Genetic differentiation in East African ethnicities and its relationship with endurance running success

André L. S. Zani¹, Mateus H. Gouveia², Marla M. Aquino³, Rodrigo Quevedo⁴, Rodrigo L. Menezes⁵, Charles Rotimi², Gerald O. Lwande⁶, Collins Ouma⁵, Ephrem Mekonnen⁶, Nelson J. R. Fagundes¹, ⁷*

¹ Postgraduate Program in Genetics and Molecular Biology, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil, ² Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, ³ Department of Genetics, Ecology and Evolution, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil, ⁴ School of Physical Education, Physical Therapy and Dance, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil, ⁵ Department of Biomedical Sciences and Technology, Maseno University, Maseno, Kenya, ⁶ Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia, ⁷ Postgraduate Program in Animal Biology, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

* nelson.fagundes@ufrgs.br

Abstract

Since the 1960s, East African athletes, mainly from Kenya and Ethiopia, have dominated long-distance running events in both the male and female categories. Further demographic studies have shown that two ethnic groups are overrepresented among elite endurance runners in each of these countries: the Kalenjin, from Kenya, and the Oromo, from Ethiopia, raising the possibility that this dominance results from genetic or/and cultural factors. However, looking at the life history of these athletes or at loci previously associated with endurance athletic performance, no compelling explanation has emerged. Here, we used a population approach to identify peaks of genetic differentiation for these two ethnicities and compared the list of genes close to these regions with a list, manually curated by us, of genes that have been associated with traits possibly relevant to endurance running in GWAS studies, and found a significant enrichment in both populations (Kalenjin, \( P = 0.048 \), and Oromo, \( P = 1.6 \times 10^{-5} \)). Those traits are mainly related to anthropometry, circulatory and respiratory systems, energy metabolism, and calcium homeostasis. Our results reinforce the notion that endurance running is a systemic activity with a complex genetic architecture, and indicate new candidate genes for future studies. Finally, we argue that a deterministic relationship between genetics and sports must be avoided, as it is both scientifically incorrect and prone to reinforcing population (racial) stereotyping.

Introduction

The ability to run long distances has played an important role in human evolution, and our species stands among the best endurance runners of all mammals [1]. In modern times, such
Endurance running success and genetic differentiation in East Africa

ability manifests itself in athletics, especially in long-distance running events, in which elite athletes can run a marathon (42.195 km) in about two hours. East African athletes have dominated these events since the 1960s, in both the male and female categories [2]. Their overrepresentation among the world’s best endurance runners is so impressive that, as of January 2021, of the 100 athletes from each sex holding the best marathon times in history, 76 women and 93 men were born in East Africa (https://www.worldathletics.org/disciplines/road-running/marathon) (Fig 1A). The vast majority of these athletes come from two countries and, more importantly, two ethnicities within those countries: the Kalenjin, from Kenya, and the Oromo, from Ethiopia [2,3]. More information about these countries and ethnicities can be found in the S1 Text.

While the regional and ethnical clustering of best-ever performing athletes may suggest a shared genetic or cultural background as important factors, no simple explanation for their dominance has emerged yet [4]. This is not surprising, considering that the success in a systemic activity such as endurance running is probably the result of a complex interplay between multiple innate and trainable traits [4]. Environmental factors suggested as relevant to the success of Kenyan and Ethiopian athletes include: (1) a favorable diet; (2) living and training at high altitude; (3) socio-economic motivation to escape poverty; and (4) the habit of running to school as children [5]. However, none of these factors alone can explain such success, as they are also present in many countries and populations that have never produced world-class athletes in this proportion [4]. Besides, even in East African populations, evidence for a causal relation in each of those cases is, at best, tenuous [5]. This suggests that either many environmental factors act in concert to favor endurance running success, that there is an innate component that needs to be accounted for, or, more likely, both.

Regarding genetic factors, only two candidate genes have been investigated in East African populations [4,5]. Yang et al. [6] studied the ACTN3 gene in both Kenyan and Ethiopian athletes while Scott et al. [7] and Ash et al. [8] investigated the ACE gene in Kenyan and Ethiopian athletes, respectively. In those three studies, the frequency of a favorable allele discovered in Eurasian runners was compared between endurance athletes and controls from their respective countries. None of them found a statistically significant difference between athletes and controls [6–8]. Similarly, differences in mitochondrial haplogroup composition have been investigated in both Ethiopian and Kenyan endurance runners under the same case-control design. While no differences have been found in Ethiopia [9], Kenyan athletes exhibited a higher frequency of haplogroup L0 and a lower frequency of haplogroup L3 [10], but this result could be due to population stratification if cases and controls do not come from the same ethnicities, for example. More recently, cohorts of athletes and controls from Kenya and Ethiopia have been used to validate 45 markers that were preliminarily associated with endurance running in a Genome-wide Association Study (GWAS) [11]. Six markers showed differences between Kenyan athletes and controls, while three different markers showed differences between Ethiopian athletes and controls. The authors conclude, also based on the results of other cohorts, that there is no evidence for a common genetic profile specific to endurance athletes [11].

Tucker et al. [4] criticized the candidate gene and case-control approaches previously applied to these populations, especially when studying highly complex traits like endurance performance. They argued that a case-control design will fail if both groups share a similar genetic favorable background but differ in training history. Instead, they proposed that the over-representation of Kalenjin and Oromo in long-distance running events should be more easily understood as a populational phenomenon. That is, these populations would have a higher frequency of many favorable alleles, which would allow the appearance of individuals with multiple favorable genotypes through the genome much more frequently than in other...
populations. When exposed to the right environmental conditions (including training), present in those countries, these individuals would then be able to reach the elite endurance runner status, accounting for the overrepresentation of East Africans among those athletes.

In the present study, we followed this reasoning by comparing genomic populational data among different ethnicities in East Africa. We hypothesize that the alleles that predispose to endurance running should be common in the general population of these ethnic groups and, consequently, that genetic factors associated with athletic success in Kalenjin and Oromo should be, at least partially, close to the genomic regions of greater differentiation in these ethnicities, enabling the identification of molecular processes that contribute to long-distance running (Fig 2). This approach has the great advantage of not relying on sampling athletes. By treating a phenotype (in this study, endurance running capacity) as a populational characteristic we can use data from public genomic datasets of genomic variation to test our hypothesis, similar to studies about the genetic basis of adaptation to high-altitude environments, for example [12,13].

Results

Using the normalized Population Branch Statistic (PBSn1) [14], we identified, for Kalenjin and Oromo, respectively, 297 and 352 genes linked to highly differentiated genomic regions (Fig 3; S1 Dataset). We submitted these lists to an enrichment analysis [15] to recover phenotypes or traits showing enriched gene-sets. After Benjamini-Hochberg false discovery rate (FDR) correction for multiple comparison, this resulted in 47 and 97 trait-related gene-sets enriched for Kalenjin and Oromo, respectively (S2 Dataset). When compared to a list of 628
endurance-relevant gene-sets (S3 Dataset) selected a priori from the GWAS catalog database [16], we found 10 matches for Kalenjin (Table 1) and 27 for Oromo (Table 2) (one-tailed Fisher’s exact test $P = 0.048$, odds ratio (OR) = 2.003, and $P = 1.6 \times 10^{-5}$, OR = 2.918, respectively; see S1 Table for a comparison of traits between these populations). When the closely related populations were considered (see Material and Methods for details), the Luhya, from Kenya, showed no significant enrichment ($P = 0.341$; OR = 1.840; 2 traits), while the Amhara, from Ethiopia, were enriched for endurance-relevant traits ($P = 0.011$; OR = 3.122; 8 traits). However, in both cases, these populations had, respectively, fewer endurance-relevant traits than its national counterparts (one-tailed Fisher’s exact test: Luhya vs. Kalenjin, $P = 0.019$, OR = 5.059; Amhara vs. Oromo, $P = 8.1 \times 10^{-4}$, OR = 3.479).

