Please cite this article in press as: Lopez et al., Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest, Current Biology (2019), https://doi.org/10.1016/j.cub.2019.07.013

Current Biology
Report

Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest

Marie Lopez,[1,2] Jeremy Choin,[1] Martin Sikora,[3] Katherine Siddle,[1,11] Christine Harmant,[1] Helio A. Costa,[4] Martin Silvert,[1,2] Patrick Mouguiama-Daouda,[5] Jean-Marie Hombert,[6] Alain Froment,[7] Sylvie Le Bomin,[8] George H. Perry,[9] Luis B. Barreiro,[10] Carlos D. Bustamante,[4] Paul Verdu,[8] Etienne Patin,[1,12,*] and Lluís Quintana-Murci[1,12,13,*]
[1]Human Evolutionary Genetics Unit, Institut Pasteur, UMR2000, CNRS, Paris 75015, France
[2]Sorbonne Universités, Ecole Doctorale Complexité du Vivant, 75005 Paris, France
[3]Centre for GeoGenetics, University of Copenhagen, 1350 Copenhagen, Denmark
[4]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA 94305, USA
[5]Laboratoire Langue, Culture et Cognition (LCC), Université Omar Bongo, 13131 Libreville, Gabon
[6]CNRS UMR 5596, Université Lumière-Lyon 2, 69007 Lyon, France
[7]Institut de Recherche pour le Développement UMR 208, Muséum National d'Histoire Naturelle, 75005 Paris, France
[8]UMR7206, Muséum National d'Histoire Naturelle, CNRS, Université Paris Diderot, Paris 75016, France
[9]Departments of Anthropology and Biology, Pennsylvania State University, University Park, PA 16802, USA
[10]Department of Medicine, The University of Chicago, Chicago, IL 60637, USA
[11]Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[12]Senior author
[13]Lead Contact
*Correspondence: etienne.patin@pasteur.fr (E.P.), quintana@pasteur.fr (L.Q.-M.)
https://doi.org/10.1016/j.cub.2019.07.013

SUMMARY

African rainforests support exceptionally high biodiversity and host the world's largest number of active hunter-gatherers [1–3]. The genetic history of African rainforest hunter-gatherers and neighboring farmers is characterized by an ancient divergence more than 100,000 years ago, together with recent population collapses and expansions, respectively [4–12]. While the demographic past of rainforest hunter-gatherers has been deeply characterized, important aspects of their history of genetic adaptation remain unclear. Here, we investigated how these groups have adapted—through classic selective sweeps, polygenic adaptation, and selection since admixture—to the challenging rainforest environments. To do so, we analyzed a combined dataset of 566 high-coverage exomes, including 266 newly generated exomes, from 14 populations of rainforest hunter-gatherers and farmers, together with 40 newly generated, low-coverage genomes. We find evidence for a strong, shared selective sweep among all hunter-gatherer groups in the regulatory region of *TRPS1*—primarily involved in morphological traits. We detect strong signals of polygenic adaptation for height and life history traits such as reproductive age; however, the latter appear to result from pervasive pleiotropy of height-associated genes. Furthermore, polygenic adaptation signals for functions related to responses of mast cells to allergens and microbes, the IL-2 signaling pathway, and host interactions with viruses support a history of pathogen-driven selection in the rainforest. Finally, we find that genes involved in heart and bone development and immune responses are enriched in both selection signals and local hunter-gatherer ancestry in admixed populations, suggesting that selection has maintained adaptive variation in the face of recent gene flow from farmers.

RESULTS

Exome Sequencing Dataset and Population Structure

African rainforest hunter-gatherers (RHGs)—historically grouped under the term "Pygmies"—live along the dense tropical rainforests of central Africa, in the western and eastern part of the Congo Basin [1–3]. Genetic studies have deeply investigated the demographic history of these groups, characterized by long-term isolation since the Upper Paleolithic and substantial admixture with neighboring Bantu-speaking farmers in the last 1,000 years [4–12]. However, their adaptive history has received less attention. Natural selection studies in RHGs have primarily focused on small adult body size as the only trait characterizing the "pygmy" phenotype [13–20], and used SNP genotyping data [14, 15, 19–21] or whole-genome/exome sequencing of a few individuals or populations [4, 6, 18, 22, 23].

To understand human genetic adaptation to the rainforest, we generated and analyzed whole-exome sequencing data (~40× coverage) for seven RHG groups from Cameroon, Gabon, and Uganda, as well as, for comparison purposes, seven sedentary groups of Bantu-speaking agriculturalists (AGRs) (Figure 1A; Table S1). After quality filters, we obtained a final dataset of 566 individuals (298 RHGs and 268 AGRs), consisting of 266 newly generated exomes that were analyzed with 300 previously reported exomes [4] (Figure S1).

Genetic differentiation among RHG groups was higher than that between RHGs and AGRs (among-RHG, F_{ST} = 0.025; among-western RHG, F_{ST} = 0.021; RHG-AGR, F_{ST} = 0.017;

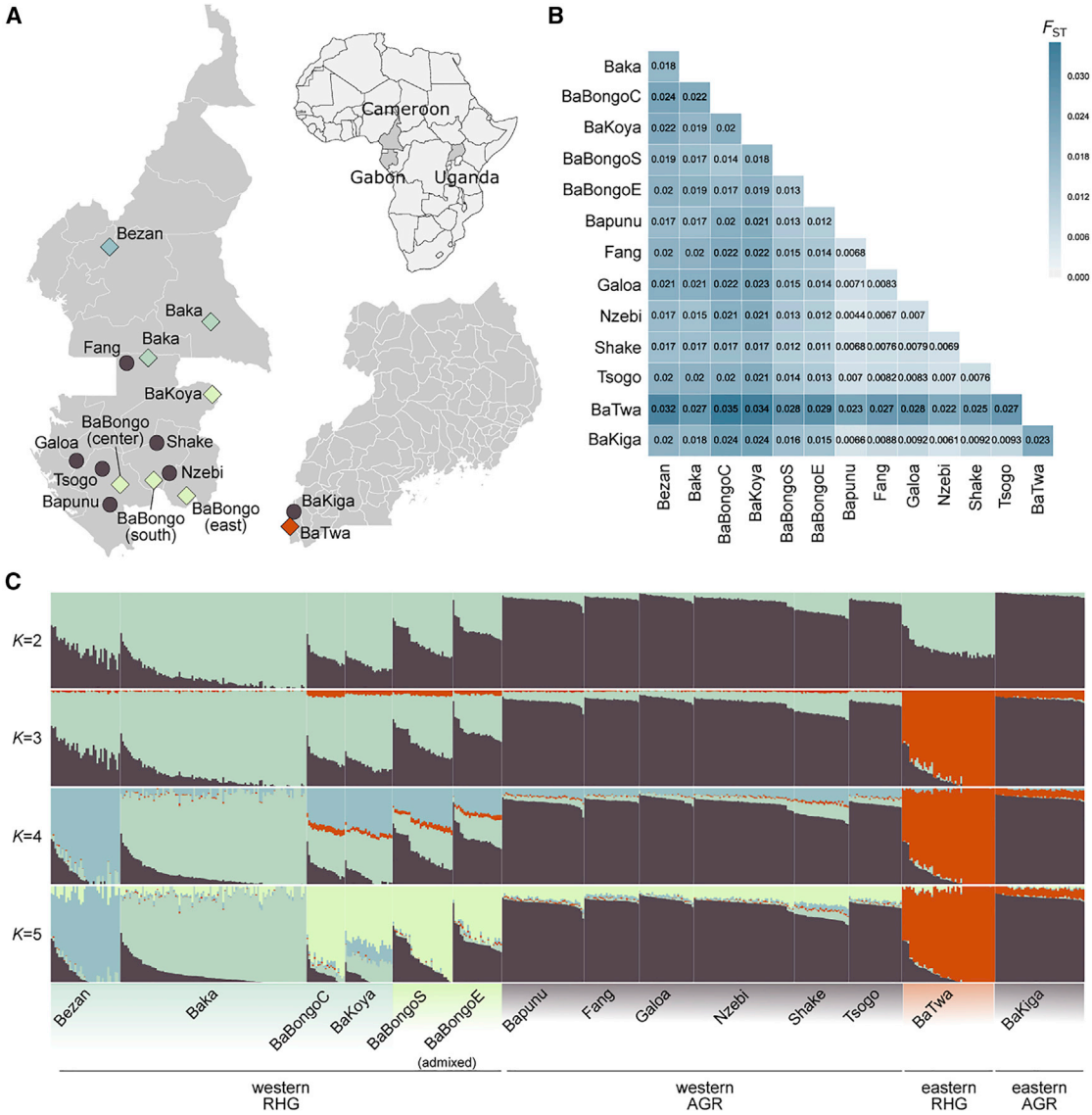

Figure 1. Location, Genetic Differentiation, and Structure of Central African Populations

(A) Geographic location of the populations analyzed. Populations of rainforest hunter-gatherers (diamonds) and neighboring farmers (circles) originating from the three countries are shown in the map of Africa. Colors indicate the dominant membership in each population, based on ADMIXTURE results (C).

(B) Levels of genetic differentiation between populations measured by pairwise F_{ST} calculated on the exome data.

(C) Cluster membership proportions estimated by ADMIXTURE on the merged exome and SNP array data. Cross-validation values were lowest at $K = 5$ clusters.

(B and C) BaBongoC, BaBongoS, and BaBongoE stand for BaBongo populations from the center, south, and east of Gabon, respectively.

See also Figure S1 and Table S1.

among-AGR, $F_{ST} = 0.007$; Figure 1B). To increase SNP density, particularly in the non-coding genome, we combined the exome data with SNP array data for the same individuals [12, 24, 25], yielding a total of 1,253,548 SNPs. When using ADMIXTURE [26] on the dataset pruned for allele frequency (MAF > 5%) and linkage disequilibrium ($r^2 < 0.5$), RHGs separated into four clusters at $K = 5$ (Figure 1C), corresponding to Bezan, Baka, BaBongo and BaKoya, and BaTwa groups. As previously observed [5, 12, 14, 24], membership proportions to the cluster assigned to AGRs were non-negligible and similar among RHG groups (~4%–9%; Table S1), with the exception of the BaBongo of east and south Gabon, who presented high AGR proportions (~43% [SD = 11%] and ~24% [SD = 17%], respectively). Membership proportions to the cluster assigned to RHGs were also non-negligible among AGRs (~10%–30%). Our results show that RHG populations are highly structured, emphasizing the importance of considering these groups separately in subsequent analyses.

Searching for Signals of Local Genetic Adaptation in Central Africans

For all natural selection analyses, we increased SNP density to 9,129,103 high-quality variants (MAF > 1%), through genotype imputation using (1) newly generated whole genomes from

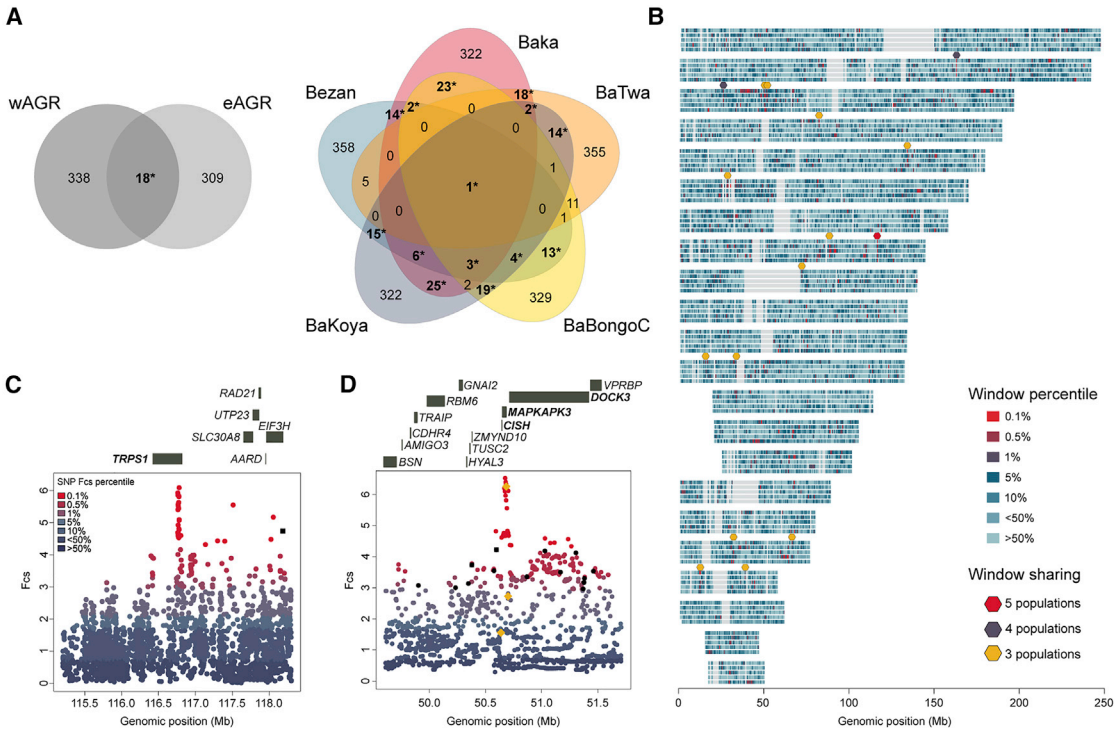

Figure 2. Shared Signals of Classic Sweeps among Rainforest Hunter-Gatherers

(A) Number of candidate windows for classic sweeps (i.e., windows with proportions of outlier SNPs among the 1% highest of the genome) common to western and eastern AGR populations (wAGR and eAGR), as well as common to RHG populations. p values obtained based on 10,000 resamples are shown: *$p < 10^{-4}$.

