Published in final edited form as:

Genes Immun. 2019 February ; 20(2): 143–157. doi:10.1038/s41435-018-0016-1.

Population genetic evidence for positive and purifying selection acting at the human *IFN- γ* locus in Africa

Michael C. Campbell^{1,2}, Lunden T. Smith¹, Jayla Harvey¹

¹Department of Biology, Howard University, Washington, DC 20059, USA

²Department of Biostatistics, Yale University, New Haven, Connecticut 06520, USA

Abstract

Despite its critical role in the defense against microbial infection and tumor development, little is known about the range of nucleotide and haplotype variation at *IFN- γ* , or the evolutionary forces that have shaped patterns of diversity at this locus. To address this gap in knowledge, we examined sequence data from the *IFN- γ* gene in 1461 individuals from 15 worldwide populations. Our analyses uncovered novel patterns of variation in distinct African populations, including an excess of high frequency-derived alleles, unusually long haplotype structure surrounding the *IFN- γ* gene, and a “star-like” genealogy of African-specific haplotypes carrying variants previously associated with infectious disease. We also inferred a deep time to coalescence of variation at *IFN- γ* (~ 0.8 million years ago) and ancient ages for common polymorphisms predating the evolution of modern humans. Taken together, these results are congruent with a model of positive selection on standing variation in African populations. Furthermore, we inferred that common variants in intron 3 of *IFN- γ* are the likely targets of selection. In addition, we observed a paucity of non-synonymous substitutions relative to synonymous changes in the exons of *IFN- γ* in African and non-African populations, suggestive of strong purifying selection. Therefore, we contend that positive and purifying selection have influenced levels of diversity in different regions of *IFN- γ* , implying that these distinct genic regions are, or have been, functionally important. Overall, this study provides additional insights into the evolutionary events that have contributed to the frequency and distribution of alleles having a role in human health and disease.

Introduction

Interferon- γ , a proinflammatory cytokine encoded by the *IFN- γ* gene on chromosome 12, has a crucial role in the fight against intracellular pathogens in humans [1–4]. To date, studies have indicated that specific polymorphisms within this gene were associated with resistance to persistent parasitic infection. For example, genetic analyses have shown that the rs1861494 T-allele was correlated with a reduced risk for hepatic fibrosis (a condition in which excessive connective tissue builds up in the liver due to chronic liver cell injury) in individuals from East Africa infected by the *Schistosomiasis mansoni* parasite [5–10]. Prior

*Michael C. Campbell, michael.campbell1@howard.edu.

Conflict of interest: The authors declare that they have no conflict of interest.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41435-018-0016-1>) contains supplementary material, which is available to authorized users.

studies have also suggested that higher IFN- γ plasma levels in individuals with the *rs1861494* T-allele provide protection against severe fibrosis [5, 7], implying a functional effect of *rs1861494* on IFN- γ expression and disease outcome. In addition, data have indicated that high levels of the IFN- γ cytokine are protective against parasitemia and severe clinical malaria caused by the *Plasmodium falciparum* protozoan in individuals living in endemic regions [8, 10–12].

However, studies have also identified several polymorphisms at *IFN- γ* , including the *rs1861494* T-allele, which are risk factors for other types of disease [13–15]. For example, prior analyses of individuals with and without tuberculosis (TB) from South Asia showed that the *rs2430561* A-allele, the *rs1861493* A-allele, and the *rs1861494* T-allele were correlated with increased susceptibility to this disease [13, 14]. A similar set of polymorphisms—specifically the *rs1861494* T-allele, *rs2069718* C-allele, and the *rs2430561* A-allele—were also associated with elevated TB risk in individuals from East Asia [16–20]. In addition, the *rs2430561* T-allele and/or the *rs1861494* T-allele have been implicated in the onset of autoimmune disorders such as systemic lupus erythematosus [21–23] and Crohn’s disease [6] primarily in non-African populations. Thus, it can be argued that *IFN- γ* acts as a “double-edged sword” in immunity, serving as an essential component of host defense against parasitic infection, while contributing to increased susceptibility to other types of disease. Moreover, it has been suggested that these opposing effects mainly depend on disease phase, timing of immune response, and cell development stage [24, 25].

In addition to these gene association studies, a resequencing analysis of 186 individuals (62 individuals from Africa, Europe, and East Asia) found evidence for positive selection at *IFN- γ* in European and African populations [26]. However, given the small number of individuals per population in this study, populations from the same continental region (e.g., Africa) were pooled together in analyses. As a result, it is unclear whether or not distinct populations within geographic regions share the same signals of selection. Given the recent evidence for local adaptation, particularly in Africa [27–32], we hypothesized that geographically and ethnically diverse populations could show distinct patterns of genetic variability and signatures of selection at *IFN- γ* . To explore this hypothesis, we analyzed sequence variation at the *IFN- γ* gene in 1461 individuals from 15 global populations, including a large set of individuals from distinct populations in Africa. Our analyses revealed striking patterns of diversity at *IFN- γ* , including an excess of high-frequency-derived alleles and unusually long haplotypes flanking common variants in African populations. We also inferred a deep coalescence time for variation at this gene and ancient ages for individual variants that predate modern human origins ~ 200,000 years ago. Taken together, these results are consistent with a model of positive selection on standing variation in Africa. We also identified variants that appear to be targets of selection in intron 3 of *IFN- γ* . In addition, we found evidence for strong purifying selection against amino acid replacement changes in the coding region of *IFN- γ* in global populations. Therefore, we argue that both positive and purifying selection have acted on different genic regions of *IFN- γ* in Africans, suggesting a complex model of evolution at this locus. Overall, this study provides additional insights into the genetic history of functional variants that contribute to susceptibility and/or resistance to disease in human populations.

Results

Patterns of nucleotide variation

We analyzed sequence variation across the 4972 base pair (bp) length of *IFN-γ* on chromosome 12 in 1461 individuals from 15 globally diverse populations (Fig. 1a). Within this gene, we identified a total of 81 biallelic single nucleotide polymorphisms (SNPs) (Fig. 1b and Supplementary Table 1). The largest number of SNPs was found in Africa (49), followed by Europe (20), East Asia (17), and South Asia (11). We also observed a higher proportion of region-specific variants in Africa (51.9%) compared with other geographic regions (which varied from 6.3% in South Asia to 16.5% in East Asia; Fig. 1b). Overall, African populations had the largest number of polymorphic loci relative to non-Africans (Table 1).

Of the 81 polymorphisms identified, 15 of them were common (i.e., with a minor allele frequency (MAF) ≥ 5%) in populations of African descent (which includes African Americans), whereas a much smaller number of common SNPs (5) were found in non-Africans (Supplementary Table 1). Furthermore, 4 of the 15 common SNPs in Africans were present in global populations (Fig. 1b and Supplementary Table 1); they are *rs2069718* (genomic coordinate 68550162, A > G), *rs1861493* (genomic coordinate 68551196, A > G), *rs1861494* (genomic coordinate 68551409, C > T), and *rs2430561* (genomic coordinate 68552522, T > A), which are associated with diseases, such as malaria, TB, and a number of autoimmune disorders (see Introduction). Our data also showed clear population differences in the MAF at these common SNPs. Specifically, we observed the highest frequency of the minor C-allele at *rs1861494* in East Asian (29.5–52.8%) and European populations (16.3–36.8%), whereas this allele was present at a much lower frequency in populations of African (10–15.6%) and South Asian (14.5–15.4%) descent. Similarly, the minor *rs1861493* G-allele occurred at high frequency in East Asians (29.5–52.8%) and in Europeans (16.3–36.8%), but was found at much lower frequency in populations of African (7.1–11.7%) and of South Asian (14.5–15.5%) descent. In contrast, the minor G-allele at *rs2069718* was observed at a relatively low frequency (7.2–16.2%) in East Asians, but ranged from 36% to 42% frequency in Africans and African Americans, from 45% to 69% in Europeans, and from 45.2% to 51% in South Asians. Similarly, the minor A-allele at *rs2430561* ranged from 7.2% to 16.2% frequency in East Asia, but was more prevalent in Europeans (35–53%), South Asians (42–46%), and in populations of African descent (10–22.2%). Overall, these results demonstrate that loci associated with risk for infectious and complex disease traits vary considerably in frequency among global populations.

