira T, Gentilini D, Jackson AU, Luan J, Randall JC, Vedantam S, Willer CJ, Winkler TW, Wood AR, Workalemahu T, Hu YJ, Lee SH, Liang L, Lin DY, Min JL, Neale BM, Thorleifsson G, Yang J *et al*: **Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture**. *Nat Genet* 2013, **45**(5):501-512.
51. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Magi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AA, Westra HJ, Winkler TW *et al*: **Defining the role of common variation in the genomic and biological architecture of adult human height**. *Nat Genet* 2014, **46**(11):1173-1186.
52. Perola M, Sammalisto S, Hiekkalinna T, Martin NG, Visscher PM, Montgomery GW, Benyamin B, Harris JR, Boomsma D, Willemsen G, Hottenga JJ, Christensen K, Kyvik KO, Sorensen TI, Pedersen NL, Magnusson PK, Spector TD, Widen E, Silventoinen K, Kaprio J, Palotie A, Peltonen L: **Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci**. *PLoS Genet* 2007, **3**(6):e97.
53. Hemani G, Yang J, Vinkhuyzen A, Powell JE, Willemsen G, Hottenga JJ, Abdellaoui A, Mangino M, Valdes AM, Medland SE, Madden PA, Heath AC, Henders AK, Nyholt DR, de Geus EJ, Magnusson PK, Ingelsson E, Montgomery GW, Spector TD, Boomsma DI, Pedersen NL, Martin NG, Visscher PM: **Inference of the Genetic Architecture Underlying BMI and Height with the Use of 20,240 Sibling Pairs**. *Am J Hum Genet* 2013, **93**(5):865-875.
54. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL: **Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits**. *PLoS Genet* 2013, **9**(5):e1003520.
55. Katzmarzyk PT, Leonard WR: **Climatic influences on human body size and proportions: ecological adaptations and secular trends**. *Am J Phys Anthropol* 1998, **106**(4):483-503.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

56. Silventoinen K, Kaprio J, Lahelma E, Koskenvuo M: **Relative effect of genetic and environmental factors on body height: differences across birth cohorts among Finnish men and women**. *Am J Public Health* 2000, **90**(4):627-630.
57. N'Diaye A, Chen GK, Palmer CD, Ge B, Tayo B, Mathias RA, Ding J, Nalls MA, Adeyemo A, Adoue V, Ambrosone CB, Atwood L, Bandera EV, Becker LC, Berndt SI, Bernstein L, Blot WJ, Boerwinkle E, Britton A, Casey G, Chanock SJ, Demerath E, Deming SL, Diver WR, Fox C, Harris TB, Hernandez DG, Hu JJ, Ingles SA, John EM *et al*: **Identification, replication, and fine-mapping of loci associated with adult height in individuals of African ancestry**. *PLoS Genet* 2011, **7**(10):e1002298.
58. Carty CL, Johnson NA, Hutter CM, Reiner AP, Peters U, Tang H, Kooperberg C: **Genome-wide association study of body height in African Americans: the Women's Health Initiative SNP Health Association Resource (SHARe)**. *Hum Mol Genet* 2012, **21**(3):711-720.
59. Froment A: **Human biology and health of African rainforest inhabitants**. In: *Hunter-gatherers of the Congo Basin: cultures, histrories, and biology of African Pygmies.* Edited by Hewlett BS, First edn. New Brunswick, New Jersey: Transaction Publishers; 2014: 117-164.
60. Lemogoum D, Ngatchou W, Janssen C, Leeman M, Van Bortel L, Boutouyrie P, Degaute JP, Van de Borne P: **Effects of hunter-gatherer subsistence mode on arterial distensibility in Cameroonian pygmies**. *Hypertension* 2012, **60**(1):123-128.
61. Dietz WH, Marino B, Peacock NR, Bailey RC: **Nutritional status of Efe pygmies and Lese horticulturists**. *Am J Phys Anthropol* 1989, **78**(4):509-518.
62. Ramirez Rozzi FV, Sardi ML: **Diversity among African pygmies**. *PLoS One* 2010, **5**(10):e13620.
63. Hiernaux J: **Long-term biological effects of human migrations from the African savanna to the equitorial forest: a case study of human adaption to a hot and wet climate.** In: *Population structure and human variation.* Edited by Harrison GA. London: Cambridge University Press; 1977: 187-217.
64. Wadsworth ME, Hardy RJ, Paul AA, Marshall SF, Cole TJ: **Leg and trunk length at 43 years in relation to childhood health, diet and family circumstances; evidence from the 1946 national birth cohort**. *Int J Epidemiol* 2002, **31**(2):383-390.
65. Buschang PH, Malina RM, Little BB: **Linear growth of Zapotec schoolchildren: growth status and yearly velocity for leg length and sitting height**. *Ann Hum Biol* 1986, **13**(3):225-234.
66. Hauspie RC, Pagezy H: **Longitudinal study of growth of African babies: an analysis of seasonal variations in the average growth rate and the effects of infectious diseases on individual and average growth patterns**. *Acta Paediatr Scand Suppl* 1989, **350**:37-43.
67. de Garine I: **Adaptation biologique et bien-être psycho-culturel**. *Bulletins et Mémoires de la Société d* 1990, **2**(2):151-174.
68. Ohenjo N, Willis R, Jackson D, Nettleton C, Good K, Mugarura B: **Health of Indigenous people in Africa**. *Lancet* 2006, **367**(9526):1937-1946.
69. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, Burtt NP, Fuchsberger C, Li Y, Erdmann J, Frayling TM, Heid IM, Jackson AU, Johnson T, Kilpelainen TO, Lindgren CM, Morris AP, Prokopenko I, Randall JC, Saxena R, Soranzo N, Speliotes EK, Teslovich TM, Wheeler E, Maguire J, Parkin M, Potter S, Rayner NW, Robertson N, Stirrups K *et al*: **The Metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits**. *PLoS Genet* 2012, **8**(8):e1002793.
70. Kilpelainen TO, Zillikens MC, Stancakova A, Finucane FM, Ried JS, Langenberg C, Zhang W, Beckmann JS, Luan J, Vandenput L, Styrkarsdottir U, Zhou Y, Smith AV, Zhao JH, Amin N, Vedantam S, Shin SY, Haritunians T, Fu M, Feitosa MF, Kumari M, Halldorsson BV, Tikkanen E, Mangino M, Hayward C, Song C, Arnold AM, Aulchenko YS, Oostra BA, Campbell H *et al*: **Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile**. *Nat Genet* 2011, **43**(8):753-760.
71. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K,

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS *et al*: **Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease**. *Nat Genet* 2011, **43**(4):333-338.

72. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R *et al*: **Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis**. *Nat Genet* 2010, **42**(7):579-589.
73. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, Thorleifsson G, Zillikens MC, Speliotes EK, Magi R, Workalemahu T, White CC, Bouatia-Naji N, Harris TB, Berndt SI, Ingelsson E, Willer CJ, Weedon MN, Luan J, Vedantam S, Esko T, Kilpelainen TO, Kutalik Z, Li S, Monda KL, Dixon AL, Holmes CC, Kaplan LM, Liang L, Min JL *et al*: **Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution**. *Nat Genet* 2010, **42**(11):949-960.
74. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Magi R, Randall JC, Vedantam S, Winkler TW, Qi L, Workalemahu T, Heid IM, Steinthorsdottir V, Stringham HM, Weedon MN, Wheeler E, Wood AR, Ferreira T, Weyant RJ, Segre AV, Estrada K, Liang L, Nemesh J, Park JH, Gustafsson S, Kilpelainen TO *et al*: **Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index**. *Nat Genet* 2010, **42**(11):937-948.
75. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y *et al*: **Biological, clinical and population relevance of 95 loci for blood lipids**. *Nature* 2010, **466**(7307):707-713.
76. Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, Willenborg C, Wright B, Chen L, Li M, Salo P, Voight BF, Burns P, Laskowski RA, Xue Y, Menzel S, Altshuler D, Bradley JR, Bumpstead S, Burnett MS, Devaney J, Doring A, Elosua R, Epstein SE, Erber W, Falchi M, Garner SF, Ghori MJ, Goodall AH, Gwilliam R *et al*: **A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium**. *Nat Genet* 2009, **41**(11):1182-1190.
77. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, Pihur V, Vollenweider P, O'Reilly PF, Amin N, Bragg-Gresham JL, Teumer A, Glazer NL, Launer L, Zhao JH, Aulchenko Y, Heath S, Sober S, Parsa A, Luan J, Arora P, Dehghan A, Zhang F, Lucas G, Hicks AA *et al*: **Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk**. *Nature* 2011, **478**(7367):103-109.
78. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Magi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T *et al*: **New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk**. *Nat Genet* 2010, **42**(2):105-116.
79. Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, Lyssenko V, Bouatia-Naji N, Dupuis J, Jackson AU, Kao WH, Li M, Glazer NL, Manning AK, Luan J, Stringham HM, Prokopenko I, Johnson T, Grarup N, Boesgaard TW, Lecoeur C, Shrader P, O'Connell J, Ingelsson E, Couper DJ, Rice K, Song K, Andreasen CH, Dina C, Kottgen A *et al*: **Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge**. *Nat Genet* 2010, **42**(2):142-148.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

80. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, Bouatia-Naji N, Langenberg C, Prokopenko I, Stolerman E, Sandhu MS, Heeney MM, Devaney JM, Reilly MP, Ricketts SL, Stewart AF, Voight BF, Willenborg C, Wright B, Altshuler D, Arking D, Balkau B, Barnes D, Boerwinkle E, Bohm B, Bonnefond A, Bonnycastle LL, Boomsma DI, Bornstein SR, Bottcher Y *et al*: **Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways**. *Diabetes* 2010, **59**(12):3229-3239.
81. Newton-Cheh C, Eijgelsheim M, Rice KM, de Bakker PI, Yin X, Estrada K, Bis JC, Marciante K, Rivadeneira F, Noseworthy PA, Sotoodehnia N, Smith NL, Rotter JI, Kors JA, Witteman JC, Hofman A, Heckbert SR, O'Donnell CJ, Uitterlinden AG, Psaty BM, Lumley T, Larson MG, Stricker BH: **Common variants at ten loci influence QT interval duration in the QTGEN Study**. *Nat Genet* 2009, **41**(4):399-406.
82. Pfeufer A, Sanna S, Arking DE, Muller M, Gateva V, Fuchsberger C, Ehret GB, Orru M, Pattaro C, Kottgen A, Perz S, Usala G, Barbalic M, Li M, Putz B, Scuteri A, Prineas RJ, Sinner MF, Gieger C, Najjar SS, Kao WH, Muhleisen TW, Dei M, Happle C, Mohlenkamp S, Crisponi L, Erbel R, Jockel KH, Naitza S, Steinbeck G *et al*: **Common variants at ten loci modulate the QT interval duration in the QTSCD Study**. *Nat Genet* 2009, **41**(4):407-414.
83. Pales L: **Contribution à l'étude anthropologique des Babinga se l'AEF**. *L'Anthropologie* 1938, **48**:503-520.
84. Olivier G: **Anthropologie physique des Négrilles du Cameroun**. *Bulletin de la Société d'Études Camerounaises* 1950, **III**:113-118.
85. Gusinde M: **Die Twiden: Pygmäen und Pygmoide im tropischen Afrika**. Vienne-Stuttgart: Braumüller; 1956.
86. Cresta M: **Contributo alla conoscenza antropologica dei Babinga**. *Ric Sci 2 Ser Pt 2 Rend B* 1964, **14**:109-130.
87. van Eijk R: **Naître et croître au Cameroun**: Université d'état Utrecht; 1986.
88. Yamauchi T, Sato H, Kawamura K: **Nutritional Status, Activity Pattern, and Dietary Intake among the Baka Hunter-gatherers in the Village Camps in Cameroon**. *African Study Monographs* 2000, **21**:67-82.
89. Koppert GJA: **Methodologie de l'enquete alimentaire**. In: *Bien manger et bien vivre : anthropologie alimentaire et développement en Afrique intertropicale du biologique au social*. Edited by Froment A, de Garine I, Binam Bikoi C, Loung JF. Paris: L'Harmattan; 1996.
90. Vansina J: **New linguistic evidence and 'the Bantu expansion'**. *J Afr Hist* 1995, **36**(2):173-195.
91. Phillipson DW: **African archaeology**, 3rd edn. Cambridge: Cambridge University Press; 2005.
92. Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, Sica L, Mouguiama-Daouda P, Comas D, Tzur S, Balanovsky O, Kidd KK, Kidd JR, van der Veen L, Hombert JM, Gessain A, Verdu P, Froment A, Bahuchet S, Heyer E, Dausset J, Salas A, Behar DM: **Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers**. *Proc Natl Acad Sci U S A* 2008, **105**(5):1596-1601.
93. Montano V, Ferri G, Marcari V, Batini C, Anyaele O, Destro-Bisol G, Comas D: **The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa**. *Mol Ecol* 2011, **20**(13):2693-2708.
94. de Filippo C, Bostoen K, Stoneking M, Pakendorf B: **Bringing together linguistic and genetic evidence to test the Bantu expansion**. *Proc Biol Sci* 2012, **279**(1741):3256-3263.
95. Destro-Bisol G, Coia V, Boschi I, Verginelli F, Caglia A, Pascali V, Spedini G, Calafell F: **The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion**. *Am Nat* 2004, **163**(2):212-226.
96. Verdu P, Leblois R, Froment A, Thery S, Bahuchet S, Rousset F, Heyer E, Vitalis R: **Limited dispersal in mobile hunter-gatherer Baka Pygmies**. *Biol Lett* 2010, **6**(6):858-861.
97. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**(2):945-959.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

98. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies**. *Genetics* 2003, **164**(4):1567-1587.
99. Paschou P, Ziv E, Burchard EG, Choudhry S, Rodriguez-Cintron W, Mahoney MW, Drineas P: **PCA-correlated SNPs for structure identification in worldwide human populations**. *PLoS Genet* 2007, **3**(9):1672-1686.
100. Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, Froment A, Bodo JM, Wambebe C, Tishkoff SA, Bustamante CD: **Genome-wide patterns of population structure and admixture in West Africans and African Americans**. *Proc Natl Acad Sci U S A* 2010, **107**(2):786-791.
101. Tupman GS: **A study of bone growth in normal children and its relationship to skeletal maturation**. *J Bone Joint Surg Br* 1962, **44-B**:42-67.
102. Seeman E: **Growth in bone mass and size--are racial and gender differences in bone mineral density more apparent than real?** *J Clin Endocrinol Metab* 1998, **83**(5):1414-1419.
103. Cardon LR, Palmer LJ: **Population stratification and spurious allelic association**. *Lancet* 2003, **361**(9357):598-604.
104. Marchini J, Cardon LR, Phillips MS, Donnelly P: **The effects of human population structure on large genetic association studies**. *Nat Genet* 2004, **36**(5):512-517.
105. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E: **Variance component model to account for sample structure in genome-wide association studies**. *Nat Genet* 2010, **42**(4):348-354.
106. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Investigators A, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S: **A versatile gene-based test for genome-wide association studies**. *Am J Hum Genet* 2010, **87**(1):139-145.
107. Bassett JH, Swinhoe R, Chassande O, Samarut J, Williams GR: **Thyroid hormone regulates heparan sulfate proteoglycan expression in the growth plate**. *Endocrinology* 2006, **147**(1):295-305.
108. Campos-Xavier AB, Martinet D, Bateman J, Belluoccio D, Rowley L, Tan TY, Baxova A, Gustavson KH, Borochowitz ZU, Innes AM, Unger S, Beckmann JS, Mittaz L, Ballhausen D, Superti-Furga A, Savarirayan R, Bonafe L: **Mutations in the heparan-sulfate proteoglycan glypican 6 (GPC6) impair endochondral ossification and cause recessive omodysplasia**. *Am J Hum Genet* 2009, **84**(6):760-770.
109. Ko J, Na M, Kim S, Lee JR, Kim E: **Interaction of the ERC family of RIM-binding proteins with the liprin-alpha family of multidomain proteins**. *J Biol Chem* 2003, **278**(43):42377-42385.
110. Oishi I, Suzuki H, Onishi N, Takada R, Kani S, Ohkawara B, Koshida I, Suzuki K, Yamada G, Schwabe GC, Mundlos S, Shibuya H, Takada S, Minami Y: **The receptor tyrosine kinase Ror2 is involved in non-canonical Wnt5a/JNK signalling pathway**. *Genes Cells* 2003, **8**(7):645-654.
111. Person AD, Beiraghi S, Sieben CM, Hermanson S, Neumann AN, Robu ME, Schleiffarth JR, Billington CJ, Jr., van Bokhoven H, Hoogeboom JM, Mazzeu JF, Petryk A, Schimmenti LA, Brunner HG, Ekker SC, Lohr JL: **WNT5A mutations in patients with autosomal dominant Robinow syndrome**. *Dev Dyn* 2010, **239**(1):327-337.
112. Kim HN, Lee EJ, Jung SC, Lee JY, Chung HW, Kim HL: **Genetic variants that affect length/height in infancy/early childhood in Vietnamese-Korean families**. *J Hum Genet* 2010, **55**(10):681-690.
113. Niikawa N, Matsuura N, Fukushima Y, Ohsawa T, Kajii T: **Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency**. *J Pediatr* 1981, **99**(4):565-569.
114. Maas NM, Van de Putte T, Melotte C, Francis A, Schrander-Stumpel CT, Sanlaville D, Genevieve D, Lyonnet S, Dimitrov B, Devriendt K, Fryns JP, Vermeesch JR: **The *C20orf133* gene is disrupted in a patient with Kabuki syndrome**. *J Med Genet* 2007, **44**(9):562-569.
115. Shprintzen RJ, Goldberg RB, Young D, Wolford L: **The velo-cardio-facial syndrome: a clinical and genetic analysis**. *Pediatrics* 1981, **67**(2):167-172.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

116. Budarf ML, Konkle BA, Ludlow LB, Michaud D, Li M, Yamashiro DJ, McDonald-McGinn D, Zackai EH, Driscoll DA: **Identification of a patient with Bernard-Soulier syndrome and a deletion in the DiGeorge/velo-cardio-facial chromosomal region in 22q11.2**. *Hum Mol Genet* 1995, **4**(4):763-766.
117. Splawski I, Timothy KW, Sharpe LM, Decher N, Kumar P, Bloise R, Napolitano C, Schwartz PJ, Joseph RM, Condouris K, Tager-Flusberg H, Priori SG, Sanguinetti MC, Keating MT: **Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism**. *Cell* 2004, **119**(1):19-31.
118. Splawski I, Timothy KW, Decher N, Kumar P, Sachse FB, Beggs AH, Sanguinetti MC, Keating MT: **Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations**. *Proc Natl Acad Sci U S A* 2005, **102**(23):8089-8096; discussion 8086-8088.
119. Yiu GK, Kaunisto A, Chin YR, Toker A: **NFAT promotes carcinoma invasive migration through glypican-6**. *Biochem J* 2011, **440**(1):157-166.
120. Shimada T, Kawai T, Takeda K, Matsumoto M, Inoue J, Tatsumi Y, Kanamaru A, Akira S: **IKK-i, a novel lipopolysaccharide-inducible kinase that is related to IkappaB kinases**. *Int Immunol* 1999, **11**(8):1357-1362.
121. Sharma S, tenOever BR, Grandvaux N, Zhou GP, Lin R, Hiscott J: **Triggering the interferon antiviral response through an IKK-related pathway**. *Science* 2003, **300**(5622):1148-1151.
122. Morford LA, Forrest K, Logan B, Overstreet LK, Goebel J, Brooks WH, Roszman TL: **Calpain II colocalizes with detergent-insoluble rafts on human and Jurkat T-cells**. *Biochem Biophys Res Commun* 2002, **295**(2):540-546.
123. Indik ZK, Hunter S, Huang MM, Pan XQ, Chien P, Kelly C, Levinson AI, Kimberly RP, Schreiber AD: **The high affinity Fc gamma receptor (CD64) induces phagocytosis in the absence of its cytoplasmic domain: the gamma subunit of Fc gamma RIIIA imparts phagocytic function to Fc gamma RI**. *Exp Hematol* 1994, **22**(7):599-606.
124. Rodrigo WW, Jin X, Blackley SD, Rose RC, Schlesinger JJ: **Differential enhancement of dengue virus immune complex infectivity mediated by signaling-competent and signaling-incompetent human Fcgamma RIA (CD64) or FcgammaRIIA (CD32)**. *J Virol* 2006, **80**(20):10128-10138.
125. Were F: **The dengue situation in Africa**. *Paediatr Int Child Health* 2012, **32 Suppl 1**:18-21.
126. Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo JM, Lema G, Fu W, Nyambo TB, Rebbeck TR, Zhang K, Akey JM, Tishkoff SA: **Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse african hunter-gatherers**. *Cell* 2012, **150**(3):457-469.
127. Shriner D, Adeyemo A, Gerry NP, Herbert A, Chen G, Doumatey A, Huang H, Zhou J, Christman MF, Rotimi CN: **Transferability and fine-mapping of genome-wide associated loci for adult height across human populations**. *PLoS One* 2009, **4**(12):e8398.
128. Turnbull C: **Wayward servants, the two worlds of the African Pygmies.**: New York: The Natural History Press; 1965.
129. Fürniss S, Bahuchet S: **Existe-t-il des instruments de musique pygmées ?** In: *Ndroje balendro, musiques, terrains et disciplines (textes offerts à Simha Arom).* Edited by Dehoux V, Fürniss S, Le Bomin S, Olivier E, H. R, Voisin F, 2 tabl., 3 partitions, 14 fig. edn. Paris: Peters-Selaf; 1995: 87-109.
130. Joiris DV: **The framework of Central African hunter-gatherers and neighbouring societies**. *African Study Monographs, Suppl* 2003, **28**:57-79.
131. Le Bomin S, Mbot J: **The musical heritage of the Gabon Bongo**. *Before Farming* 2012(1):1-19.
132. Weiner JS, Lourie JA: **Practical human biology**. London: Academic Press; 1981.
133. Weir BS: **Genetic data analysis II**. Sunderland, MA: Sinauer; 1996.
134. Pemberton TJ, Wang C, Li JZ, Rosenberg NA: **Inference of unexpected genetic relatedness among individuals in HapMap Phase III**. *Am J Hum Genet* 2010, **87**(4):457-464.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

