

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCÊNCIAS

BACHARELADO EM CIÊNCIAS BIOLÓGICAS

ANDRÉ LUÍS DA SILVA ZANI

**A GENÉTICA DA MARATONA: DIFERENCIAÇÃO GENÉTICA EM ETNIAS DO
LESTE DA ÁFRICA E SUA RELAÇÃO COM O SUCESSO NA CORRIDA DE
LONGA DISTÂNCIA**

Porto Alegre

2019

ANDRÉ LUÍS DA SILVA ZANI

**A GENÉTICA DA MARATONA: DIFERENCIAÇÃO GENÉTICA EM ETNIAS DO
LESTE DA ÁFRICA E SUA RELAÇÃO COM O SUCESSO NA CORRIDA DE
LONGA DISTÂNCIA**

Trabalho de Conclusão de curso apresentado como requisito parcial para obtenção do título de Bacharel em Ciências Biológicas com ênfase em Genética na Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Nelson J. R. Fagundes

Porto Alegre

2019

Resumo

Humanos têm uma capacidade incrível de correr longas distâncias. Esta capacidade é levada ao extremo nos esportes, em provas como a maratona, onde atletas correm 42,195 km em pouco mais de duas horas. Desde a década de 1960, atletas do Leste da África, principalmente do Quênia e da Etiópia, têm dominado tais competições. Estudos demográficos, em cada um desses países, mostram que duas etnias são super-representadas entre os corredores de longa distância de elite: os Kalenjin, no Quênia, e os Oromo, na Etiópia. Embora tanto fatores biológicos como culturais devam ter parte nesse sucesso, nenhuma explicação simples foi obtida ao se olhar para a história de vida ou loci classicamente associados à corrida de longa distância. O objetivo deste estudo é, ao investigar as possíveis bases genéticas dessa super-representação, contribuir para um melhor entendimento dos processos moleculares e fisiológicos, que, em última instância, resultam em sucesso na corrida de longa distância. Para tanto, começamos por hipotetizar que múltiplos loci devem predispor indivíduos das etnias Kalenjin e Oromo a serem a super-representados entre atletas de elite de provas de corrida de longa distância, e que esses loci devem ser altamente diferenciados em tais populações para explicar o nível de dominância observado. Dessa forma, procuramos os picos genômicos de diferenciação genética nessas duas etnias, utilizando dados públicos de frequência alélica de 1.152.000 polimorfismos de nucleotídeo único (SNPs, em inglês) para compara-las com outras populações, através da Estatística de Ramo Populacional (PBS, em inglês). Uma vez detectados os picos de diferenciação, submetemos as listas completas genes a eles associados à análise de conjuntos de genes, a fim de verificar o enriquecimento de genes relacionados a processos biológicos e fenótipos que possam estar associados à corrida de longa distância. Além disso, investigamos na literatura a função molecular dos genes mais diferenciados dentro dessas listas obtidas, para averiguar a existência de alguma relação com a corrida de longa distância. Nossos resultados revelam o enriquecimento de genes relacionados a muitas características que afetam o desempenho na corrida de longa distância, como traços antropométricos, função pulmonar, desenvolvimento e homeostase do sistema circulatório, metabolismo energético e homeostase do íon cálcio. Além disso, foram recuperados genes que participam das mesmas vias metabólicas que loci já associados a corrida de longa distância. Esses resultados, além de apontar genes candidatos a serem investigados em futuros estudos, reforçam a noção da corrida de resistência como uma atividade complexa e sistêmica.

Palavras-chave: Corrida de resistência, Genética do Esporte, Quênia, Etiópia.

Nota à Banca: No Trabalho de Conclusão de Curso em Ciências Biológicas, quando há a inclusão de um manuscrito em preparação, o mesmo deve ser precedido de uma introdução geral. O manuscrito em preparação será apresentado no formato de submissão de uma *Report* para a revista *Current Biology*. Foram realizadas algumas alterações em relação ao formato de submissão da editora para facilitar a leitura por parte da Banca.

Introdução Geral

A capacidade de correr longas distâncias (acima de 5km) é uma característica marcante dos seres humanos, sendo única entre primatas, e surpreendente mesmo quando comparada com a de outros mamíferos cursoriais [1, 2]. Isso se deve a uma série de adaptações associadas à dissipação de calor, estabilização do corpo, conservação de energia, capacidade respiratória, entre outras, que, ao que tudo indica, surgiram no gênero *Homo* e estão intimamente ligadas à evolução de nossa espécie [1-3]. Existem duas hipóteses principais, não mutuamente excludentes, para explicar por que tal incrível capacidade pode ter sido selecionada. A primeira delas deriva da provável estratégia de forrageamento carniceira de hominídeos ancestrais, na qual a capacidade de correr, ou pelo menos deslocar-se no ambiente, por longas distâncias permitiria chegar mais rapidamente a carcaças, antes de outros animais, como hienas ou chacais, especialmente nas horas mais quentes do dia, quando esses carnívoros estariam abrigados do calor [2-4]. A segunda hipótese se relaciona à caça de persistência, na qual, em climas quentes, o caçador perseguiria sua presa mantendo-a em movimento por um longo período, até que ela colapsasse devido à hipertermia, permitindo o abate a curta distância, algo fundamental antes da invenção de armas de médio/longo alcance, como arcos e lanças de arremesso [1-4]. Tal técnica já foi registrada inclusive em algumas populações humanas atuais da África, América do Norte e Austrália [4]. Essas estratégias teriam permitido aos primeiros humanos enriquecer sua dieta em proteínas e gorduras, tornando possível a manutenção de um cérebro maior e mais energeticamente custoso [5, 6].

Para o humano moderno, com todo seu aparato tecnológico, parece haver pouca vantagem adaptativa na corrida de resistência. A caça de persistência, por exemplo, embora ainda praticada, é rara mesmo entre tribos de caçadores-coletores, devido ao advento de estratégias energeticamente menos custosas, como a caça com armadilhas, com arco e flecha ou com auxílio de cães, por exemplo [7, 8]. Apesar disso, ainda existe um espaço claramente dedicado a celebrar essa incrível capacidade: os esportes. Nas provas de resistência do atletismo, atletas atingem feitos impressionantes, tais como percorrer 42,195 km em pouco mais de duas horas, no caso de maratonas. Tal percurso é praticado desde as Olimpíadas de 1896 [9], inspirado na história de Fidípides, que, em 490 a.C., teria corrido da cidade de Maratona até Atenas para avisar da vitória dos gregos sobre os persas, morrendo logo em seguida [9, 10]. Desde os anos 1960, tais provas têm sido dominadas por atletas do Leste da África, principalmente do Quênia e da Etiópia [11]. Dos 50 atletas com melhores tempos de maratona

da história, na categoria masculina, 49 são do Quênia ou da Etiópia (20 quenianos e 29 etíopes). Já na categoria feminina, das 50 atletas com melhores tempos, 39 são do Quênia ou da Etiópia (17 quenianas e 22 etíopes), como listado pela Associação Internacional de Federações de Atletismo em 29 de junho 2019. Parece haver não apenas um fator de nacionalidade associado a esse sucesso, mas também um componente étnico. Estudos mostram que entre os corredores de resistência de elite desses países existe uma super-representação das etnias Kalenjin, no Quênia, e Oromo, na Etiópia, que é tanto maior quanto mais longa é a distância da prova considerada [11, 12]. Tal prevalência étnica parece indicar que não apenas fatores ambientais, mas também genéticos estão associados a esse sucesso.

De acordo com último censo nacional do Quênia, de 2009, os Kalenjin representavam cerca de 13% da população, sendo a terceira etnia em número no país, atrás dos Kikuyu (17%) e dos Luhya (14%) [13]. Os Kalenjin diferem de forma crucial dessas outras duas populações, de origem banta, por falarem o idioma Kalenjin, de origem nilótica [11], com as línguas oficiais do Quênia sendo o Inglês e o banto Swahili [14]. Quanto à ancestralidade genética, o componente Sub-Sahariano predominante nos Kalenjin é o Nilótico, embora também possuam uma parcela considerável de ascendência Banta e Cushítica, além de um pequeno percentual de ancestralidade Eurasiática [15, 16]. Já a etnia Oromo, segundo censo de 2007, é a mais numerosa na Etiópia, contando com cerca de 34% da população, seguida pelas etnias Amhara (27%) e Somali (6%) [17]. Os Oromo são falantes da língua de mesmo nome, de origem Cushítica, enquanto o idioma oficial da Etiópia é o Amharico, de origem Semítica [12]. Tanto Oromo quanto Amhara possuem ancestralidade Sub-Sahariana Cushítica, e ambas também apresentam uma grande fração de ancestralidade Eurasiática [15, 16].

Apesar de tanto fatores biológicos como culturais poderem ter um papel em explicar o sucesso dessas etnias na corrida de longa distância, nenhuma explicação simples surgiu ao se olhar para sua história de vida ou a loci classicamente associados a performance atlética [18, 19]. Scott *et al.* [20] investigaram a prevalência do alelo I do gene da enzima conversora de angiotensina (ACE), associado à corrida de resistência em populações Eurasiáticas, nos maratonistas de elite do Quênia, em relação a população geral do país, mas nenhuma diferença estatisticamente significativa foi encontrada. Já Yang *et al.* [21] buscaram o mesmo tipo de associação para outro polimorfismo ligado à corrida de resistência, o R577X do gene da alfa-actinina 3 (ACTN3), desta vez usando amostras de corredores e controles tanto do Quênia quanto da Etiópia, mas também houve diferenças genéticas significativas entre os grupos. Partidários da ideia que apenas fatores ambientais implicam no sucesso na corrida de resistência

poderiam argumentar que esses resultados indicam que o componente genético é menos importante nessa questão. Por exemplo, tanto no Quênia quanto na Etiópia também existe uma super-representação de indivíduos que relataram irem correndo para a escola [11, 12]. Além disso, ambas as populações vivem em altitudes moderadas e têm um forte incentivo de ascensão socioeconômica relacionado ao sucesso no esporte [18].

Entretanto, como bem apontado por Tucker *et al.* [19], dois pontos precisam ser levados em consideração. O primeiro é que nenhum dos fatores ambientais apontados como causas do sucesso na corrida de resistência é exclusivo do Leste da África. Pobreza, altitude e/ou motivação socioeconômica favorável existem em muitos países e grupos populacionais que nunca produziram atletas de elite nessa categoria, muito menos nessa proporção. O segundo ponto é que, considerando a complexidade do fenótipo em questão, é pouco provável que apenas uma variante, ao invés de uma interação entre vários genes e processos fisiológicos, seja capaz de explicar tal diferença no sucesso na corrida. Como argumentado por Tucker *et al.* [19], o sucesso na corrida de longa distância deve ser encarado como um fenômeno populacional, ou seja, as etnias em questão teriam frequências elevadas de diversos alelos, de efeito pequeno, favoráveis à corrida, permitindo o surgimento de muitos indivíduos com genótipos favoráveis em diversos loci ao longo do genoma. Dessa forma, não apenas estudos olhando para uma única variante, mas também estudos intrapopulacionais, comparando corredores (casos) e não corredores (controles) da mesma etnia terão um poder estatístico muito pequeno, pois as frequências alélicas muito provavelmente são semelhantes entre os grupos. Isso é reforçado pelos próprios resultados de Scott *et al.* [20] e Yang *et al.* [21], já que em ambos os casos os autores encontraram o alelo favorável em elevada frequência mesmo entre os controles. Além disso, são comuns os casos de atletas do Leste da África que atingiram a elite pouco tempo depois de começar a treinar. Ou seja, um estudo poderia classificar o mesmo indivíduo como caso ou como controle, com alguns anos de diferença. Assim, é possível que muitos não corredores, tanto Kalenjin como Oromo, caso submetidos a treinamento, atingissem níveis consideráveis de sucesso no esporte. Assim, a maior diferença não seria observada quando comparássemos corredores com não corredores dessas etnias, mas sim quando comparássemos corredores Kalenjin e Oromo com corredores do resto do mundo [19]. Portanto, novas hipóteses sobre o papel de fatores genéticos associados ao sucesso na corrida de resistência poderiam se beneficiar de comparações populacionais em nível genômico que investigassem as diferenças entre essas etnias e as demais.