Thus, genes close to highly differentiated genomic regions are enriched for endurance-relevant traits in Kalenjin and Oromo, and both have more endurance-relevant gene-sets than its national pairs. The 10 endurance-relevant gene-sets for Kalenjin and the 27 for Oromo comprised 49 and 96 candidate genes (S2 Dataset). For each population comparison (see Material and Methods for details), we also selected the genes in the five most differentiated genomic regions, highlighting seven candidate genes for Kalenjin: TRAM2-AS1, WWOX, ARHGEF1, GATAD1, AMOTL1, GSTO1, and MRAP2; and eight for Oromo: LTBP1, COL5A2, MIR6797, ADAMTS9-AS2, EXOC4, EXOC5, EXOC6B, and LIPE-AS1 whose biological function is related to the endurance-relevant gene-sets. These functions affect anthropometric/biomechanical traits (LTBP1, TRAM2-AS1, COL5A2, WWOX and MIR6797), lung function (WWOX), blood pressure and the circulatory system (ADAMTS9-AS2, ARHGEF1, GATAD1, AMOTL1, and GSTO1), glucose metabolism (EXOC4, EXOC5, EXOC6B and WWOX), fatty acids metabolism (MRAP2 and LIPE-AS1) and calcium homeostasis (WWOX, MIR6797, MRAP2, GSTO1).

**Discussion**

In this study, we used publicly available genomic data to tackle the hypothesis that genetic variation could be associated with endurance running in East African populations [4], but there
are limitations. Our approach identified highly differentiated genomic regions irrespective of whether they are relevant to endurance running or not. Thus, we had to restrict our analysis to traits whose relationship with endurance running is well documented. Given the widespread pleiotropy affecting complex phenotypes, meaningful biological associations may have gone unnoticed. On the other hand, further studies are also necessary to corroborate and clarify the specific association between these candidate genes and endurance running in different human populations. While many of the biological associations discussed below may look speculative, they serve as a starting point for investigating the possible contribution of these genes to endurance running.

There was an abundance of anthropometric-related traits, such as height, waist circumference, and waist-to-hip ratio, among the enriched sets for both Kalenjin and Oromo. Ekstervicz et al. [17] found that Marathon finishing-time was positively correlated with upper limb length, torso length, hip width, and waist-hip ratio in a sample of Kenyan endurance runners. More generally, anthropometric features affect stride length, movement stability, air and ground resistance [1,17,18]. LTBP1, one of the most differentiated genes for Oromo, participates in the molecular pathway associated with those phenotypes. LTBP1 binds to TGF-β1,
facilitating its export to the extracellular matrix of bone cells, where it plays a key role in chondrocyte maturation, mineralization, and bone remodeling [19,20].

The structure of bones and tendons also plays an important role in physical activity, including long-distance running. Less flexibility and greater stiffness are observed in the lower limb’s tendons of long-distance runners [21,22], and lower overall body flexibility is associated with running economy by increasing body stability and enhancing the use of elastic energy [23]. Two genes highly differentiated in Kalenjin and Oromo: TRAM2-AS1 and COL5A2, respectively, are associated with the synthesis and organization of type I collagen, the main protein in ligaments and tendons [24]. TRAM2-AS1 encodes an antisense RNA against TRAM2, which participates in type I collagen’s biosynthesis [25]. Given the importance of antisense RNAs in gene expression regulation [26], TRAM2-AS1 may modulate the synthesis of type I collagen. Likewise, COL5A2 encodes, together with COL5A1 and COL5A3, one of the three alpha chains that form type V collagen, which on its turn regulates the assembly and structure of type I collagen fibrils [27,28]. Mutations in COL5A2 or COL5A1 account for over 90% of the cases of the classic Ehlers-Danlos syndrome, characteristic for joint hypermobility [29]. Two polymorphisms in COL5A1 (rs12722, C/T, and rs71746744, -/AGGG) have been directly associated with performance in long-distance running, with individuals carrying the T/T or AAGG/AGGG genotypes being considerably faster and less flexible [30,31]. Variants in COL5A2 may have a similar effect to COL5A1, considering that these genes encode different subunits of the same protein and that mutations in both have the same clinical outcome [29].

Long-distance running has also been associated with bone mineral density, which is increased in the legs and reduced in the vertebrae of runners compared to controls [32–34]. We found enriched gene-sets related to bone mineral density for both populations. In addition to LTBP1, previously discussed, other highly differentiated genes that could affect this trait are WWOX, shared by both populations, MIR6797 in Oromo, and TRAM2-AS1 in Kalenjin. The products encoded by these genes interact with the transcription factor RUNX2, the major regulator of bone development [35]. The first two act as negative regulators of RUNX2 expression,
controlling osteoblast differentiation [36,37]. Indeed, WWOX deficient mice show a considerable delay in skeletal development and mineralization, including reduced blood calcium levels [36]. RUNX2 also regulates the expression of TRAM2 in bone cells, affecting the availability of type I collagen in the bone extracellular matrix [38].

Together with anthropometric traits, physiological processes are central to endurance running [39]. In general, good endurance runners will have a high capacity to consume oxygen (high maximum oxygen uptake—VO$_2$max, high capacity to mobilize oxygen and energy), but will consume less oxygen to run at intermediate speeds (good running economy, low energy requirement) [23]. The efficiency of the aerobic metabolism required for maintaining high intensity physical activity depends on several distinct physiological steps: absorbing oxygen from the air, transporting it through the bloodstream to the skeletal muscle, mobilizing energy

Table 2. Enriched gene-sets, in Oromo, for “endurance relevant” traits in GWAS Catalog (P<0.05 after FDR correction). The number of genes in Oromo’s list is shown for each trait (Genes in set). Gene’s names can be found in S2 Dataset.

| GWAS catalog gene-set                                      | Oromo          |
|-----------------------------------------------------------|---------------|
|                                                           | Genes in set  | Genes in set (total) | P-value |
| Diastolic blood pressure                                   | 22            | 648                   | <0.001  |
| Hand grip strength                                         | 9             | 156                   | <0.001  |
| Pulse pressure                                             | 18            | 690                   | <0.001  |
| Systolic blood pressure                                    | 19            | 783                   | 0.001   |
| Body mass index                                            | 24            | 1197                  | 0.001   |
| Waist circumference                                        | 6             | 73                    | 0.002   |
| Height                                                     | 14            | 529                   | 0.003   |
| Lung function (FVC)                                        | 8             | 178                   | 0.004   |
| Waist-to-hip ratio adjusted for BMI                        | 11            | 350                   | 0.004   |
| Lung function (FEV1)                                       | 5             | 65                    | 0.006   |
| Peak expiratory flow                                       | 6             | 115                   | 0.009   |
| Waist circumference adjusted for BMI in active individuals | 5             | 83                    | 0.014   |
| Body fat distribution (arm fat ratio)                      | 6             | 129                   | 0.014   |
| Atrial fibrillation                                        | 8             | 238                   | 0.014   |
| Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction) | 5             | 85                    | 0.014   |
| Lung function (FEV1/FVC)                                   | 7             | 184                   | 0.015   |
| Bone mineral density (hip)                                 | 4             | 55                    | 0.021   |
| Hypertension                                               | 5             | 100                   | 0.023   |
| Hip circumference adjusted for BMI                         | 5             | 101                   | 0.024   |
| Waist-hip ratio                                            | 4             | 60                    | 0.025   |
| Lean body mass                                             | 2             | 7                     | 0.028   |
| Heel bone mineral density                                  | 15            | 834                   | 0.032   |
| Vigorous physical activity                                 | 2             | 8                     | 0.032   |
| Waist-to-hip ratio adjusted for body mass index            | 5             | 114                   | 0.034   |
| FEV1                                                       | 6             | 175                   | 0.036   |
| Hip circumference                                          | 4             | 73                    | 0.039   |
| Waist circumference adjusted for body mass index           | 5             | 126                   | 0.044   |