135. Bahuchet S, Guillaume H: **Aka-farmer relations in the Northwest Congo Basin.** Cambridge/Paris: Cambridge University Press/M S H; 1982.
136. Boehnke M, Cox NJ: **Accurate inference of relationships in sib-pair linkage studies**. *Am J Hum Genet* 1997, **61**(2):423-429.
137. Epstein MP, Duren WL, Boehnke M: **Improved inference of relationship for pairs of individuals**. *Am J Hum Genet* 2000, **67**(5):1219-1231.
138. R Development Core Team: **R: A language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing; 2010.
139. Jakobsson M, Rosenberg NA: **CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure**. *Bioinformatics* 2007, **23**(14):1801-1806.
140. Rosenberg NA: **DISTRUCT: a program for the graphical display of population structure**. *Mol Ecol Notes* 2004, **4**:137-138.
141. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness**. *Nat Genet* 2006, **38**(2):203-208.
142. Fisher SRA: **The correlation between relatives on the supposition of Mendelian inheritance.** *Transactions of the Royal Society of Edinburgh* 1918, **52**:399–433.
143. Ober C, Abney M, McPeek MS: **The genetic dissection of complex traits in a founder population**. *Am J Hum Genet* 2001, **69**(5):1068-1079.
144. Balding DJ, Nichols RA: **A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity**. *Genetica* 1995, **96**(1-2):3-12.
145. Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Raney BJ, Pohl A, Malladi VS, Li CH, Lee BT, Learned K, Kirkup V, Hsu F, Heitner S, Harte RA, Haeussler M, Guruvadoo L, Goldman M, Giardine BM, Fujita PA, Dreszer TR, Diekhans M, Cline MS, Clawson H, Barber GP *et al*: **The UCSC Genome Browser database: extensions and updates 2013**. *Nucleic Acids Res* 2013, **41**(Database issue):D64-69.
146. Hintze JL, Nelson RD: **Violin plots: A box plot-density trace synergism.** *Am Stat* 1998, **52**(2):181-184.
147. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ: **LocusZoom: regional visualization of genome-wide association scan results**. *Bioinformatics* 2010, **26**(18):2336-2337.
148. The International HapMap Consortium: **A second generation human haplotype map of over 3.1 million SNPs**. *Nature* 2007, **449**(7164):851-861.
149. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: **A general framework for estimating the relative pathogenicity of human genetic variants**. *Nat Genet* 2014, **46**(3):310-315.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

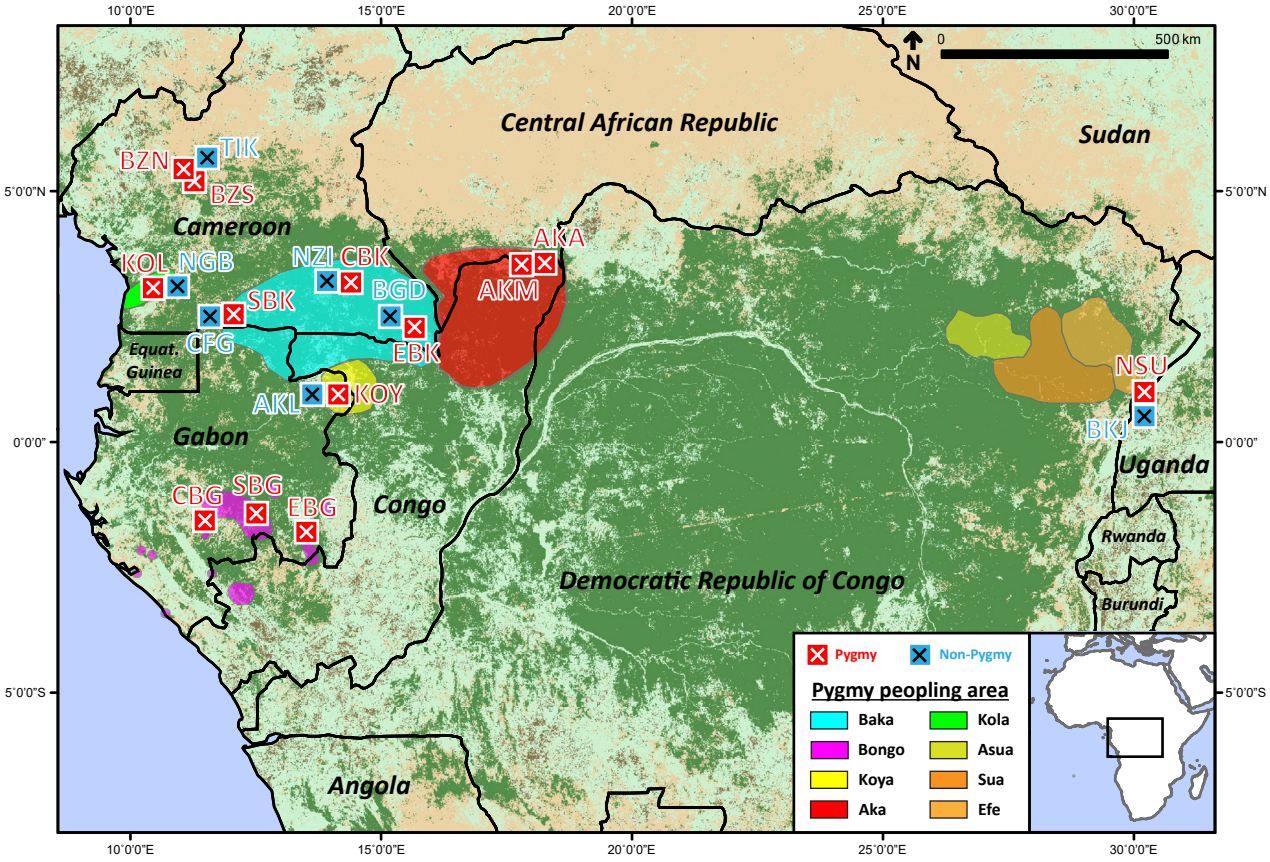

Figure 1. Sampling locations. Pygmy/non-Pygmy categorization was assessed in ethnographic field work, relying on numerous cultural criteria that do not include adult height (see **Material and Methods**). Population codes follow **Table 1**. Population ranges were inferred from ethnographic field work [7].