Considerando a super-representação das etnias Kalenjin e Oromo na elite mundial de corredores de maratona, esse projeto tem por objetivo testar se as regiões genômicas com maior diferenciação genética em uma amostra da população geral de Kalenjin e Oromo - quando comparadas a etnias relacionadas e que não parecem ter tanto sucesso esportivo em provas de longa distância - podem ser associados a processos biológicos ou fenótipos relacionados a corrida de resistência. Em caso afirmativo, isso permitirá propor novos genes ligados a corrida de resistência, assim como processos que podem ser importantes para esse tipo de atividade. A produção de tal conhecimento poderá contribuir de forma significativa para áreas como genética do esporte, fisiologia do exercício, evolução humana e até mesmo para a área médica, no entendimento de doenças relacionadas à mobilidade ou à resposta ao exercício [22].

Hipóteses de Trabalho e Objetivos

Hipóteses de Trabalho

- 1) Existem fatores genéticos que predisõem indivíduos das etnias Kalenjin e Oromo a estarem super-representados no grupo de atletas de elite das modalidades de corrida de longa distância.
- 2) As regiões genômicas mais diferenciadas na comparação das etnias alvo e seus contrastes locais devem conter genes associados ao sucesso dessas etnias na corrida de longa distância.
- 3) Tais genes, em Kalenjin e Oromo, podem participar vias moleculares distintas que acabam afetando o mesmo fenótipo (corrida de longa distância), representando um caso de convergência.

Objetivo Geral

O objetivo geral é contribuir para uma melhor compreensão dos processos moleculares e fisiológicos que podem culminar no sucesso em corrida de longa distância em etnias do leste da África

Objetivos Específicos

- Estabelecer a lista de regiões genômicas mais diferenciadas em Kalenjin, utilizando diferentes contrastes populacionais.
- Estabelecer a lista de regiões genômicas mais diferenciadas em Oromo, utilizando diferentes contrastes populacionais.
- Estabelecer o enriquecimento de fenótipos com a lista de regiões genômicas mais diferenciadas em cada uma das populações alvo.
- Estabelecer o enriquecimento de processos biológicos com a lista de regiões genômicas mais diferenciadas em cada uma das populações alvo.
- Avaliar a relação entre os processos biológicos e fenótipos recuperados e o sucesso na corrida de longa distância.

Manuscrito em Preparação

A Genética da Maratona: Diferenciação Genética em Etnias do Leste da África e sua Relação com o Sucesso na Corrida de Longa Distância

André Luís da Silva Zani^{1,3*}, Mateus Henrique Gouveia^{2*}, Marla Mendes de Aquino^{2*} e Nelson Jurandi Rosa Fagundes^{1*}

¹Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 91501-970, Brasil.

²Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, 31270-901, Brasil.

³Contato Principal

*Contato: zani.andre@yahoo.com.br (A.L.S.Z.), mateushgbio@gmail.com (M.H.G.), marlamendesdeaquino@gmail.com (M.M.A.), nelson.fagundes@ufrgs.br (N.J.R.F.)

Resumo

Atletas de elite competindo em provas de longa distância conquistam feitos impressionantes, tais como percorrer 42,195 km em pouco mais de duas horas, no caso da maratona. Desde a década de 1960, atletas do Leste da África, principalmente do Quênia e da Etiópia, têm dominado tais provas [1]. Estudos demográficos mostram que, em cada um desses países, duas etnias são super-representadas entre os corredores de resistência de elite: os Kalenjin, no Quênia, e os Oromo, na Etiópia [1, 2]. Embora tanto fatores biológicos como culturais certamente estejam relacionados a tal sucesso, nenhuma explicação simples surgiu ao se olhar para a história de vida ou loci classicamente associados a performance atlética [3, 4]. O objetivo desse estudo é contribuir para um melhor entendimento dos processos moleculares e fisiológicos, que, em última instância, resultam em sucesso na corrida de longa distância. Nós começamos por hipotetizar que múltiplos fatores genéticos predisõem indivíduos das etnias Kalenjin e Oromo a serem a super-representados entre atletas de elite de provas de corrida de

longa distância. Dessa forma, identificamos os genes próximos aos picos genômicos de diferenciação genética nessas duas etnias, bem como os processos e fenótipos a eles associados. Então, os comparamos com a literatura, em busca de conexões com a corrida de resistência. Encontramos diversos genes altamente diferenciados relacionados, a características antropométricas, aos sistemas circulatório e respiratório, ao metabolismo energético e a homeostase do íon cálcio. Esses resultados reforçam a ideia da corrida de resistência como uma atividade complexa e sistêmica, bem como apontam genes candidatos para futuros estudos

Palavras-chave: Corrida de resistência, Leste da África, Kalenjin, Oromo, PBS.

Resultados e Discussão

A capacidade de correr longas distâncias, conhecida como corrida de resistência, parece ter desempenhado um importante papel na evolução humana [5-7]. De fato, humanos estão entre os melhores corredores de resistência de todos os mamíferos [6, 7]. Nas provas de longa distância do atletismo, como no caso da maratona, por exemplo, atletas de elite correm uma distância de 42,195 km em pouco mais de duas horas. Desde a década de 1960, atletas do leste da África, principalmente do Quênia e da Etiópia, têm dominado essas competições [1]. Dos 100 atletas com os melhores tempos de maratona na história, como listado pela Associação Internacional de Federações de Atletismo em 29 de junho 2019, 72 e 93 nasceram no leste da África, considerando mulheres e homens, respectivamente (Figura 1A). Estudos demográficos mostram que, em cada um desses países, duas etnias são super-representadas entre os corredores de resistência de elite: os Kalenjin no Quênia e os Oromo na Etiópia [1, 2]. A etno-especificidade desse fenômeno levanta a hipótese de que fatores genéticos ou culturais possam ser a causa principal de tal dominância. Entretanto, nenhuma explicação simples surgiu ao se olhar para a história de vida ou loci classicamente associados a performance atlética [3, 4]. Porém, como apontado por Tucker *et al.* [4], a incapacidade de se detectar uma associação com marcadores genéticos pode se dever às limitações das abordagens de gene candidato e estudo caso-controle ao se estudar traços altamente complexos e influenciados pelo ambiente, como a performance atlética. Dessa forma, ele propõe que a super-representação das etnias Kalenjin e Oromo nessas modalidades poderia ser mais facilmente compreendida se estudada como um fenômeno populacional. No presente estudo, usamos essa estratégia para comparar dados

genômicos populacionais entre diferentes etnias do leste da África. Nossa hipótese é que parte dos fatores genéticos associados ao sucesso atlético em Kalenjin e Oromo devem estar entre as regiões genômicas de maior diferenciação comparando diferentes etnias, possibilitando a identificação de processos moleculares que contribuem com a corrida de longa distância.

As regiões genômicas mais diferenciadas em Kalenjin e Oromo foram identificadas através da Estatística de Ramo Populacional normalizada (*normalized Population Branch Statistic, PBSn1*), que indica a diferenciação genética específica em uma população alvo (Kalenjin ou Oromo), em relação a outras duas populações de referência, representadas aqui por duas etnias do leste da África (Luhya, do Quênia, e Amhara, da Etiópia) e duas populações compostas representando o oeste da África e a Eurásia (Figura 1B, *Star Methods*). Devido às diferenças de ancestralidade genética entre populações [8] foram realizados três cálculos de PBSn1 para cada etnia, retendo os 0,1% valores mais extremos (ver *Star Methods* para mais detalhes), resultando em uma lista final de 377 e 322 genes diferenciados para Kalenjin e Oromo [S3], respectivamente, anotados em uma janela de 5 kb em relação à posição do SNP de maior PBSn1 da janela (Figura 2).

A relação entre as regiões genômicas mais diferenciadas e a performance na corrida de resistência foi estabelecida de duas formas: 1) os cinco genes associados às janelas de maior PBSn1 de cada comparação (Figura 2) tiveram sua função molecular investigada na literatura; 2) além disso, as listas de genes diferenciados foram submetidas a análises de enriquecimento de conjuntos de genes, resultando em 130 e 193 conjuntos enriquecidos para Kalenjin e Oromo [S1, S2], respectivamente, filtrando-se processos moleculares e fatores fisiológicos ou antropométricos associados à economia de corrida (Tabela 1).

Traços antropométricos desempenham um papel fundamental, na medida em que determinam a biomecânica de corrida através de características como tamanho de passada, estabilidade do movimento e resistência oferecida pelo ar e chão [9, 10]. Eksterowicz *et al.* [9] encontraram correlações positivas entre o tempo de maratona e o comprimento dos membros superiores, o comprimento do torso, a largura do quadril e a razão cintura-quadril, para uma amostra de corredores de resistência quenianos. Além disso, membros inferiores mais longos reduzem o custo energético da corrida [11], e corredores de resistência de elite têm estatura média menor do que atletas de elite de outros esportes de resistência [12]. Interessantemente, nossa análise retornou conjuntos enriquecidos para altura, circunferência da cintura, e razão

cintura-quadril, tanto para Kalenjin quanto para Oromo, assim como conjuntos para formato ou circunferência do quadril para Kalenjin e Oromo, respectivamente.

Um dos genes mais diferenciado para Oromo, *LTBP1*, pode fornecer um indício sobre o mecanismo molecular parcialmente associado a esses fenótipos. *LTBP1* é uma proteína que se liga ao Fator de Transformação do Crescimento Beta 1 (TGF- β 1), auxiliando na sua exportação para a matriz extracelular de células ósseas e posterior estocagem na forma latente [13, 14]. Uma das funções do TGF- β 1, uma vez ativado na matriz extracelular, é estimular a maturação inicial e inibir a mineralização dos condrócitos na placa de crescimento, no processo de ossificação endocondral, além de ter um papel essencial na remodelação óssea [13, 15]. Assim, é possível que polimorfismos no gene *LTBP1* influenciem a quantidade disponível de TGF- β 1 durante o desenvolvimento, impactando a formação de estruturas ósseas e, portanto, as características antropométricas do indivíduo.

Além das medidas corporais, a estrutura de ossos e tendões desempenha um papel importante em todos os tipos de atividade física, não sendo diferente para a corrida de longa distância. Falando primeiramente em tendões, é observada uma menor flexibilidade e maior rigidez nos tendões dos membros inferiores, especialmente do joelho, de corredores de longa distância [16, 17], com uma menor flexibilidade corpórea geral associada à redução do custo energético da corrida [18]. Tais resultados reforçam a ideia de que uma maior rigidez em alguns grupos de tendões e ligamentos aumenta a estabilidade do movimento e o aproveitamento de energia elástica [19].