FEV1 = Forced expiratory volume in 1 second.
FVC = Forced vital capacity.
BMI = Body mass index.

https://doi.org/10.1371/journal.pone.0265625.t002
reserves, and performing muscle contraction [23]. Lung function, representing the first step of this pathway has been correlated with performance in endurance running trials [40,41], and was improved in rats artificially selected for endurance running performance [42], as expected from the considerably high heritability shown by VO\textsubscript{2}max ($h^2 \geq 0.56$) [43]. We found enriched gene-sets for lung function in both Kalenjin and Oromo. WWOX, present in the gene lists of both populations, has been associated with lung function in GWAS [44,45], case-control, and family-based studies [46], even though the molecular mechanism is not fully understood.

The next step, oxygen transport, is highly dependent on blood pressure (BP), on the amount of available hemoglobin, and tissue vascularization. Diastolic BP has been identified as a predictor of endurance performance [47,48], and endurance runners frequently have left ventricular hypertrophy [49]. Oromo shows enriched gene-sets related to BP and cardiac conduction. One of its highly differentiated genes, ADAMTS9-AS2, encodes the antisense RNA of the ADAMTS9, which codes for a metalloproteinase essential for the normal development and homeostasis of the heart and arteries [50]. Unlike Oromo, Kalenjin does not have enriched gene-sets associated with BP. However, four of its highly differentiated genes (ARHGFE1, GATAD1, AMOTL1, and GSTO1) participate in biological processes related to BP. ARHGFE1 is activated by angiotensin II in arterial smooth muscle cells, leading to increased BP [51]. ACE codes for the enzyme that produces angiotensin II, and has been associated with endurance running success in other studies [52], though not in Ethiopian runners [8]. GATAD1 and AMOTL1 seem to be more important in cardiac muscle development, but may affect BP indirectly. GATAD1 is expressed in ventricular myocytes, and mutations in this gene cause dilated cardiomyopathy, a disease characterized by excessive enlargement of cardiac ventricles [53]. AMOTL1 has been associated with the enlargement and proliferation of cardiomyocytes, cardiac hypertrophy [54], and angiogenesis [55,56]. Finally, GSTO1 downregulates the activity of the cardiac ryanodine channel RYR2, which releases calcium from the sarcoplasmic reticulum into the cytoplasm to perform muscle contraction [57]. Interestingly, RYR2 is in the Oromo gene list—in the gene sets associated with lung function (S2 Dataset), and has been associated with increased VO\textsubscript{2}max trainability in Europeans [58].

Most metabolic energy for endurance running comes from aerobic glycolysis in the mitochondria [59]. Both Kalenjin and Oromo have enriched gene-sets associated with glucose metabolism: “Glucose homeostasis traits” and “Vigorous physical activity”, respectively. In the Oromo, three genes encoding proteins of the exocyst complex were linked to high differentiation regions (EXOC4, EXOC5, and EXOC6B). The exocyst is essential in the insulin-induced transport to the cell membrane of the main glucose transporter (GLUT4) in the muscle (skeletal and cardiac) and adipose tissue [60,61]. The knockout of either of those genes reduces the entry of glucose into adipocytes and skeletal muscle cells considerably [60–62]. Within the cell, WWOX binds to HIF1α and modulates aerobic glycolysis by inhibiting the activation of genes that induce aerobic metabolism. WWOX deficient cells show an increase in HIF1α levels and activity, as well as an increase in glucose uptake [63]. HIF1A, the gene encoding HIF1α, has already been directly associated with elite long-distance runner status [64].

On the other hand, the use of fatty acids in energy metabolism is linked to the ability to maintain physical activity for a longer period while preserving the systemic glucose levels [65]. Again, both populations had lipid metabolism enriched gene-sets. In the Oromo list, LIPE-AS1 encodes the antisense RNA for LIPE, which codes for the main regulator of lipolysis in adipocytes, responsible for releasing fatty acids from stored triglycerides. Mutations in this gene are directly related to several metabolic diseases [66]. In the Kalenjin’s highly differentiated genes, MRAP2 codes for an accessory protein that modulates the activity of melanocortin receptors, in particular MC4R [67,68]. MC4R regulates physiological processes related to energy
metabolism, having been associated with body mass index in African and European populations [69]. Mutations in MRAP2 and MC4R cause obesity in both humans and mice [68,70,71]. Mc4r knocked-out (MC4RKO) mice are obese, but have a specific metabolic profile, with lower heart rate, lower lean body mass, lower muscle strength, lower bone density, and lower performance in endurance running [71].

Finally, muscle contraction in both skeletal and cardiac muscles depends on the calcium ion release from the sarcoplasmic reticulum into the cytoplasm [72]. Both Kalenjin and Oromo had many calcium-related enriched gene-sets (S4 Dataset), which were not observed for other ions like K+, Na+, Cl− or Mg2+. Many genes discussed previously, like WWOX, MIR6797, MRAP2, GSTO1, and RYR2 affect calcium homeostasis. Considering the system GSTO1/RYR2, high intrinsic aerobic exercise capacity in mice was associated with greater contraction amplitude in cardiomyocytes, an increased peak of calcium release, and increased expression of RYR2 [73]. Besides, endurance training in rats induced higher contractibility and Ca2+ sensibility in cardiac muscle cells [74].

The fact that most genes discussed here have not been previously associated with endurance running may be due to differences in study design. Indeed, an important assumption of our study is that interpopulation comparisons may be more informative to explain the overrepresentation of East African ethnicities among elite runners [4]. On the other hand, former studies have relied on intrapopulation comparisons [11,75]. While case-control studies may help to validate or refute our findings, designing a proper case-control study is challenging [4].

Also, despite the overwhelming genetic diversity in Africa [76], most studies about genetics and athletic performance were either performed only in Eurasian populations [11,75], or used African populations to test for genetic associations originally discovered in Eurasians [6–8,75]. Remarkably, for many genes discussed here, a direct molecular interaction exists with genes previously associated with endurance running (e.g. COL5A2 and COL5A1, ARHGEF1, and ACE, and WWOX and HIF1A). This indicates that, while the same pathways are important for endurance running in different ethnicities, different genes may be more important in a population-specific context. RYR2 and its regulator, GSTO1, being highly differentiated in Oromo and Kalenjin, respectively, further highlights this point. The overrepresentation of Eurasian populations in GWAS studies [77] may also explain, at least in part, why we found gene enrichment more often in Oromo, than in Kalenjin, both overall and for endurance relevant traits, given the much higher Eurasian ancestry in the former. Our results also reinforce the idea of endurance running as a complex and systemic activity [78]. Several genes (such as WWOX, MRAP2, TRAM-AS1, and GSTO1) seem to affect more than one trait, while all traits seem to be influenced by multiple genes, in a complex many-to-many relationship [79] that is also dependent on environmental and developmental processes. This view strengthens the criticism towards direct-to-consumer genetic tests to inform “genetic predisposition” for endurance sports, especially when used as a tool for prospecting children and young athletes for specific modalities [80].