(B) Genome-wide map of classic sweep signals in RHG groups. The autosomes of each of the five RHG populations (from top to bottom: Bezan, Baka, lowly admixed BaBongo, BaKoya, and BaTwa) are shown. Colored dots indicate genomic regions that are common to at least three RHG populations.

(C) Selective sweep signal at the locus containing the *TRPS1* gene (chr8:116702422-116802422) in the Baka RHGs.

(D) Selective sweep signal at the locus containing *CISH*, *MAPKAPK3*, and *DOCK3* genes (chr3:50610197-50710197 and chr3:50660197-50760197) in the BaTwa RHGs.

(C and D) Dot colors indicate SNP F_{CS} percentiles, black squares indicate non-synonymous mutations, and black dots indicate eQTLs (q value < 0.005) [33]. eQTLs of *MAPKAPK3* (rs107457 and rs9879397) and *DOCK3* (rs12629788) are shown as yellow diamonds. Not all genes of the genomic region are shown for convenience.

See also Figures S2 and S3 and Data S1.

20 RHG Baka and 20 AGR Nzébi from Gabon (5–6× coverage) and (2) the 1000 Genomes Phase 3 panel [27] (STAR Methods; Figure S1). We focused on the five RHG populations presenting the lowest average levels of AGR ancestry and analyzed the highly admixed RHG groups differently (see Recent Genetic Adaptation of Admixed Rainforest Hunter-Gatherers). To identify signals of strong sweeps, we searched for variants with both high allele frequency and extended haplotype homozygosity in RHGs, relative to AGRs (STAR Methods). Genome-wide ranks of PBS [28] and XP-EHH [29] were combined into a Fisher's score (F_{CS}), and to reduce false positives, candidate regions were defined as 100-kb windows with the 1% highest proportion of outlier SNPs of the genome.

We first scanned the genomes of AGR populations (Figure S2), the evolutionary history of whom is well characterized [24, 29–32]. We found 18 candidate regions for positive selection in both western and eastern AGRs, while only ~3.5 were expected to be shared if candidate loci were false positives (10,000 random samples; resampling $p < 10^{-4}$) (Figure 2A; Data S1). Among candidates, we replicated, for example, the signal encompassing the *LARGE* gene, involved in Lassa virus infectivity [34]. These results provide evidence that the genomic regions detected by our approach are enriched in true signals.

A Strong, Shared Selective Sweep at *TRPS1* across All Hunter-Gatherer Groups

Our search for sweeps in RHGs identified candidates that were shared by RHG groups more than expected by chance (resampling $p < 10^{-4}$) (Figure 2A; Data S1). Remarkably, we identified a single genomic region that exhibits sweep signals in all RHG populations, but not in AGRs (Figures 2A–2C and S3). This region lies upstream of the 5′ UTR of *TRPS1*, which encodes a transcription factor (TF) with multiple pleiotropic effects, including skeletal development and inflammatory T_H17 cell differentiation [35–37]. The six variants presenting the highest frequency differences between RHGs and AGRs (Data S1) define a 5,777 bp region that contains a primate-specific THE1B endogenous retrovirus sequence, known to control the expression of nearby genes [38]. Given the high expression of *TRPS1* in monocytes [39], we analyzed published RNA sequencing (RNA-seq) data from monocytes of individuals of central African ancestry to test if candidate variants affect *TRPS1* expression

[40]. A highly differentiated variant that falls within the THE1B fragment was associated with increased expression of a short, non-canonical *TRPS1* transcript upon immune stimulation (rs111351287; regression $p = 5 \times 10^{-6}$). These findings suggest that the most robust signal of adaptation to the African rainforest can be ascribed to *TRPS1*, possibly in relation with variation in morphological and/or immunological traits.

Detection of Other Classic Sweep Signals in Rainforest Hunter-Gatherers

Other selective sweep signals were specific to a smaller number of RHG groups (Figure S2; Data S1). These include the known 150-kb region encompassing *CISH*, *MAPKAPK3*, and *DOCK3* [6, 14], which we show here to be shared among western and eastern RHGs (Baka, BaKoya, and BaTwa). We searched the GTEx database [33] for regulatory variation at these genes (eQTLs) and found two *cis*-eQTLs for *MAPKAPK3* (rs107457 and rs9879397), one for *DOCK3* (rs12629788), and none for *CISH* (Data S1). Selection scores at these eQTLs were among the highest of the region, particularly for *MAPKAPK3* (Figure 2D), which affects hepatitis C virus (HCV) infectivity [41].

We also detected two contiguous regions at the *IFIH1* locus [18], which present strong enrichments in selection scores that are shared by all western RHG groups. Candidate variation at this locus (rs12479043) controls the expression of the nearby *FAP* gene [33], which regulates fibroblast and myofibroblast growth and wound healing during chronic inflammation [42]. We also identified two windows—shared by Bezan, Baka, and BaKoya—encompassing *RASGEF1B*, whose expression is induced in macrophages by lipopolysaccharide, a membrane component of Gram-negative bacteria [43]. Finally, we found a window in the Bezan, BaBongo, and BaKoya that overlaps *PITX1*, recently identified as a selection candidate in RHGs [22]. PITX1 modulates the core development of limb [44], is associated with height variation [45], acts as an early TF in the developing pituitary gland [46], and regulates interferon-α virus induction [47]. These results support the hypothesis that development and immunity are key traits in local adaptation to the rainforest.

Evidence for Polygenic Selection Favoring the "Pygmy" Phenotype

Given the polygenic nature of most adaptive traits [48, 49], we searched for evidence of polygenic adaptation focusing on 12 candidate quantitative traits. These include height, body mass index, skin pigmentation, life history traits, and immune cell counts, the genetic architectures of which have been extensively studied [50]. We compared the distribution of mean F_{CS} scores in non-overlapping, 100-kb genomic windows containing trait-associated SNPs to that of randomly sampled windows, accounting for SNP density, LD levels, and background selection (STAR Methods). Stature-related traits showed the most significant polygenic selection signals, in all RHG groups (adjusted $p < 0.05$) while being non-significant in AGRs (Figure 3A). Life-history traits related to reproduction also exhibited selection signals in various RHG groups, consistent with the proposed adaptive nature of early reproduction in RHGs [51, 52]. Furthermore, we replicated selection signals for cardiovascular traits in the BaTwa (adjusted $p < 0.001$) [23]. Notably, we found significant signals in "Leukocyte count" in the Baka and the BaBongo (adjusted $p < 0.05$), suggesting polygenic adaptation related to immunity.

We next examined whether signals of polygenic selection could result from pleiotropy; e.g., advantageous height-associated variants affect other correlated traits [49]. Using the UK Biobank dataset [50], we computed the genetic correlations from LD-score regressions between "Standing height" and the remaining traits, and found significant correlations for eight of them (STAR Methods; Data S1). For these, we repeated the analysis after excluding windows associated with "Standing height" or "Comparative height at age 10," and the significance of selection signals was lost or dramatically reduced (Figures 3B and S4). Conversely, when excluding windows associated with non-height traits (e.g., reproduction-related traits), we found that "Standing height" was still significant in four RHG populations (adjusted $p < 0.05$) (Figure 3C). These results show that height has been an adaptive trait in RHGs, resulting in spurious polygenic selection signals for other correlated traits because of pleiotropy.

Evidence of Pervasive Pathogen-Driven Selection in the Equatorial Rainforest

We further investigated genomic signatures of polygenic adaptation, by searching for excesses in mean F_{CS} among windows related to 5,354 gene ontology (GO) terms [53] (STAR Methods). We detected 38 terms that were significant in at least three RHG groups, but not in AGRs (Figure 3D; Data S1). Among these, we found positive regulation of "mast cell degranulation" and "the phosphatidylinositol 3-kinase (PI3K) pathway" (false discovery rate [FDR] $p < 5\%$). Recognition by mast cells of allergens and antigens induces degranulation, a process mediated by the PI3K pathway that results in inflammation and allergy [54]. Enrichments were also found in the IL-2 signaling pathway, which activates the PI3K pathway and regulates immune tolerance [55]. All enrichments remained significant after removing windows associated with height (FDR $p < 5\%$), excluding potential pleiotropic effects. To gain further insights into pathogen-driven selection, we next focused on 1,553 innate immunity genes (IIGs) [56] and 1,257 genes encoding virus-interacting proteins (VIPs) [57]. We found significant enrichments in selection signals for both gene sets in RHGs, but not in AGRs, in particular for VIPs interacting with double-stranded DNA (dsDNA) and single-stranded RNA (ssRNA) viruses (FDR $p < 5\%$; Table S2; Data S1). These results collectively support the notion that pathogens have been a major driver of local adaptation in the African rainforest.

Recent Genetic Adaptation of Admixed Rainforest Hunter-Gatherers

To search for evidence of recent selection in RHG since their admixture with AGRs, we focused on the highly admixed BaBongo (Figure 1C) and performed local ancestry inference with RFMix [58], using as putative parental populations western RHG and AGR individuals with the lowest AGR and RHG membership proportions, respectively (STAR Methods). Six contiguous windows on chromosome 1 showed both evidence of selection (i.e., top 1% of the proportion of outlier SNPs) and an excess of RHG local ancestry (i.e., higher than the genome-wide average + 2 SD) in admixed RHG (Figures 4A and S2; Data S1). Among the