A principal proteína formadora de ligamentos e tendões é o colágeno do tipo I [20]. Dois genes associados a síntese e organização desse tipo de colágeno estão altamente diferenciados para Kalenjin e Oromo, respectivamente: *TRAM2-AS1* e *COL5A2*. O primeiro codifica um RNA antissenso do gene *TRAM2*, cuja proteína participa da via de translocação das cadeias formadoras do colágeno do tipo I para o retículo endoplasmático durante sua síntese [21]. Considerando a importância dos RNAs antissenso na regulação gênica [22], é plausível propor que *TRAM2-AS1* module a expressão de *TRAM2* e, assim, a síntese de colágeno do tipo I. Já o gene *COL5A2* codifica, juntamente com os genes *COL5A1* e *COL5A3*, uma das três cadeias alfa que compõe o colágeno do tipo V [23, 24], proteína que regula a montagem e estrutura das fibrilas de colágeno do tipo I [25, 26]. Mutações no *COL5A2* ou *COL5A1* são responsáveis por mais de 90% dos casos da Síndrome de Ehlers-Danlos na sua forma clássica, cujo um dos principais sintomas é a hipermobilidade articular [27]. Dois polimorfismos no *COL5A1*

(rs12722, C/T e rs71746744, -/AGGG) já foram diretamente associados com performance na corrida de longa distância, com indivíduos carregando os genótipos TT ou AGGG/AGGG sendo consideravelmente mais rápidos e inflexíveis [28-30]. Considerando que *COL5A1* e *COL5A2* codificam subunidades diferentes da mesma proteína estrutural, e que mutações em ambos têm as mesmas consequências clínicas, é possível que variantes no *COL5A2* tenham o mesmo efeito.

A corrida de longa distância também está associada ao aumento da densidade mineral óssea, nas pernas, bem como redução, nas vértebras, de corredores em comparação à indivíduos controle [31-33]. Nossos resultados indicam um conjunto enriquecido de genes relacionados à densidade mineral óssea, tanto em Kalenjin quanto em Oromo. Além do já descrito *LTBP1*, os genes altamente diferenciados, em Oromo, *MIR6797*, e em Kalenjin, *WFOX* (também presente na lista de Oromo), *TRAM2-AS1* e *MRAP2* poderiam afetar essa característica. Em especial, os três primeiros interagem com o fator de transcrição RUNX2, o grande maestro do desenvolvimento ósseo [34]. Foi demonstrado que o microRNA miR-6797-5p, transcrito a partir de *MIR6797*, atua como um regulador negativo da expressão de RUNX2, inibindo a diferenciação de osteoblastos [35]. De forma similar, a Oxirredutase Contendo Domínio WW (*WFOX*), proteína descrita inicialmente como supressora tumoral, inibe a capacidade ativadora de RUNX2 ao se ligar fisicamente a ele, funcionando como um supressor dos genes osteogênicos ativados por este. Essa função parece ser fundamental para a formação óssea. Camundongos *WFOX* deficientes (*Wfox*^{-/-}) morrem prematuramente, atraso considerável no desenvolvimento esquelético, baixa mineralização dos ossos e redução nos níveis de cálcio circulante [36, 37]. O fator RUNX2 também regula a expressão do já citado gene *TRAM2* em células ósseas, uma vez que o colágeno do tipo I é o principal componente orgânico da matriz extracelular óssea [38].

Por sua vez, *MRAP2* é um gene cujo produto proteico modula a atividade dos receptores de melanocortina, em especial o Receptor de Melanocortina 4 (MC4R) [39-41]. O MC4R regula processos fisiológicos relacionados ao metabolismo energético, tendo sido associado a índice de massa corporal em populações africanas e europeias [42]. Mutações nos genes *MRAP2* e *MC4R* causam quadros de obesidade tanto em humanos como em camundongos [43-45]. Braun *et al.* [44] mostraram que camundongos com o *Mc4r* nocauteado (MC4RKO) apresentavam um quadro de obesidade distinto da obesidade induzida por dieta. Camundongos MC4RKO apresentavam uma menor taxa de batimentos cardíacos, levada massa magra, força muscular e, especialmente, densidade óssea, mostrando um menor desempenho na corrida de resistência.

Apesar da grande importância das características morfológicas, os fatores fisiológicos também são fundamentais para a corrida de resistência [46]. A principal via envolvida nesse processo é a de metabolização aeróbica de energia, que determina a quantidade de energia disponível e o tempo pelo qual uma atividade física pode ser mantida sem que haja o acúmulo excessivo de subprodutos indesejados da respiração anaeróbica [19]. A eficácia de tal via é o somatório da performance de uma série de etapas fisiológicas distintas: a eficiência com a qual o indivíduo capta o oxigênio do ar, transporta esse oxigênio para o músculo via corrente sanguínea, mobiliza as reservas energéticas para serem oxidadas e, finalmente, utiliza a energia produzida para a contração muscular. O produto dessas etapas, somado a eficiência biomecânica da corrida, é normalmente medido como consumo máximo de oxigênio (VO_{2max}) e o consumo de oxigênio para correr a uma determinada velocidade submáxima (economia de corrida). Os melhores corredores de resistência serão, no geral, aqueles com capacidade de consumir mais oxigênio (elevado VO_{2max} , alta capacidade de mobilizar oxigênio e energia), mas que consumirão menos oxigênio para correr a velocidades intermediárias (boa economia de corrida, baixa necessidade energética) [19].

Começaremos pela primeira etapa dessa via. A captação de oxigênio nos pulmões, medida como função pulmonar, já foi diretamente correlacionada com o desempenho em provas de resistência de diferentes durações [47, 48]. Além disso, Kirkton *et al.* [49] demonstraram que a seleção artificial para resistência na corrida em ratos também seleciona uma melhor função pulmonar, revelando não somente a importância de tal traço para a corrida de resistência, como a forte influência de fatores genéticos sobre ele. Isso vai de acordo com a noção de que o VO_{2max} possui uma herdabilidade considerável ($h^2 \cong 0.56$) [50, 51]. Na mesma linha, nossa análise mostra um número considerável de conjuntos enriquecidos para função pulmonar, tanto em Kalenjin quanto em Oromo, assim como um conjunto para desenvolvimento do sistema respiratório em Kalenjin. Interessantemente, o já citado gene *WWOX*, presente na lista de ambas as populações, vem sendo associado a capacidade respiratória em uma série de estudos de associação genômica (GWAS) [52-57], embora seu mecanismo de ação sobre esse fenótipo ainda não esteja bem compreendido.

Após o oxigênio ser absorvido, ele precisa ser transportado até o músculo pela corrente sanguínea. A eficiência desse transporte vai depender, entre outros fatores, da pressão sanguínea, da quantidade de hemoglobina disponível, determinada pelo tamanho e concentração de hemácias, e da vascularização do tecido. De fato, a pressão sanguínea

diastólica já foi identificada como um preditor do desempenho em algumas provas de resistência [58, 59]. Além disso, corredores de resistência apresentam hipertrofia do ventrículo cardíaco esquerdo [60], relacionada à pressão sanguínea, e, frequentemente, hiperplasia da medula óssea vermelha [61]. Também é sabido que o exercício é um potente indutor da hematopoiese [62, 63] e da angiogênese [64-66], e que a vascularização da musculatura está diretamente relacionada ao VO_{2max} [67].

A pressão sanguínea parece ser especialmente importante em Oromo, já que a lista de genes nessa população apresenta um enriquecimento para conjuntos associados diretamente a essa característica, assim como condução cardíaca. Um dos genes altamente diferenciados nessa etnia, o *ADAMTS9-AS2*, codifica o RNA antissenso do gene *ADAMTS9*, cuja proteína, uma metaloprotease, é essencial para o desenvolvimento normal e homeostase do coração e artérias [68, 69]. Tanto Kalenjin quanto Oromo apresentam enriquecimento do conjunto associado ao desenvolvimento do sistema circulatório e, embora Kalenjin, ao contrário de Oromo, não apresente enriquecimento de conjuntos gênicos associados à pressão sanguínea, quatro dos seus genes altamente diferenciados (*ARHGEF1*, *GATADI*, *GSTO1* e *AMOTLI*) influenciam essa característica. Isso pode apontar uma diferença de contribuição individual de genes para uma mesma característica entre as populações. O gene *ARHGEF1* codifica o fator de troca de guanina RhoA que é ativado pela Angiotensina II nas células musculares lisas arteriais, levando ao aumento da pressão [70, 71]. O gene que codifica para a Enzima Conversora de Angiotensina I (ACE), que produz a Angiotensina II, já foi associado com a corrida de resistência em diversos estudos [72-74], embora nenhuma associação tenha sido encontrada em corredores etíopes [75]. Por sua vez, *GATADI* codifica a Proteína 1 Contendo Domínio de Dedo de Zinco GATA, expressa em miócitos ventriculares. Mutações nesse gene causam cardiomiopatia dilatada, uma doença caracterizada pelo aumento excessivo dos ventrículos cardíacos [76, 77]. Já *GSTO1* codifica uma glutathione transferase que modula a atividade dos receptores de rianodina (RyR). Os RyRs são os canais de cálcio do retículo sarcoplasmático, responsáveis por liberar o cálcio do retículo no citoplasma e assim, realizar a contração muscular. Foi demonstrado que a proteína GSTO1 regula negativamente a atividade do receptor de rianodina do músculo cardíaco (RyR2) [78], cujo gene, *RyR2*, está presente na lista de Oromo. Por fim, o gene *AMOTLI* codifica uma proteína similar a Angiomotina que está associada ao aumento e proliferação de cardiomiócitos e hipertrofia cardíaca [79], assim como ao processo de angiogênese, por controlar a adesão entre células endoteliais [80, 81].

Além destes genes, em nossas comparações existe um gene altamente diferenciado em ambas as populações, o *RPS19*, que tem associação causal com a anemia de Diamond-Blackfan [82, 83], podendo, portanto, afetar a hematopoiese de uma forma geral. Entretanto, a etiologia da anemia de Diamond-Blackfan parece estar associada a defeitos amplos na síntese proteica na medula óssea [84, 85]. Além disso, não foram encontradas diferenças nos níveis de hemoglobina entre corredores de resistência quenianos e de outras etnias [86]. Como os genes *RPS19* e *MIR6797* são muito próximos, e uma vez que ambas as populações possuem outros genes com alto valor de PBSn1 que, assim como o *MIR6797*, interagem com o fator RUNX2, nós hipotetizamos que o sinal associado ao *RPS19* é na verdade um sinal associado ao *MIR6797*, que não foi recuperado em algumas comparações devido ao tamanho conservador de nossa janela de anotação.

Quando o oxigênio atinge a musculatura, ele precisa encontrar reservas energéticas mobilizadas para oxidar e, assim, produzir energia. A principal via de produção de energia na corrida de resistência será a via glicolítica aeróbica, na mitocôndria [87], com o consumo de carboidratos, antes e durante o exercício, diretamente associado a melhora na performance [88]. De fato, foi demonstrado que a dieta de corredores de resistência quenianos consiste majoritariamente de carboidratos [89]. Por outro lado, a utilização de gordura como substrato energético, ao invés da glicose, está ligada à capacidade de manter a atividade física por mais tempo, ao preservar os níveis sistêmicos de glicose [90]. Tanto Kalenjin quanto Oromo possuem conjuntos enriquecidos associados ao metabolismo energético, respectivamente, os conjuntos “Traços de Homeostase de Glicose” e “Atividade Física Vigorosa”, que contém genes ligados ao metabolismo de glicose.