Because we rely on a general measurement of population differentiation, our approach is also unable to test if endurance running evolved as an adaptation or should be seen as a byproduct of neutral demographic processes. Another possibility is that adaptation to high-altitude favors some East African populations in endurance events, even though the direct relationship between being born in high-altitude and increased endurance performance is controversial [81]. We found only one gene, RYR2, in the Oromo list, that has been previously associated with high-altitude adaptation in studies involving East African (Amhara) populations [82], and another, LIPE-AS1, whose target, LIPE, was shown to affect survival rates in Drosophila exposed to low oxygen conditions [83]. However, while the Amhara are usually considered as a “model population” for studying high-altitude adaptation, they are not...
overrepresented among Ethiopian elite endurance athletes [2]. Finally, there were no enriched GWAS traits associated with high-altitude adaptation or hypoxia in neither Kalenjin nor Oromo. Taken together, it seems unlikely that high-altitude adaptation in East Africans is the major driver of endurance running success in these populations.

Finally, we would like to emphasize that genetic predisposition does not mean predestination, and success in sports should not be taken as a racial (or regional) stereotype (even if a putatively “positive” one). First, we must be very careful to avoid reinforcing the horrific racist ideas from the late 19th century that, among other things, antagonized athleticism and intellectual ability [84]. A recent discussion in the US about racial stereotyping of black quarterbacks in American Football, for example, revealed that black athletes have been perceived as more “physical” and less “mental” than their white peers [85,86]. These associations are not only scientifically incorrect, but also ethically unacceptable. Second, as we have just emphasized, the genetics of complex traits is far from deterministic. Even though we restricted our analysis to genetic factors that may influence long-distance running, environmental, socio-cultural, and motivational factors must never be ignored [87,88]. Even if we understand that some populations have a higher frequency of alleles predisposing it to a specific phenotype, such as “long-distance running performance”, assuming that individuals from these populations adhere to the phenotype is an example of the ecological fallacy. Obviously, most Kalenjin and Oromo are not elite long-distance runners, and may never become one. Conversely, individuals from different ethnic backgrounds may become elite runners, such as the Olympic gold medalist Miruts Yifter, an Ethiopian long-distance athlete who was not from Oromo ethnicity, among a myriad of other examples. More than exploring the bases of long-distance running, this study illustrates the beauty of human genetic diversity and some of its fascinating physiological potentials.

**Materials and methods**

**Basic assumptions and design**

Our study assumes a “populational effect” for the dominance of some East African ethnicities in endurance running events [4]. That is, we expect that some of the more differentiated loci for Kalenjin and Oromo may be relevant to their endurance running success. As pointed out in the introduction, because we hypothesize that the alleles that predispose to endurance running should be common in the general population of these ethnic groups, there is no need to have athletes (or only athletes) among sampled individuals. This allows us to use data from public genomic datasets of genomic variation (see below). Again, our study is not the first to consider a phenotype as a populational characteristic. For example, studies about the genetic basis of altitude adaptation in humans often adopt this strategy, looking for peaks of genomic differentiation between closely related populations in high-altitude vs. lowland, regardless of possible intrapopulation individual differences in the response to high-altitude [12,13].

For estimating the peaks of genomic differentiation, we used the Population Branch Statistic (PBS) [12], in its standardized version, PBSn1, as described by Malaspina et al. [14]. This statistic, which is based on the classical $F_{ST}$ statistic, uses allele frequency data to estimate the degree of genetic differentiation specific to a population of interest, or focal population, in relation to two reference populations, generally one closely related to the focal population and another more distantly related [12,13]. It is important to note that, although widely used in adaptation studies, PBSn1 can be used to measure genetic differentiation regardless of the evolutionary processes causing differentiation (natural selection, genetic drift, admixture, etc.). This occurs because $F_{ST}$, which is the basis of PBS (and PBSn1) calculation, is affected by several evolutionary processes [89]. When neutral evolutionary processes can be accounted for,
PBS (and PBSn1) then becomes a measurement of genetic differentiation caused by adaptive processes (thus indicating natural selection) [90]. In this study, we do not make any assumptions about the evolutionary processes predisposing the Kalenjin and Oromo ethnic groups to endurance running dominance in sports events. In other words, we do not control for demographic history of neither population because we assume that both neutral and adaptive processes may have affected highly differentiated genomic regions associated with traits affecting endurance running.

Similarly, when the studied populations are admixed for different ancestries, differences in allele frequencies between ancestry components may affect genetic differentiation statistics. This must be corrected if the aim is to detect signals of natural selection [82,91]. However, when neutral processes are concerned, genomic admixture is a genuine process of population differentiation. In this study, we used different sets of closely and distantly related populations to account for admixture in the focal populations (see below for further details).

**Genomic datasets and population comparisons**

We obtained curated genotyped SNP data made available by the African Genome Variation Project (AGVP), for 1,152,000 single nucleotide polymorphisms (SNPs) across the genome of 1,481 individuals from 18 ethnolinguistic groups from Sub-Saharan Africa using the Illumina Omni2.5 chip [76]. The data was subjected to the same quality control procedures employed by Gurdasani et al. [76], and only genotyped SNPs were considered (i.e. no imputation was performed). To access the diversity of other African and Eurasian populations used in the calculation of PBSn1, we also included equivalent data obtained from the 1000 Genomes Project database [92]. All data used in this study came from databases that were assembled respecting the ethical considerations elaborated by relevant research committees, both nationally (for the countries involved) and internationally.

Following demographic studies for Kenyan and Ethiopian athletes [2,3], we selected the Kalenjin (K, n = 100) and the Oromo (O, n = 26) as the focal populations in this study. These populations have differing levels of Eurasian ancestry (7.3–10.99% for Kalenjin, 43.62–50.82% for Oromo), and considerable diversity in their sub-Saharan African ancestry. Kalenjin has a greater contribution of Nilotic, followed by Horn of Africa and Southeast Bantu ancestries, while Oromo has a greater contribution of Horn of Africa ancestry [76]. We used two populations as “closely related”: the Luhy (L, n = 74, from Kenya) and the Amhara (A, n = 42, from Ethiopia), which were selected by being in the same countries as the focal populations (and theoretically, under the same socioeconomic policies), as well as for representing Bantu and Horn of Africa ancestries, respectively. For the distantly related population we built a metapopulation from a set of seven ethnic groups (Wolof, Mandinka, Jola, Fula, Ga-Adangbe, Yoruba, and Igbo, n = 618) representing Central and Western Africa (WA), in opposition to East Africa (Fig 1B). We also built a second metapopulation to be used as distantly related by combining six populations to represent Eurasia (EUR) from the 1000 Genomes Project (Individuals residing in the United States with European ancestry (CEU), Tuscan, Finnish, British, Iberian, and Individuals residing in the United States with Indian ancestry (GIH), n = 569). We used a Eurasian contrast as distantly related because the exclusive use of an African contrast could over-represent loci with high Eurasian ancestry in the focal populations due to Eurasian admixture. All procedures involving allele frequency calculation or population merging were performed in Plink 1.9 [93].

For the PBSn1 calculations, we used a focal population, a closely related and a distantly related population for all possible combinations, except for the two combinations that would include the Amhara and Eurasia. Those combinations were excluded because the Amhara
have the highest level of Eurasian ancestry (47.82–54.70%) [76], and spurious results can emerge in PBSn1 when the two reference populations are closer to each other than to the focal population [14]. On the other hand, thanks to this level of mixing, the Amhara themselves act as good controls for Eurasian ancestry. Overall, we performed six comparisons, three for each focal population (focal population x closely related; distantly related): 1) K x L; WA, 2) K x L; EUR, 3) K x A; WA, 4) O x L; WA, 5) O x L; EUR, 6) O x A; WA. We estimated the PBSn1 values using in-house scripts in R [94]. We used sliding windows of 20 SNPs moving every five SNPs (thus, with overlapping of 15 SNPs), ensuring a homogeneous coverage of the genome and reducing the effect of individual SNPs. This resulted in a total 230,396 windows across the genome with a median size of 36,289 bp. We calculated the PBSn1 value for each SNP and then generated the average value of PBSn1 for the whole window. To check if the strategy of having multiple PBSn1 comparisons would effectively result in distinct ancestry components in different comparisons, we performed a discriminant analysis of principal components (DAPC) [95] for each of the six populations comparisons. This analysis was performed in the “adegenet” package [96] considering the populations as a priori clusters. We filtered the dataset for linkage disequilibrium and retained, based on principal component (PC) loadings, the first 400 PCs for the estimation of the discriminant function (S1 Fig).