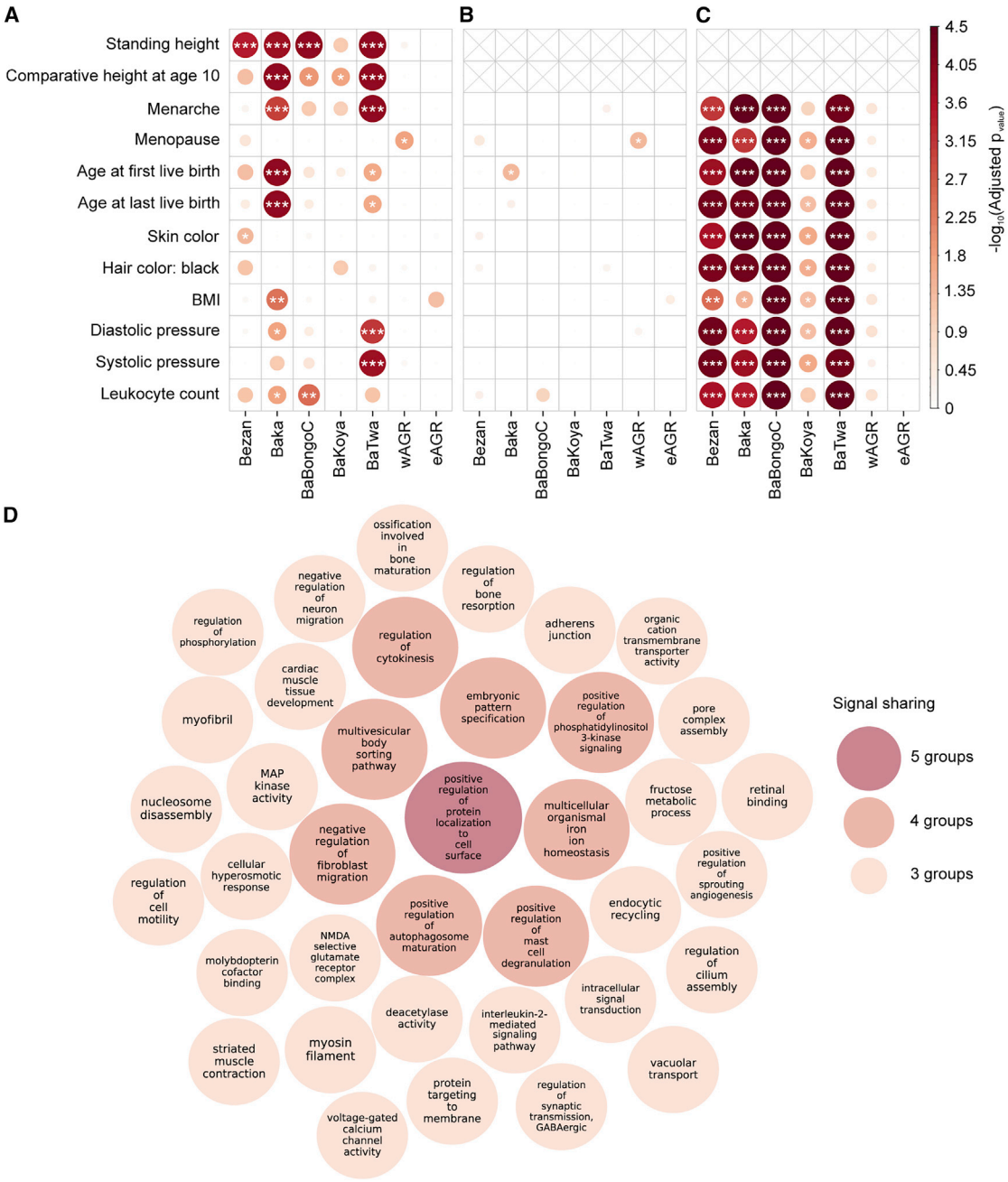

Figure 3. Signals of Polygenic Selection in African Rainforest Hunter-Gatherers

(A) Signals of polygenic selection for 12 candidate quantitative traits, based on higher mean F_{CS} of trait-associated windows relative to genome-wide expectations.

(B) Signals of polygenic selection for the candidate quantitative traits, based on higher mean F_{CS} of trait-associated windows relative to genome-wide expectations, after removing windows associated with "Standing height" and "Comparative height at age 10." Loss of significance was not explained by the reduced number of windows tested (Figure S4).

(C) Signals of polygenic selection for "Standing height," based on higher mean F_{CS} of trait-associated windows relative to genome-wide expectations, after removing windows associated with each of the remaining quantitative traits.

(A–C) Color gradient and circle sizes are proportional to $-\log_{10}$(adjusted p) with adjusted *p < 0.05, **p < 0.01, and ***p < 0.001. Multiple testing corrections were performed using the Benjamini-Hochberg method. wAGR and eAGR stand for western and eastern AGR groups. Signals were generally stronger in Baka and BaTwa RHGs, probably because of their larger sample size.

(D) Gene Ontology (GO) terms enriched in selection scores (FDR p < 5%) in RHG, but not in AGR, populations, considering the window mean F_{CS} as selection score. Circle color and size indicate the number of RHG populations that show significant evidence of polygenic selection for a given GO term.

See also Figure S4, Table S2, and Data S1.

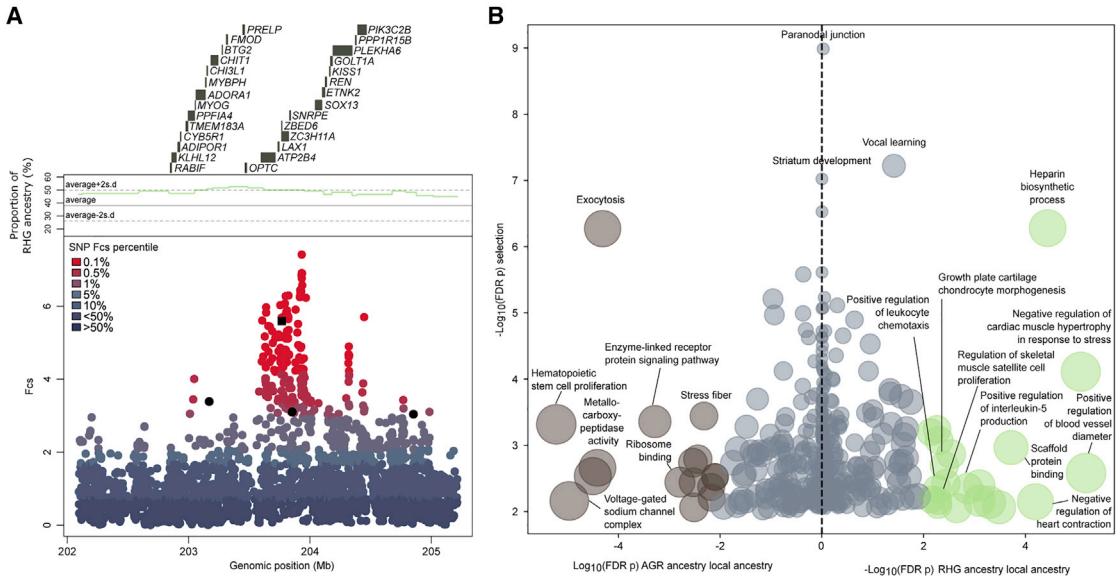

Figure 4. Selection Signals in Highly Admixed Rainforest Hunter-Gatherers

(A) Selective sweep signal and average local RHG ancestry at the chr1:203564464-203764464 locus in the highly admixed RHG BaBongo. Dot colors indicate SNP F_{CS} percentiles, the black square indicates the non-synonymous variant (rs6697388) at *ZBED6*, and black dots indicate eQTLs (q value < 0.005) [33].

(B) GO terms enriched in both local ancestry in the highly admixed RHG BaBongo, and selection scores in each of the two putative parental populations, with respect to the rest of the genome (FDR p < 5%). Green (brown) dots indicate GO terms enriched in both western RHG (western AGR) local ancestry and selection scores in parental western RHG (western AGR) populations (FDR p < 5%). Enrichments were assessed using the Mann-Whitney-Wilcoxon rank-sum test. See also Data S1.

strongest candidate variants, we found a non-synonymous mutation (rs6697388) in *ZBED6*, which encodes a TF that controls muscle growth through *IGF2* repression [59]. *ZBED6* is located within the intron of the *ZC3H11A* gene, whose product is required for the efficient growth of several nuclear-replicating viruses [60]. The rs6697388 G allele (p.Leu391Arg) is present at the highest frequency in admixed BaBongo (51%), with lower frequencies in parental RHG (42%) and AGR (15%) groups. With respect to the strong, shared selective sweep detected at *TRPS1* (Figure 2C), the locus also presented selection signals in the BaBongo but no excess of RHG or AGR ancestry (Figures S2 and S3), suggesting weaker or no positive selection at *TRPS1* since admixture.

Finally, we searched for evidence of polygenic selection since admixture, by testing for excesses in AGR or RHG local ancestry in genomic windows related to GO terms in the admixed BaBongo (STAR Methods). We found 21 GO terms that were enriched in both RHG local ancestry and selection signals in the parental RHGs (Figure 4B; Data S1), an overlap that was significantly larger than expected (7.3% versus 4.7%, χ^2 test, p = 0.042). These terms were mostly related to cardiac and skeletal development and immune functions, and included "heparin biosynthetic process," which participates in mast cell-mediated immune and inflammatory responses [61], echoing the signals detected for "mast cell degranulation" in weakly admixed RHGs (Figure 3D). We also found 16 GO terms that were enriched in both AGR local ancestry and selection signals in the parental AGRs (Figure 4B; Data S1), including stem cell proliferation, exocytosis, and muscle composition. Together, these results support further the notion that heart and bone development as well as immune responses have been an important substrate of selection in RHGs, before and after their admixture with neighboring farmers.

DISCUSSION

Here we present the first exome-based survey of multiple geographically dispersed groups of African rainforest hunter-gatherers, with the aim of investigating how populations have adapted to the challenging habitats of the equatorial rainforest. Because positive selection often targets regulatory regions [62], we combined the exome dataset with SNP array data, to cover both genic and intergenic regions. In doing so, we found evidence of a unique, strong sweep that is shared by all RHG groups, targeting the regulatory region of *TRPS1*, mutations in which can cause growth retardation, distinctive craniofacial features [63], and hypertrichosis [64]. Furthermore, the transcription factor TRPS1 regulates STAT3, a mediator of inflammation and immunity [65], and RUNX2, controlling facial features and viral clearance [66, 67]. Interestingly, *TRPS1* has been recently shown to carry signals of archaic introgression in western Africans [68]. Functional studies should help determine the adaptive nature—developmental and/or immune-related—of variation at this locus, which possibly introgressed from extinct African hominins [18, 68, 69].

This study also extends previous findings of a sweep targeting the *CISH-MAPKAPK3-DOCK3* region [6, 14], by delineating *MAPKAPK3* as the most likely target. *MAPKAPK3* expression is regulated by two eQTLs that are among the strongest candidates for positive selection at the locus in RHG populations. MAPKAPK3 directly interacts with HCV and regulates cell infectivity [41]. A lower prevalence of HCV infection has been reported in RHG, with respect to AGR [70, 71]. Our results strengthen the evolutionary importance of the *CISH-MAPKAPK3-DOCK3* region in both western and eastern RHGs, and pinpoint

MAPKAPK3 variation as a putative, additional risk factor for HCV infection in Africans.

Our analyses provide robust evidence for polygenic selection of height, which we replicate in various RHG groups. Importantly, our results are not affected by biased genome-wide association study (GWAS) summary statistics due to partial control for population stratification, which can result in spurious polygenic selection signals [72, 73]. Our approach tests for the co-localization of selection signals and trait-associated genes; thus, it does not depend on effect size estimates and does not assume that associated variants are the same across populations. More generally, polygenic selection of height is unlikely to result from sexual selection [74] but from genetic adaptation to equatorial forest environments [75]. Our study sheds new light onto the debated adaptive nature of height, and supports that the early reproductive age of RHGs is not the cause of their small body size, as previously suggested [51, 52]. Instead, our results suggest that directional selection of height has resulted in changes in life-history traits because of pervasive pleiotropy of height-associated genes.

We also found signals of polygenic selection in RHGs at functions related to the IL-2 pathway, the sensing of allergens and microbes, and interactions with dsDNA and ssRNA viruses. Interestingly, higher seropositivity for more than 30 viruses has been reported in the BaTwa from Uganda, with respect to AGRs, particularly for dsDNA viruses [76]. That we also found an excess of RHG ancestry related to heparin biosynthesis, interleukin production, and leukocyte chemotaxis in highly admixed RHGs suggests preferential retention of RHG variation at immune-related functions. This finding supports a long-standing history of adaptation of RHGs to high pathogen pressures. This contrasts with a study in southern Africa, which reported a low exposure and adaptation to pathogens of hunter-gatherers of the Kalahari Desert, except for those who recently came in contact with other populations [77].

Collectively, our analyses uncover height, development, and immune response as iconic adaptive traits of African RHGs. It is interesting to note that the PI3K signaling pathway—under polygenic selection in four RHG populations—modulates inflammatory responses [78], body energy homeostasis [79, 80], and insulin secretion [81]. Several studies have highlighted the reciprocal relationship between proinflammatory cytokines and the regulation of the growth hormone through the IGF-1 axis [82]. It is thus tempting to speculate that pleiotropic effects between developmental growth and immunity could have further participated in the "pygmy" phenotype. Epidemiological work on the infectious disease burden in hunter-gatherers should increase our understanding of how historical high pathogen-driven selection has contributed to the reduced stature characterizing populations of the rainforest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.07.013.