A etnia Oromo, em especial, apresenta uma alta diferenciação em genes que codificam proteínas do complexo do exocisto (*EXOC4*, *EXOC5* e *EXOC6B*), responsável por direcionar vesículas intracelulares para regiões específicas da membrana citoplasmática [91]. Esse complexo é essencial para o metabolismo de glicose, uma vez que é o responsável pelo transporte induzido pela insulina do GLUT4, principal transportador de glicose do músculo esquelético, cardíaco e tecido adiposo, para a membrana plasmática [92]. Variantes próximas ao *EXOC4* estão associadas a diabetes tipo 2 [93] e o nocaute desse gene reduz consideravelmente a entrada de glicose nos adipócitos [92]. O mesmo efeito foi observado em células do músculo esquelético através do nocaute de *EXOC5* ou da adição de inibidor do exocisto [94]. Da mesma forma, o nocaute de *EXOC6B* inibe a translocação do GLUT4 em

adipócitos [95]. Dentro da célula, foi demonstrado que o gene *WFOX* modula a glicólise aeróbica, ao se ligar fisicamente ao fator de transcrição Fator 1 Induzido por Hipóxia ($HIF1\alpha$) e inibir ativação de genes indutores da glicólise por este. Dessa forma, células *WFOX* deficientes apresentam elevação nos níveis e atividade do $HIF1\alpha$, assim como aumento na captação de glicose [96-98]. O gene que codifica o $HIF1\alpha$, o *HIF1A*, já foi diretamente associado com o status de corredor de resistência de elite [99]. Por fim, um gene altamente diferenciado em Oromo é *LIPE-AS1*, que codifica o RNA antissenso do gene da Lipase Sensível a Hormônio (*LIPE*). A *LIPE* é a principal reguladora da lipólise em adipócitos, liberando ácidos graxos a partir dos triglicerídeos armazenados. Mutações nesse gene estão diretamente relacionadas a quadros de lipomatose, miopatia, elevados níveis de glicose no sangue e diabetes tipo 2 [100-102].

Uma vez que energia é produzida, seu destino final nesse processo será sua utilização na contração muscular, tanto dos músculos estriados esqueléticos quanto do cardíaco. Um dos protagonistas do processo de contração muscular é o íon Ca^{2+} , cuja liberação do retículo sarcoplasmático leva à contração muscular [103]. Tanto a etnia Kalenjin quanto a etnia Oromo apresentaram enriquecimento na lista de genes para conjuntos de processos relacionados ao cálcio, o que não foi observado para outros íons como K^+ , Na^+ , Cl^- ou Mg^{2+} . Ademais, o já citado gene *GSTO1*, da lista de Kalenjin, além de regular negativamente o canal de cálcio do músculo cardíaco (*RYR2*), regula positivamente a versão expressa no músculo esquelético (*RYR1*) [78], podendo ter um papel duplo na corrida de resistência. Já foi demonstrado que ratos com maior capacidade intrínseca na corrida de resistência possuem cardiomiócitos com maior amplitude de contração, pico de liberação de cálcio (correlacionado à performance) e expressão de *RYR2* em relação a ratos com baixa capacidade intrínseca na corrida de resistência [104]. Além disso, foi observado que o treinamento em ratos induz uma maior capacidade de contração e sensibilidade ao Ca^{2+} nessas mesmas células [105]. Esses resultados reforçam a importância da regulação do cálcio intracelular no tecido muscular durante a atividade física prolongada.

Por que os genes aqui apresentados não foram identificados em outros estudos buscando as bases genéticas da corrida de resistência? Como grande parte destes estudos é realizada dentro da mesma população [106], é possível que tais genes não sejam detectados por estarem ligados a diferenças interpopulacionais de desempenho, com as diferenças intrapopulacionais sendo explicadas por outros loci ou fatores ambientais, como treinamento. Além disso, estudos

genômicos deste tipo foram realizados apenas em populações eurasiáticas [106, 107], enquanto é sabido que populações africanas possuem um grande número de variantes raras ou exclusivas [108]. O fato de termos recuperados genes que interagem diretamente com genes classicamente associados a corrida de resistência, como no caso dos genes *COL5A2*, *ARHGEF1* e *WWOX* em oposição aos genes *COL5A1*, *ACE* e *HIF1A*, pode indicar que as mesmas vias são importantes para corredores de resistência de diferentes etnias, porém com diferentes contribuições de cada gene, dependendo da etnia. Em conjunto, esses resultados reforçam a ideia da corrida de resistência como uma atividade complexa e sistêmica. De fato, o aparecimento de genes que influenciam mais de uma característica, como *WWOX*, *MRAP2*, *TRAM-AS1* e *GSTO1* mostram a importância de genes pleiotrópicos para esse tipo de traço. Mesmo com diversos genes altamente diferenciados nessas populações associados a características que sabidamente afetam o desempenho na corrida, mais estudos são necessários para esclarecer a associação específica dos genes candidatos aqui apontados com a corrida de resistência, assim como, caracterizá-los em diferentes populações humanas. É importante destacar que nossa discussão aqui tem de se limitar a processos cuja relação com a corrida de resistência está bem documentada na literatura, pois as populações são diferenciadas em diversos aspectos não necessariamente ligado a corrida. Dessa forma, conforme nosso conhecimento sobre fisiologia do esporte avança, novas associações podem ser encontradas nos resultados deste estudo.

Finalmente, como autores, gostaríamos de reforçar que o sucesso nos esportes não deve ser tomado como um estereótipo, assim como predisposição genética não significa predestinação. A genética nunca deve ser vista de forma determinística, especialmente para traços complexos, onde fatores ambientais e socioculturais sempre desempenharão um papel importante em populações humanas. O propósito desse trabalho, mais do que explorar as bases biológicas da corrida de longa distância, é ilustrar a beleza da diversidade genética humana, especialmente na África, cuja riqueza genética e cultural tende a ser vista de maneira simplificada e homogeneizada.

Agradecimentos

Gostaríamos de agradecer à Bruna Santos e Gabriela Cybis pelas valiosas contribuições, bem como aos integrantes do Laboratório de Genética Médica e Evolução do Departamento de Genética da Universidade Federal do Rio Grande do Sul por todo o suporte oferecido.

Contribuição dos Autores

N.J.R.F. concebeu o estudo. A.L.S.Z. e N.J.R.F. contribuíram igualmente para o desenho metodológico. M.H.G. auxiliou no desenho metodológico e codificou os dados. M.M.A. auxiliou na construção dos scripts utilizados na análise dos dados. A.L.S.Z. analisou os dados. A.L.S.Z. e N.J.R.F. escreveram a versão preliminar do artigo.

Declaração de Interesses

Os autores declaram não haver conflito de interesses.

Figuras e Tabelas

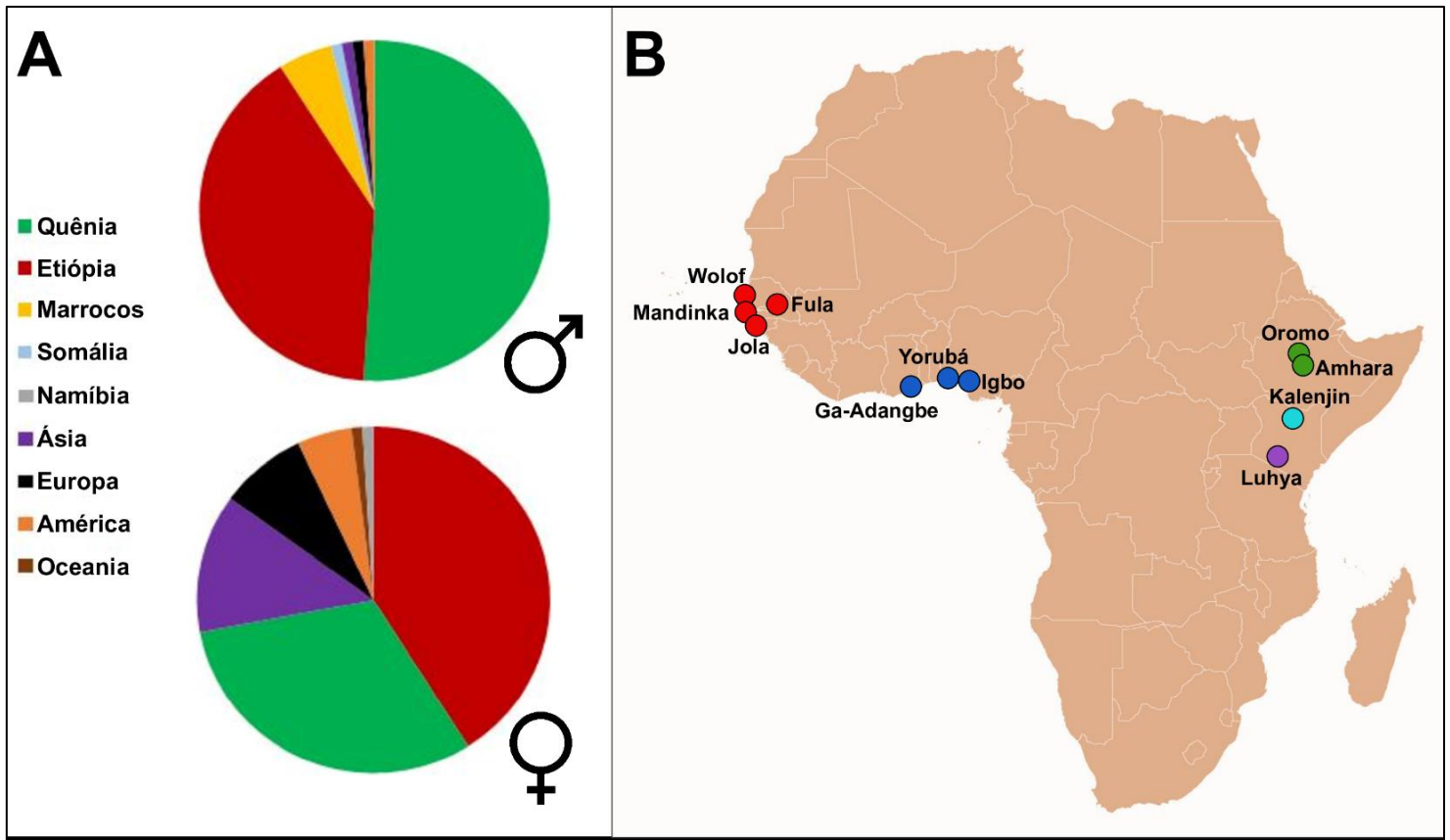
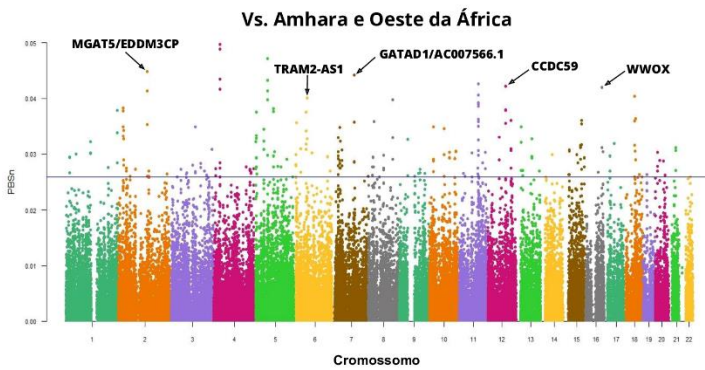
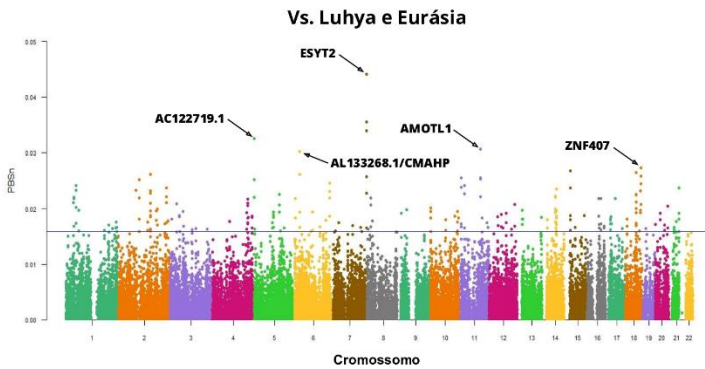
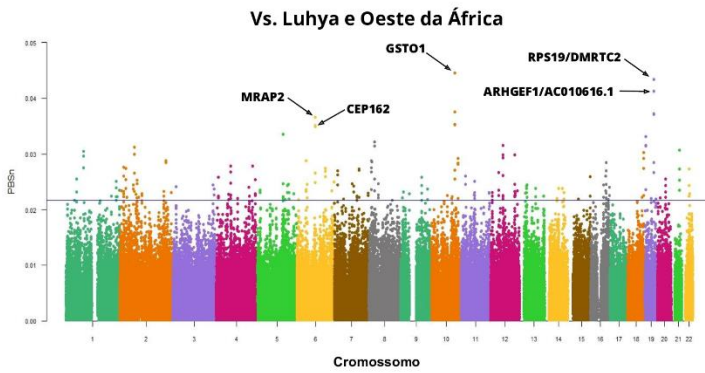


Figura 1. A) Gráfico de pizza dos 100 melhores maratonistas de todos os tempos, homens e mulheres, por local de nascimento, como listado pela Associação Internacional de Federações de Atletismo em 29 de junho 2019. Note a grande fração correspondente ao Quênia e a Etiópia, em ambos os sexos. B) Grupos etno-linguísticos africanos utilizados neste estudo.