When there is an adaptive hypothesis about the phenotype of interest, demographic simulations can be used to determine an empirical threshold above which the PBSn1 values are indicative of an adaptive process [13]. However, as previously stated, in this study we avoided any premise about evolutionary processes, assuming that purely demographic factors, such as drift and gene flow, may cause high PBSn1 values in loci associated with endurance running. In this case, it is impossible to establish an empirical significance value for PBSn1 from demographic simulations. Thus, following similar studies [13,97], we retained, from each comparison, the 0.1% windows with the highest PBSn1 values for further analysis.

**Annotation and statistical analysis**

We used the Ensembl platform [98] to annotate the genes in the 5 Kb neighborhood of the SNP with the highest PBSn1 of each window in the top 0.1% PBSn1 values. We chose a short window size compared to other studies [82,99] to avoid including genes distantly linked to the highest peaks of genetic differentiation in the final gene list, as it could affect the enrichment analysis of gene-sets and endurance-relevant traits. Merging the lists from the three comparisons, we obtained a final list of highly differentiated genes for each focal population (S1 Dataset). We performed a gene-set analysis in the FUMA-GWAS platform [15], using the GENE2FUNC process, to determine which traits [16] and biological processes [100] were enriched for genes present in the original list (S2 and S4 Datasets) considering a \( P < 0.05 \) after Benjamini-Hochberg False Discovery Ratio (FDR) correction. FUMA-GWAS performs the analysis using hypergeometric tests [15] to test for the enrichment of genes on our lists in gene-set from databases (GWAS-catalog [16] for traits and GO: Biological Process in MSigDB [100] for biological processes). The FDR correction is performed per data source of tested gene sets, using the number of gene sets in that data source (5,246 for GWAS-catalog and 7,658 for GO: Biological Process).

Our study design does not allow the specific detection of genes associated with endurance running success. However, we expect that genes associated with endurance running success will be among highly differentiated genomic regions in general. Conversely, many gene-sets with no biological relation with endurance running will show enrichment in the analysis.

As a “proof-of-concept”, we also reasoned that our approach should be able to detect genetic signals of other traits for which differences among populations are known. Ethiopian
populations have lighter skin pigmentation compared to Kenyan populations. Therefore, the comparison (O × L; WA) should be enriched for gene-sets associated with skin pigmentation. Indeed, we found significant enrichment ($P = 0.019$) for “Skin pigmentation traits” (genes BNC2 and FANCA). This example also illustrates the benefits of not controlling for admixture in some settings, since it is likely this difference would not emerge if only the sub-Saharan ancestry in the Oromo and Luhya were considered, given that it is probably associated with the large Eurasian ancestry in Ethiopian populations.

Following this, we tested if gene-sets associated with endurance-relevant traits would be overrepresented among enriched gene-sets. First, we selected from the full list of 5,246 traits from the GWAS Catalog database [16] all traits that could be relevant in endurance running. Four of the authors (A.L.S.Z., R.Q.C., R.L.M., and N.J.R.F) made the selection of these traits independently, based on a general literature for sports physiology (e.g. [101,102]). Traits selected at least twice composed the final list of endurance relevant traits (S3 Dataset). We compared the list of 628 endurance-relevant traits with the list of GWAS traits obtained from enriched gene-sets to test, for each population, if the number of matches was higher than expected by chance using Fisher’s exact test in R [94]. To further test the robustness of our results, we inverted the focal population and used the same comparisons and procedures to test if Luhya and Amhara were enriched for endurance-relevant traits. We note, however, that hypothesis does not require that these populations were not enriched for these traits, even though we would expect that these populations must either 1) show no significant enrichment, or 2) that the number of traits is lower compared to its national pair. Finally, the molecular functions of the five genes associated with the highest PBSn1 values from each comparison were also investigated for plausible biological associations with traits affecting endurance running.

**Sensitivity analyses**

We also performed some sensitivity analyses to test the robustness of our results to parameter changes. We varied window size (original value, 20 SNPs with a 15 SNPs overlap; alternative value: 30 SNPs with a 20 SNPs overlap), the percentage of top-PBSn1 windows retained (original value 0.01%; alternative value 0.05%), and the reference list for endurance-relevant traits (original criteria: a trait had to be flagged by at least two of the four authors (2A) that checked the list of GWAS traits; alternative criteria: a trait had to be flagged by at least three authors (3A)). This alternative list of endurance-relevant traits is shown as S5 Dataset. We explored all combinations of parameters, and repeated the enrichment analysis of endurance-relevant traits for both Oromo and Kalenjin. The results of the sensitivity analyses can be found in S6 Dataset. Overall, the parameter changes had little impact on the analysis results. For Oromo, all parameter combinations resulted in statistically significant enrichment, with ORs varying from 1.67 ($P = 3.0 \times 10^{-4}$) to 3.09 ($P = 0.008$). For Kalenjin, six parameter combinations resulted in statistically significant enrichment, with ORs varying from 1.47 ($P = 0.012$) to 2.70 ($P = 0.034$). However, the two combinations including 30 SNPs windows and 0.1% of windows retained resulted in non-significant values (2A, OR = 1.84, $P = 0.211$; 3A, OR = 2.03 $P = 0.280$), possibly because the number of total sets (and hence the number of endurance-related sets) was too small, reducing the statistical power of Fisher’s exact test. All endurance-related gene-sets detected in the sensitivity analyses are shown in S7 Dataset.

**Supporting information**

S1 Fig. Discriminant analysis of principal components (DAPC) for the six population comparisons performed.

(DOCX)
S1 Table. Enriched gene-sets for “endurance-relevant” traits in GWAS Catalog (P<0.05 after FDR correction) for both Kalenjin and Oromo.

(S1 Text. Further information about the studied populations.

(S1 Dataset. List of the genes associated with the 0.1% most differentiated windows, by comparison, for Kalenjin and Oromo.

(S2 Dataset. List of all enriched gene-sets (including those not endurance-relevant), in GWAS Catalog, considering all retained genes for Kalenjin and Oromo.

(S3 Dataset. List of all gene-sets in GWAS Catalog selected as “endurance-relevant”, according with the literature, by at least 2 authors (see Material and Methods for details).

(S4 Dataset. List of all enriched gene-sets (including those not endurance-relevant), in GO biological processes, considering all retained genes for Kalenjin and Oromo.

(S5 Dataset. List of all gene-sets in GWAS Catalog selected as “endurance-relevant”, according with the literature, by at least 3 authors (see Material and Methods for details).

(S6 Dataset. Results for the sensitivity analyses (see Material and Methods for details).

(S7 Dataset. Enriched gene-sets for “endurance-relevant” traits in GWAS Catalog (P<0.05 after FDR correction) for both Kalenjin and Oromo, considering the parameters used in the sensitivity analyses.

Acknowledgments
We are grateful to Juliano A. Boquett, Tâbita Hünemeier and Alice Tagliani-Ribeiro for useful comments in early stages of the study. We would also like to thank the three anonymous reviewers, whose comments helped to greatly improve this work.