ACKNOWLEDGMENTS

We thank all the participants for providing the DNA samples used in this study. We also thank Guillaume Laval and Maxime Rotival for useful discussions, Muh-Ching Yee for laboratory assistance, and the "Paléogénomique et Génétique Moléculaire" Platform from the MnHn at the *Musée de l'Homme* for assistance in sample preparation. This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique, the "Histoire du Génome des Populations Humaines Gabonaises" project (Institut Pasteur/Republic of Gabon), and an Agence Nationale de la Recherche grant "AGRHUM" (ANR-14-CE02-0003-01) to L.Q.-M. M.L. was supported by the Fondation pour la Recherche Médicale (FDT20170436932), and J.C. by the INCEPTION program and the "Ecole Doctorale FIRE - Programme Bettencourt."

AUTHOR CONTRIBUTIONS

E.P. and L.Q.-M. conceived and supervised the study. M.L. conducted all the analyses and analyzed the data, with contributions from J.C., M. Silvert, and E.P. C.H. performed laboratory work. M. Sikora, K.S., H.C., and C.D.B. generated and/or analyzed whole-genome data. P.M.-D., J.-M.H., A.F., S.L.B., G.H.P., L.B.B., and P.V. assembled the samples. M.L., E.P., and L.Q.-M. wrote the manuscript, with contributions from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 14, 2019
Revised: June 26, 2019
Accepted: July 4, 2019
Published: August 8, 2019

REFERENCES

1. Hewlett, B.S. (2014). Hunter-Gatherers of the Congo Basin: Cultures, Histories and Biology of African Pygmies (Transaction Publishers).
2. Perry, G.H., and Verdu, P. (2017). Genomic perspectives on the history and evolutionary ecology of tropical rainforest occupation by humans. Quat. Int. *448*, 150–157.
3. Bahuchet, S. (2012). Changing language, remaining pygmy. Hum. Biol. *84*, 11–43.
4. Lopez, M., Kousathanas, A., Quach, H., Harmant, C., Mouguiama-Daouda, P., Hombert, J.M., Froment, A., Perry, G.H., Barreiro, L.B., Verdu, P., et al. (2018). The demographic history and mutational load of African hunter-gatherers and farmers. Nat Ecol Evol *2*, 721–730.
5. Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., Théry, S., Froment, A., Le Bomin, S., Gessain, A., Hombert, J.M., et al. (2009).

Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. Curr. Biol. *19*, 312–318.

6. Hsieh, P., Veeramah, K.R., Lachance, J., Tishkoff, S.A., Wall, J.D., Hammer, M.F., and Gutenkunst, R.N. (2016). 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. *26*, 279–290.
7. Patin, E., Laval, G., Barreiro, L.B., Salas, A., Semino, O., Santachiara-Benerecetti, S., Kidd, K.K., Kidd, J.R., Van der Veen, L., Hombert, J.M., et al. (2009). Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. PLoS Genet. *5*, e1000448.
8. Batini, C., Lopes, J., Behar, D.M., Calafell, F., Jorde, L.B., van der Veen, L., Quintana-Murci, L., Spedini, G., Destro-Bisol, G., and Comas, D. (2011). Insights into the demographic history of African Pygmies from complete mitochondrial genomes. Mol. Biol. Evol. *28*, 1099–1110.
9. Veeramah, K.R., Wegmann, D., Woerner, A., Mendez, F.L., Watkins, J.C., Destro-Bisol, G., Soodyall, H., Louie, L., and Hammer, M.F. (2012). An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. Mol. Biol. Evol. *29*, 617–630.
10. Aimé, C., Laval, G., Patin, E., Verdu, P., Ségurel, L., Chaix, R., Hegay, T., Quintana-Murci, L., Heyer, E., and Austerlitz, F. (2013). Human genetic data reveal contrasting demographic patterns between sedentary and nomadic populations that predate the emergence of farming. Mol. Biol. Evol. *30*, 2629–2644.
11. Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., Sica, L., Mouguiama-Daouda, P., Comas, D., Tzur, S., et al. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. Proc. Natl. Acad. Sci. USA *105*, 1596–1601.
12. Patin, E., Siddle, K.J., Laval, G., Quach, H., Harmant, C., Becker, N., Froment, A., Régnault, B., Lemée, L., Gravel, S., et al. (2014). The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. Nat. Commun. *5*, 3163.
13. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., and Pritchard, J.K. (2009). Signals of recent positive selection in a worldwide sample of human populations. Genome Res. *19*, 826–837.
14. Jarvis, J.P., Scheinfeldt, L.B., Soi, S., Lambert, C., Omberg, L., Ferwerda, B., Froment, A., Bodo, J.M., Beggs, W., Hoffman, G., et al. (2012). Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. PLoS Genet. *8*, e1002641.
15. Migliano, A.B., Romero, I.G., Metspalu, M., Leavesley, M., Pagani, L., Antao, T., Huang, D.W., Sherman, B.T., Siddle, K., Scholes, C., et al. (2013). Evolution of the pygmy phenotype: evidence of positive selection fro genome-wide scans in African, Asian, and Melanesian pygmies. Hum. Biol. *85*, 251–284.
16. Becker, N.S., Verdu, P., Georges, M., Duquesnoy, P., Froment, A., Amselem, S., Le Bouc, Y., and Heyer, E. (2013). The role of GHR and IGF1 genes in the genetic determination of African pygmies' short stature. Eur. J. Hum. Genet. *21*, 653–658.
17. Pemberton, T.J., Verdu, P., Becker, N.S., Willer, C.J., Hewlett, B.S., Le Bomin, S., Froment, A., Rosenberg, N.A., and Heyer, E. (2018). A genome scan for genes underlying adult body size differences between Central African hunter-gatherers and farmers. Hum. Genet. *137*, 487–509.
18. Lachance, J., Vernot, B., Elbers, C.C., Ferwerda, B., Froment, A., Bodo, J.M., Lema, G., Fu, W., Nyambo, T.B., Rebbeck, T.R., et al. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell *150*, 457–469.
19. Mendizabal, I., Marigorta, U.M., Lao, O., and Comas, D. (2012). Adaptive evolution of loci covarying with the human African Pygmy phenotype. Hum. Genet. *131*, 1305–1317.
20. Perry, G.H., Foll, M., Grenier, J.C., Patin, E., Nédélec, Y., Pacis, A., Barakatt, M., Gravel, S., Zhou, X., Nsobya, S.L., et al. (2014). Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. Proc. Natl. Acad. Sci. USA *111*, E3596–E3603.
21. Amorim, C.E., Daub, J.T., Salzano, F.M., Foll, M., and Excoffier, L. (2015). Detection of convergent genome-wide signals of adaptation to tropical forests in humans. PLoS ONE *10*, e0121557.
22. Fan, S., Kelly, D.E., Beltrame, M.H., Hansen, M.E.B., Mallick, S., Ranciaro, A., Hirbo, J., Thompson, S., Beggs, W., Nyambo, T., et al. (2019). African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. Genome Biol. *20*, 82.
23. Bergey, C.M., Lopez, M., Harrison, G.F., Patin, E., Cohen, J.A., Quintana-Murci, L., Barreiro, L.B., and Perry, G.H. (2018). Polygenic adaptation and convergent evolution on growth and cardiac genetic pathways in African and Asian rainforest hunter-gatherers. Proc. Natl. Acad. Sci. USA *115*, E11256–E11263.
24. Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., Laval, G., Perry, G.H., Barreiro, L.B., Froment, A., et al. (2017). Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. Science *356*, 543–546.
25. Fagny, M., Patin, E., MacIsaac, J.L., Rotival, M., Flutre, T., Jones, M.J., Siddle, K.J., Quach, H., Harmant, C., McEwen, L.M., et al. (2015). The epigenomic landscape of African rainforest hunter-gatherers and farmers. Nat. Commun. *6*, 10047.
26. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
27. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
28. Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science *329*, 75–78.
29. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al.; International HapMap Consortium (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913–918.
30. Grossman, S.R., Andersen, K.G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D.J., Griesemer, D., Karlsson, E.K., Wong, S.H., et al.; 1000 Genomes Project (2013). Identifying recent adaptations in large-scale genomic data. Cell *152*, 703–713.
31. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al.; International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.
32. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. Nature *517*, 327–332.
33. Battle, A., Brown, C.D., Engelhardt, B.E., and Montgomery, S.B.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

34. Andersen, K.G., Shylakhter, I., Tabrizi, S., Grossman, S.R., Happi, C.T., and Sabeti, P.C. (2012). Genome-wide scans provide evidence for positive selection of genes implicated in Lassa fever. Philos. Trans. R. Soc. Lond. B Biol. Sci. *367*, 868–877.

35. Fantauzzo, K.A., and Christiano, A.M. (2012). Trps1 activates a network of secreted Wnt inhibitors and transcription factors crucial to vibrissa follicle morphogenesis. Development *139*, 203–214.

36. Wuelling, M., Kaiser, F.J., Buelens, L.A., Braunholz, D., Shivdasani, R.A., Depping, R., and Vortkamp, A. (2009). Trps1, a regulator of chondrocyte proliferation and differentiation, interacts with the activator form of Gli3. Dev. Biol. *328*, 40–53.

37. Yosef, N., Shalek, A.K., Gaublomme, J.T., Jin, H., Lee, Y., Awasthi, A., Wu, C., Karwacz, K., Xiao, S., Jorgolli, M., et al. (2013). Dynamic regulatory network controlling TH17 cell differentiation. Nature *496*, 461–468.

38. Dunn-Fletcher, C.E., Muglia, L.M., Pavlicev, M., Wolf, G., Sun, M.A., Hu, Y.C., Huffman, E., Tumukuntala, S., Thiele, K., Mukherjee, A., et al. (2018). Anthropoid primate-specific retroviral element THE1B controls expression of CRH in placenta and alters gestation length. PLoS Biol. *16*, e2006337.

39. Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of genetic polymorphisms on human immune cell gene expression. Cell *175*, 1701–1715.e16.

40. Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic adaptation and Neandertal admixture shaped the immune system of human populations. Cell *167*, 643–656.e17.

41. Ngo, H.T., Pham, L.V., Kim, J.W., Lim, Y.S., and Hwang, S.B. (2013). Modulation of mitogen-activated protein kinase-activated protein kinase 3 by hepatitis C virus core protein. J. Virol. *87*, 5718–5731.

42. Tillmanns, J., Hoffmann, D., Habbaba, Y., Schmitto, J.D., Sedding, D., Fraccarollo, D., Galuppo, P., and Bauersachs, J. (2015). Fibroblast activation protein alpha expression identifies activated fibroblasts after myocardial infarction. J. Mol. Cell. Cardiol. *87*, 194–203.

43. Andrade, W.A., Silva, A.M., Alves, V.S., Salgado, A.P., Melo, M.B., Andrade, H.M., Dall'Orto, F.V., Garcia, S.A., Silveira, T.N., and Gazzinelli, R.T. (2010). Early endosome localization and activity of RasGEF1b, a toll-like receptor-inducible Ras guanine-nucleotide exchange factor. Genes Immun. *11*, 447–457.

44. Nemec, S., Luxey, M., Jain, D., Huang Sung, A., Pastinen, T., and Drouin, J. (2017). *Pitx1* directly modulates the core limb development program to implement hindlimb identity. Development *144*, 3325–3335.

45. Rüeger, S., McDaid, A., and Kutalik, Z. (2018). Evaluation and application of summary statistic imputation to discover new height-associated loci. PLoS Genet. *14*, e1007371.

46. Szeto, D.P., Rodriguez-Esteban, C., Ryan, A.K., O'Connell, S.M., Liu, F., Kioussi, C., Gleiberman, A.S., Izpisúa-Belmonte, J.C., and Rosenfeld, M.G. (1999). Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. Genes Dev. *13*, 484–494.

47. Island, M.L., Mesplede, T., Darracq, N., Bandu, M.T., Christeff, N., Djian, P., Drouin, J., and Navarro, S. (2002). Repression by homeoprotein pitx1 of virus-induced interferon a promoters is mediated by physical interaction and trans repression of IRF3 and IRF7. Mol. Cell. Biol. *22*, 7120–7133.