PBSn1 para Kalenjin



PBSn1 para Oromo

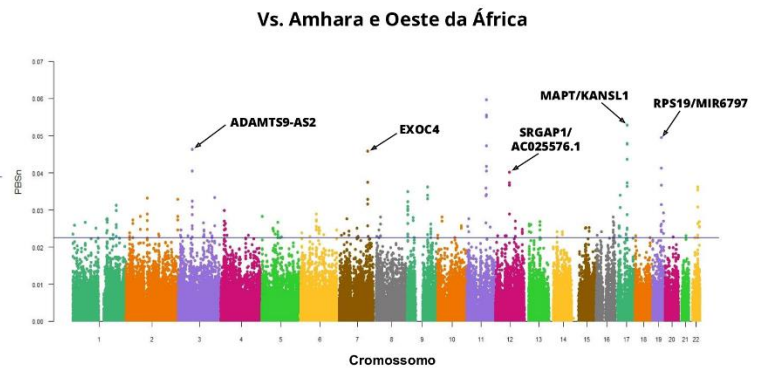
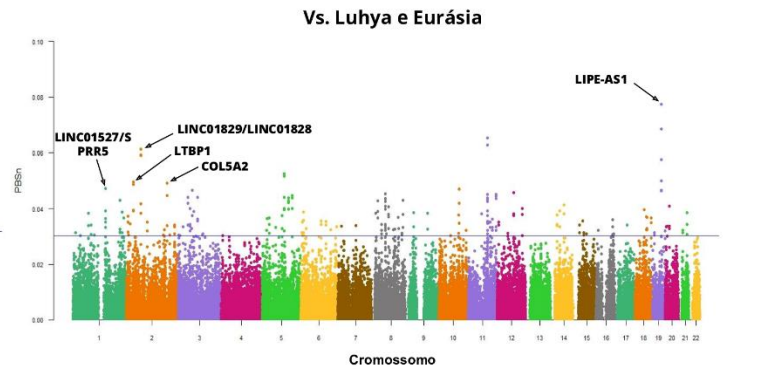
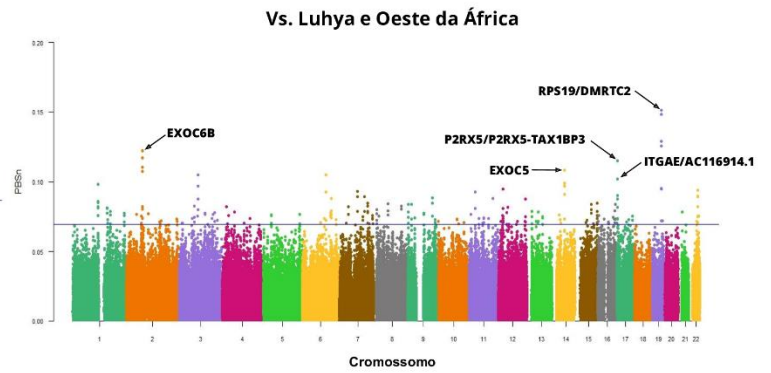


Figura 2. *Manhattan plots* dos valores de PBSn1 para as etnias Kalenjin e Oromo quando comparadas com cada par de populações de referência. Cada ponto representa uma janela de 20 SNPs. A linha azul demarca os 0,1% SNPs mais diferenciados. Observe que a escala não é a mesma para todas as comparações. As janelas não intergênicas com os 5 maiores valores de PBSn1, de cada comparação, têm seus genes associados mostrados. Para lista completa de genes, consultar Tabela S3.

Tabela 1. Categorias enriquecidas ($P < 0,05$; FDR) para Kalenjin e Oromo possivelmente associados à corrida de resistência. Os conjuntos de genes foram agrupados em categorias mais amplas em relação à base FUMA-GWAS. Categorias em negrito e itálico têm conjuntos em comum entre Kalenjin e Oromo. Para a lista completa de conjuntos, consultar a Tabela S1 e a Tabela S2. N, número de conjuntos em cada categoria; Ref, referências que justificam a inclusão desta categoria.

Kalenjin					
Categoria GWAS catalog	N	Ref	Cartegoria GO_biological_process	N	Ref
<i>Altura</i>	1	11, 12	GO_CALCIUM_DEPENDENT_CELL_CELL_ADHESION	1	103-105
<i>Circunferência da cintura</i>	1	9	<i>GO_CIRCULATORY_SYSTEM_DEVELOPMENT</i>	1	58-67, 72-74
<i>Densidade mineral óssea</i>	1	31-33, 44	<i>GO_REGULATION_OF_CALCIUM_MEDIATED_SIGNALING</i>	1	103-105
Forma do quadril	1	9	GO_RESPIRATORY_SYSTEM_DEVELOPMENT	1	47-49
<i>Função pulmonar</i>	4	47-49			
Homeostase de glicose	1	87-90, 99			
<i>Razão cintura-quadril</i>	2	9			
Oromo					
Categoria GWAS catalog	N	Ref	Cartegoria GO_biological_process	N	Ref
<i>Altura</i>	1	11, 12	GO_CALCIUM_ION_TRANSPORT	2	103-105
Atividade física vigorosa	1	87-90, 99	GO_CALCIUM_MEDIATED_SIGNALING	2	103-105
<i>Circunferência da cintura</i>	7	9	GO_CARDIAC_CONDUCTION	1	58-60, 104, 105
Circunferência do quadril	2	9	<i>GO_CIRCULATORY_SYSTEM_DEVELOPMENT</i>	1	58-67, 72-74
<i>Densidade mineral óssea</i>	2	31-33, 44	GO_REGULATION_OF_CALCIUM_ION_TRANSPORT	10	103-105
<i>Função pulmonar</i>	5	47-49	<i>GO_REGULATION_OF_CALCIUM_MEDIATED_SIGNALING</i>	1	103-105
Pressão sanguínea	6	58-60, 72-74			
<i>Razão cintura-quadril</i>	6	9			

STAR Methods

Contato Principal e Disponibilidade de Materiais

Para obtenção de quaisquer informações adicionais, por favor, contatar o Contato Principal André Luís da Silva Zani, zani.andre@yahoo.com.br.

Detalhes do Método

Uma vez que nosso trabalho parte de uma “hipótese populacional”, isto é, de que de fatores genéticos mais diferenciados em Kalenjin ou Oromo podem ter relação com o sucesso na corrida de resistência, a amostra de indivíduos que contribuíram com dados genético não é composta exclusivamente por corredores, justamente porque os fatores genéticos que predisporiam à corrida de resistência são comuns na população geral dessas etnias. De fato, nosso trabalho não é o primeiro a considerar um fenótipo como uma característica populacional. Estudos de adaptação à altitude em humanos frequentemente adotam essa estratégia, comparando populações que se sabe estarem adaptadas às restrições impostas pela altitude com populações semelhantes de terras baixas, em busca de picos de diferenciação que possam ser a base genética para tal adaptação, independentemente de possíveis variações individuais intraespecíficas na resposta à altitude [109, 110].

De maneira semelhante a esses estudos, os picos de diferenciação genômica populacional foram estimados a partir da estatística de tamanho de ramo populacional (*Population Branch Statistics* – PBS) [109], em sua versão normalizada, o PBSn1, como descrito por Malaspina *et al.* [111]. Tal estatística se utiliza de dados de frequência alélica para verificar o grau de diferenciação de uma população de interesse, ou população focal, em relação a outras duas populações de referência, em geral uma mais próxima da população focal e outra mais distantemente relacionada (muitas vezes referidas como populações “*closely related*” e “*distantly related*”, respectivamente). Dependendo das populações de referência escolhidas, também é possível filtrar diferentes componentes de ancestralidade, uma vez que fatores em comum com as populações relacionadas não irão contribuir com um valor alto para PBSn1. É importante notar que, embora muito amplamente usado em estudos sobre adaptação, PBSn1 pode ser utilizado para mesurar diferenciação genética independentemente da origem de tal diferenciação (seleção natural, deriva genética, miscigenação, etc.). Dessa forma, ao escolhê-

lo, evitamos fazer qualquer suposição sobre a causa da possível divergência populacional que geraria a predisposição a corrida de resistência nas etnias Kalenjin e Oromo.

Obtivemos as informações de frequência alélica das populações de interesse a partir de dados públicos disponibilizados pelo Projeto de Variação Genômica Africana (*African Genome Variation Project* - AGVP) [8], que genotipou 1.152.000 polimorfismos de nucleotídeo único (SNPs, em inglês) ao longo do genoma de 1481 indivíduos de 18 grupos etno-linguísticos da África Subsaariana. Para acessar a diversidade de outras populações africanas, assim como do resto do planeta, utilizada no cálculo do PBSn1, também incluímos dados equivalentes obtidos do banco de dados do Projeto 1000 Genomas [112]. Todos os dados utilizados fazem parte de bancos de dados públicos que foram montados respeitando as considerações éticas elaboradas pelos comitês de pesquisa relevantes, tanto em âmbito nacional (para os países envolvidos) quanto internacional.

Para determinar o nível de diferenciação genética ao longo do genoma utilizamos uma estratégia baseada em janelas deslizantes (*sliding windows*). Nessa estratégia, cada janela tem um tamanho de 20 SNPs, com sobreposição de 15 SNPs entre cada uma, garantindo uma cobertura homogênea do genoma e reduzindo o efeito de SNPs individuais. Calculamos um valor de PBSn1 para cada SNP e depois geramos o valor médio para o PBSn1 como representante do PBSn1 da janela. As populações focais foram os Kalenjin (n=100) e os Oromo (n=26). A fim de filtrar diferentes componentes ancestralidade, escolhemos dois contrastes locais (*"closely related"*) e dois contrastes gerais (*"distantly related"*) como referências. Os contrastes locais estão representados pelos Luhya (n=74), do Quênia, e os Amhara (n=42), da Etiópia, que foram selecionados por se encontrarem nos mesmos países das populações focais (e teoricamente, sob as mesmas políticas socioeconômicas), assim como por representarem ótimos filtros para ancestralidade banta e cushítica, respectivamente. Os outros dois grupos de referência foram construídos de forma a representar o Oeste da África (em contrapartida ao Leste), com 7 etnias (Wolof, Mandinka, Jola, Fula, Ga-Adangbe, Yorubá e Igbo, n=618), e a Eurásia, com 6 populações do Projeto 1000 Genomas (Indivíduos residentes nos Estados Unidos com ancestralidade europeia - *CEU*, Toscanos, Finlandeses, Britânicos, Ibéricos, e Indivíduos residentes nos Estados Unidos com ancestralidade indiana - *GIH*, n=569). As populações africanas utilizadas estão representadas na Figura 1B. Optamos pela utilização tanto de um contraste eurasiático quanto de um contraste do Oeste da África pois foi demonstrado que as populações de interesse têm um considerável nível de ancestralidade eurasiática, entre 7,3 e 10,99% para Kalenjin e entre 43,62 e 50,82% para Oromo [8]. A utilização exclusiva de

um contraste Africano poderia super-representar loci com alta ancestralidade Eurasiática nas populações focais do estudo.