Finally, nucleotide diversity statistics for *IFN-γ* indicated a lower number of average pairwise differences per nucleotide between sequences (θ_π) compared with the number segregating sites (θ_w) in African and African American populations; specifically, estimates of θ_π ranged from 4.0×10^{-4} to 5.0×10^{-4} , whereas θ_w estimates occurred from 7.4×10^{-4} to 9.9×10^{-4} (Table 1). Conversely, estimates of θ_w were similar to θ_π estimates in non-Africans (ranging from 2.7×10^{-4} to 4.1×10^{-4} for θ_π and from 2.7×10^{-4} to 5.5×10^{-4} for θ_w). In addition, we examined the entire 501 bp coding region of *IFN-γ* and found that θ_π and θ_w were both zero for most populations (with the exception of the Mandinka from The

Gambia) due to the near-total absence of polymorphic sites within exons (Fig. 1c and Table 1).

Haplotype variation and relationships at *IFN-γ*

We analyzed phased haplotype data from *IFN-γ* and identified 75 distinct haplotypes in global populations (Supplementary Table 2). Of these variants, four common haplotypes, H3, H16, H28, and H32, accounted for 89.7–91.4% and 95.2–97.6% of the haplotypes in European and South Asian populations, respectively, whereas three common haplotypes (H3, H16, and H32) comprised more than 95% of the haplotypes in East Asians. Notably, the H28 haplotype was completely absent in our Chinese and Japanese samples (Supplementary Table 3). In contrast, populations of African descent have a larger set of common haplotypes (that includes H3, H16, H28, and H32), which accounted for 82.5–88.3% of the haplotypes present (Supplementary Table 4). Overall, low-frequency haplotypes (i.e., with < 5% frequency) comprised a small fraction of the total variation observed in African (11.7–17.5%) and non-African (2.4–10.3%) populations. In addition, estimates of haplotype diversity were higher in Africans and African Americans compared with non-African populations (Table 1).

A closer examination of H3, H16, H28, and H32 also revealed that these haplotypes differed from one another at four common sites, namely *rs2069718*, *rs1861493*, *rs1861494*, and *rs2430561* (corresponding to genomic coordinates 68550162, 68551196, 68551409, and 68552522, respectively). For example, the H3 and H28 haplotypes both carry the ancestral A-allele at *rs1861493* and the derived T-allele at *rs1861494*, but differ from one another by a single nucleotide at *rs2069718* (specifically, the ancestral A-allele at *rs2069718* was found on H3, whereas the derived G-allele at this site was present on H28).

We also reconstructed the genealogical relationships among global haplotypes and found that haplotype variation grouped into four distinct clusters (arbitrarily labeled I, II, III, and IV; Fig. 2). Each cluster consisted of low-frequency haplotypes radiating from a high-frequency haplotype in a “star-like” pattern congruent with models of positive or purifying selection [33]. Of particular interest is the observation that in clusters II and III most of the low-frequency haplotypes were specific to Africans and radiated around the common H3 and H28 haplotypes (both carry the ancestral *rs1861493* A-allele and the derived *rs1861494* T-allele) (Fig. 2). The low-frequency haplotypes radiating from H16 and H32 in clusters I and IV, respectively, were mostly specific to non-African populations (Fig. 2). Furthermore, the H16 and H32 haplotypes are highly divergent from another; specifically, they do not share any of the same alleles at *rs2069718*, *rs1861493*, *rs1861494*, and *rs2430561* (Fig. 2).

In addition, we did not observe extensive reticulations among haplotypes in the genealogy, suggesting that little recombination has occurred at *IFN-γ*. This inference of low levels of recombination was further supported by our pairwise linkage disequilibrium (LD) analysis, which showed strong allelic associations in African and non-African populations based on the D' statistic [34] (Supplementary Figure 1).

Tests of neutrality

To determine whether or not patterns of variation fit a neutral model of evolution, we applied the Fay and Wu's $H(H)$ statistic, which compares the number of high- and intermediate-frequency variants [35], to sequence data from global populations (Table 1). The Tajima's D (D_T) statistic was also calculated to test for a departure from neutrality in populations based on a comparison of intermediate- and low-frequency variants [36] (Table 1). As ancestral human populations have experienced changes in size within the last 100,000 years [37–39], which can skew the observed allele frequency spectrum, we estimated the significance of H and D_T under different simulated models of population growth generated using the ms software [40]. This analysis showed that Africans and African Americans exhibited significantly negative H statistics at $IFN-\gamma$ ($P < 0.05$) under models of population growth indicative of positive selection (Table 1 and Supplementary Table 6). By contrast, H values did not significantly deviate from neutral expectation in our non-African populations under any simulated model of growth (Table 1 and Supplementary Table 6). In addition, although observed D_T estimates were consistently negative in African and African American populations, we only detected a significant departure from neutral expectation under a growth model in the Mende from West Africa (Table 1 and Supplementary Table 7). Given that the D_T statistic does not distinguish between ancestral and derived states, and is influenced mostly by intermediate- and low-frequency variation [35], D_T may not be sensitive enough to detect selection in most of the populations of African descent based on the observed allele frequency spectrum at $IFN-\gamma$ (which consists of a number of high-frequency alleles). In addition, D_T values were primarily negative in non-African populations but they did not reach statistical significance (Table 1).

To further assess whether variation in the $IFN-\gamma$ protein-coding region is evolving neutrally, we calculated K_a/K_s ratios for the coding region using orthologous human–chimpanzee, human–gorilla, human–orangutan, human–dog, and human–mouse sequence pairs (Supplementary Table 8). K_a is the number of non-synonymous substitutions per non-synonymous site, whereas K_s is the number of synonymous substitutions per synonymous site [41, 42]. In this classic test, K_a/K_s ratios > 1 , $= 1$, and < 1 indicate positive selection, neutral evolution, and purifying selection within genes, respectively. Our analyses showed that K_a/K_s ratios ranged from 0.0833 to 0.4897 in our between-species comparisons, and were significantly < 1 ($P < 0.05$) based on the Fisher's exact test (Supplementary Table 8). We also performed a sliding window analysis of K_a/K_s across the coding region (using human–chimpanzee orthologous sequence pairs) to test for local variation in this ratio and found that K_a/K_s ranged from 0 to 0.36 (Supplementary Figure 3). Overall, our results indicated a striking deficit of amino acid substitutions relative to synonymous changes consistent with a model of purifying selection.

In addition, we applied the McDonald–Kreitman test [43] using polymorphisms from human populations and divergence data from the chimpanzee, to test for a departure from neutrality. This analysis showed that the relative number of polymorphism and divergence at synonymous and non-synonymous sites was not significantly different ($P > 0.05$; Table 1) in any of our populations. However, further examination of the coding region revealed a near

absence of intraspecific polymorphisms in the exons of *IFN-γ* in all populations (Table 1), indicative of purifying selection [44].