48. Pritchard, J.K., Pickrell, J.K., and Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. *20*, R208–R215.

49. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: from polygenic to omnigenic. Cell *169*, 1177–1186.

50. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.

51. Migliano, A.B., Vinicius, L., and Lahr, M.M. (2007). Life history trade-offs explain the evolution of human pygmies. Proc. Natl. Acad. Sci. USA *104*, 20216–20219.

52. Walker, R., Gurven, M., Hill, K., Migliano, A., Chagnon, N., De Souza, R., Djurovic, G., Hames, R., Hurtado, A.M., Kaplan, H., et al. (2006). Growth rates and life histories in twenty-two small-scale societies. Am. J. Hum. Biol. *18*, 295–311.

53. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

54. Kim, M.S., Rådinger, M., and Gilfillan, A.M. (2008). The multiple roles of phosphoinositide 3-kinase in mast cell biology. Trends Immunol. *29*, 493–501.

55. Malek, T.R., and Castro, I. (2010). Interleukin-2 receptor signaling: at the interface between tolerance and immunity. Immunity *33*, 153–165.

56. Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.L., Patin, E., and Quintana-Murci, L. (2016). Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. Am. J. Hum. Genet. *98*, 5–21.

57. Enard, D., and Petrov, D.A. (2018). Evidence that RNA viruses drove adaptive introgression between Neanderthals and modern humans. Cell *175*, 360–371.e13.

58. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288.

59. Younis, S., Schönke, M., Massart, J., Hjortebjerg, R., Sundström, E., Gustafson, U., Björnholm, M., Krook, A., Frystyk, J., Zierath, J.R., and Andersson, L. (2018). The ZBED6-IGF2 axis has a major effect on growth of skeletal muscle and internal organs in placental mammals. Proc. Natl. Acad. Sci. USA *115*, E2048–E2057.

60. Younis, S., Kamel, W., Falkeborn, T., Wang, H., Yu, D., Daniels, R., Essand, M., Hinkula, J., Akusjärvi, G., and Andersson, L. (2018). Multiple nuclear-replicating viruses require the stress-induced protein ZC3H11A for efficient growth. Proc. Natl. Acad. Sci. USA *115*, E3808–E3816.

61. Humphries, D.E., Wong, G.W., Friend, D.S., Gurish, M.F., Qiu, W.T., Huang, C., Sharpe, A.H., and Stevens, R.L. (1999). Heparin is essential for the storage of specific granule proteases in mast cells. Nature *400*, 769–772.

62. Kudaravalli, S., Veyrieras, J.B., Stranger, B.E., Dermitzakis, E.T., and Pritchard, J.K. (2009). Gene expression levels are a target of recent natural selection in the human genome. Mol. Biol. Evol. *26*, 649–658.

63. Momeni, P., Glöckner, G., Schmidt, O., von Holtum, D., Albrecht, B., Gillessen-Kaesbach, G., Hennekam, R., Meinecke, P., Zabel, B., Rosenthal, A., et al. (2000). Mutations in a new gene, encoding a zinc-finger protein, cause tricho-rhino-phalangeal syndrome type I. Nat. Genet. *24*, 71–74.

64. Fantauzzo, K.A., Tadin-Strapps, M., You, Y., Mentzer, S.E., Baumeister, F.A., Cianfarani, S., Van Maldergem, L., Warburton, D., Sundberg, J.P., and Christiano, A.M. (2008). A position effect on TRPS1 is associated with Ambras syndrome in humans and the Koala phenotype in mice. Hum. Mol. Genet. *17*, 3539–3551.

65. Hillmer, E.J., Zhang, H., Li, H.S., and Watowich, S.S. (2016). STAT3 signaling in immunity. Cytokine Growth Factor Rev. *31*, 1–15.

66. Adhikari, K., Fuentes-Guajardo, M., Quinto-Sánchez, M., Mendoza-Revilla, J., Camilo Chacón-Duque, J., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R.B., Pérez, G.M., et al. (2016). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat. Commun. *7*, 11616.

67. Chopin, M., Preston, S.P., Lun, A.T.L., Tellier, J., Smyth, G.K., Pellegrini, M., Belz, G.T., Corcoran, L.M., Visvader, J.E., Wu, L., and Nutt, S.L. (2016). RUNX2 mediates plasmacytoid dendritic cell egress from the bone marrow and controls viral immunity. Cell Rep. *15*, 866–878.

68. Durvasula, A., and Sankararaman, S. (2019). Recovering signals of ghost archaic introgression in African populations. bioRxiv. https://doi.org/10.1101/285734.

69. Hsieh, P., Woerner, A.E., Wall, J.D., Lachance, J., Tishkoff, S.A., Gutenkunst, R.N., and Hammer, M.F. (2016). Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. Genome Res. *26*, 291–300.

70. Foupouapouognigni, Y., Mba, S.A., Betsem à Betsem, E., Rousset, D., Froment, A., Gessain, A., and Njouom, R. (2011). Hepatitis B and C virus infections in the three Pygmy groups in Cameroon. J. Clin. Microbiol. *49*, 737–740.

71. Kowo, M.P., Goubau, P., Ndam, E.C., Njoya, O., Sasaki, S., Seghers, V., and Kesteloot, H. (1995). Prevalence of hepatitis C virus and other blood-borne viruses in Pygmies and neighbouring Bantus in southern Cameroon. Trans. R. Soc. Trop. Med. Hyg. *89*, 484–486.

72. Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., Boyle, E.A., Zhang, X., Racimo, F., Pritchard, J.K., and Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. eLife *8*, e39725.

73. Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., Chiang, C.W., Hirschhorn, J., Daly, M.J., Patterson, N., et al. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife *8*, e39702.

74. Becker, N., Touraille, P., Froment, A., Heyer, E., and Courtiol, A. (2012). Short stature in African pygmies is not explained by sexual selection. Evol. Hum. Behav. *33*, 615–622.

75. Perry, G.H., and Dominy, N.J. (2009). Evolution of the human pygmy phenotype. Trends Ecol. Evol. *24*, 218–225.

76. Harrison, G.F., Sanz, J., Boulais, J., Mina, M.J., Grenier, J.-C., Leng, Y., Dumaine, A., Yotova, V., Bergey, C.M., Elledge, S.J., et al. (2019). Natural selection contributed to immunological differences between human hunter-gatherers and agriculturalists. bioRxiv. https://doi.org/10.1101/487207.

77. Owers, K.A., Sjödin, P., Schlebusch, C.M., Skoglund, P., Soodyall, H., and Jakobsson, M. (2017). Adaptation to infectious disease exposure in indigenous Southern African populations. Proc. Biol. Sci. *284*, 20170226.

78. Hawkins, P.T., and Stephens, L.R. (2015). PI3K signalling in inflammation. Biochim. Biophys. Acta *1851*, 882–897.

79. Hopkins, B.D., Pauli, C., Du, X., Wang, D.G., Li, X., Wu, D., Amadiume, S.C., Goncalves, M.D., Hodakoski, C., Lundquist, M.R., et al. (2018). Suppression of insulin feedback enhances the efficacy of PI3K inhibitors. Nature *560*, 499–503.

80. Fruman, D.A., Chiu, H., Hopkins, B.D., Bagrodia, S., Cantley, L.C., and Abraham, R.T. (2017). The PI3K pathway in human disease. Cell *170*, 605–635.

81. Odegaard, J.I., and Chawla, A. (2013). Pleiotropic actions of insulin resistance and inflammation in metabolic homeostasis. Science *339*, 172–177.

82. Smith, T.J. (2010). Insulin-like growth factor-I regulation of immune function: a potential therapeutic target in autoimmune diseases? Pharmacol. Rev. *62*, 199–236.

83. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

84. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.

85. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

86. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

87. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.

88. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.

89. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M.; Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

90. Klopfenstein, D.V., Zhang, L., Pedersen, B.S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., et al. (2018). GOATOOLS: a Python library for Gene Ontology analyses. Sci. Rep. *8*, 10872.

91. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.

92. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. *6*, e1001025.

93. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 1–33.

94. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

95. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., et al.; International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.

96. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

97. Cavalli-Sforza, L.L. (1986). African Pygmies (Academic Press).

98. Diamond, J.M. (1991). Anthropology. Why are pygmies small? Nature *354*, 111–112.

99. Shea, B.T., and Bailey, R.C. (1996). Allometry and adaptation of body proportions and stature in African pygmies. Am. J. Phys. Anthropol. *100*, 311–340.

100. Froment, A. (2001). Hunter-gatherers: an interdisciplinary perspective. In Hunter-Gatherers: An Interdisciplinary Perspective, L.R.-C. Panter-Brick, ed. (Cambridge University Press), pp. 239–266.

101. Becker, N.S., Verdu, P., Hewlett, B., and Pavard, S. (2010). Can life history trade-offs explain the evolution of short stature in human pygmies? A response to Migliano et al. (2007). Hum. Biol. *82*, 17–27.

102. International HapMap Consortium (2005). A haplotype map of the human genome. Nature *437*, 1299–1320.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
Nextera Rapid Capture Expanded Exome kit	Illumina	Cat#FC-140-1006
HumanOmniExpress-24 v1.1 DNA Analysis Kit	Illumina	N/A
Deposited Data		
Exome and whole-genome sequencing	This paper	EGAS00001003722
Software and Algorithms		
PLINK v1.9	[83]	http://www.cog-genomics.org/plink/1.9/
KING v1.4	[84]	http://people.virginia.edu/~wc9c/KING/history.htm
ADMIXTURE	[26]	http://software.genetics.ucla.edu/admixture/download.html
BWA v.0.7.7	[85]	http://bio-bwa.sourceforge.net/
Picard Tools v.1.94	N/A	http://broadinstitute.github.io/picard
GATK v3.5	[86]	https://software.broadinstitute.org/gatk/download/
SHAPEIT2	[87]	http://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
IMPUTE v.2	[88]	http://mathgen.stats.ox.ac.uk/impute/impute_v2.1.0.html
LDSC	[89]	https://data.broadinstitute.org/alkesgroup/LDSCORE/
GOATOOLS	[90]	https://github.com/tanghaibao/goatools
RFMIX v1.5.4	[58]	https://sites.google.com/site/rfmixlocalancestryinference/
BEAGLE	[91]	https://faculty.washington.edu/browning/beagle/beagle.html
GERP++	[92]	http://mendel.stanford.edu/SidowLab/downloads/gerp/

LEAD CONTACT AND MATERIALS AVAILABILITY

This study did not generate new unique reagents. Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Lluís Quintana-Murci (quintana@pasteur.fr).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sample collection

Sampling consisted in human saliva or blood from 157 rainforest hunter-gatherers and 120 farmers from western and eastern central Africa (Figure S1), including 208 males and 69 females. Informed consent was obtained from all participants in this study, which was overseen by the institutional review board of Institut Pasteur (2011-54/IRB/8), the *Comité National d'Ethique du Gabon* (0016/2016/SG/CNE), the University of Chicago (IRB 16986A) and Makerere University, Kampala, Uganda (IRB 2009-137). The 277 new samples collected for exome sequencing were analyzed together with 317 exomes of central Africans from Lopez et al. 2018 [4] and 101 Europeans from Quach et al. 2016 [40] (Table S1).