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

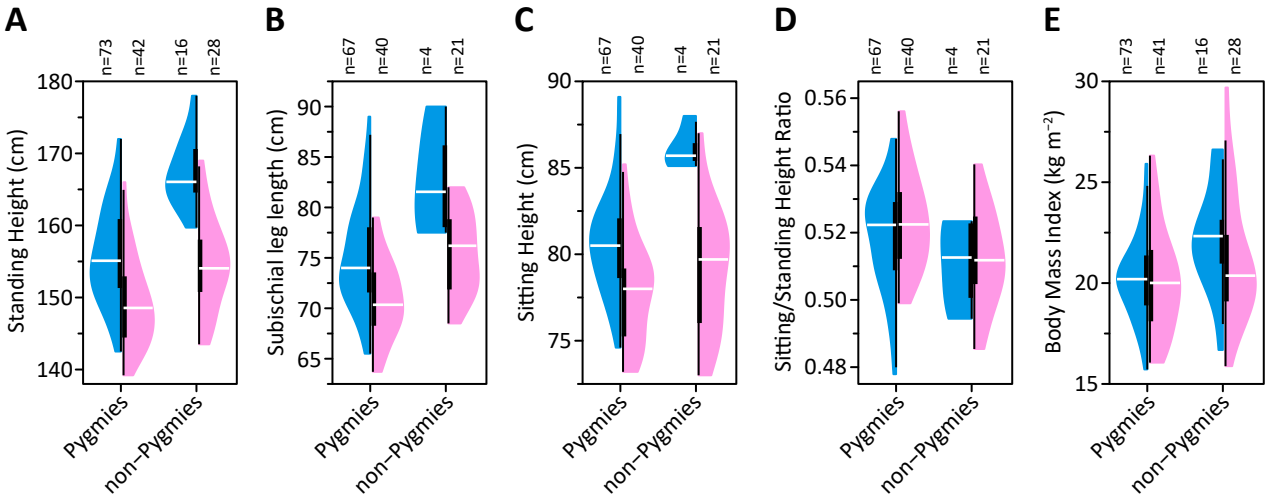

Figure 2. Distributions of standing height, sitting/standing height ratios, and BMI in Pygmies and non-Pygmies. Split violin plot [146] representations of the distributions of (**A**) standing heights, (**B**) subischial leg lengths, (**C**) sitting heights, (**D**) sitting/standing height ratios, and (**E**) BMI shown separately for males (blue) and females (pink). The number (*n*) of males and females with data available for each measurement is provided above each plot. The mean and SD of each group are reported in **Table S1** (Additional File 1). Each ‘‘violin’’ contains a vertical black line (25%–75% range) and a horizontal white line (median), with the width depicting a 90º-rotated kernel density trace.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

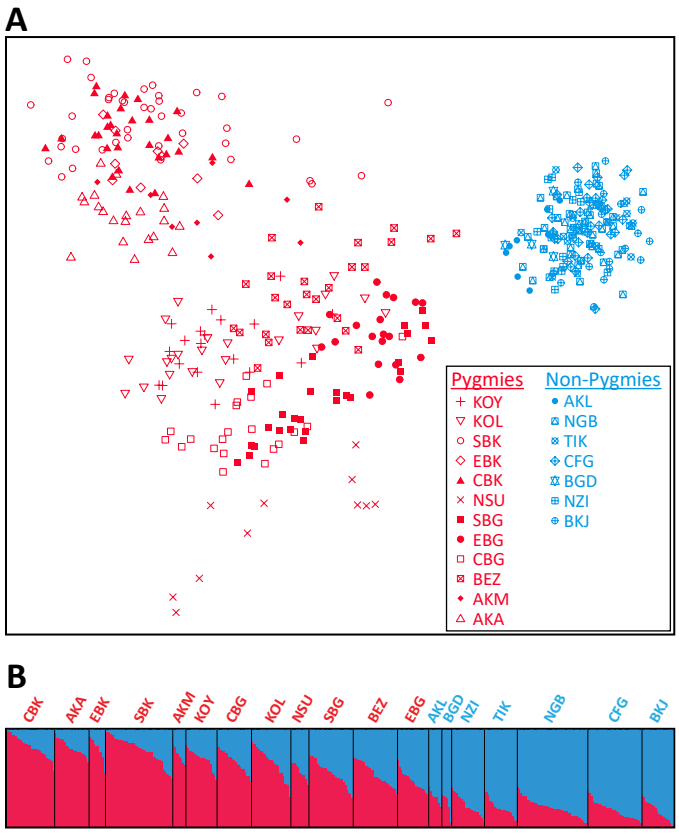

Figure 3. Genetic diversity and admixture patterns. (A) A two-dimensional multidimensional scaling (MDS) representation of allele-sharing distances (ASD) for 406 unrelated individuals in the genetic dataset. The first dimension is plotted on the horizontal axis, and the second dimension on the vertical axis. The Spearman ρ between pairwise Euclidean distances in the MDS plot and ASD matrix is 0.478. **(B)** Population structure inferred by *STRUCTURE* at K=2 for 406 unrelated individuals. Each individual is shown as a thin vertical line partitioned into two components representing inferred genotype membership proportions in the two clusters. Black vertical lines separate individuals from different populations. The most salient population structure observation is the Pygmy/non-Pygmy distinction at K=2.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