Extended haplotype homozygosity

To further interrogate *IFN-γ* for signatures of selection, we plotted the decay of haplotype homozygosity (using the extended haplotype homozygosity (EHH) statistic) with increasing distance from core SNPs (represented by *rs2069718*, *rs1861493*, *rs1861494*, and *rs2430561*). Our results showed long-range haplotype homozygosity, extending ~ 200 kb on chromosomes carrying the ancestral *rs1861493* A-allele relative to chromosomes with the derived *rs1861493* T-allele in Africans and African Americans. Similarly, we detected long EHH, spanning the same genomic distance on chromosomes carrying the derived T-allele at *rs1861494* in these same populations (Fig. 3). By contrast, we observed a faster decay of haplotype homozygosity on chromosomes with the derived and ancestral alleles at core SNPs *rs2069718* and *rs2430561* in Africans and African Americans (Fig. 3 and Supplementary Figure 2). Furthermore, our results showed no difference in the length of haplotype homozygosity on chromosomes with the derived or ancestral alleles at any of the four core SNPs in non-African populations (Fig. 3 and Supplementary Figure 2).

We also characterized long-range LD across the ~ 134 Mb region of chromosome 12 for each population using the integrated haplotype score (*iHS*) statistic [45]. Specifically, we calculated standardized *iHS* scores by normalizing raw values using the selscan software [46]. Based on this analysis, we did not observe outlier scores (i.e., standardized *iHS* > 2) for SNPs within *IFN-γ* in African populations (Supplementary Table 9). It is important to note, however, that the *iHS* statistic is most powerful for detecting recent selection when an advantageous allele has swept to moderate frequency (65–80%) [45]. Given that the *rs1861493* A-allele and the *rs1861494* T-allele range in frequency from 84% to 93% (Supplementary Table 1) in Africans and African Americans, we may not have sufficient power to detect long-range haplotypes flanking alleles at these core SNPs using this statistic [45, 47, 48]. In addition, *IFN-γ* did not show extreme *iHS* signals in any of the non-African populations consistent with the results from our EHH analyses. Specifically, standardized *iHS* scores for the core polymorphic sites did not fall outside the 95% cutoff of the empirical distribution of *iHS* values (Supplementary Table 9).

Population differentiation

To investigate patterns of genetic differentiation, we calculated mean F_{ST} among geographic regions (namely Africa, South Asia, East Asia, and Europe) using all of the polymorphic sites at *IFN-γ*. Global F_{ST} was estimated to be 0.11459, similar to previous F_{ST} estimates for autosomal loci, which typically range from 0.10 to 0.16 [49]. Further analyses of population differentiation showed that F_{ST} at most individual SNPs within *IFN-γ* were below 0.065, whereas F_{ST} values for *rs2069718*, *rs1861493*, *rs1861494*, and *rs2430561* were 0.171903, 0.11925, 0.0736783, and 0.171039, respectively (Supplementary Table 1). However, F_{ST} for these common SNPs were not outliers in the empirical distribution of F_{ST} values calculated for ~ 7.5 million SNPs that were randomly selected from across the genome (Supplementary Table 1).

Estimated age of *IFN-γ* alleles

Using a coalescent-based method [50], we constructed a gene tree showing the mutation history of variation at *IFN-γ* (Fig. 4). We also estimated the time to the most recent common ancestor (TMRCA) of the tree and the ages of individual polymorphisms (in years $\pm$ SD) (Fig. 4 and Supplementary Table 10). Our results showed a deep coalescence time for the origin of variation at *IFN-γ*; the mean TMRCA was 763204.16 years ago (ya) $\pm$ 232493.66 years, in general agreement with the TMRCA estimated in a prior study [26]. We also estimated the age for each variant of interest: *rs2069718* (484007.98 ya $\pm$ 91706.77 years), *rs1861493* (302292.70 ya $\pm$ 101047.28 years), *rs1861494* (240305.71 ya $\pm$ 69544.30 years), and *rs2430561* (203792.83 ya $\pm$ 74214.56 years) (Fig. 4 and Supplementary Table 8). We also inferred a more recent age of origin for most of the other polymorphisms within this gene (Fig. 4 and Supplementary Table 10).

Discussion

Evidence for recent selection in Africa

Our analyses uncovered striking patterns of diversity at *IFN-γ* in distinct Africans and African Americans, which have not previously been reported for this gene. In particular, we detected significantly negative Fay and Wu's *H* statistics, indicating an excess of high-frequency-derived alleles. Furthermore, our data revealed long-range EHH extending $\sim$ 200 kb on chromosomes with the ancestral *rs1861493* A-allele and long EHH spanning the same distance on chromosomes with the derived *rs1861494* T-allele. The “star-like” phylogeny of African-specific haplotypes surrounding the common H3 and H28 haplotypes (which carry the *rs1861493* A-allele and the *rs1861494* T-allele, respectively) also indicates that these haplotypes had undergone expansion in Africa. Overall, these above results are congruent with a model of positive selection in populations of African origin. Moreover, distinct African and African American populations exhibited similar signatures of selection. In addition, our analyses suggest that the *rs1861493* A-allele and the *rs1861494* T-allele are the likely targets of selection in these populations [51].

Of further interest, *rs1861493* and *rs1861494* are located in intron 3 of *IFN-γ*, raising the possibility that these non-coding loci have regulatory functions. Indeed, previous studies have demonstrated that the *rs1861494* T-allele has a greater affinity for binding to inducible transcription factors, such as nuclear factor- κ B (κ -light-chain-enhancer of activated B cells) and RU49 (zinc-finger transcription factor RU49), relative to the *rs1861494* C-allele, resulting in an upregulation of *IFN-γ* expression [6, 52]. Given that *rs1861493* and *rs1861494* are in strong LD, it is even possible that alleles at these loci could interact epistatically to influence gene transcription. However, further investigation into the potential interaction between alleles at *rs1861493* and *rs1861494* is needed. Nevertheless, based on our data and current information in the literature, we argue that the *rs1861493* A-allele and/or the *rs1861494* T-allele are genetic adaptations in Africa that play a role in gene function.

Although the precise environmental pressures driving adaptive evolution at *IFN-γ* are not known, we suggest that pathogens that diminish reproductive potential and that have a long-

standing relationship with *Homo sapiens* are likely the underlying causes of the signals of selection at this gene. One such pathogen is the *Plasmodium falciparum* parasite causing severe malaria, which is responsible for approximately one million deaths per year in Africa alone, representing one of the most life-threatening diseases on the continent [53, 54]. *P. falciparum* is an obligate pathogen (i.e., an organism that must infect a host in order to survive) that is estimated to have emerged in Africa between 10,000 and 300,000 years ago [55–57]. A prior study also showed that higher *IFN-γ* expression was associated with lower levels of malaria parasitemia and mild disease in affected individuals from Ghana [11, 12]. The impact of malaria on patterns of variation at a number of loci in Africans has been widely documented in the literature [27, 49, 57–59]. One classic example is the β -globin (*HBB*) gene on chromosome 11, which contains the HbS, HbC, and HbE alleles associated with resistance to *P. falciparum* malaria that occur at high frequency in malaria-exposed populations in West Africa [49, 54]. Another well-known example is the FY*O-null allele (a loss-of-function mutation) at the Duffy locus on chromosome 1, which underwent an ancient selective sweep to high frequency or to fixation in western and central Africa due to *Plasmodium vivax* malarial infection [58]. We cannot rule out the possibility, however, that the genetic adaptations at *IFN-γ* may have evolved in response to other types of pathogen, such as *Schistosoma* and *Leishmania* causing schistosomiasis and leishmaniasis, respectively, which are also fatal in Africa [60–62]. Thus, adaptive alleles at the *IFN-γ* gene could have risen to high frequency in Africans due to exposure to one or several types of obligate pathogen [5, 7, 63].

By contrast, our analyses also revealed that the allele frequency spectrum and haplotype data in non-Africans were not consistent with the patterns of diversity expected under a model of positive selection. In particular, Fay and Wu's *H* values did not deviate from neutrality and we did not observe extensive EHH around core SNPs at *IFN-γ* in these populations. Instead, we suggest that the high frequency of alleles at *IFN-γ* associated with disease susceptibility/resistance and their widespread geographic distribution outside of Africa likely resulted from the population bottleneck that occurred during the migration of ancestral humans from East Africa to other regions of the globe within the last 100,000 years.