METHOD DETAILS

Exome Sequencing

Sample libraries were prepared with the Nextera Rapid Capture Expanded Exome Kit, which delivers 62Mb of genomic content per individual, including exons, untranslated regions and microRNAs, and were sequenced on Illumina HiSeq2500 machines. Using the GATK Best Practices recommendations [93], pairs of 101-bp reads were mapped onto the human reference genome (GRCh37) with Burrows-Wheeler Aligner (BWA) version 0.7.7 [85], using 'bwa mem -M -t 4 -R', and reads duplicating the start position of another read were marked as duplicates with Picard Tools version 1.94 (http://broadinstitute.github.io/picard/), using 'MarkDuplicates'. We used GATK version 3.5 [86] for base quality score recalibration ('Base Recalibrator'), insertion/deletion (indel) realignment ('IndelRealigner'), and SNP and indel discovery ('Haplotype Caller') for each sample. Individual variant files were combined with 'GenotypeGVCFs' and filtered with 'VariantQualityScoreRecalibration'. We used high confidence variants from the 1000G Phase 1 and HapMap 3 projects [94, 95] as VQSR training callsets, and applied a tranche sensitivity threshold of 99.5%. From the 947,523 sites detected, we removed indels as well as SNPs that (i) were located on the sex chromosomes,

(ii) were not biallelic, (iii) were monomorphic in our total sample, (iv) had a depth of coverage < 5 ×, (v) had a genotype quality score (GQ) < 20, (vi) presented missingness > 15%, and (vii) presented a Hardy-Weinberg test $p < 10^{-6}$ in at least one of population. As criteria to remove low-quality samples, we required a total genotype missingness < 15% (21 excluded samples). In addition, we checked for unexpectedly high or low heterozygosity values, suggesting high levels of inbreeding or DNA contamination, and excluded 3 individuals presenting heterozygosity levels 4 SD higher than their population average. We thus retained exome data for 671 individuals, with an average depth of coverage after duplicate removal of 38 × (SD: 9 ×), ranging from 25 × to 95 ×. The application of these quality-control filters resulted in a final dataset of 682,468 SNPs (Figure S1), of which 107,621 SNPs were polymorphic only in the 268 newly-sequenced individuals.

SNP Array Data

In addition to exome sequencing, we retrieved the genotyping data of the same 671 individuals from Quach et al. 2016 [40], Patin et al. 2014 [12], Patin et al. 2017 [24] and Fagny et al. 2015 [25] (Figure S1; Table S1). We removed SNPs located on the X and Y chromosomes, problematic genotype clustering profiles (i.e., Illumina GenTrain score < 0.35) or with call rate < 95%. We kept 599,559 SNPs common to different genotyping SNP arrays. We removed a total of 53 C/G or A/T SNPs to prevent misaligned SNPs, and excluded a total 5 additional SNPs that were under Hardy-Weinberg disequilibrium in at least one of the populations ($p < 10^{-6}$) using PLINK [96], leading to a final dataset of 559,501 SNPs.

We applied additional filters on the genotyping dataset of the 671 individuals retained for exome sequencing. We removed two individuals with heterozygosity levels higher or lower than the population mean ± 4 SD Although related individuals were avoided during the sampling and for exome sequencing (based on published SNP array data) [5, 12, 17, 24, 25], we sought to exclude possibly remaining pairs of cryptically related individuals. Indeed, RHG populations are small isolated communities, where individuals can be related to many others. We considered that two individuals were strongly (cryptically) related if they presented a first-degree relationship (kinship coefficient > 0.177), as inferred by KING [84]. Following this criterion, only one individual was removed. Additionally, we removed another individual who did not present any first-degree relatedness but was related in second-degree to many others. After removing these two individuals, the dataset included 77 and 232 pairs of second-degree (kinship coefficient > 0.0884) and third-degree (kinship coefficient > 0.0442) related RHG individuals, respectively. The application of these quality-control filters resulted in a final genotyping dataset of 667 individuals and 599,501 SNPs (Figure S1).

Merging Exome and SNP Array Data

Before merging the genotyping array and the exome data from the 667 high-quality individuals in common, we flipped alleles for 8,393 SNPs with incompatible allelic states, and removed 9 SNPs with alleles that remained incompatible after allele flipping from the genotyping dataset. The total concordance rate was evaluated on 28,403 SNPs common to both datasets. The concordance rates for each of the 667 individuals exceeded 98%, confirming an absence of errors during DNA sample processing. The entire genotyping and exome datasets (599,492 and 682,468 SNPs, respectively) were then merged, yielding a final dataset of 1,253,548 SNPs for 667 individuals, 566 of whom were African farmers or hunter-gatherers (Figure S1).

Whole-Genome Sequencing

We generated whole genomes of 20 RHG Baka and 20 AGR Nzébi of Gabon, which were also part of the exome and SNP array datasets. All the samples were processed using the paired-end library preparation protocol from Illumina. Libraries were sequenced on Illumina HiSeq 2000 machines at the Stanford Center for Genomics and Personalized Medicine. 101-pb reads were aligned to the human reference genome (GRCh37) using BWA [85], followed by base quality recalibration and realignment around known indels with GATK [86]. Genotyping was carried out across all 40 individuals jointly using GATK 'UnifiedGenotyper', and called variants were stratified into variant quality tranches using 'VariantQualityScoreRecalibration' tool (VQSR) from GATK. SNPs with a VQSR tranche > 99.0 were considered as confidently called. Genotype calls were refined and improved based on LD using BEAGLE [91], yielding a final dataset of 17,687,206 variants (Figure S1). All individuals presented very low rates of missing values ranging from 0.5% to 4%, and a mean depth of coverage of 6.5 × (ranging from 4 × to 13 ×).

Imputation of SNP Array and Exome Data

Before imputation, we phased the data with SHAPEIT2 using 100 states, 20 MCMC main steps, 7 burnin and 8 pruning steps [87]. SNPs and allelic states were then aligned with the 1000 Genomes Project imputation reference panel (Phase 3 [27]), referred to as 'reference panel 1', as well as the 40 whole genomes of Baka RHG and Nzébi AGR of Gabon, referred to as 'reference panel 2' (Figure S1). We removed from the reference panels SNPs with MAF < 1%, SNPs with C/G or A/T alleles and 414,679 multiallelic SNPs in the reference panel 1. We evaluated the allelic concordance between the two reference panels and excluded 9,649 additional sites from the reference panel 2, yielding to final datasets of 11,501,018 SNPs in the reference panel 1 and 14,252,666 SNPs in the reference panel 2.

Genotype imputation was performed with IMPUTE v.2 [88] considering 1-Mb windows and both reference panels simultaneously, with the '-merge_ref_panels' option. We used genotype calls instead of genotype probabilities, which are not handled by downstream programs, and considered as confident genotype calls genotypes with posterior probability > 0.8. Of the 13,092,258 SNPs obtained after imputation, we removed SNPs that: (i) presented an information metric < 0.8, (ii) had a duplicate, (iii) presented a call rate < 95%, and (iv) were monomorphic. The final imputed dataset included 10,262,236 SNPs, and 9,129,103 after filtering

SNPs with MAF < 1%. To evaluate imputation accuracy, we estimated correlation coefficients r^2 between true genotypes (i.e., obtained by Illumina genotyping array or exome sequencing) and imputed genotypes for the same SNPs (i.e., obtained by artificially removing genotyped SNPs from the data before imputation and then imputing them). The average correlation coefficient across all genotyped SNPs with information metric > 0.8 were 0.86 and 0.85 for reference panels 1 and 2, respectively, showing that our quality filters ensure to keep accurately imputed SNPs for further analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome Scans for Selective Sweeps

Genomic regions candidate for positive selection were detected in seven populations of RHG (Bezan, Baka, BaBongo of central Gabon, BaKoya, BaBongo of south and east Gabon and BaTwa) and two populations of AGR (western and eastern AGR), with an outlier approach that considers two interpopulation statistics: PBS (Population Branch Score [28]), and XP-EHH [29]. We combined these scores into a Fisher's score (F_{CS}) equal to the sum, over the two statistics, of $-\log_{10}$(rank of the statistic for a given SNP/number of SNPs). Interpopulation statistics require a reference population, and PBS statistics an outgroup population. We performed separate scans of classic sweeps for each population, using Europeans as outgroup, and different reference populations: western AGR for each western RHG population, eastern AGR for eastern RHG, pooled western RHG for western AGR, and eastern RHG for eastern AGR. PBS was calculated for each SNP using AMOVA-based F_{ST} values computed with home-made scripts (available upon request). The derived allele of each SNP was defined based on the 6-EPO alignment. XP-EHH was computed in 100-kb sliding windows with a 50-kb pace, with home-made scripts (available upon request). Only SNPs with a derived allele frequency (DAF) between 10% and 90% were analyzed further. XP-EHH scores were normalized in 40 separate bins of DAF. An outlier SNP was defined as a SNP with an F_{CS} among the 1% highest of the genome. A putatively selected genomic region was defined as a 100-kb window presenting a proportion of outlier SNPs among the 1% highest of all windows, in five bins of SNP numbers. Windows containing less than 50 SNPs were discarded as well as 500-kb regions around gaps, to avoid biases in the outlier enrichment scores.

Polygenic Selection of Complex Traits

We retrieved the results of the Genome Wide Association studies from UK BIOBANK (round 2, http://www.nealelab.is/uk-biobank/) of 12 complex traits that we selected as candidates for adaptation of RHG, based on previous hypotheses from biological anthropology studies [51, 73, 97–101]. Our genomic dataset was split into non-overlapping 100-kb windows. We considered a window as associated with a trait if it included a SNP with a genome-wide significant association with this trait ($P_{assoc}<5 \times 10^{-8}$). We computed for each genomic window, associated or not with the trait, the average F_{CS}, the proportion of conserved SNP positions based on GERP scores > 2 [92], and the recombination rate using the combined HapMap genetic map [102], to account for the confounding effects of background selection.

In order to test for polygenic selection, we generated a null distribution by randomly sampling *x* windows (*x* being the number of windows associated with a tested trait) among windows with a similar number of SNPs, proportion of GERP > 2 sites and recombination rate observed in the trait-associated windows. We then calculated the average of the mean of the F_{CS} across the *x* resampled windows. We resampled 100,000 sets of *x* windows for each trait. To test for significance, we computed a resampling *P*-value by calculating the proportion of resampled windows which mean F_{CS} was higher than that observed for the tested trait. All *P*-values for polygenic adaptation were then adjusted for multiple testing by the Benjamini-Hochberg method, to account for the number of traits tested, and traits with an adjusted $p < 0.05$ were considered as candidates for polygenic selection.

To test if polygenic selection signals are due to pleiotropy of height-associated genes, we first estimated genetic correlations between candidate traits from LD-score regression using the ldsc tool [89]. We used precomputed European LD-scores (https://data.broadinstitute.org/alkesgroup/LDSCORE/). *P*-values were corrected for multiple testing using the Bonferroni correction, and adjusted *P*-values < 0.05 were considered as significant.

To correct for pleiotropy for each trait genetically correlated with height, we removed windows significantly associated with 'Standing Height' and 'Comparative height at age 10' in both windows associated with the candidate trait and resampled windows. Similarly, we re-tested for polygenic adaptation on ''Standing height'' and ''Comparative height at age 10'' associated regions using the same approach, but by removing all trait-associated windows, except height-associated windows. To test if loss of significance was due to a decrease in power, we down-sampled the number of tested trait-associated windows to the same number as after removing height-associated windows. We down-sampled a 100 times trait-associated windows, and estimated a hundred *P*-values as described above. We finally compared the distribution of the 100 obtained *P*-values with the estimated *P*-value (non-adjusted for multiple testing) both before and after removing height-associated windows.

Polygenic Selection of Gene Ontologies

To detect enrichment of F_{CS} scores in sets of genes corresponding to a given biological pathway, we compared the distributions of F_{CS} between genes that were part of the gene ontology (GO) term tested, relative to the rest of the genes of the genome, using a Mann-Whitney-Wilcoxon rank-sum test. To limit the effect of clusters of genes on the enrichment calculation, we assigned to each 100-kb non-overlapping genomic window both a GO term, based on the presence of at least one gene from the corresponding term, and a mean F_{CS} score. We tested if mean F_{CS} of windows assigned to a given GO term were different from genome-wide expectations, accounting for multiple testing. We restricted the enrichment analysis to 5,354 GO terms with levels comprised between

levels 3 and 7 [53], using the python library goatools [90], and that include at least 5 genes. We examined a total of 15,503 windows and determined *P*-values corresponding to 5% and 1% of false discoveries, FDR p = 9.24×10^{-3} and FDR p = 4.03×10^{-4}, respectively, by randomly resampling *y* genes (*y* being sampled from the distribution of the number of genes assigned to each GO term). We also studied additional gene sets, including 1,553 manually-curated genes involved in innate immunity [56] and 1,257 genes encoding proteins known to have physical interactions with multiple families of viruses [57].