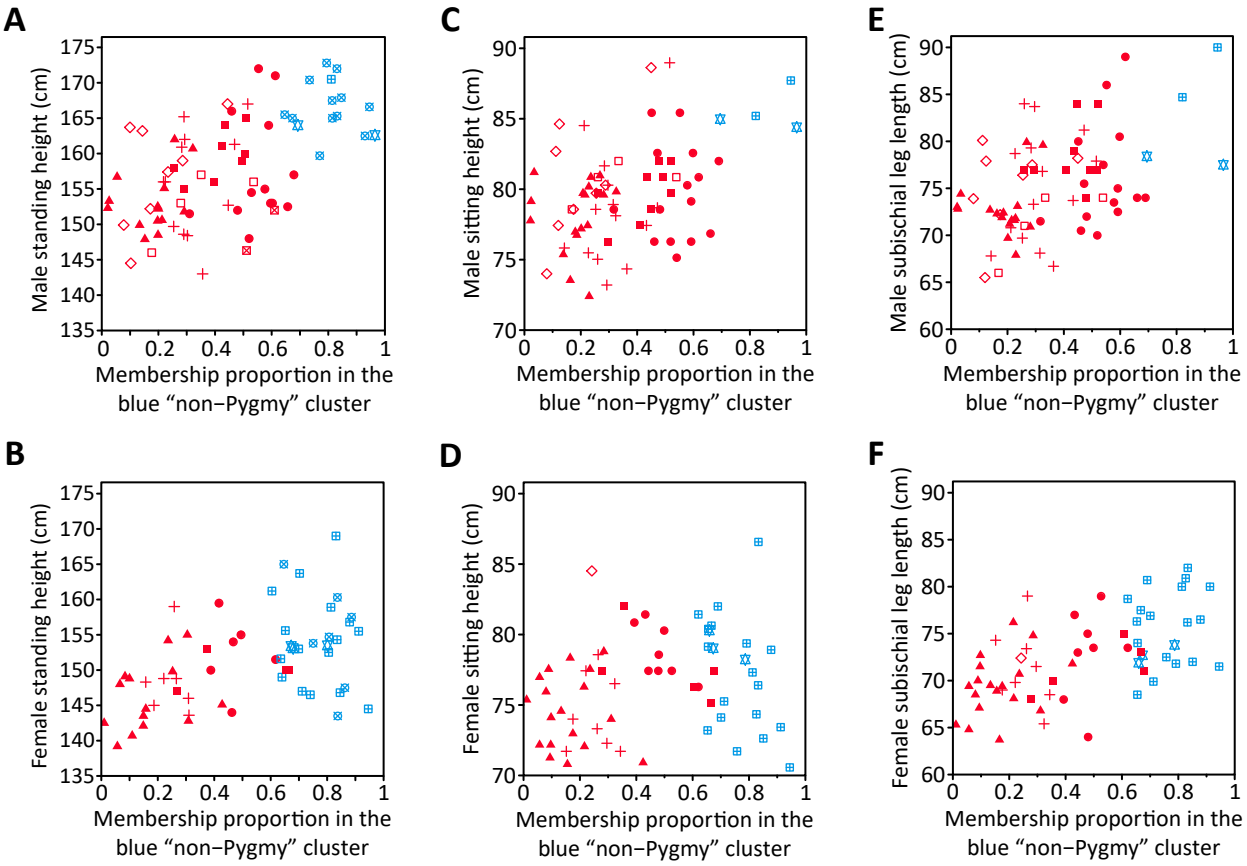

Figure 4. Relationship between non-Pygmy admixture and adult height-related traits. Scatterplots are shown of membership proportions in the blue "non-Pygmy" *STRUCTURE* cluster at K=2 (**Figure 3B**) and standing height for (**A**) 76 males (r=0.585, P=3×10^{-8}) and (**B**) 57 females (r=0.485, P=3×10^{-5}), sitting height for (**C**) 71 males (r=0.431, P=8.90×10^{-5}) and (**D**) 57 females (r=0.226, P=0.040), and subischial leg length for (**E**) 71 males (r=0.475, P=1.41×10^{-5}) and (**F**) 57 females (r=0.574, P=6.53×10^{-7}). Only individuals included in the genetic dataset with a measurement for that trait were included in each comparison. Individuals are indicated by the symbols in **Figure 3A**.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

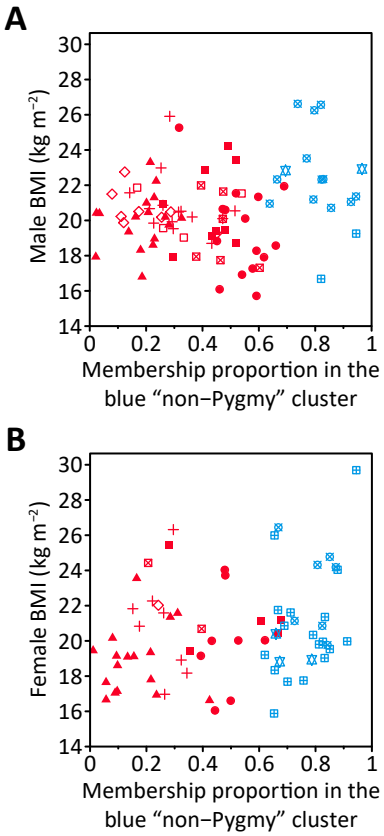

Figure 5. Relationship between non-Pygmy admixture and body mass index. Scatterplots are shown of membership proportions in the blue "non-Pygmy" *STRUCTURE* cluster at *K*=2 (**Figure 3B**) and BMI for (**A**) 89 males (r=0.166, P=0.060) and (**B**) 69 females (r=0.262, P=0.015) with both standing height and body weight measurements available in the genetic dataset were included in the comparison. Individuals are indicated by the symbols in **Figure 3A**.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