Author Contributions
Conceptualization: André L. S. Zani, Nelson J. R. Fagundes.

Data curation: André L. S. Zani, Mateus H. Gouveia, Rodrigo Quevedo, Rodrigo L. Menezes, Charles Rotimi, Nelson J. R. Fagundes.

Formal analysis: André L. S. Zani.

Resources: Mateus H. Gouveia, Marla M. Aquino, Charles Rotimi.

Supervision: Nelson J. R. Fagundes.

Visualization: André L. S. Zani.
References

1. Bramble DM, Lieberman DE. Endurance running and the evolution of Homo. Nature. 2004; 432: 345–352. https://doi.org/10.1038/nature03052 PMID: 15549097

2. Scott RA, Georgiades E, Wilson RH, Goodwin WH, Wolde B, Pitsiladis YP. Demographic characteristics of elite Ethiopian endurance runners. Med Sci Sports Exerc. 2003; 35: 1727–1732. https://doi.org/10.1249/01.MSS.0000093535.85254.89 PMID: 14523311

3. Onywera VO, Scott RA, Boit MK, Pitsiladis YP. Demographic characteristics of elite Kenyan endurance runners. J Sports Sci. 2006; 24: 415–422. https://doi.org/10.1080/02640410500189033 PMID: 16492605

4. Tucker R, Santos-Concejero J, Collins M. The genetic basis for elite running performance. Br J Sports Med. 2013; 47: 545–549. https://doi.org/10.1136/bjsports-2013-092408 PMID: 2366980

5. Wilber RL, Pitsiladis YP. Kenyan and Ethiopian distance runners: what makes them so good?. Int J Sports Physiol Perform. 2012; 7: 92–102. https://doi.org/10.1123/ijssp.7.2.92 PMID: 22634972

6. Yang N, MacArthur DG, Wolde B, Onywera VO, Boit MK, Lau SYM-A, et al. The ACTN3 R577X polymorphism in East and West African athletes. Med Sci Sports Exerc. 2007; 39: 1985–1988. https://doi.org/10.1249/mss.0b013e31814844c9 PMID: 17986906

7. Scott RA, Moran C, Wilson RH, Onywera V, Boit MK, Goodwin WH, et al. No association between Angiotensin Converting Enzyme (ACE) gene variation and endurance athlete status in Kenyans. Comp Biochem Physiol A Mol Integr Physiol. 2005; 141: 169–175. https://doi.org/10.1016/j.cbpa.2005.05.001 PMID: 15950509

8. Ash GI, Scott RA, Deason M, Dawson TA, Wolde B, Bekele Z, et al. No association between ACE gene variation and endurance athlete status in Ethiopians. Comp Biochem Physiol A Mol Integr Physiol. 2011; 43: 590–597. https://doi.org/10.1249/MSS.0b013e3181f07bd6 PMID: 20798657

9. Scott RA, Wilson RH, Goodwin WH, Moran CN, Georgiades E, Wolde B, et al. Mitochondrial DNA lineages of elite Ethiopian athletes. Comp Biochem Physiol B Biochem Mol Biol. 2005; 140: 497–503. https://doi.org/10.1016/j.cbpb.2004.11.014 PMID: 15694598

10. Scott RA, Fuku N, Onywera VO, Boit M, Wilson RH, Tanaka M, et al. Mitochondrial haplogroups associated with elite Kenyan athlete status. Med Sci Sports Exerc. 2009; 41: 123–128. https://doi.org/10.1249/MSS.0b013e31813813a2 PMID: 19092698

11. Rankinen T, Fuku N, Wolfarth B, Wang G, Sarzynski MA, Alexeev DG, et al. No evidence of a common DNA variant profile specific to world class endurance athletes. PLoS One. 2016; 11: e0147330. https://doi.org/10.1371/journal.pone.0147330 PMID: 26824906

12. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010; 329: 75–78. https://doi.org/10.1126/science.1190371 PMID: 20595611

13. Jacovas VC, Couto-Silva CM, Nunes K, Lemes RB, de Oliveira MZ, Salzano FM, et al. Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. Sci Rep. 2018; 8: 12733. https://doi.org/10.1038/s41598-018-31100-6 PMID: 30143708

14. Malaspinas A-S, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, et al. A genomic history of Aboriginal Australia. Nature. 2016; 538: 207–214. https://doi.org/10.1038/nature18299 PMID: 27854914

15. Watanabe K, Taskesen E, van Bochoven A, Posthumus D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017; 8: 1826. https://doi.org/10.1038/s41467-017-01261-5 PMID: 29184056

16. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017; 45: D896–D901. https://doi.org/10.1093/nar/gkw1133 PMID: 2789670

17. Ekslerowicz J, Napierala M, Żukow W. How the Kenyan runner's body structure affects sports results. Hum Mov. 2016;17. https://doi.org/10.1515/humo-2016-0002
18. Fletcher JR, MacIntosh BR. Running economy from a muscle energetics perspective. Front Physiol. 2017; 8: 433. https://doi.org/10.3389/fphys.2017.00433 PMID: 28690549

19. Pedrozo HA, Schwartz Z, Mokeyev T, Ormoy A, Xin-Sheng W, Bonewald LF, et al. Vitamin D3 metabolites regulate LTB1P and latent TGF-β1 expression and latent TGF-β1 incorporation in the extracellular matrix of chondrocytes. J Cell Biochem. 1999; 72: 151–165. PMID: 10025676

20. Tang Y, Wu X, Lei W, Pang L, Wan C, Shi Z, et al. TGF-β1-induced migration of bone mesenchymal stem cells couples bone resorption with formation. Nat Med. 2009; 15: 757–765. https://doi.org/10.1038/nm.1979 PMID: 19584867

21. Kubo K, Tabata T, Ikebukuro T, Igashiki K, Yata H, Tsunoda N. Effects of mechanical properties of muscle and tendon on performance in long distance runners. Eur J Appl Physiol. 2010; 110: 507–514. https://doi.org/10.1007/s00421-010-1528-1 PMID: 20535616

22. Kubo K, Miyazaki D, Shimojou S, Tsunoda N. Relationship between elastic properties of tendon structures and performance in long distance runners. Eur J Appl Physiol. 2015; 115: 1725–1733. https://doi.org/10.1007/s00421-015-3156-2 PMID: 25813019

23. Saunders PU, Pyne DB, Telford RD, Hawley JA. Factors affecting running economy in trained distance runners. Sports Med. 2004; 34: 465–485. https://doi.org/10.2165/00007256-200434070-00005 PMID: 15233599

24. Stepień-Slodkowska M, Ficek K, Eider J, Leonska-Duniec A, Maciejewska-Karłowska A, Sawczuk M, et al. The +1245g/t polymorphisms in the collagen type 1 alpha 1 (col1a1) gene in polish skiers with anterior cruciate ligament injury. Biol Sport. 2013; 30: 57–60. https://doi.org/10.5604/20831862.1029823 PMID: 24744467

25. Stefanovic B, Stefanovic L, Schnabl B, Bataller R, Brenner DA. TRAM2 protein interacts with endoplasmic reticulum Ca2+ pump Serca2b and is necessary for collagen type I synthesis. Mol Cell Biol. 2004; 24: 1758–1768. https://doi.org/10.1128/MCB.24.4.1758-1768.2004 PMID: 14749390

26. Pellechano V, Steinmetz LM. Gene regulation by antisense transcription. Nat Rev Genet. 2013; 14: 880–893. https://doi.org/10.1038/nrg3594 PMID: 24217315

27. Andrikopoulos K, Liu X, Keene DR, Jaenisch R, Ramirez F. Targeted mutation in the col5a2 gene reveals a regulatory role for type V collagen during matrix assembly. Nat Genet. 1995; 9: 31–36. https://doi.org/10.1038/ng0195-31 PMID: 7704020