Evidence for purifying selection in global populations

In addition, the reduced rate of fixation of amino acid substitutions (as measured by K_a/K_s ratios) and the absence of non-synonymous polymorphisms in the exons of *IFN-γ* in African and non-African populations suggest strong pressure (i.e., strong purifying selection) against variants that lead to amino acid substitutions. This finding is consistent with results from a prior study that showed a similar absence of amino acid replacement polymorphisms at *IFN-γ* in human populations [26, 64, 65]. Prior studies also have indicated that purifying selection has acted on a number of immunity genes and that *IFN-γ* has one of the most intense selective constraints on amino acid variation of all the immune genes [64]. This extreme evolutionary constraint at *IFN-γ* suggests that conserved proteins, encoded by exon sequences, have key functions that are performed either at an optimal or near-optimal level. In fact, it has been proposed that protein conservation at *IFN-γ* allows cells that have a role in immunity to recognize invariant protein motifs across a broad range of pathogens [51], enhancing the immune system's capacity to mount an effective response against a number of

invading microorganisms. Therefore, functional constraints may be imposed on protein-altering mutations that could potentially diminish the ability of IFN- γ cytokines to detect conserved motifs in pathogens. In addition, given the evidence for positive selection in intron 3 in populations of African descent, we argue that both purifying and positive selection have acted on loci in distinct genic regions of *IFN- γ* (i.e., coding and non-coding regions, respectively), suggesting a complex model of evolution at this gene.

Timing of common variants in African populations

Our GENETREE analysis yielded a deep coalescence time for variation at *IFN- γ* and an ancient TMRCA for alleles at *rs2069718*, *rs1861493*, *rs1861494* and *rs2430561* that predated the emergence of anatomically modern humans ~ 200,000 years ago. The widespread distribution of alleles at these sites across globally diverse populations also suggests that these common polymorphisms likely existed before human population divergence, further supporting our GENETREE results.

Typically, when mutations are selectively advantageous in a classic hard sweep model, the ages of these beneficial alleles are relatively young (with all coalescence events occurring in a short time in a population) [66–69]. That is to say, the TMRCA at a selected site is shorter than the time since the onset of selection pressure [67]. However, in the present study, we estimated a deep age of coalescence for common variants that are inferred to be under selection (specifically, the *rs1861493* A- and *rs1861494* T-alleles). Moreover, it appears that both the ancestral *rs1861493* A-allele and the derived *rs1861494* T-allele rose to high frequency on different haplotypes (H3 and H28). These genetic patterns (that is, the deep coalescent time and the occurrence of selected alleles on different haplotype backgrounds) are congruent with a scenario of positive selection from standing variation (a soft sweep). In this model of selection, if alleles are neutral or nearly neutral when they arise, they can persist in a population for a long time (with their frequency governed by drift) and when these pre-existing alleles become selectively advantageous in response to environmental change they quickly rise in frequency, resulting in a relatively deep TMRCA. Furthermore, if these adaptive alleles occur on different haplotypes, these haplotypes can also sweep through a population at the same time [66, 67]. Under this framework, multiple haplotypes with polymorphisms that are selectively advantageous can become frequent in populations.

Notably, although positive selection on standing variation is expected to increase allele frequency differences among populations [70, 71], we did not observe unusually high levels of genetic differentiation at common SNPs in our data. One possible explanation for this unexpected result is as follows: as both ancestral and derived alleles at common sites existed in humans in Africa before their geographic expansion across the globe and because these alleles also occur at appreciable frequency in most populations outside of Africa (likely due to the bottleneck event that accompanied the out of Africa exodus ~ 60,000–80,000 years ago), it is feasible that we would not necessarily observe large among-population differences in allele frequency, and thus not detect larger than expected F_{ST} values at common polymorphisms within this gene.

Implications for human biology and biomedicine

Pathogens are arguably among the strongest selective pressures to influence levels and patterns of variation in modern human populations [57, 72]. Environmental and cultural changes have exposed populations to infectious organisms together with human colonization of new environments, increased population density, and closer contact with animal disease vectors, such as domesticated animals [57]. As a result, during human history genetic and phenotypic adaptations evolved primarily as countermeasures to ongoing exposure to pathogens causing severe disease. In the case of IFN- γ , important roles of the IFN- γ protein include the activation of macrophages and the regulation of cytokine expression at other genes, such as *IL-10*, in response to pathogenic infection [73]. Therefore, selection for variation that upregulates IFN- γ protein production could potentially lead to greater resistance and/or tolerance to pathogenic organisms.

Although genetic adaptations are generally argued to be protective against infectious disease, there is increasing evidence that, in some cases, positively selected mutations are associated with increased risk for immune dysfunction, such as autoimmunity and inflammation [74]. For example, a resequencing analysis inferred that positive selection in response to pathogen-induced selective pressure increased the frequency of the V1059M variant at the *NALP1* locus in Europeans [75, 76]; however, data also showed that selection increased the frequency of the linked L155H variant, which has been implicated in the onset of autoimmune disorders, such as Addison's disease and vitiligo [77, 78]. Furthermore, studies have found that loci associated with common inflammatory disorders are enriched for signals of positive selection [53, 79], implying a connection between population genetic forces and disease risk. Thus, genetic adaptations that are protective against infection could potentially increase the risk for common complex disorders, representing a biological cost associated with infectious disease resistance. In terms of IFN- γ , we hypothesize that alleles correlated with increased susceptibility to autoimmune disorders exist at high frequency in humans due to recent selection for variation that provides protection against infectious agents in Africa. Given that these alleles exist in non-Africans, it is likely to be that they were present in the ancestors of contemporary non-African populations who migrated across the globe from Africa within the last 100,000 years.

Evolutionary methods, thus, can be highly informative for identifying variants that have a role in human adaptation and complex disease. An enhanced understanding of the interplay between evolution, genes, environment, and disease can also have clinically important implications, such as the development of new therapies for improved disease management that mimic, magnify or mitigate the physiological effects of resistance alleles.

Concluding remarks

Although the field of immunogenetics research has advanced over the last several years, further investigation is needed in order to understand the full range of genetic and environmental variables influencing infectious disease risk in human populations. In particular, additional resequencing analyses of genome-wide data together with disease phenotypes in diverse Africans (who are still highly under-represented in genomic studies) and in non-Africans will be important for discovering novel variants that contribute to

disease susceptibility and/or resistance. Future research that also incorporates environmental variables, such as diet, soil contaminants and climate, into analyses will be better able to identify non-genetic determinants of disease. Overall, these integrative studies will be informative for elucidating the relative impact of genetics, environment, and population-level phenomena on infectious and complex disease risk in modern humans.

Subjects and Methods

Population samples

We analyzed the 4972 base pair length of the *IFN- γ* gene on chromosome 12 in 1461 unrelated individuals in the 1000 Genomes Project (Phase 3) originating from four geographic regions: [1] Africa consisted of 99 Esan (ESN, Nigeria), 113 Mandinka (GWD; The Gambia), 54 Mende (MSL; Sierra Leone), 107 Yoruba (YRI; Nigeria), 99 Luhya (LWK; Kenya), 60 African Americans (ASW; United States); [2] South Asia encompassed 84 Bengali (BEB; Bangladesh), 103 Gurjarati (GIH, India); [3] East Asia included 103 Han (CHB, Beijing), 105 Han (CHS; southern China), 104 Japanese (JPT; Japan); and [4] Europe comprised 90 Great British (GBR; Great Britain), 99 Finnish (FIN; Finland), 107 Iberian (IBS; Spain), 104 Toscani Italians (TSI; Italy) [80]. Although we grouped African Americans with indigenous Africans above, we did not combine this recently admixed population with Africans when we examined genetic differences between geographic regions (for example, between Africa and East Asia, or between Africa and Europe).