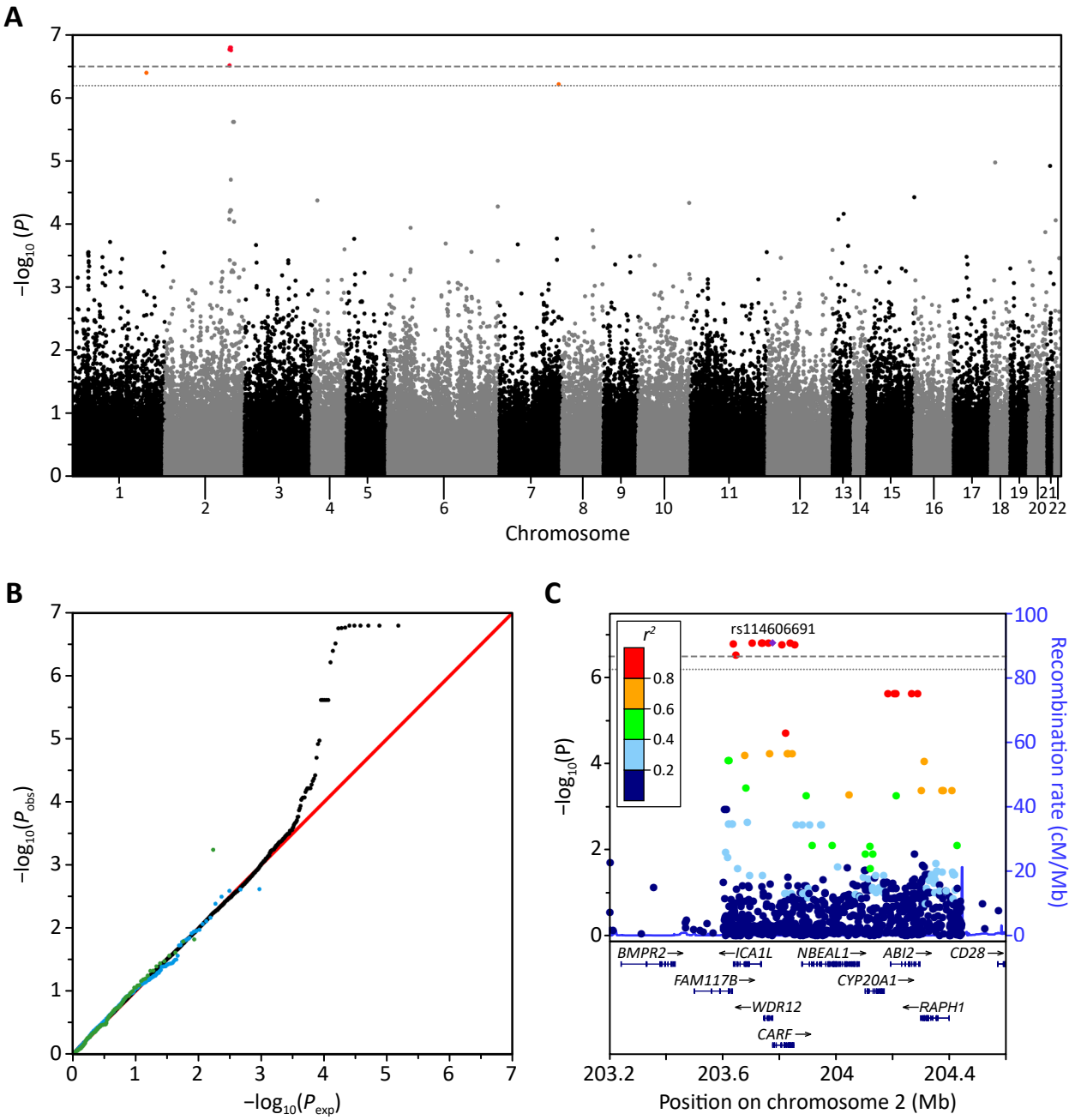

Figure 6. Association signals for Pygmy/non-Pygmy categorization. (**A**) Manhattan, and (**B**) quantile–quantile plots of the 153,798 autosomal SNPs in the genetic dataset. In **A**, the gray horizontal dashed line indicates the significance level for a Bonferroni correction at the 5% significance level while the dotted gray line indicates the significance level for a Bonferroni correction at the 10% significance level. In **B**, the identity line is shown in red, while the 126 SNPs identified by Wood *et al*. as significantly associated with standing height variation in Europeans [51] are plotted separately in green and the 949 SNPs listed in the Cardio-MetaboChip manifest [69] as associated with standing height variation are plotted separately in blue. (**C**) LocusZoom plot [147] of the genomic region surrounding the ten significant SNPs (**Table 2**). Top, -log$_{10}$-transformed *P* values at individual SNPs colored by the strength of their correlation (r^2) with the most significant SNP in the region (purple diamond), and HapMap Phase 2 recombination rates [148] depicted by the blue line. Bottom, gene locations in release hg19 of the UCSC database [145].

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

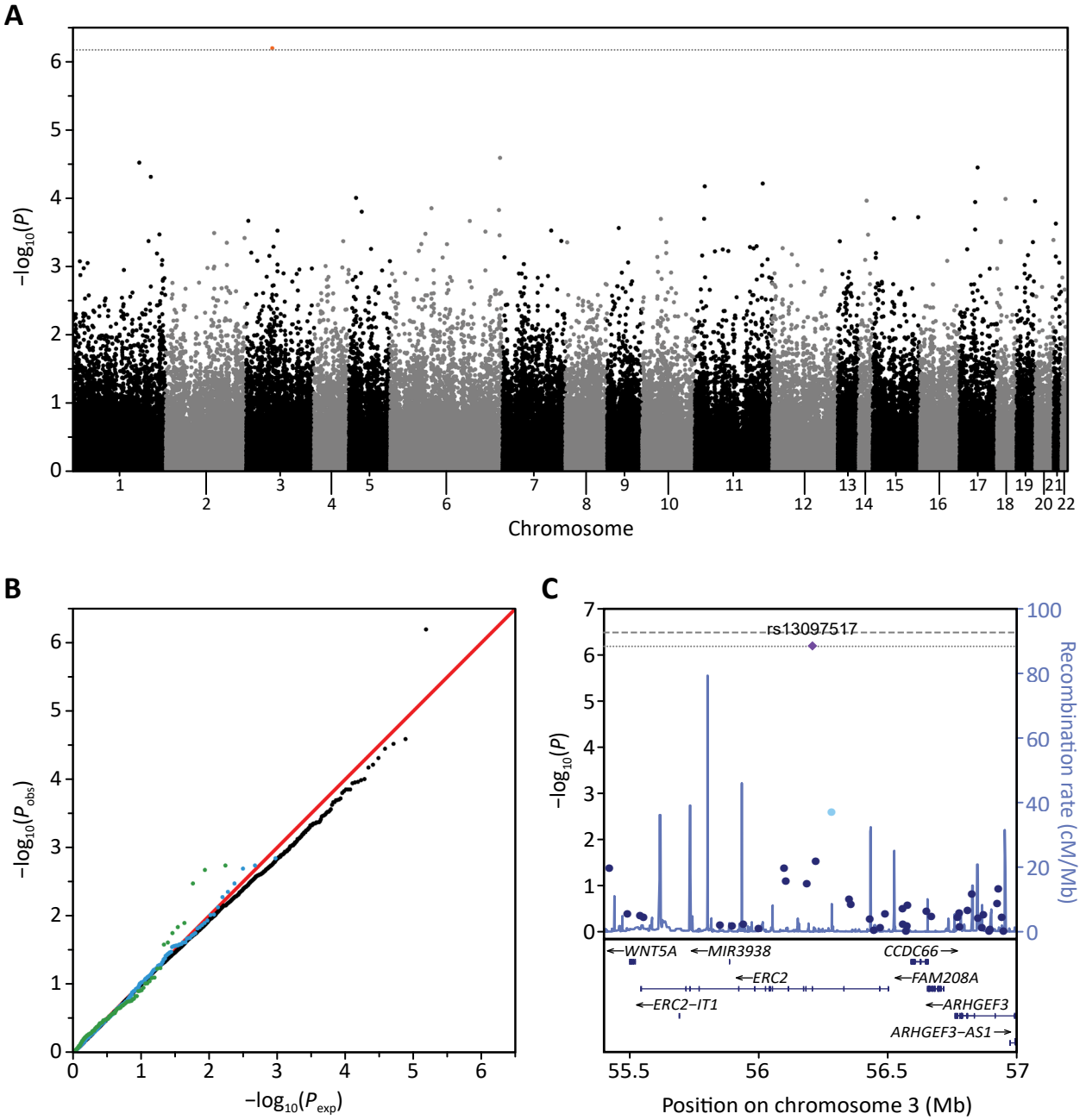

Figure 7. Association signals for adult sitting/standing height ratio. Manhattan (**A**) and quantile–quantile (**B**) plots of the 153,798 autosomal SNPs in the genetic dataset restricted to the 132 individuals with measurements for both standing and sitting height (Additional File 1: **Table S1**). (**C**) LocusZoom plot [147] of the genomic region surrounding the single marginally significant SNP (rs13097517, P=6.35×10^{-7}). The figure follows the same format as **Figure 6**.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Populations and their sample sizes.

Category	Population		Country	Trait data	Sample size	
	Code	Name			Total	Genetic dataset
Pygmies	AKA	Aka	CAR	No	26	21 (M=14; F=7)
	AKM	Aka (Mbati)	CAR	No	17	8 (M=6; F=2)
	CBK	Central/Eastern Baka	Cameroon	Yes	36	29 (M=14; F=15)
	EBK	Southeastern Baka	Cameroon	Yes	11	10 (M=10; F=0)
	SBK	Southern Baka	Cameroon	No	50	41 (M=16; F=25)
	BZN[a]	Northern Bezan	Cameroon	Yes	27	27 (M=12; F=15)
	BZS[a]	Southern Bezan	Cameroon	Yes	48	
	CBG	Central Bongo	Gabon	Yes	24	21 (M=14; F=7)
	EBG	Eastern Bongo	Gabon	Yes	30	19 (M=13; F=6)
	SBG	Southern Bongo	Gabon	Yes	34	27 (M=13; F=14)
	KOY	Koya	Gabon	Yes	25	19 (M=12; F=7)
	KOL	Kola	Cameroon	No	31	24 (M=18; F=6)
	NSU	Nsua (Efe)	Uganda	No	17	11 (M=6; F=5)
	Total				**376**	**257 (M=148; F=109)**
Non-Pygmies	AKL	Akele	Gabon	No	12	8 (M=7; F=1)
	CFG	Fang	Cameroon	No	36	33 (M=19; F=14)
	NZI	Nzime	Cameroon	Yes	23	20 (M=3; F=17)
	NGB	Ngumba	Cameroon	No	58	43 (M=20; F=23)
	TIK	Tikar	Cameroon	Yes	20	20 (M=12; F=8)
	BGD	Bangando	Cameroon	Yes	6	6 (M=3; F=3)
	BKJ	Konjo	Uganda	No	27	19 (M=8; F=11)
	Total				**182**	**149 (M=72; F=77)**

Numbers of males (M) and females (F) appear in parentheses.