28. Imamura Y, Scott C, Greenspan S. The pro-α3 (V) collagen chain complete primary structure, expression domains in adult and developing tissues, and comparison to the structures and expression domains of the other types V and XI procollagen chains. J Biol Chem. 2000; 275: 8749–8759. https://doi.org/10.1074/jbc.275.12.8749 PMID: 10722718

29. Malfait F, Francocomano C, Byers P, Belmont J, Berglund B, Black J, et al. The 2017 International classification of the Ehlers-Danlos syndromes. Am J Med Genet C Semin Med Genet. 2017; 175: 8–26. https://doi.org/10.1002/ajmg.c.31552 PMID: 28306229

30. Brown JC, Miller C-J, Posthumus M, Collins M. The COL5A1 gene, ultra-marathon running performance, and range of motion. Int J Sports Physiol Perform. 2011; 6: 485–496. https://doi.org/10.1123/ijssp.6.4.485 PMID: 21934170

31. Abrahams S, Posthumus M, Collins M. A polymorphism in a functional region of the COL5A1 gene: association with ultraendurance-running performance and joint range of motion. Int J Sports Physiol Perform. 2014; 9: 583–590. https://doi.org/10.1123/ijssp.2013-0222 PMID: 24085259

32. Morel J, Combe B, Francisco J, Bernard J. Bone mineral density of 704 amateur sportsmen involved in different physical activities. Osteoporos Int. 2001; 12: 152–157. https://doi.org/10.1007/s001980170148 PMID: 11303716

33. Hind K, Truscott JG, Evans JA. Low lumbar spine bone mineral density in both male and female endurance runners. Bone. 2006; 39: 880–885. https://doi.org/10.1016/j.bone.2006.03.012 PMID: 16682267

34. Tam N, Santos-Concejero J, Tucker R, Lamberts RP, Micklefield LK. Bone health in elite Kenyan runners. J Sports Sci. 2017; 36: 1–6.

35. Schroeder TM, Jensen ED, Westendorf JJ. Runx2: a master organizer of gene transcription in developing and maturing osteoblasts. Birth Defects Res C Embryo Today. 2005; 75: 213–225. https://doi.org/10.1002/bdr.20043 PMID: 16187316

36. Aqilani RI, Hassan MQ, de Bruin A, Hagan JP, Volinia S, Palumbo T, et al. The WWOX tumor suppressor is essential for postnatal survival and normal bone metabolism. J Biol Chem. 2008; 283: 21629–21639. https://doi.org/10.1074/jbc.M800855200 PMID: 18487609

37. Arumugam B, Vishal M, Shreyas S, Malavika D, Rajpriya V, He Z, et al. Parathyroid hormone-stimulation of Runx2 during osteoblast differentiation via the regulation of Inc-SUPT3H-1:16 (RUNX2-AS1:32) and miR-6797-5p. Biochimie. 2019; 158: 43–52. https://doi.org/10.1016/j.bioch.2018.12.006 PMID: 30562548
38. Pregizer S, Barski A, Gersbach CA, García AJ, Frenkel B. Identification of novel Runx2 targets in osteoblasts: cell type-specific BMP-dependent regulation of Tram2. J Cell Biochem. 2007; 102: 1458–1471. https://doi.org/10.1002/jcb.21366 PMID: 17486635

39. Morgan DW, Craib M. Physiological aspects of running economy. Med Sci Sports Exerc. 1992; 24: 456–461. PMID: 1560743

40. Warren GL, Cureton KJ, Sparling PB. Does lung function limit performance in a 24-hour ultramarathon? Respir Physiol. 1989; 78: 253–263. https://doi.org/10.1016/0034-5687(89)90057-1 PMID: 2609032

41. Pringle E, Latin R, Berg K. The Relationship Between 10 KM Running Performance And Pulmonary Function. J Exerc Physiol. 2005; 8: 22–28.

42. Kirkton SD, Howlett RA, Gonzalez NC, Giuliano PG, Britton SL, Koch LG, et al. Continued artificial selection for running endurance in rats is associated with improved lung function. J Appl Physiol. 2009; 106: 1810–1818. https://doi.org/10.1152/japplphysiol.90419.2008 PMID: 19299574

43. Miyamoto-Mikami E, Zempo H, Fuku N, Kikuchi N, Miyachi M, Murakami H. Heritability estimates of endurance-related phenotypes: A systematic review and meta-analysis. Scand J Med Sci Sports. 2018; 28: 834–845. https://doi.org/10.1111/sms.12958 PMID: 28801974

44. Artigas MS, Loth DW, Wain LV, Gharib SA, Obeidat M, Wu D, Chen J, et al. Genome-wide association and selection for running endurance in rats is associated with improved lung function. J Appl Physiol. 2009; 106: 1810–1818. https://doi.org/10.1152/japplphysiol.90419.2008 PMID: 19299574

45. Loth DW, Soler Artigas M, Gharib SA, Wain LV, Francescini N, Koch B, et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. Nat Genet. 2011; 43: 1082–1090. https://doi.org/10.1038/ng.941 PMID: 21946350

46. Karjalainen J, Mäntysaari M, Viitasalo M, Kujala U. Left ventricular mass, geometry, and filling in endurance athletes: association with exercise blood pressure. J Appl Physiol. 1990; 59: 443–449. https://doi.org/10.1007/BF02388626 PMID: 2303049

47. Kern CB, Wessels A, McGarity J, Dixon LJ, Alston E, Argraves WS, et al. Reduced versican cleavage due to Adams9 haploinsufficiency is associated with cardiac and aortic anomalies. Matrix Biol. 2010; 29: 304–316. https://doi.org/10.1016/j.matbio.2010.01.005 PMID: 20096780

48. Gratzte G, Rudnicki R, Urban W, Mayer H, Schloegl A, Skrabal F. Hemodynamic and autonomic changes induced by Ironman: prediction of competition time by blood pressure variability. J Appl Physiol. 2005; 99: 1728–1735. https://doi.org/10.1152/japplphysiol.00487.2005 PMID: 16002770

49. Xie C, Chen X, Qiu F, Zhang L, Wu D, Chen J, et al. The role of WWOX polymorphisms on COPD susceptibility and pulmonary function traits in Chinese: a case-control study and family-based analysis. Sci Rep. 2016; 6.

50. Tanaka K, Takeshima N, Kato T, Niihata S, Ueda K. Critical determinants of endurance performance in middle-aged and elderly endurance runners with heterogeneous training habits. Eur J Appl Physiol Occup Physiol. 1990; 59: 443–449. https://doi.org/10.1007/BF02388626 PMID: 2303049

51. Guilluy C, Brégeon J, Touramaniantz G, Rolli-Derenthal M, Retailleau K, Loufrani L, et al. The Hippo effector Yap1 downstream of Fat4 to restrict heart growth. Nat Commun. 2017; 4: 585–594. https://doi.org/10.1101/NCIRGENETICS.111.961052 PMID: 21965549

52. Ragni CV, Diguet N, Le Garrec J-F, Novotova M, Resende TP, Pop S, et al. Amotl1 mediates sequestration of the Hippo effector Yap1 downstream of Fat4 to restrict heart growth. Nat Commun. 2017; 8: 14582. https://doi.org/10.1038/s41467-017-01458-2 PMID: 28239148

53. Zheng Y, Vertuani S, Nyström S, Audebert S, Meijer I, Tegebratt T, et al. Angiomotin-like protein 1 controls endothelial polarity and junction stability during sprouting angiogenesis. Circ Res. 2009; 105: 260–270. https://doi.org/10.1161/CIRCRESAHA.109.195156 PMID: 19590046