Nucleotide variation

We extracted variant calls (from variant call format files; build GRCh 37 human assembly) for the *IFN- γ* gene located on chromosome 12 in the 1000 Genomes dataset using the vcftools software [81]. In the present study, we defined MAF as the second most frequent allele at a given polymorphic site in pooled populations (i.e., 2922 chromosomes). SNPs were also broadly classified as common or rare based on MAF. In particular, SNPs with an MAF > 5% were classified as common, whereas SNPs with an MAF < 5% were categorized as rare.

Tests of neutrality

Fay and Wu's $H(H)$ [34] and Tajima's $D(D_T)$ [36] statistics were calculated for *IFN- γ* . Statistical significance of observed H and D_T values was determined by comparing these observed statistics to the theoretical distributions of H and D_T values, generated using the ms program [40], to identify outlier values in the extreme tails of these distributions ($P = 0.05$). For Africans, we simulated a 2-, 4-, 6-, 8-, and 10-fold increase in population size (from an initial population size of 10,000 individuals) starting at 70,000 years ago (kya) until the present (Supplementary Tables 6 and 7) [37–39, 82]. For non-Africans, we simulated several demographic models involving a population bottleneck (from an initial population size of 10,000 individuals and decreasing to 2000 individuals) at 60 kya, corresponding to the approximate time modern humans migrated out of Africa [49], followed by a 10-, 20-, and 40- fold increase in population size starting at 50 kya until the present [37, 38] (Supplementary Tables 6 and 7). For all simulations, we used a generation time of 20 years.

To further analyze patterns of diversity, we calculated K_a/K_s (or dN/dS) ratios using orthologous human and non-human protein-coding sequences (specifically, human–chimpanzee, human–gorilla, human–orangutan, human–dog, and human–mouse sequence pairs) using the KaKs_Calculator v. 1.2 software. K_a (or dN) is the number of non-synonymous substitutions per non-synonymous site, whereas K_s (or dS) is the number of synonymous substitutions per synonymous site. For this analysis, a human protein-coding consensus sequence, derived from our 1461 individuals, was compared to non-human orthologous sequences downloaded from the Ensembl genome browser (<https://useast.ensembl.org/index.html>, last accessed on 13 January 2018). In this classical test, $K_a/K_s = 1$ indicates a given gene is evolving neutrally, whereas $K_a/K_s < 1$ and $K_a/K_s > 1$ signify purifying selection and an accelerated fixation of amino acid changes, respectively [41, 42]. Significant deviation from neutrality was determined using the Fisher's exact test implemented in KaKs_Calculator v. 1.2 [42]. We also performed a sliding window analysis of K_a/K_s (using human–chimpanzee orthologous sequence pairs) to identify localized departures from neutrality using DnaSP [83]. In this analysis, overlapping windows with a length of 50 bases and a step size of 25 bases (default settings in DnaSP [83]) were analyzed across the entire coding region of *IFN-γ* for each population and the results were presented as plots in Supplementary Figure 3.

In addition, we applied the McDonald–Kreitman test to our data. This statistical test compares the observed number of non-synonymous and synonymous polymorphic sites within species to the observed number of non-synonymous and synonymous divergent sites between species in a 2×2 contingency table. Under neutrality, the ratio of non-synonymous to synonymous fixed substitutions between species should be the same as the ratio of non-synonymous to synonymous polymorphism within species. This test of neutrality was performed using the DnaSP software [83].

Haplotype analysis

We examined fully phased haplotypes from *IFN-γ* in 1461 individuals in the 1000 Genomes Project. The genealogical relationships among haplotypes were inferred using the median-joining algorithm implemented in the Network 4.5 program [84]. The resulting topology was the tree with the minimum number of changes among all possible maximum parsimony trees from a given dataset. Haplotype diversity (*h*-diversity) was also calculated using the DnaSP software [83].

Extended haplotype homozygosity

We characterized long-range LD patterns on chromosome 12 for each population by calculating the EHH statistic [85]. In particular, we measured the decay of LD for a given core SNP by calculating EHH for the core SNP and the surrounding SNPs on chromosome 12 in the order of increasing distance, starting with the core SNP ($EHH = 1$) followed by the next closest neighboring SNPs on either side of the core SNP using the selscan package [46]. We focused our EHH analysis on four core SNPs (*rs2069718*, *rs1861493*, *rs1861494*, and *rs2430561*) that were previously associated with infectious and complex disease risk. In addition, we calculated *iHS* [45], which quantifies the amount of EHH on chromosomes with the ancestral allele versus chromosomes with the derived allele. In this method, *iHS* is

not reported at core sites with a $MAF < 0.05$ [45]. To identify outlier values, the unstandardized scores for $> 330,000$ SNPs across the ~ 134 Mb region of chromosome 12 were normalized in frequency bins with the norm program implemented in selscan [46]. This normalization step was performed using the default parameter of 100 frequency bins. The resulting standardized $|zHS|$ scores are roughly distributed normally with mean = 0 and SD = 1. SNPs with a standardized $|zHS| > 2$ represent the most extreme 5% of scores.

Linkage disequilibrium

We performed pairwise comparisons of LD between SNPs in each population using Haploview version 4.2 [86]. LD was measured with D' [34], which calculates the non-random association between alleles at SNP loci based on allele frequency. To maximize the statistical power to detect a linear dependence between SNP loci, we only included SNPs with a MAF above 1% in this analysis. The color of each square signifies the strength of the relationship between SNP alleles. For example, bright red squares indicate statistically significant LD between a given pair of SNP loci ($D' = 1$; logarithm of odds (LOD) > 2), whereas shades of pink/white squares indicate little evidence of LD ($D' = 1$; LOD < 2). Purple squares indicate high LD but with little statistical support (low LOD) [86]. The bold triangles in the plots also indicate strong blocks of LD between SNP markers; specifically, the first and last markers in a block are in strong LD with all intermediate markers. However, the intermediate markers are not necessarily in LD with each other [86].

Population differentiation

To quantify genetic differentiation among global populations, average F_{ST} , derived from sequence data, was calculated using the Weir and Cockerham method implemented in vcftools [81, 87]. We also computed among-population F_{ST} at individual polymorphic sites across *IFN- γ* and at ~ 7.5 million SNPs randomly selected from across the genome. The observed F_{ST} estimates for *IFN- γ* SNPs were then compared with the empirical distribution of F_{ST} values derived from genome-wide SNPs in order to identify outlier values (> 95 th percentile of the distribution). F_{ST} at individual *IFN- γ* sites and F_{ST} at the ~ 7.5 million individual sites across the genome were calculated using the same 15 populations.

Age estimates of mutations

We used a coalescent-based approach implemented in GENETREE [50] to estimate the following: (1) the expected TMRCA for the gene tree and (2) the expected ages of individual *IFN- γ* polymorphisms. As the first step in these calculations, we estimated the mutation rate (μ) by solving for the equation $\mu = D/2T$, where D is the average number of pairwise difference between human and chimpanzee sequences and T is an estimate of their divergence time (i.e., we used 6×10^6 years). We applied the GENETREE software to our human sequence data to obtain the maximum likelihood estimate (MLE) of θ over 1,000,000 runs [50]. Using our estimate of μ and the MLE of θ , we calculated the effective population size (N_e) parameter, which was inferred to be 21,228, according to the formula $N_e = \theta/4\mu$.

Then, using the MLE of θ as a parameter, TMRCA of the gene tree and the ages of individual mutations were estimated from the weighted average of simulated ages over 100,000 independent runs [50]. The resulting TMRCA estimates are given in units of N_e .

generations. To convert these estimates into years, ages in units of N_e generations were multiplied by 2, a generation time of 20 years and N_e (specifically, units of N_e generations $\times 2 \times 20 \times 21,228$). GENETREE assumes an infinite alleles model and no recombination [50]. Haplotypes and/or sites that were not compatible with these conditions were removed from the analysis (we mainly removed haplotypes that were present as singletons in pooled global populations).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the Center for Computational Biology and Bioinformatics (CCBB) at Howard University for providing cluster computational resources for this project. This work was supported by the Start-up Funds of Howard University to M.C. Campbell.