Local Ancestry Inference

To perform local ancestry inference in the genomes of the highly-admixed BaBongo RHG from south and east Gabon, we first constituted putative parental populations that were representative of RHG and AGR ancestry. We considered as the parental AGR population, 163 individuals with less than 20% of their ancestry assigned to the RHG component, based on the ADMIXTURE analysis at $K = 5$. Likewise, we considered as the parental RHG population, 101 individuals with less than 5% AGR ancestry. The genomes of the highly-admixed BaBongo were decomposed into segments of RHG or AGR ancestry with RFMix v.1.5.4 [58], including two EM steps. We excluded 2-Mb regions from the telomeres of each chromosome. Based on RFMix ancestry estimations, the mean AGR ancestry was 94% [SD = 1.6%] in the parental AGR population, 62% [SD = 5.9%] in the highly-admixed BaBongo, and 27% [SD = 3.7%] in the parental RHG population. These ancestry proportions were highly correlated with ADMIXTURE membership proportions at $K = 2$ (Pearson's correlation coefficient $R^2 = 0.99$). We then searched for excesses in RHG or AGR ancestry in pathways by assigning ancestry proportions to 100-kb windows across the genome, with the same approach used for GO enrichments.

DATA AND CODE AVAILABILITY

The newly generated exomes (n = 277) and genomes (n = 40) of central African rainforest hunter-gatherers and agriculturalists have been deposited in the European Genome-phenome Archive (EGA). The accession number for the newly generated data reported in this paper is EGA: EGAS00001003722. Data accessibility is restricted to academic research on human genetic history and adaptation. Exome sequencing data for the remaining, previously published samples are available under accession codes EGA: EGAS00001002457 and EGA: EGAS00001001895.

Current Biology, Volume *29*

Supplemental Information

Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest

Marie Lopez, Jeremy Choin, Martin Sikora, Katherine Siddle, Christine Harmant, Helio A. Costa, Martin Silvert, Patrick Mouguiama-Daouda, Jean-Marie Hombert, Alain Froment, Sylvie Le Bomin, George H. Perry, Luis B. Barreiro, Carlos D. Bustamante, Paul Verdu, Etienne Patin, and Lluís Quintana-Murci

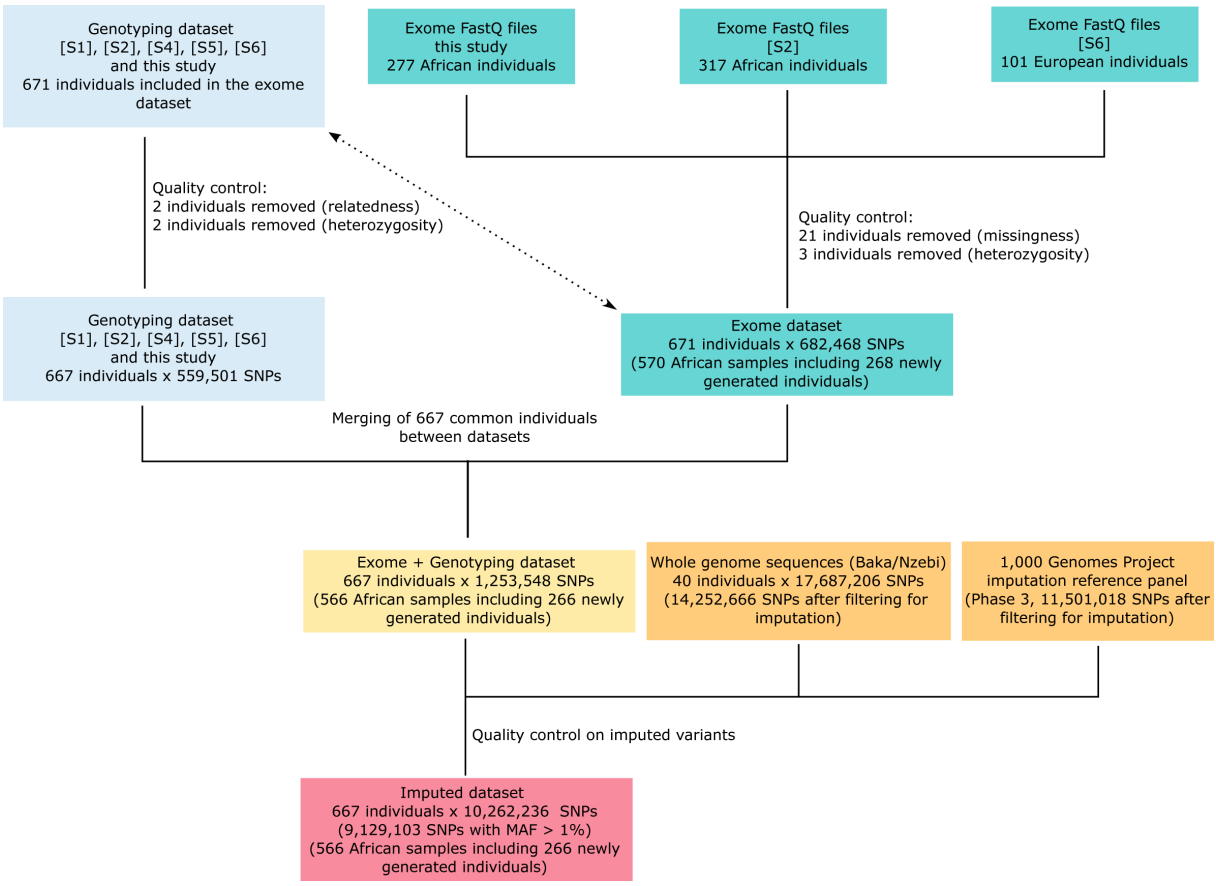

Figure S1. Summary of the Data Processing Performed in this Study. Related Figure 1.
The arrow indicates the correspondence between individuals analyzed in both datasets.

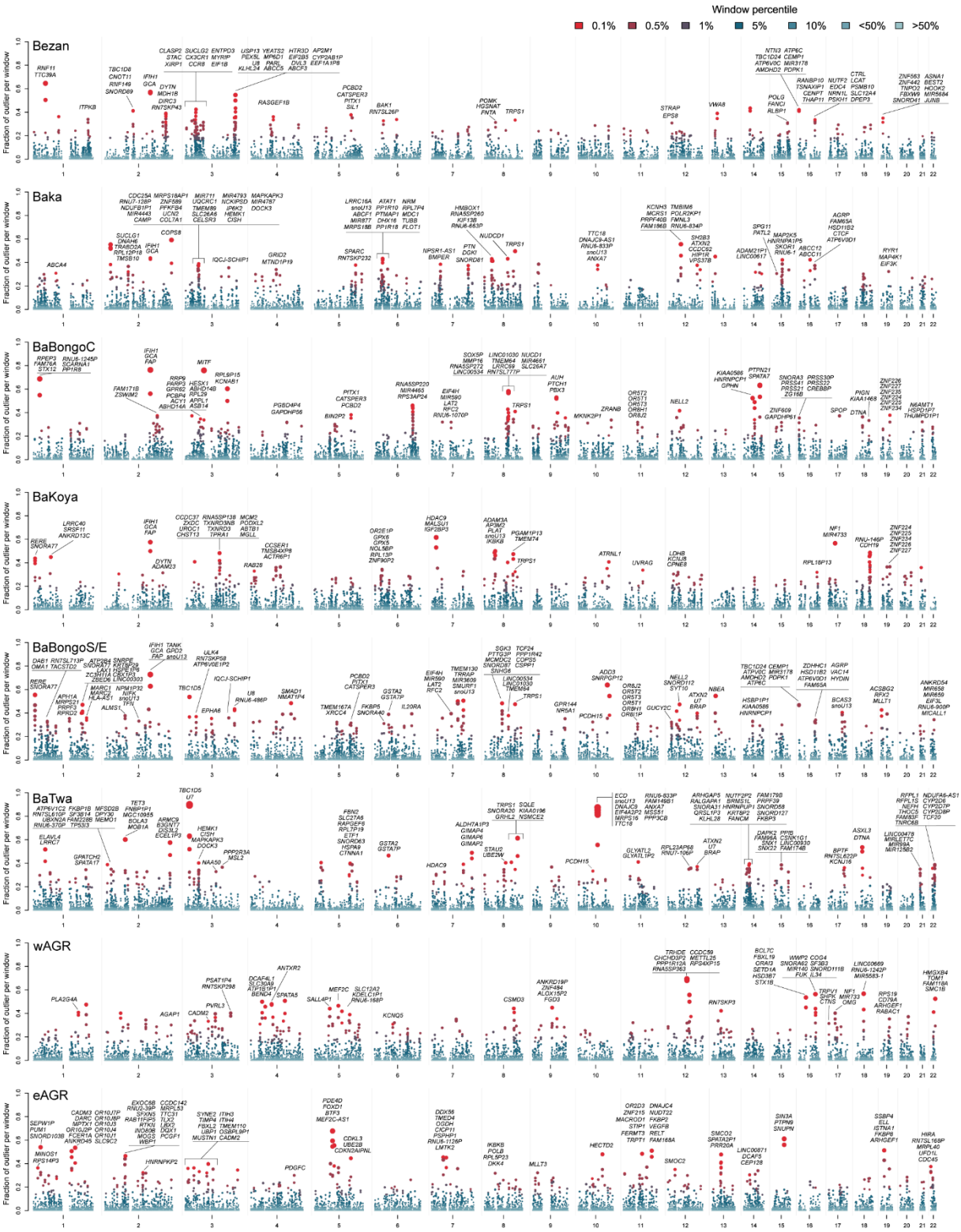

Figure S2. Genome-Wide Signals of Classic Sweeps in Central African Populations. Related to Figure 2.

Proportions of outlier SNPs (i.e., F_{CS} in the top 1% of the empirical distribution) in 100-kb windows along the genome of RHG and AGR populations. Gene names are shown for candidate windows with a proportion of F_{CS} outlier SNPs > 30%.

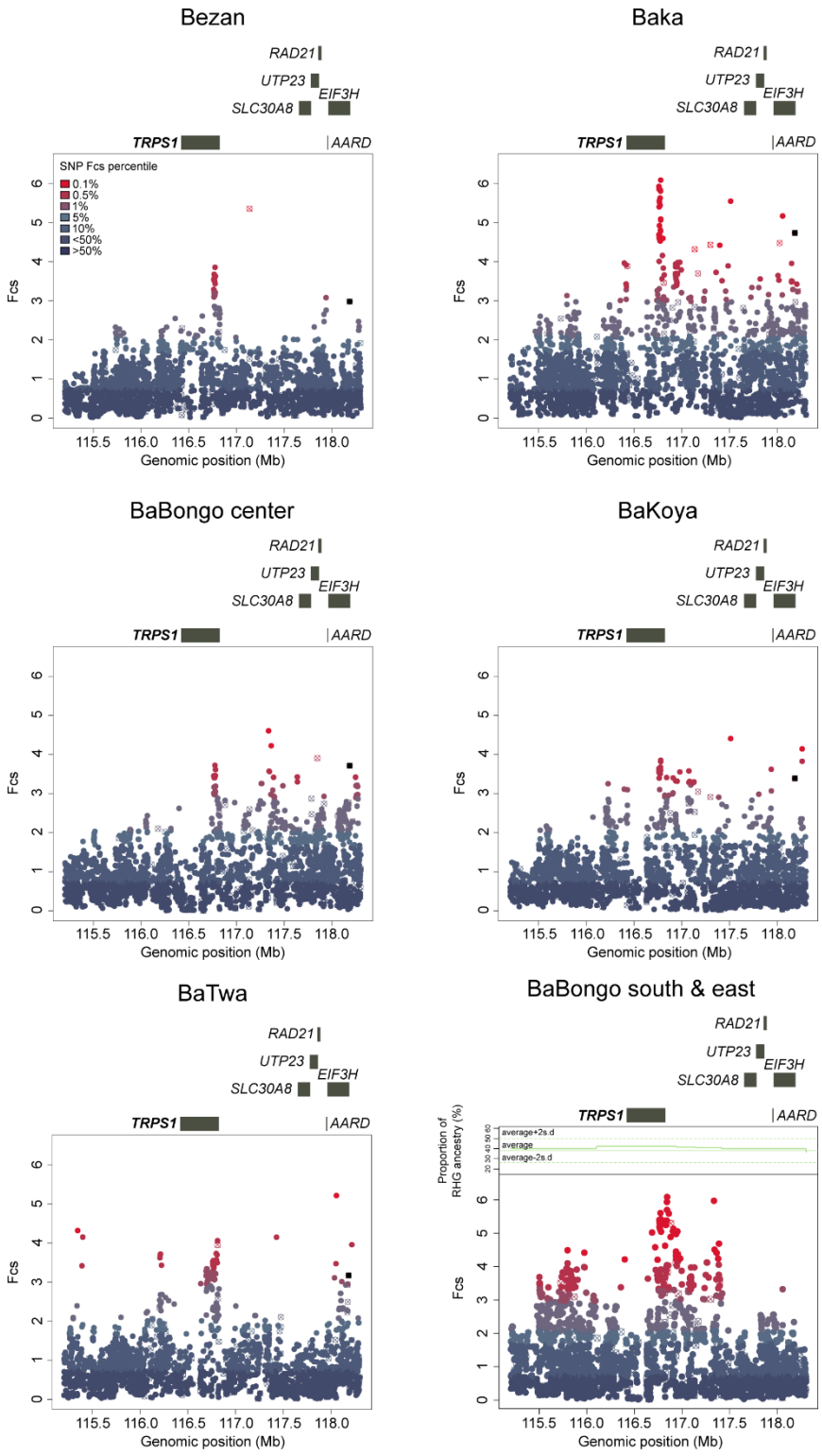

Figure S3. Selective Sweep Signal at the *TRPS1* Locus in African Rainforest Hunter-Gatherers. Related to Figure 2C.