Realizamos os cálculos do PBSn1 utilizando sempre uma população focal, uma contrapartida local e uma composição geral, em todas as combinações possíveis, a exceção das duas combinações que incluiriam os Amhara e a Eurásia, resultando em seis comparações. Essas combinações foram excluídas pois os Amhara possuem um nível muito elevado de mistura eurasiática (de 47,82 a 54,70%), e, no cálculo do PBSn1, resultados espúrios podem ser gerados quando as duas populações de referência têm mais proximidade entre si do que com a população de interesse [111]. Por outro lado, graças a esse nível de mistura, os Amhara por si só já funcionam como bons filtros para a ancestralidade eurasiática.

Quando há uma hipótese adaptativa sobre o fenótipo em questão, simulações demográficas podem ser usadas para determinar um valor empírico a partir do qual valores de PBSn1 são indicativos de um processo adaptativo. Nesse estudo, evitamos qualquer hipótese adaptativa para o fenótipo de interesse, e, portanto, assumimos que seja possível que fenômenos puramente demográficos como deriva genética e fluxo gênico estejam associados a PBSn1 altos em loci genômicos associados com corrida de resistência. Nesse caso, é impossível estabelecer um valor empírico de significância para os valores de PBSn1 a partir de simulações demográficas. Dessa forma, seguindo a lógica de trabalhos semelhantes [110, 113] retivemos, para cada comparação, as 0,1% janelas com maior valor de PBSn1 para análises posteriores.

Os genes em uma vizinhança de até 5 Kb do SNP de maior PBSn1 de cada janela foram anotados, e a partir das três comparações foram elaboradas uma lista final de genes altamente diferenciados para cada população focal. As listas de genes assim obtidas foram submetidas a Análise de Conjuntos de Genes, a fim averiguar quais categorias de fenótipos ou processos biológicos estão enriquecidos para os genes presentes nas listas (ou seja, que possuem, para cada categoria, mais genes na lista do que seria esperado ao acaso). Uma vez que partimos de comparações interpopulacionais amplas, diversos processos biológicos sem relação biológica (a priori) com corrida de resistência, podem estar enriquecidas nas análises. A seleção de processos biológicos ou fenótipos de potencial relevância para corrida de resistência foi feita com base em trabalhos antropométricos e fisiológicos sobre economia de corrida [9-12, 16-19, 28-33, 47-49, 58-67, 72-75, 86-90, 99, 103-107]. Além disso, as funções moleculares dos 5 genes de maior valor de PBSn1 de cada comparação foram investigadas detalhadamente na literatura em busca de associações biologicamente plausíveis com corrida de resistência.

Quantificação e Análises Estatísticas

Os cálculos de frequência alélica foram feitos no software Plink [114], que foram utilizados para a estimativa do PBSn1, tanto por SNP quanto por janela, por scripts próprios desenvolvidos em linguagem R [115]. A anotação dos genes foi feita com base na plataforma Ensembl [116]. As Análises de Conjuntos de Genes foi feita através da plataforma Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS) [117]. Foram considerados significantes apenas valores de enriquecimento com $P < 0,05$ após correção para testes múltiplos por *False Discovery Ratio* (FDR). As categorias de fenótipos foram recuperadas a partir da base de dados GWAS Catalog [118], enquanto as categorias de processos biológicos associados às ontologias gênicas (*Gene Ontology biological processes*) foram recuperados a partir da base de dados MsigDB [119].

Disponibilidade de Dados e Códigos

Todos os dados analisados neste artigo podem ser requisitados do European Genome-phenome Archive, através do link www.ebi.ac.uk/ega/dacs/EGAC00001000237, e do Projeto 1000 Genomas, através do link <https://www.internationalgenome.org/data/>. Os scripts utilizados para o cálculo dos valores de PBSn1 podem ser obtidos entrando em contato com o Contato Principal.

Tabela de Recursos-chave

REAGENTE ou RECURSO	FONTE	IDENTIFICADOR
Dados Depositados		
Dados de SNPs de Populações Africanas	Projeto de Variação Genômica Africana	www.ebi.ac.uk/ega/dacs/EGAC00001000237
Dados de SNPs de Populações Eurasiáticas	Projeto 1000 Genomas	https://www.internationalgenome.org/data/
Fenótipos de estudos de GWAS	Base de dados GWAS Catalog	https://www.ebi.ac.uk/gwas/
Ontologias Gênicas	Base de dados MsigDB	http://software.broadinstitute.org/gsea/msigdb/index.jsp
Softwares e Algoritimos		
Software Plink	Purcell <i>et al.</i> [114]	https://www.cog-genomics.org/plink2/
Software R	R Core Team [118]	https://cran.r-project.org/bin/windows/base/
Scripts em na linguagem R	Esse estudo	zani.andre@yahoo.com.br

Material Suplementar

Tabelas S1 e S2: listas de conjuntos enriquecidos ou não, em inglês, para Kalenjin e Oromo, respectivamente. Os conjuntos $P < 0,05$ após a correção por FDR estão em negrito. Destes, conjuntos associados a corrida de resistência estão destacados em amarelo.

Tabela S3: Lista dos genes associados as 0,1% janelas mais diferenciadas, por comparação, para Kalenjin e Oromo.

As tabelas podem ser acessadas através do link:

<https://drive.google.com/drive/folders/1E4EVGOc6mENBQLMRyyVfwKOscbkZTftU?usp=sharing>

Referências

1. Onywera, V., Scott, R., Boit, M., and Pitsiladis, Y. (2006). Demographic characteristics of elite Kenyan endurance runners. *Journal Of Sports Sciences* 24, 415-422.
2. Scott, R., Georgiades, E., Wilson, R., Goodwin, W., Wolde, B., and Pitsiladis, Y. (2003). Demographic Characteristics of Elite Ethiopian Endurance Runners. *Medicine & Science In Sports & Exercise* 35, 1727-1732.
3. Wilber, R., and Pitsiladis, Y. (2012). Kenyan and Ethiopian Distance Runners: What Makes Them so Good?. *International Journal Of Sports Physiology And Performance* 7, 92-102.
4. Tucker, R., Santos-Concejero, J., and Collins, M. (2013). The genetic basis for elite running performance. *British Journal Of Sports Medicine* 47, 545-549.
5. Carrier, D., Kapoor, A., Kimura, T., Nickels, M., Scott, E., So, J., and Trinkaus, E. (1984). The Energetic Paradox of Human Running and Hominid Evolution [and Comments and Reply]. *Current Anthropology* 25, 483-495.
6. Bramble, D., and Lieberman, D. (2004). Endurance running and the evolution of Homo. *Nature* 432, 345-352.
7. Lieberman, D., and Bramble, D. (2007). The Evolution of Marathon Running. *Sports Medicine* 37, 288-290.
8. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M., and Choudhury, A. et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature* 517, 327-332.
9. Eksterowicz, J., Napierała, M., and Żukow, W. (2016). How the Kenyan Runner's Body Structure Affects Sports Results. *Human Movement* 17, 8-14.
10. Fletcher, J., and MacIntosh, B. (2017). Running Economy from a Muscle Energetics Perspective. *Frontiers In Physiology* 8, 433.
11. Steudel-Numbers, K., Weaver, T., and Wallscheffler, C. (2007). The evolution of human running: Effects of changes in lower-limb length on locomotor economy. *Journal Of Human Evolution* 53, 191-196.
12. Sleivert, G., and Rowlands, D. (1996). Physical and Physiological Factors Associated with Success in the Triathlon. *Sports Medicine* 22, 8-18.
13. Pedrozo, H., Schwartz, Z., Mokeyev, T., Ornoy, A., Xin-Sheng, W., Bonewald, L., Dean, D., and Boyan, B. (1999). Vitamin D3 metabolites regulate LTBP1 and latent TGF- β 1 expression and latent TGF- β 1 incorporation in the extracellular matrix of chondrocytes. *Journal Of Cellular Biochemistry* 72, 151-165.
14. Unsöld, C., Hyytiäinen, M., Bruckner-Tuderman, L., and Keski-Oja, J. (2001). Latent TGF- β binding protein LTBP-1 contains three potential extracellular matrix interacting domains. *Journal Of Cell Science* 114, 187-197.

15. Tang, Y., Wu, X., Lei, W., Pang, L., Wan, C., Shi, Z., Zhao, L., Nagy, T., Peng, X., and Hu, J. et al. (2009). TGF- β 1-induced migration of bone mesenchymal stem cells couples bone resorption with formation. *Nature Medicine* *15*, 757-765.
16. Kubo, K., Tabata, T., Ikebukuro, T., Igarashi, K., Yata, H., and Tsunoda, N. (2010). Effects of mechanical properties of muscle and tendon on performance in long distance runners. *European Journal Of Applied Physiology* *110*, 507-514.
17. Kubo, K., Miyazaki, D., Shimoju, S., and Tsunoda, N. (2015). Relationship between elastic properties of tendon structures and performance in long distance runners. *European Journal Of Applied Physiology* *115*, 1725-1733.
18. Gleim, G., Stachenfeld, N., and Nicholas, J. (1990). The influence of flexibility on the economy of walking and jogging. *Journal Of Orthopaedic Research* *8*, 814-823.
19. Saunders, P., Pyne, D., Telford, R., and Hawley, J. (2004). Factors Affecting Running Economy in Trained Distance Runners. *Sports Medicine* *34*, 465-485.
20. Stępień-Słodkowska, M., Ficek, K., Eider, J., Leońska-Duniec, A., Maciejewska-Karłowska, A., Sawczuk, M., Zarębska, A., Jastrzębski, Z., Grenda, A., and Kotarska, K. et al. (2013). The +1245G/T polymorphisms in the collagen type I alpha 1 (COL1A1) gene in polish skiers with anterior cruciate ligament injury. *Biology Of Sport* *30*, 57-60.
21. Stefanovic, B., Stefanovic, L., Schnabl, B., Bataller, R., and Brenner, D. (2004). TRAM2 Protein Interacts with Endoplasmic Reticulum Ca²⁺ Pump Serca2b and Is Necessary for Collagen Type I Synthesis. *Molecular And Cellular Biology* *24*, 1758-1768.
22. Pelechano, V., and Steinmetz, L. (2013). Gene regulation by antisense transcription. *Nature Reviews Genetics* *14*, 880-893.
23. Birk, D., Fitch, M., Babiarz, P., Doane, J., and Linsenmayer, F. (1990). Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. *Journal of cell science* *95*, 649-657.
24. Imamura, Y., Scott, C., and Greenspan, S. (2000). The pro- α 3 (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. *Journal of Biological Chemistry* *275*, 8749-8759.
25. Andrikopoulos, K., Liu, X., Keene, D., Jaenisch, R., and Ramirez, F. (1995). Targeted mutation in the col5a2 gene reveals a regulatory role for type V collagen during matrix assembly. *Nature Genetics* *9*, 31-36.
26. Richards, A., Martin, S., Nicholls, A., Harrison, J., Pope, F., and Burrows, N. (1998). A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type II. *Journal Of Medical Genetics* *35*, 846-848.
27. Malfait, F., Francomano, C., Byers, P., Belmont, J., Berglund, B., Black, J., Bloom, L., Bowen, J., Brady, A., and Burrows, N. et al. (2017). The 2017 international classification of the Ehlers-Danlos syndromes. *American Journal Of Medical Genetics Part C: Seminars In Medical Genetics* *175*, 8-26.