[a]The two Bezan populations were merged into a combined BEZ population in the 406UNRELAT dataset.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. 12 SNPs significantly associated with Pygmy/non-Pygmy categorization.

Chr	Position (bp)	SNV ID	Alleles[a]		Closest Gene			Scaled C-Score[c]	β	P[d]
			Ref	Alt	Name	Dist. (bp)	Location			
2	203,637,779	rs115739423	**T**	C	*FAM117B*	665	Downstream	**11.48**	0.186	$\mathbf{1.69\times10^{-7}}$
2	203,647,419	rs115742265	**C**	T	*ICA1L*	–	Intronic	0.195	0.181	$\mathbf{3.02\times10^{-7}}$
2	203,704,661	rs74845106	**A**	G	*ICA1L*	–	Intronic	3.986	0.189	$\mathbf{1.57\times10^{-7}}$
2	203,737,851	rs75394200	**A**	G	*WDR12*	1,143	Downstream	1.190	0.189	$\mathbf{1.57\times10^{-7}}$
2	203,741,115	rs77148329	**T**	C	*WDR12*	–	3'-UTR	0.219	0.189	$\mathbf{1.57\times10^{-7}}$
2	203,762,323	rs114606691	**C**	G	*WDR12*	–	Intronic	**11.46**	0.189	$\mathbf{1.57\times10^{-7}}$
2	203,776,533	rs115474205	**T**	G	*CARF*	–	Intronic	8.407	0.189	$\mathbf{1.57\times10^{-7}}$
2	203,809,373	rs76849941	**C**	T	*CARF*	–	Intronic[b]	8.519	0.188	$\mathbf{1.72\times10^{-7}}$
2	203,839,087	rs78815956	**C**	T	*CARF*	–	Synonymous	**15.93**	0.189	$\mathbf{1.57\times10^{-7}}$
2	203,855,310	rs7561827	**T**	A	*CARF*	4,250	Downstream	5.831	0.188	$\mathbf{1.74\times10^{-7}}$
1	206,647,787	rs1539243	T	**C**	*IKBKE*	–	5'-UTR	1.271	-0.102	3.96×10^{-7}
7	150,594,658	rs12703105	**C**	A	*ABP1*	36,066	Downstream	2.700	0.134	6.01×10^{-7}

[a]Ancestral allele is shown in bold.

[b]Located within a CTCF binding site (Ensembl Regulatory Feature ENSR00001046703).

[c]Scaled C-scores ≥10 (among highest 10% of all observed scores [149]) are shown in **bold**.

[d]*P*-values in **bold** are significant after Bonferroni correction at α=5% otherwise significant at α=10%.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Seven genes significantly associated with Pygmy/non-Pygmy categorization.

Gene	Genomic position			Number of SNPs	*VEGAS* P_{perm}
	Chr	Begin (bp)	End (bp)		
DLGAP1	18	3,496,029	4,455,266	165	$<10^{-7}$
GPC6	13	93,879,077	95,060,273	111	$<10^{-7}$
MACROD2	20	13,976,145	16,033,841	183	$<10^{-7}$
CACNA1C	12	2,162,415	2,807,115	76	1.00×10^{-7}
PPP2R2B	5	145,969,066	146,461,083	24	5.00×10^{-7}
GP1BB	22	19,704,742	19,712,297	6	6.00×10^{-7}
CLDN10	13	96,085,852	96,232,010	10	1.50×10^{-6}

Gene names shown in **bold** were also significantly associated with adult height-based traits (**Table 3**).

For each gene, begin and end positions were defined using the union of the ranges of all isoforms.

bioRxiv preprint doi: https://doi.org/10.1101/187369; this version posted September 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4. 19 genes significantly associated with adult height.

Gene	Genomic position			Number of SNPs	*VEGAS* P_{perm}[c]			
	Chr	Begin (bp)	End (bp)		Standing height	Sitting height	Sit/stand height ratio	Subischial leg length
GPC6	13	93,879,077	95,060,273	111	**$<10^{-7}$**	**$<10^{-7}$**	**$<10^{-7}$**	**$<10^{-7}$**
DLGAP1	18	3,496,029	4,455,266	165	**$<10^{-7}$**	**$<10^{-7}$**	**$<10^{-7}$**	**$<10^{-7}$**
MACROD2	20	13,976,145	16,033,841	183	**$<10^{-7}$**	**$<10^{-7}$**	**$<10^{-7}$**	**3.00×10^{-7}**
CACNA1C	12	2,162,415	2,807,115	76	2.46×10^{-5}	**9.00×10^{-7}**	**$<10^{-7}$**	**3.00×10^{-7}**
DSCAM	21	41,384,342	42,219,039	75	**$<10^{-7}$**	0.001	**$<10^{-7}$**	**$<10^{-7}$**
DPYD	1	97,543,299	98,386,615	30	0.085	1.00×10^{-5}	**$<10^{-7}$**	**$<10^{-7}$**
ERC2	3	55,542,335	56,502,391	36	0.001	0.021	**2.00×10^{-7}**	**8.00×10^{-7}**
XXYLT1	3	194,789,012	194,991,895	21	**$<10^{-7}$**	**$<10^{-7}$**	0.161	0.003
MAGI2[a]	7	77,646,373	79,082,890	132	**3.00×10^{-7}**	**$<10^{-7}$**	0.009	0.002
NRG1	8	31,497,267	32,622,558	87	2.33×10^{-5}	1.83×10^{-5}	**$<10^{-7}$**	**$<10^{-7}$**
FOXN3	14	89,622,515	90,085,494	63	**9.00×10^{-7}**	**9.00×10^{-7}**	7.82×10^{-4}	1.83×10^{-4}
RDH14	2	18,735,988	18,770,846	10	**$<10^{-7}$**	0.003	0.073	0.003
NT5C1B	2	18,735,988	18,770,846	10	**$<10^{-7}$**	0.003	0.071	0.003
NPPA	1	11,900,375	11,907,673	618	1.67×10^{-5}	**$<10^{-7}$**	0.002	0.001
PACRG	6	163,148,163	163,736,524	48	0.002	**$<10^{-7}$**	3.70×10^{-6}	0.001
TNFRSF6B	20	62,289,162	62,330,051	24	0.012	**$<10^{-7}$**	3.00×10^{-5}	0.019
ZNF32	10	44,124,264	44,170,147	24	0.049	0.035	**1.50×10^{-6}**	5.61×10^{-5}
SHANK2[b]	11	70,313,960	70,935,808	27	2.72×10^{-4}	0.031	0.003	**$<10^{-7}$**
CCDC169	13	36,742,344	36,871,992	8	0.009	0.535	1.20×10^{-4}	**$<10^{-7}$**

[a]Associated with adult standing height variation in Europeans [112].

[b]Within 1 Mb of genomic regions identified by Perry *et al*. as associated with the short stature phenotype of East Central African Batwa Pygmies [34].

[c]Permutation *P*-values shown in **bold** are significant after Bonferroni correction (α=1%).

Gene names shown in **bold** were also significantly associated with Pygmy/non-Pygmy categorization (**Table 3**).

For each gene, begin and end positions were defined using the union of the ranges of all isoforms.