References

1. Schoenborn JR, Wilson CB. Regulation of interferon-gamma during innate and adaptive immune responses. *Adv Immunol.* 2007;96:41–101. [PubMed: 17981204]
2. Newport MJ, Finan C. Genome-wide association studies and susceptibility to infectious diseases. *Brief Funct Genom.* 2011;10:98–107.
3. Schroder K, Hertzog PJ, Ravasi T, Hume DA. Interferon-gamma: an overview of signals, mechanisms and functions. *J Leukoc Biol.* 2004;75:163–89. [PubMed: 14525967]
4. Zaidi MR, Merlino G. The two faces of interferon-gamma in cancer. *Clin Cancer Res.* 2011;17:6118–24. [PubMed: 21705455]
5. Chevillard C, Moukoko CE, Elwali NE, Bream JH, Kouriba B, Argiro L, et al. IFN-gamma polymorphisms (IFN-gamma + 2109 and IFN-gamma + 3810) are associated with severe hepatic fibrosis in human hepatic schistosomiasis (*Schistosoma mansoni*). *J Immunol.* 2003;171:5596–601. [PubMed: 14607968]
6. Gonsky R, Deem RL, Landers CJ, Haritunians T, Yang S, Targan SR. IFNG rs1861494 polymorphism is associated with IBD disease severity and functional changes in both IFNG methylation and protein secretion. *Inflamm Bowel Dis.* 2014;20:1794–801. [PubMed: 25171510]
7. Henri S, Chevillard C, Mergani A, Paris P, Gaudart J, Camilla C, et al. Cytokine regulation of periportal fibrosis in humans infected with *Schistosoma mansoni*: IFN-gamma is associated with protection against fibrosis and TNF-alpha with aggravation of disease. *J Immunol.* 2002;169:929–36. [PubMed: 12097398]
8. King T, Lamb T. Interferon-gamma: the Jekyll and Hyde of malaria. *PLoS Pathog.* 2015;11:e1005118. [PubMed: 26426121]
9. D’Ombrain MC, Robinson LJ, Stanicic DI, Taraika J, Bernard N, Michon P, et al. Association of early interferon-gamma production with immunity to clinical malaria: a longitudinal study among Papua New Guinean children. *Clin Infect Dis.* 2008;47:1380–7. [PubMed: 18947328]
10. Gun SY, Claser C, Tan KS, Renia L. Interferons and interferon regulatory factors in malaria. *Mediat Inflamm.* 2014;2014:243713.
11. Dodo D, Omer FM, Todd J, Akanmori BD, Koram KA, Riley EM. Absolute levels and ratios of proinflammatory and anti-inflammatory cytokine production in vitro predict clinical immunity to *Plasmodium falciparum* malaria. *J Infect Dis.* 2002;185:971–9. [PubMed: 11920322]
12. Inoue S, Niikura M, Mineo S, Kobayashi F. Roles of IFN-gamma and gammadelta T cells in protective immunity against blood-stage malaria. *Front Immunol.* 2013;4:258. [PubMed: 24009610]

13. Ansari A, Hasan Z, Dawood G, Hussain R. Differential combination of cytokine and interferon-gamma + 874 T/A polymorphisms determines disease severity in pulmonary tuberculosis. *PLoS One*. 2011;6:e27848. [PubMed: 22140472]
14. Abhimanyu BM, Jha P, Indian Genome Variation C. Footprints of genetic susceptibility to pulmonary tuberculosis: cytokine gene variants in north Indians. *Indian J Med Res*. 2012;135:763–70. [PubMed: 22771610]
15. Lio D, Marino V, Serauto A, Gioia V, Scola L, Crivello A, et al. Genotype frequencies of the + 874T- > A single nucleotide polymorphism in the first intron of the interferon-gamma gene in a sample of Sicilian patients affected by tuberculosis. *Eur J Immunogenet*. 2002;29:371–4. [PubMed: 12358843]
16. Lee SW, Chuang TY, Huang HH, Lee KF, Chen TT, Kao YH, et al. Interferon gamma polymorphisms associated with susceptibility to tuberculosis in a Han Taiwanese population. *J Microbiol Immunol Infect*. 2015;48:376–80. [PubMed: 24529854]
17. Shen C, Jiao WW, Feng WX, Wu XR, Xiao J, Miao Q, et al. IFNG polymorphisms are associated with tuberculosis in Han Chinese pediatric female population. *Mol Biol Rep*. 2013;40:5477–82. [PubMed: 23737189]
18. Hu Y, Wu L, Li D, Zhao Q, Jiang W, Xu B. Association between cytokine gene polymorphisms and tuberculosis in a Chinese population in Shanghai: a case-control study. *BMC Immunol*. 2015;16:8. [PubMed: 25887222]
19. He J, Wang J, Lei D, Ding S. Analysis of functional SNP in ifng/ifngr1 in Chinese Han population with tuberculosis. *Scand J Immunol*. 2010;71:452–8. [PubMed: 20500698]
20. Sun Y, Lu Y, Li T, Xie L, Deng Y, Li S, et al. Interferon gamma + 874T/A polymorphism increases the risk of hepatitis virus-related diseases: evidence from a meta-analysis. *PLoS ONE*. 2015;10:e0121168. [PubMed: 25939029]
21. Al-Kholy W, Elsaid A, Sleem A, Fathy H, Elshazli R, Settin A. TNF-alpha - 308 G>A and IFN-gamma + 874 A>T gene polymorphisms in Egyptian patients with lupus erythematosus. *Meta Gene*. 2016;9:137–41. [PubMed: 27331019]
22. Kim K, Cho SK, Sestak A, Namjou B, Kang C, Bae SC. Interferon-gamma gene polymorphisms associated with susceptibility to systemic lupus erythematosus. *Ann Rheum Dis*. 2010;69:1247–50. [PubMed: 19919944]
23. Lee YH, Bae SC. Association between interferon-gamma + 874 T/A polymorphism and susceptibility to autoimmune diseases: a meta-analysis. *Lupus*. 2016;25:710–8. [PubMed: 26703437]
24. Ottum PA, Arellano G, Reyes LI, Iruretagoyena M, Naves R. Opposing roles of interferon-gamma on cells of the central nervous system in autoimmune neuroinflammation. *Front Immunol*. 2015;6:539. [PubMed: 26579119]
25. Salih MA, Fakiola M, Abdelraheem MH, Younis BM, Musa AM, ElHassan AM, et al. Insights into the possible role of IFNG and IFNGR1 in Kala-azar and Post Kala-azar Dermal Leishmaniasis in Sudanese patients. *BMC Infect Dis*. 2014;14:662. [PubMed: 25466928]
26. Manry J, Laval G, Patin E, Fornarino S, Tichit M, Bouchier C, et al. Evolutionary genetics evidence of an essential, nonredundant role of the IFN-gamma pathway in protective immunity. *Hum Mutat*. 2011;32:633–42. [PubMed: 21448974]
27. Ko WY, Kaercher KA, Giombini E, Marcatili P, Froment A, Ibrahim M, et al. Effects of natural selection and gene conversion on the evolution of human glycophorins coding for MNS blood polymorphisms in malaria-endemic African populations. *Am J Hum Genet*. 2011;88:741–54. [PubMed: 21664997]
28. Ko WY, Rajan P, Gomez F, Scheinfeldt L, An P, Winkler CA, et al. Identifying Darwinian selection acting on different human APOL1 variants among diverse African populations. *Am J Hum Genet*. 2013;93:54–66. [PubMed: 23768513]
29. Ranciaro A, Campbell MC, Hirbo JB, Ko WY, Froment A, Anagnostou P, et al. Genetic origins of lactase persistence and the spread of pastoralism in Africa. *Am J Hum Genet*. 2014;94:496–510. [PubMed: 24630847]

30. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet.* 2007;39:31–40. [PubMed: 17159977]
31. Campbell MC, Ranciaro A, Zinshteyn D, Rawlings-Goss R, Hirbo J, Thompson S, et al. Origin and differential selection of allelic variation at TAS2R16 associated with salicin bitter taste sensitivity in Africa. *Mol Biol Evol.* 2014;31:288–302. [PubMed: 24177185]
32. Crawford NG, Kelly DE, Hansen MEB, Beltrame MH, Fan S, Bowman SL, et al. Loci associated with skin pigmentation identified in African populations. *Science.* 2017;358:pii: eaan8433. [PubMed: 29025994]
33. Bamshad M, Wooding SP. Signatures of natural selection in the human genome. *Nat Rev Genet.* 2003;4:99–111. [PubMed: 12560807]
34. Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics.* 1964;49:49–67. [PubMed: 17248194]
35. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics.* 2000;155:1405–13. [PubMed: 10880498]
36. Tajima F, Nei M. Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol.* 1984;1:269–85. [PubMed: 6599968]
37. Marth GT, Czabarka E, Murvai J, Sherry ST. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics.* 2004;166:351–72. [PubMed: 15020430]
38. Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci USA.* 2005;102:18508–13. [PubMed: 16352722]
39. Cox MP, Morales DA, Woerner AE, Sozanski J, Wall JD, Hammer MF. Autosomal resequence data reveal Late Stone Age signals of population expansion in sub-Saharan African foraging and farming populations. *PLoS ONE.* 2009;4:e6366. [PubMed: 19641603]
40. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. *Genetics.* 1992;132:583–9. [PubMed: 1427045]
41. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. *Annu Rev Genet.* 2013;47:97–120. [PubMed: 24274750]
42. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom Proteom Bioinform.* 2006;4:259–63.
43. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature.* 1991;351:652–4. [PubMed: 1904993]
44. Oleksyk TK, Smith MW, O'Brien SJ. Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond B Biol Sci.* 2010;365:185–205. [PubMed: 20008396]
45. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4:e72. [PubMed: 16494531]
46. Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol.* 2014;31:2824–7. [PubMed: 25015648]
47. Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. Properties of different selection signature statistics and a new strategy for combining them. *Hered (Edinb).* 2015;115:426–36.
48. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 2007;5:e171. [PubMed: 17579516]
49. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genom Hum Genet.* 2008;9:403–33.
50. Griffiths RC, Tavaré S. Ancestral inference in population genetics. *Stat. Sci* 1994;9:307–319.
51. Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proc Natl Acad Sci USA.* 2009;106:7073–8. [PubMed: 19359493]
52. Kumar A, Ghosh B. A single nucleotide polymorphism (A → G) in intron 3 of IFN γ gene is associated with asthma. *Genes Immun.* 2008;9:294–301. [PubMed: 18385742]

53. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet.* 2010;11:17–30. [PubMed: 19953080]
54. Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet.* 2005;77:171–92. [PubMed: 16001361]
55. Hughes AL, Verra F. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc Biol Sci.* 2001;268:1855–60. [PubMed: 11522206]
56. Tanabe K, Mita T, Jombart T, Eriksson A, Horibe S, Palacpac N, et al. *Plasmodium falciparum* accompanied the human expansion out of Africa. *Curr Biol.* 2010;20:1283–9. [PubMed: 20656209]
57. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. *Nat Rev Genet.* 2014;15:379–93. [PubMed: 24776769]
58. McManus KF, Taravella AM, Henn BM, Bustamante CD, Sikora M, Cornejo OE. Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS Genet.* 2017;13:e1006560. [PubMed: 28282382]
59. Gomez F, Tomas G, Ko WY, Ranciaro A, Froment A, Ibrahim M, et al. Patterns of nucleotide and haplotype diversity at ICAM-1 across global human populations with varying levels of malaria exposure. *Hum Genet.* 2013;132:987–99. [PubMed: 23609612]
60. Sirugo G, Hennig BJ, Adeyemo AA, Matimba A, Newport MJ, Ibrahim ME, et al. Genetic studies of African populations: an overview on disease susceptibility and response to vaccines and therapeutics. *Hum Genet.* 2008;123:557–98. [PubMed: 18512079]
61. Adenowo AF, Oyinloye BE, Ogunyinka BI, Kappo AP. Impact of human schistosomiasis in sub-Saharan Africa. *Braz J Infect Dis.* 2015;19:196–205. [PubMed: 25636189]
62. Al-Salem W, Herricks JR, Hotez PJ. A review of visceral leishmaniasis during the conflict in South Sudan and the consequences for East African countries. *Parasit Vectors.* 2016;9:460. [PubMed: 27549162]
63. Russell SB, Smith JC, Huang M, Trupin JS, Williams SM. Pleiotropic effects of immune responses explain variation in the prevalence of fibroproliferative diseases. *PLoS Genet.* 2015;11: e1005568. [PubMed: 26540410]
64. Manry J, Laval G, Patin E, Fornarino S, Itan Y, Fumagalli M, et al. Evolutionary genetic dissection of human interferons. *J Exp Med.* 2011;208:2747–59. [PubMed: 22162829]
65. Wang Z, Zhong M, Fu M, Dou T, Bian Z. Evidence of positive selection at signal peptide region of interferon gamma. *Biosci Biotechnol Biochem.* 2014;78:588–92. [PubMed: 25036954]
66. Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol.* 2013;28:659–69. [PubMed: 24075201]
67. Hermisson J, Pennings PS. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol Evol.* 2017;8:700–16.
68. Kelley JL. Systematic underestimation of the age of selected alleles. *Front Genet.* 2012;3:165. [PubMed: 22969791]
69. Fu W, Akey JM. Selection and adaptation in the human genome. *Annu Rev Genom Hum Genet.* 2013;14:467–89.
70. Peter BM, Huerta-Sanchez E, Nielsen R. Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genet.* 2012;8:e1003011. [PubMed: 23071458]
71. Barrett RD, Schluter D. Adaptation from standing genetic variation. *Trends Ecol Evol.* 2008;23:38–44. [PubMed: 18006185]
72. Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R. Signatures of environmental genetic adaptation pin-point pathogens as the main selective pressure through human evolution. *PLoS Genet.* 2011;7:e1002355. [PubMed: 22072984]
73. Stahl JL, Cook EB, Graziano FM, Barney NP. Differential and cooperative effects of TNFalpha, IL-1beta, and IFNgamma on human conjunctival epithelial cell receptor expression and chemokine release. *Invest Ophthalmol Vis Sci.* 2003;44:2010–5. [PubMed: 12714637]
74. Sironi M, Clerici M. The hygiene hypothesis: an evolutionary perspective. *Microbes Infect.* 2010;12:421–7. [PubMed: 20178858]

75. Vasseur E, Boniotto M, Patin E, Laval G, Quach H, Manry J, et al. The evolutionary landscape of cytosolic microbial sensors in humans. *Am J Hum Genet.* 2012;91:27–37. [PubMed: 22748209]
76. Cagliani R, Sironi M. Pathogen-driven selection in the human genome. *Int J Evol Biol.* 2013;2013:204240. [PubMed: 23533945]
77. Magitta NF, Boe Wolff AS, Johansson S, Skinningsrud B, Lie BA, Myhr KM, et al. A coding polymorphism in NALP1 confers risk for autoimmune Addison's disease and type 1 diabetes. *Genes Immun.* 2009;10:120–4. [PubMed: 18946481]
78. Jin Y, Mailloux CM, Gowan K, Riccardi SL, LaBerge G, Bennett DC, et al. NALP1 in vitiligo-associated multiple autoimmune disease. *N Engl J Med.* 2007;356:1216–25. [PubMed: 17377159]
79. Raj T, Kuchroo M, Replogle JM, Raychaudhuri S, Stranger BE, De Jager PL. Common risk alleles for inflammatory diseases are targets of recent positive selection. *Am J Hum Genet.* 2013;92:517–29. [PubMed: 23522783]
80. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491:56–65. [PubMed: 23128226]
81. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8. [PubMed: 21653522]
82. Gay J, Myers S, McVean G. Estimating meiotic gene conversion rates from population genetic data. *Genetics.* 2007;177:881–94. [PubMed: 17660532]
83. Librado P, Rozas J. DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25:1451–2. [PubMed: 19346325]
84. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999;16:37–48. [PubMed: 10331250]
85. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature.* 2002;419:832–7. [PubMed: 12397357]
86. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21:263–5. [PubMed: 15297300]
87. Weir CC. Estimating F-statistics for the analysis of population structure. *Evolution.* 1984;38:1358–70. [PubMed: 28563791]

Fig. 1.