28. Posthumus, M., Schwelanus, M., and Collins, M. (2011). The COL5A1 Gene: a novel marker of endurance running performance. *Medicine & Science In Sports & Exercise* 43, 584-589.
29. Brown, J., Miller, C., Posthumus, M., Schwelanus, M., and Collins, M. (2011). The COL5A1 Gene, Ultra-Marathon Running Performance, and Range of Motion. *International Journal Of Sports Physiology And Performance* 6, 485-496.
30. Abrahams, S., Posthumus, M., and Collins, M. (2014). A Polymorphism in a Functional Region of the COL5A1 Gene: Association With Ultraendurance-Running Performance and Joint Range of Motion. *International Journal Of Sports Physiology And Performance* 9, 583-590.
31. Morel, J., Combe, B., Francisco, J., and Bernard, J. (2001). Bone Mineral Density of 704 Amateur Sportsmen Involved in Different Physical Activities. *Osteoporosis International* 12, 152-157.
32. Hind, K., Truscott, J., and Evans, J. (2006). Low lumbar spine bone mineral density in both male and female endurance runners. *Bone* 39, 880-885.
33. Tam, N., Santos-Concejero, J., Tucker, R., Lamberts, R., and Micklesfield, L. (2018). Bone health in elite Kenyan runners. *Journal Of Sports Sciences* 36, 1-6.
34. Schroeder, T., Jensen, E., and Westendorf, J. (2005). Runx2: A master organizer of gene transcription in developing and maturing osteoblasts. *Birth Defects Research Part C: Embryo Today: Reviews* 75, 213-225.
35. Arumugam, B., Vishal, M., Shreya, S., Malavika, D., Rajpriya, V., He, Z., Partridge, N., and Selvamurugan, N. (2019). Parathyroid hormone-stimulation of Runx2 during osteoblast differentiation via the regulation of lnc-SUPT3H-1:16 (RUNX2-AS1:32) and miR-6797-5p. *Biochimie* 158, 43-52.
36. Aqeilan, R., Hassan, M., de Bruin, A., Hagan, J., Volinia, S., Palumbo, T., Hussain, S., Lee, S., Gaur, T., and Stein, G. et al. (2008). The WWOX Tumor Suppressor Is Essential for Postnatal Survival and Normal Bone Metabolism. *Journal Of Biological Chemistry* 283, 21629-21639.
37. Del Mare, S., Kurek, K., Stein, G., Lian, J., and Aqeilan, R. (2011). Role of the WWOX tumor suppressor gene in bone homeostasis and the pathogenesis of osteosarcoma. *American journal of cancer research* 1, 585.
38. Pregizer, S., Barski, A., Gersbach, C., García, A., and Frenkel, B. (2007). Identification of novel Runx2 targets in osteoblasts: Cell type-specific BMP-dependent regulation of Tram2. *Journal Of Cellular Biochemistry* 102, 1458-1471.
39. Chan, L., Webb, T., Chung, T., Meimaridou, E., Cooray, S., Guasti, L., Chapple, J., Egertova, M., Elphick, M., and Cheetham, M. et al. (2009). MRAP and MRAP2 are bidirectional regulators of the melanocortin receptor family. *Proceedings Of The National Academy Of Sciences* 106, 6146-6151.
40. Sebag, J., Zhang, C., Hinkle, P., Bradshaw, A., and Cone, R. (2013). Developmental Control of the Melanocortin-4 Receptor by MRAP2 Proteins in Zebrafish. *Science* 341, 278-281.
41. Liu, T., Elmquist, J., and Williams, K. (2013). Mrap2: An Accessory Protein Linked to Obesity. *Cell Metabolism* 18, 309-311.

42. Kang, S., Chiang, C., Palmer, C., Tayo, B., Lettre, G., Butler, J., Hackett, R., Adeyemo, A., Guiducci, C., and Berzins, I. et al. (2010). Genome-wide association of anthropometric traits in African- and African-derived populations. *Human Molecular Genetics* *19*, 2725-2738.
43. Yeo, G., Farooqi, I., Aminian, S., Halsall, D., Stanhope, R., and O'Rahilly, S. (1998). A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nature Genetics* *20*, 111-112.
44. Braun, T., Orwoll, B., Zhu, X., Levasseur, P., Szumowski, M., Nguyen, M., Buxsein, M., Klein, R., and Marks, D. (2012). Regulation of Lean Mass, Bone Mass, and Exercise Tolerance by the Central Melanocortin System. *Plos ONE* *7*, e42183.
45. Asai, M., Ramachandrapa, S., Joachim, M., Shen, Y., Zhang, R., Nuthalapati, N., Ramanathan, V., Strohlic, D., Ferket, P., and Linhart, K. et al. (2013). Loss of Function of the Melanocortin 2 Receptor Accessory Protein 2 Is Associated with Mammalian Obesity. *Science* *341*, 275-278.
46. Morgan, D., and Craib, M. (1992). Physiological aspects of running economy. *Medicine & Science In Sports & Exercise* *24*, 456-461.
47. Warren, G., Cureton, K., and Sparling, P. (1989). Does lung function limit performance in a 24-hour ultramarathon?. *Respiration Physiology* *78*, 253-263.
48. Pringle, E., Latin, R., and Berg, K. (2005). The Relationship Between 10 KM Running Performance And Pulmonary Function. *Journal of Exercise Physiology Online* *8*, 22-28.
49. Kirkton, S., Howlett, R., Gonzalez, N., Giuliano, P., Britton, S., Koch, L., Wagner, H., and Wagner, P. (2009). Continued artificial selection for running endurance in rats is associated with improved lung function. *Journal Of Applied Physiology* *106*, 1810-1818.
50. Schutte, N., Nederend, I., Hudziak, J., Bartels, M., and de Geus, E. (2016). Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiological Genomics* *48*, 210-219.
51. Miyamoto-Mikami, E., Zempo, H., Fuku, N., Kikuchi, N., Miyachi, M., and Murakami, H. (2017). Heritability estimates of endurance-related phenotypes: A systematic review and meta-analysis. *Scandinavian Journal Of Medicine & Science In Sports* *28*, 834-845.
52. Artigas, M., Loth, D., Wain, L., Gharib, S., Obeidat, M., Tang, W., Zhai, G., Zhao, J., Smith, A., and Huffman, J. et al. (2011). Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nature Genetics* *43*, 1082-1090.
53. Hancock, D., Artigas, M., Gharib, S., Henry, A., Manichaikul, A., Ramasamy, A., Loth, D., Imboden, M., Koch, B., and McArdle, W. et al. (2012). Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function. *Plos Genetics* *8*, e1003098.
54. Loth, D., Artigas, M., Gharib, S., Wain, L., Franceschini, N., Koch, B., Pottinger, T., Smith, A., Duan, Q., and Oldmeadow, C. et al. (2014). Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nature Genetics* *46*, 669-677.
55. Lutz, S., Cho, M., Young, K., Hersh, C., Castaldi, P., McDonald, M., Regan, E., Mattheisen, M., DeMeo, D., and Parker, M. et al. (2015). A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genetics* *16*.

56. Kichaev, G., Bhatia, G., Loh, P., Gazal, S., Burch, K., Freund, M., Schoech, A., Pasaniuc, B., and Price, A. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *The American Journal Of Human Genetics* *104*, 65-75.
57. Shrine, N., Guyatt, A., Erzurumluoglu, A., Jackson, V., Hobbs, B., Melbourne, C., Batini, C., Fawcett, K., Song, K., and Sakornsakolpat, P. et al. (2019). New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nature Genetics* *51*, 481-493.
58. Tanaka, K., Takeshima, N., Kato, T., Niihata, S., and Ueda, K. (1990). Critical determinants of endurance performance in middle-aged and elderly endurance runners with heterogeneous training habits. *European Journal Of Applied Physiology And Occupational Physiology* *59*, 443-449.
59. Gratze, G., Rudnicki, R., Urban, W., Mayer, H., Schlögl, A., and Skrabal, F. (2005). Hemodynamic and autonomic changes induced by Ironman: prediction of competition time by blood pressure variability. *Journal Of Applied Physiology* *99*, 1728-1735.
60. Karjalainen, J., Mäntysaari, M., Viitasalo, M., and Kujala, U. (1997). Left ventricular mass, geometry, and filling in endurance athletes: association with exercise blood pressure. *Journal Of Applied Physiology* *82*, 531-537.
61. Shellock, F., Morris, E., Deutsch, A., Mink, J., Kerr, R., and Boden, S. (1992). Hematopoietic bone marrow hyperplasia: high prevalence on MR images of the knee in asymptomatic marathon runners. *American Journal Of Roentgenology* *158*, 335-338.
62. Baker, J., De Lisio, M., and Parise, G. (2011). Endurance exercise training promotes medullary hematopoiesis. *The FASEB Journal* *25*, 4348-4357.
63. Marycz, K., Mierzejewska, K., Śmieszek, A., Suszynska, E., Malicka, I., Kucia, M., and Ratajczak, M. (2016). Endurance Exercise Mobilizes Developmentally Early Stem Cells into Peripheral Blood and Increases Their Number in Bone Marrow: Implications for Tissue Regeneration. *Stem Cells International* *2016*, 1-10.
64. Steiner, S., Niessner, A., Ziegler, S., Richter, B., Seidinger, D., Pleiner, J., Penka, M., Wolzt, M., Huber, K., and Wojta, J. et al. (2005). Endurance training increases the number of endothelial progenitor cells in patients with cardiovascular risk and coronary artery disease. *Atherosclerosis* *181*, 305-310.
65. Chinsomboon, J., Ruas, J., Gupta, R., Thom, R., Shoag, J., Rowe, G., Sawada, N., Raghuram, S., and Arany, Z. (2009). The transcriptional coactivator PGC-1 α mediates exercise-induced angiogenesis in skeletal muscle. *Proceedings Of The National Academy Of Sciences* *106*, 21401-21406.
66. Olfert, I., Howlett, R., Wagner, P., and Breen, E. (2010). Myocyte vascular endothelial growth factor is required for exercise-induced skeletal muscle angiogenesis. *American Journal Of Physiology-Regulatory, Integrative And Comparative Physiology* *299*, R1059-R1067.
67. Saltin, B., Henriksson, J., Nygaard, E., Andersen, P., and Jansson, E. (1977). Fiber Types And Metabolic Potentials Of Skeletal Muscles In Sedentary Man And Endurance Runners. *Annals Of The New York Academy Of Sciences* *301*, 3-29.