Local genomic signals of classic sweeps at the candidate windows containing the *TRPS1* gene (chr8:116702422-116802422) in all RHG populations. Dot colors indicate SNP F_{CS} percentiles, black squares indicate non-synonymous mutations and circled crosses indicate non-imputed SNPs. Average local RHG ancestry is shown for the admixed BaBongo of south and east Gabon.

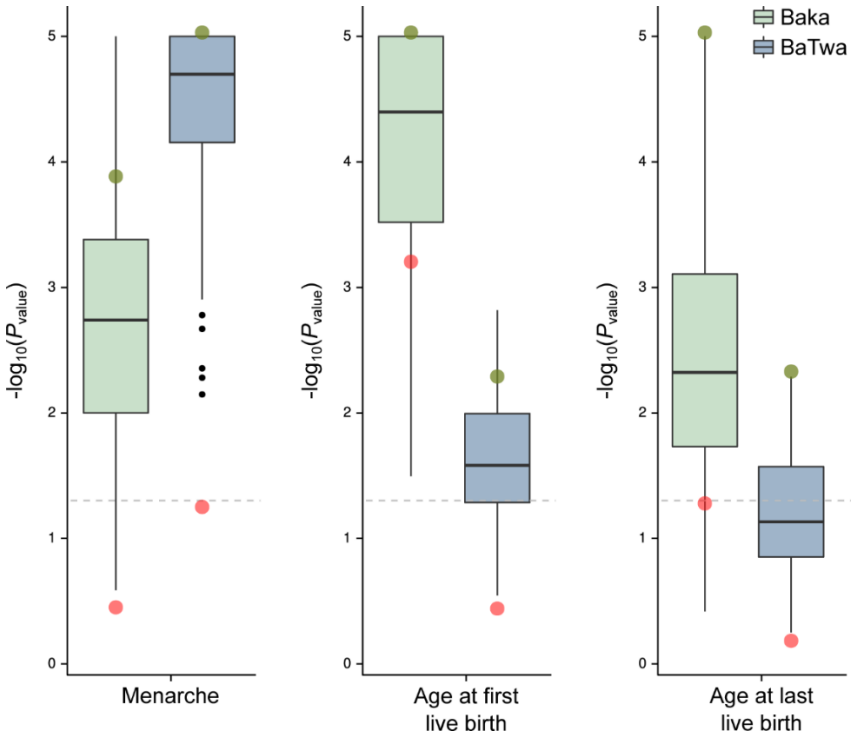

Figure S4. Significance of Tests for Polygenic Selection when Accounting for Pleiotropy or Reduced Number of Genomic Windows. Related to Figure 3A-C.
Red dots indicate $-\log_{10}$(non-adjusted *P*) *when accounting* for pleiotropy (i.e., after excluding height-associated windows). Green dots indicate $-\log_{10}$(non-adjusted *P*) *when not accounting* for pleiotropy. Boxplots correspond to$-\log_{10}$(non-adjusted *P*) of 100 random samples of *x* trait-associated windows, where *x* is the number of windows associated to the trait tested, *when accounting* for pleiotropy. The grey dashed line indicates the significance threshold $-\log_{10}(0.05)$. When red points are below both the dashed line and box plots, this indicates that the significant signals of polygenic selection are no longer significant, because of our correction for pleiotropy, and not because of the reduced number of windows.

Group	Population	Country	N	Reference for SNP array data (accession number)	Reference for exome sequencing (accession number)	Mean AGR ancestry	SD of AGR ancestry	Minimum AGR ancestry	Maximum AGR ancestry
wAGR	Tsogo	Gabon	29	[S1] (EGAS00001002078)	This study	80.7%	1.8%	77.4%	84.3%
wAGR	Galoa	Gabon	30	[S1] (EGAS00001002078)	This study	86.80%	3.2%	78.6%	92.5%
wAGR	Shake	Gabon	30	[S1] (EGAS00001002078)	This study	67%	2.90%	59.6%	71.3%
wAGR	Fang	Gabon	31	[S1] (EGAS00001002078)	This study	85.3%	1.1%	82.3%	87.8%
wAGR	Bapunu	Gabon	44	[S1] (EGAS00001002078)	[S2] (EGAS00001002457)	82.5%	3.6%	66.4%	88.2%
wAGR	Nzébi	Gabon	55	[S1] (EGAS00001002078)	[S2] (EGAS00001002457)	82.7%	3.1%	73%	87.7%
eAGR	BaKiga	Uganda	49	[S3] (EGAS00001000908)	[S2] (EGAS00001002457)	88.6%	1.9%	84.5%	92.8%
wRHG	Bezan	Cameroon	38	[S3] (EGAS00001000605)	This study	9.5%	13.3%	0%	45.5%
wRHG	BaBongo (center)	Gabon	21	This study	This study	9.4%	9.1%	0%	39.8%
wRHG	BaBongo (east)	Gabon	27	[S4] (EGAS00001000605)	This study	43.3%	11.2%	31.3%	82.8%
wRHG	BaBongo (south)	Gabon	33	[S4] (EGAS00001000605)	This study	24.3%	17.4%	0%	59.1%
wRHG	BaKoya	Gabon	26	[S1] (EGAS00001002078)	This study	4.1%	5.5%	0%	20.8%
wRHG	Baka	Cameroon/ Gabon	72/ 30	[S5] (EGAS00001001066) [S4] (EGAS00001000605)	[S2] (EGAS00001002457) This study	8.1%	10.7%	0%	51.4%
eRHG	BaTwa	Uganda	51	[S3] (EGAS00001000908)	[S2] (EGAS00001002457)	8.7%	12.2%	0%	43.5%
EUR	Belgian	Belgium	101	[S6] (EGAS00001001895)	[S6] (EGAS00001001895)	NA	NA	NA	NA

Table S1. Population description, sample size, and AGR ancestry proportions of the final dataset of 667 individuals. Related to Figure 1.

Ancestry proportions were estimated in AGR and RHG populations with ADMIXTURE at *K*=5 clusters.

Immune traits	Bezan	Baka	BaKoya	BaBongoC	BaBongoS/E	BaTwa	wAGR	eAGR
All II genes	2.13×10^{-2}	6.44×10^{-2}	0.324	$\mathbf{2.95\times10^{-3}}$	3.51×10^{-2}	0.674	3.33×10^{-2}	0.210
Adaptors	0.224	$\mathbf{2.54\times10^{-3}}$	0.034	9.25×10^{-3}	0.359	0.574	0.752	0.310
Regulators	0.140	0.497	0.166	1.84×10^{-2}	$\mathbf{3.36\times10^{-3}}$	$\mathbf{4.81\times10^{-3}}$	0.675	0.731
Secondary receptors	$\mathbf{2.26\times10^{-3}}$	$\mathbf{1.42\times10^{-4}}$	0.682	5.05×10^{-2}	0.144	0.697	0.985	0.487
Signal transducers	0.382	0.074	0.623	$\mathbf{1.55\times10^{-3}}$	2.53×10^{-2}	0.925	0.399	0.387
Sensors	$\mathbf{3.89\times10^{-4}}$	0.424	0.338	$\mathbf{5.99\times10^{-3}}$	0.507	0.157	$\mathbf{8.57\times10^{-3}}$	0.350
Transcription factors	0.536	0.931	0.930	0.450	0.496	0.362	1.67×10^{-2}	0.758
Accessory molecules	0.904	0.777	0.946	0.894	0.548	0.991	0.063	0.287
Effectors	0.250	0.302	0.140	0.410	0.699	0.418	0.387	0.395
Uncharacterized	0.413	0.154	0.047	0.939	0.678	0.906	0.134	3.09×10^{-2}
All VIP genes	0.179	0.199	0.083	$\mathbf{6.85\times10^{-4}}$	1.17×10^{-2}	0.236	0.193	0.323
dsDNA	0.268	$\mathbf{8.88\times10^{-3}}$	$\mathbf{3.93\times10^{-3}}$	2.46×10^{-2}	4.22×10^{-2}	0.073	0.648	0.787
ssRNA	0.182	0.339	0.084	$\mathbf{4.21\times10^{-4}}$	$\mathbf{7.98\times10^{-3}}$	0.562	0.336	0.316
ssDNA	0.919	0.093	$\mathbf{2.31\times10^{-3}}$	1.29×10^{-2}	0.429	0.619	0.592	0.736
dsDNART	0.365	0.985	0.710	0.382	0.152	0.459	0.111	0.569
ssRNART	0.213	0.496	0.864	0.162	0.258	0.069	0.118	0.135

Table S2. Polygenic Selection Signals for Immune-Related Traits in Central Africans. Related to Figure 3D.

Evidence for polygenic selection across 1,553 innate immunity (II) and 1,257 viral interacting protein (VIP) genes, based on their enrichment in high F_{CS} selection scores (FDR P<5%; in bold), relative to genome-wide expectations. Families of viral interacting proteins include host genes interacting with: double-stranded DNA virus (dsDNA), double-stranded DNA retrovirus (dsDNART), single-stranded DNA virus (ssDNA), single-stranded RNA virus (ssRNA) and single-stranded RNA retrovirus (ssRNART).

Supplemental References

S1. Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., Laval, G., Perry, G.H., Barreiro, L.B., Froment, A., et al. (2017). Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. Science *356*, 543-546.

S2. Lopez, M., Kousathanas, A., Quach, H., Harmant, C., Mouguiama-Daouda, P., Hombert, J.M., Froment, A., Perry, G.H., Barreiro, L.B., Verdu, P., et al. (2018). The demographic history and mutational load of African hunter-gatherers and farmers. Nat Ecol Evol *2*, 721-730.

S3. Perry, G.H., Foll, M., Grenier, J.C., Patin, E., Nedelec, Y., Pacis, A., Barakatt, M., Gravel, S., Zhou, X., Nsobya, S.L., et al. (2014). Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. Proc Natl Acad Sci U S A *111*, E3596-3603.

S4. Patin, E., Siddle, K.J., Laval, G., Quach, H., Harmant, C., Becker, N., Froment, A., Regnault, B., Lemee, L., Gravel, S., et al. (2014). The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. Nat Commun *5*, 3163.

S5. Fagny, M., Patin, E., MacIsaac, J.L., Rotival, M., Flutre, T., Jones, M.J., Siddle, K.J., Quach, H., Harmant, C., McEwen, L.M., et al. (2015). The epigenomic landscape of African rainforest hunter-gatherers and farmers. Nat Commun *6*, 10047.

S6. Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell *167*, 643-656 e617.