Distribution of nucleotide diversity. **a** Gene structure for *IFN-γ* (5' to 3'). The rs identifier and location of common SNPs, previously associated with disease susceptibility/resistance, are also indicated in the figure. **b** Venn diagram showing the number of region-specific and shared polymorphisms within and among continental populations, respectively. Each colored oval represents a distinct geographic region (i.e., Africa, South Asia, Europe, and East Asia). Because we are focused on continental differences here, we did not include our African American sample in this specific analysis. The counts in the intersecting portions of the Venn diagram represent the number of shared polymorphisms among regions, whereas the counts in the non-intersecting portions of diagram are specific to a given geographic region. The percentages in parentheses are the counts divided by the total number of polymorphisms identified at *IFN-γ*. **c** This bar plot indicates the proportion of polymorphisms found in introns, exons, 5'-UTR (5'-untranslated region) and the 3'-UTR (3'-untranslated region) of *IFN-γ*.

Fig. 2.

Network of global haplotype relationships. Genealogical relationships of haplotypes were reconstructed using the maximum parsimony criterion implemented in Network v. 5.0 [84]. Circles represent haplotypes and the size of the circles is proportional to the number of chromosomes with a given haplotype. Colors within each haplotype represent the proportion of individuals within a population or geographic region with that particular haplotype. The lines are the mutational changes between haplotypes. The small red dots represent haplotypes that were not observed in our data set (i.e., they were unsampled) but are inferred to exist. The nucleotide variants present at *rs2069718*, *rs1861493*, *rs1861494*, and *rs2430561* (ordered 3' to 5'), which define a given haplotype cluster, are present at the top of the genealogy. The dashed rectangular boxes surround haplotypes that are inferred to be under selection; underlined nucleotides are the inferred targets of selection. We also indicated the chimpanzee haplotype with a red arrow and included the nucleotides present within this haplotype at *rs2069718*, *rs1861493*, *rs1861494*, and *rs2430561*.

Fig. 3.

EHH plots for selected African and non-African populations. Plots show the decay of haplotype homozygosity with increasing distance from core SNPs (represented here by *rs1861493* and *rs1861494*) located in the *IFN- γ* gene. The x axis is the chromosome position of SNPs spanning chromosome 12 and the y axis is the probability that two chromosomes are homozygous at all SNPs for the entire interval from the core region to distance x . $EHH = 0$ means all extended haplotypes are different, whereas $EHH = 1$ indicates that all extended haplotypes are the same. In the plots, the blue line represents the decay of homozygosity of chromosomes carrying the ancestral allele at the core, while the orange line signifies the decay of homozygosity on chromosomes carrying the derived core allele

Fig. 4.

Inferred TMRCA of global variation and individual ages of mutations. A gene tree describing the mutation history of polymorphisms in the *IFN-γ* gene was constructed using a coalescent-based method implemented in GENETREE [50]. The TMRCA of the resulting tree and the ages of individual mutations were estimated from the weighted average of simulated TMRCA over 100,000 independent runs. The mean TMRCA and corresponding SD are given in Supplementary Table 10. The scale (given in years) is located along the vertical axis of the figure. Dots on internal and external branches represent mutations. More specifically, red dots indicate variants that arose after the origin of anatomically modern *Homo sapiens* ~ 200,000 years ago (delineated by the dashed horizontal line), whereas black dots signify variants that predate the emergence of modern *H. sapiens*. The black dots with the SNP identifiers (from *dbSNP*) are common SNPs that are shared among populations and are associated with disease susceptibility and/or resistance

Table 1

Nucleotide diversity statistics and tests of neutrality

Population	2N	S	Pi (θ_{π})	Theta (θ_W)	D_T	H	h	h-diversity	D _N	D _S	P _N	P _S	P-value (MK test)
African American (ASW)	120	20	0.00041	0.00075	-1.27958	-3.85686 ^a	15	0.814	1	3	0	0	NA
Esan (ESN)	198	23	0.00046	0.00079	-1.15516	-3.25981 ^a	18	0.843	1	3	0	0	NA
Mandinka (GWD)	226	27	0.00048	0.00091	-1.29260	-2.80401 ^a	20	0.831	1	3	0	1	$P = 1.000$
Mende (MSL)	162	28	0.00050	0.00099	-1.42569 ^a	-3.14240 ^a	21	0.866	1	3	0	0	NA
Luhya (LWK)	198	19	0.00040	0.00079	-1.34884	-4.25842 ^a	19	0.876	1	3	0	0	NA
Yoruba (YRI)	214	22	0.00045	0.00074	-1.04979	-3.34360 ^a	17	0.834	1	3	0	0	NA
Bengali (BEB)	168	9	0.00032	0.00032	0.01621	-0.79498	10	0.677	1	3	0	0	NA
Gujarati (GIH)	206	7	0.00031	0.00027	0.31458	-0.98129	8	0.667	1	3	0	0	NA
Finnish (FIN)	198	9	0.00041	0.00034	0.48338	0.11198	10	0.746	1	3	0	0	NA
Great British (GBR)	180	7	0.00039	0.00024	1.29579	-0.57728	7	0.694	1	3	0	0	NA
Iberian (IBS)	214	10	0.00038	0.00034	0.26467	-0.76486	9	0.738	1	3	0	0	NA
Toscani (TSI)	214	12	0.00034	0.00041	-0.40609	-1.24979	12	0.736	1	3	0	0	NA
Han Chinese in Beijing (CHB)	206	10	0.00030	0.00034	-0.29118	0.10703	10	0.629	1	3	0	0	NA
Southern Han Chinese (CHS)	210	10	0.00029	0.00034	-0.34324	-0.06015	8	0.617	1	3	0	0	NA
Japanese (JPT)	208	8	0.00027	0.00027	-0.05623	0.16499	7	0.579	1	3	0	0	NA

Tajima's $D(D_T)$ and Fay and Wu's $H(H)$ were calculated for $IFN-\gamma$. Statistical significance was determined by comparing the observed D_T and H estimates for each population to simulated D_T and H statistics under different scenarios of demographic growth (Supplementary Tables 2 and 3). $2N$ is the number of gene copies analyzed in each population. S is number of segregating sites; θ_{π} is the average number of pairwise nucleotide differences per site. θ_W is Watterson's estimate of diversity based on segregating sites; h is the number of haplotypes and h -diversity is the corresponding haplotype diversity for each population. The table also reports synonymous polymorphic sites (PS), replacement polymorphic sites (PN), synonymous divergent sites (DS), and replacement divergent sites (DN). Statistical significance for the McDonald-Kreitman (MK) test was determined using Fisher's exact test (NA means that the contingency table could not be computed).

^aSignificance at $P < 0.05$.