54. Choi K-S, Choi H-J, Lee J-K, Im S, Zhang H, Jeong Y, et al. The endothelial E3 ligase HECW2 promotes endothelial cell junctions by increasing AMOTL1 protein stability via K63-linked ubiquitination. Cell Signal. 2016; 28: 1642–1651. https://doi.org/10.1016/j.cellsig.2016.07.015 PMID: 27498087

55. Pregizer S, Barski A, Gersbach CA, García AJ, Frenkel B. Identification of novel Runx2 targets in osteoblasts: cell type-specific BMP-dependent regulation of Tram2. J Cell Biochem. 2007; 102: 1458–1471. https://doi.org/10.1002/jcb.21366 PMID: 17486635

56. Coelho J, Vertuani S, Nyström S, Audebert S, Meijer I, Tegebratt T, et al. Amotl1 mediates sequestration of the Hippo effector Yap1 downstream of Fat4 to restrict heart growth. Nat Commun. 2017; 8: 14582. https://doi.org/10.1038/s41467-017-01458-2 PMID: 28239148

57. Kern CB, Wessels A, McGarity J, Dixon LJ, Alston E, Argraves WS, et al. Reduced versican cleavage due to Adams9 haploinsufficiency is associated with cardiac and aortic anomalies. Matrix Biol. 2010; 29: 304–316. https://doi.org/10.1016/j.matbio.2010.01.005 PMID: 20096780

58. Guilluy C, Brégeon J, Touramaniantz G, Rolli-Derenthal M, Retailleau K, Loufrani L, et al. The Hippo effector Yap1 downstream of Fat4 to restrict heart growth. Nat Commun. 2017; 4: 585–594. https://doi.org/10.1101/NCIRGENETICS.111.961052 PMID: 21965549

59. Ragni CV, Diguet N, Le Garrec J-F, Novotova M, Resende TP, Pop S, et al. Amotl1 mediates sequestration of the Hippo effector Yap1 downstream of Fat4 to restrict heart growth. Nat Commun. 2017; 8: 14582. https://doi.org/10.1038/s41467-017-01458-2 PMID: 28239148

60. Zheng Y, Vertuani S, Nyström S, Audebert S, Meijer I, Tegebratt T, et al. Angiomotin-like protein 1 controls endothelial polarity and junction stability during sprouting angiogenesis. Circ Res. 2009; 105: 260–270. https://doi.org/10.1161/CIRCRESAHA.109.195156 PMID: 19590046

61. Choi K-S, Choi H-J, Lee J-K, Im S, Zhang H, Jeong Y, et al. The endothelial E3 ligase HECW2 promotes endothelial cell junctions by increasing AMOTL1 protein stability via K63-linked ubiquitination. Cell Signal. 2016; 28: 1642–1651. https://doi.org/10.1016/j.cellsig.2016.07.015 PMID: 27498087

62. Dufilhaut A, Gage P, Curtis S, Chelvanayagam G, Board P. The glutathione transferase structural family includes a nuclear chloride channel and a ryanodine receptor calcium release channel modulator. J Biol Chem. 2001; 276: 3319–3323. https://doi.org/10.1074/jbc.M007874200 PMID: 11035031
78. Joyner MJ. Genetic approaches for sports performance: How far away are we? Sports Med. 2019; 49: 199–204.
79. Koonin EV. The logic of chance: The nature and origin of biological evolution. Upper Saddle River, NJ: Financial Times Prentice Hall; 2011.
80. Webborn N, Williams A, McNamee M, Bouchard C, Pitsiladis Y, Ahmetov I, et al. Direct-to-consumer genetic testing for predicting sports performance and talent identification: Consensus statement. Br J Sports Med. 2015; 49: 1486–1491. https://doi.org/10.1136/bjsports-2015-095343 PMID: 26582191
81. Bönig D. Altitude and Hypoxia Training—A Short Review. Int J Sports Med. 1997; 18: 565–570. https://doi.org/10.1055/s-2007-972682 PMID: 9443586
82. Huerta-Sanchez E, Degiorgio M, Pagani L, Tarekegn A, Ekong R, Antao T, et al. Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. Mol Biol Evol. 2013; 30: 1877–1888. https://doi.org/10.1093/molbev/msr089 PMID: 23666210
83. Udpa N, Ronen R, Zhou D, Liang J, Stobdan T, Appenzeller O, et al. Whole genome sequencing of Ethiopian highlanders reveals conserved hypoxia tolerance genes. Genome Biol. 2014; 15; R36. https://doi.org/10.1186/gb-2014-15-2-r36 PMID: 24555826
84. Dennis RM. Social Darwinism, scientific racism, and the metaphysics of race. J Negro Educ. 1995; 64: 243.
85. Mercurio E, Filak VF. Roughing the passer: The framing of black and white quarterbacks prior to the NFL draft. Howard J Commun. 2010; 21: 56–71.
86. Ferrucci P, Tandoc EC. Race and the deep ball: Applying stereotypes to NFL quarterbacks. Int J Sport Communication. 2017; 10: 41–57.
87. Jarvie G. The promise and possibilities of running in and out of east Africa.” in East African Running: Towards a Cross-Disciplinary Perspective. Pitsiladis Y, Bale J, Sharp C, Noakes T, editors. 2007. pp. 24–39.
88. Jarvie G, Sikes M. Running as a resource of hope? Voices from Eldoret. Rev Afr Polit Econ. 2012; 39: 629–644.
89. Holsinger K, Weir B. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet. 2009; 10: 639–650. https://doi.org/10.1038/nrg2611 PMID: 19687804
90. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev Genet. 2013; 47: 97–120. https://doi.org/10.1146/annurev-genet-111212-133526 PMID: 24274750
91. Vicuña L, Fernández MI, Vial C, Valdebenito P, Chaparro E, Espinoza K, et al. Adaptation to extreme environments in an admixed human population from the Atacama Desert. Genome Biol Evol. 2019; 11: 2468–2479. https://doi.org/10.1093/gbe/evx172 PMID: 31384924
92. The 1000 Genomes Project Consortium, Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, et al. A global reference for human genetic variation. Nature. 2015; 526: 68–74. https://doi.org/10.1038/nature15393 PMID: 26432245
93. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–575. https://doi.org/10.1086/519795 PMID: 17701901
94. R Core Team. R: A language and environment for statistical computing. 2018. Available: https://www.r-project.org/.
95. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010; 11: 94. https://doi.org/10.1186/1471-2156-11-94 PMID: 20950446
96. Jombart T, Ahmed I. adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011; 27: 3070–3071. https://doi.org/10.1093/bioinformatics/btr521 PMID: 21926124
97. Amorim CE, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM, et al. Genetic signature of natural selection in first Americans. Proc Natl Acad Sci U S A. 2017; 114: 2195–2199. https://doi.org/10.1073/pnas.1620541114 PMID: 28193867
98. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018; 46: D754–D761. https://doi.org/10.1093/nar/gkx1098 PMID: 29159590
99. Lindo J, Huerta-Sánchez E, Nakagome S, Rasmussen M, Petzelt B, Mitchell J, et al. A time transect of exomes from a Native American population before and after European contact. Nat Commun. 2016; 7: 13175. https://doi.org/10.1038/ncomms13175 PMID: 27845766
100. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011; 27: 1739–1740. https://doi.org/10.1093/bioinformatics/btr260 PMID: 21546393
101. Kenney WL, Wilmore JH, Costill DL, L D. Physiology of Sport and Exercise. Human Kinetics Publishers; 2015.

102. Powers SK, Howley ET. Exercise physiology: Theory and application to fitness and performance. 10th ed. Columbus, OH: McGraw-Hill Education; 2017.