68. Jungers, K., Le Goff, C., Somerville, R., and Apte, S. (2005). Adamts9 is widely expressed during mouse embryo development. *Gene Expression Patterns* 5, 609-617.
69. Kern, C., Wessels, A., McGarity, J., Dixon, L., Alston, E., Argraves, W., Geeting, D., Nelson, C., Menick, D., and Apte, S. (2010). Reduced versican cleavage due to Adamts9 haploinsufficiency is associated with cardiac and aortic anomalies. *Matrix Biology* 29, 304-316.
70. Guilluy, C., Brégeon, J., Toumaniantz, G., Rolli-Derkinderen, M., Retailleau, K., Loufrani, L., Henrion, D., Scalbert, E., Bril, A., and Torres, R. et al. (2010). The Rho exchange factor Arhgef1 mediates the effects of angiotensin II on vascular tone and blood pressure. *Nature Medicine* 16, 183-190.
71. Carbone, M., Brégeon, J., Devos, N., Chadeuf, G., Blanchard, A., Azizi, M., Pacaud, P., Jeunemaître, X., and Loirand, G. (2015). Angiotensin II Activates the RhoA Exchange Factor Arhgef1 in Humans. *Hypertension* 65, 1273-1278.
72. Myerson, S., Hemingway, H., Budget, R., Martin, J., Humphries, S., and Montgomery, H. (1999). Human angiotensin I-converting enzyme gene and endurance performance. *Journal Of Applied Physiology* 87, 1313-1316.
73. Amir, O., Amir, R., Yamin, C., Attias, E., Eynon, N., Sagiv, M., Sagiv, M., and Meckel, Y. (2007). The ACE deletion allele is associated with Israeli elite endurance athletes. *Experimental Physiology* 92, 881-886.
74. Tobina, T., Michishita, R., Yamasawa, F., Zhang, B., Sasaki, H., Tanaka, H., Saku, K., and Kiyonaga, A. (2010). Association between the angiotensin I-converting enzyme gene insertion/deletion polymorphism and endurance running speed in Japanese runners. *The Journal Of Physiological Sciences* 60, 325-330
75. Ash, G., Scott, R., Deason, M., Dawson, T., Wolde, B., Bekele, Z., Teka, S., and Pitsiladis, Y. (2011). No Association between ACE Gene Variation and Endurance Athlete Status in Ethiopians. *Medicine & Science In Sports & Exercise* 43, 590-597.
76. Theis, J., Sharpe, K., Matsumoto, M., Chai, H., Nair, A., Theis, J., de Andrade, M., Wieben, E., Michels, V., and Olson, T. (2011). Homozygosity Mapping and Exome Sequencing Reveal GATAD1 Mutation in Autosomal Recessive Dilated Cardiomyopathy. *Circulation: Cardiovascular Genetics* 4, 585-594.
77. Yang, J., Shah, S., Olson, T., and Xu, X. (2016). Modeling GATAD1-Associated Dilated Cardiomyopathy in Adult Zebrafish. *Journal Of Cardiovascular Development And Disease* 3, 6.
78. Dulhunty, A., Gage, P., Curtis, S., Chelvanayagam, G., and Board, P. (2000). The Glutathione Transferase Structural Family Includes a Nuclear Chloride Channel and a Ryanodine Receptor Calcium Release Channel Modulator. *Journal Of Biological Chemistry* 276, 3319-3323.
79. Ragni, C., Diguët, N., Le Garrec, J., Novotova, M., Resende, T., Pop, S., Charon, N., Guillemot, L., Kitasato, L., and Badouel, C. et al. (2017). Amotl1 mediates sequestration of the Hippo effector Yap1 downstream of Fat4 to restrict heart growth. *Nature Communications* 8.

80. Zheng, Y., Vertuani, S., Nyström, S., Audebert, S., Meijer, I., Tegnebratt, T., Borg, J., Uhlén, P., Majumdar, A., and Holmgren, L. (2009). Angiomotin-Like Protein 1 Controls Endothelial Polarity and Junction Stability During Sprouting Angiogenesis. *Circulation Research* *105*, 260-270.
81. Choi, K., Choi, H., Lee, J., Im, S., Zhang, H., Jeong, Y., Park, J., Lee, I., Kim, Y., and Kwon, Y. (2016). The endothelial E3 ligase HECW2 promotes endothelial cell junctions by increasing AMOTL1 protein stability via K63-linked ubiquitination. *Cellular Signalling* *28*, 1642-1651.
82. Gazda, H., Zhong, R., Long, L., Niewiadomska, E., Lipton, J., Ploszynska, A., Zaucha, J., Vlachos, A., Atsidaftos, E., and Viskochil, D. et al. (2004). RNA and protein evidence for haplo-insufficiency in Diamond-Blackfan anaemia patients with RPS19 mutations. *British Journal Of Haematology* *127*, 105-113.
83. Campagnoli, M., Ramenghi, U., Armiraglio, M., Quarello, P., Garelli, E., Carando, A., Avondo, F., Pavesi, E., Fribourg, S., and Gleizes, P. et al. (2008). RPS19 mutations in patients with Diamond-Blackfan anemia. *Human Mutation* *29*, 911-920.
84. Flygare, J., Aspesi, A., Bailey, J., Miyake, K., Caffrey, J., Karlsson, S., and Ellis, S. (2006). Human RPS19, the gene mutated in Diamond-Blackfan anemia, encodes a ribosomal protein required for the maturation of 40S ribosomal subunits. *Blood* *109*, 980-986.
85. Idol, R., Robledo, S., Du, H., Crimmins, D., Wilson, D., Ladenson, J., Bessler, M., and Mason, P. (2007). Cells depleted for RPS19, a protein associated with Diamond Blackfan Anemia, show defects in 18S ribosomal RNA synthesis and small ribosomal subunit production. *Blood Cells, Molecules, And Diseases* *39*, 35-43.
86. Prommer, N., Thoma, S., Quecke, L., Gutekunst, T., Volzke, C., Wachsmuth, N., Niess, A., and Schmidt, W. (2010). Total Hemoglobin Mass and Blood Volume of Elite Kenyan Runners. *Medicine & Science In Sports & Exercise* *42*, 791-797.
87. Holloszy, J., and Booth, F. (1976). Biochemical Adaptations to Endurance Exercise in Muscle. *Annual Review Of Physiology* *38*, 273-291.
88. Wilber, R., and Moffatt, R. (1992). Influence of Carbohydrate Ingestion on Blood Glucose and Performance in Runners. *International Journal Of Sport Nutrition* *2*, 317-327.
89. Fudge, B., Westerterp, K., Kiplamai, F., Onywera, V., Boit, M., Kayser, B., and Pitsiladis, Y. (2006). Evidence of negative energy balance using doubly labelled water in elite Kenyan endurance runners prior to competition. *British Journal Of Nutrition* *95*, 59-66.
90. Fan, W., Waizenegger, W., Lin, C., Sorrentino, V., He, M., Wall, C., Li, H., Liddle, C., Yu, R., and Atkins, A. et al. (2017). PPAR δ Promotes Running Endurance by Preserving Glucose. *Cell Metabolism* *25*, 1186-1193.e4.
91. Polgar, N., and Fogelgren, B. (2017). Regulation of Cell Polarity by Exocyst-Mediated Trafficking. *Cold Spring Harbor Perspectives In Biology* *10*, a031401.

92. Inoue, M., Chiang, S., Chang, L., Chen, X., and Saltiel, A. (2006). Compartmentalization of the Exocyst Complex in Lipid Rafts Controls Glut4 Vesicle Tethering. *Molecular Biology Of The Cell* 17, 2303-2311.
93. Laramie, J., Wilk, J., Williamson, S., Nagle, M., Latourelle, J., Tobin, J., Province, M., Borecki, I., and Myers, R. (2008). Polymorphisms near EXOC4 and LRGUK on chromosome 7q32 are associated with Type 2 Diabetes and fasting glucose; The NHLBI Family Heart Study. *BMC Medical Genetics* 9.
94. Fujimoto, B., Young, M., Carter, L., Pang, A., Corley, M., Fogelgren, B., and Polgar, N. (2019). The exocyst complex regulates insulin-stimulated glucose uptake of skeletal muscle cells. *American Journal Of Physiology-Endocrinology And Metabolism* 317, E957-E972.
95. Sano, H., Peck, G., Blachon, S., and Lienhard, G. (2015). A potential link between insulin signaling and GLUT4 translocation: Association of Rab10-GTP with the exocyst subunit Exoc6/6b. *Biochemical And Biophysical Research Communications* 465, 601-605.
96. Abu-Remaileh, M., and Aqeilan, R. (2014). Tumor suppressor WWOX regulates glucose metabolism via HIF1 α modulation. *Cell Death & Differentiation* 21, 1805-1814.
97. Abu-Remaileh, M., Seewaldt, V., and Aqeilan, R. (2014). WWOX loss activates aerobic glycolysis. *Molecular & Cellular Oncology* 2, e965640.
98. Choo, A., O'Keefe, L., Lee, C., Gregory, S., Shaukat, Z., Colella, A., Lee, K., Denton, D., and Richards, R. (2015). Tumor suppressor WWOX moderates the mitochondrial respiratory complex. *Genes, Chromosomes And Cancer* 54, 745-761.
99. Döring, F., Onur, S., Fischer, A., Boulay, M., Pérusse, L., Rankinen, T., Rauramaa, R., Wolfarth, B., and Bouchard, C. (2010). A common haplotype and the Pro582Ser polymorphism of the hypoxia-inducible factor-1 α (HIF1A) gene in elite endurance athletes. *Journal Of Applied Physiology* 108, 1497-1500.
100. Albert, J., Yerges-Armstrong, L., Horenstein, R., Pollin, T., Sreenivasan, U., Chai, S., Blaner, W., Snitker, S., O'Connell, J., and Gong, D. et al. (2014). Null Mutation in Hormone-Sensitive Lipase Gene and Risk of Type 2 Diabetes. *New England Journal Of Medicine* 370, 2307-2315.
101. Zechner, R., and Langin, D. (2014). Hormone-Sensitive Lipase Deficiency in Humans. *Cell Metabolism* 20, 199-201.
102. Zolotov, S., Xing, C., Mahamid, R., Shalata, A., Sheikh-Ahmad, M., and Garg, A. (2016). Homozygous LIPE mutation in siblings with multiple symmetric lipomatosis, partial lipodystrophy, and myopathy. *American Journal Of Medical Genetics Part A* 173, 190-194.
103. Cheng, H., Lederer, W., and Cannell, M. (1993). Calcium sparks: elementary events underlying excitation-contraction coupling in heart muscle. *Science* 262, 740-744.
104. Prímola-Gomes, T., Campos, L., Lauton-Santos, S., Balthazar, C., Guatimosim, S., Capettini, L., Lemos, V., Coimbra, C., Soares, D., and Carneiro-Júnior, M. et al. (2009). Exercise capacity is related to calcium transients in ventricular cardiomyocytes. *Journal Of Applied Physiology* 107, 593-598.

105. Wisløff, U., Loennechen, J., Falck, G., Beisvag, V., Currie, S., Smith, G., and Ellingsen, Ø. (2001). Increased contractility and calcium sensitivity in cardiac myocytes isolated from endurance trained rats. *Cardiovascular research* *50*, 495-508.
106. Moir, H., Kemp, R., Folkerts, D., Spendiff, O., Pavlidis, C., and Opara, E. (2019). Genes and Elite Marathon Running Performance: A Systematic Review. *Journal of sports science & medicine* *18*, 559-568.
107. Rankinen, T., Fuku, N., Wolfarth, B., Wang, G., Sarzynski, M., Alexeev, D., Ahmetov, I., Boulay, M., Cieszczyk, P., and Eynon, N. et al. (2016). No Evidence of a Common DNA Variant Profile Specific to World Class Endurance Athletes. *PLOS ONE* *11*, e0147330.
108. Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C., Prado-Martinez, J., Bouman, H., Abascal, F., Haber, M., and Tachmazidou, I. et al. (2019). Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* *179*, 984-1002.e36.
109. Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z., Pool, J., Xu, X., Jiang, H., Vinckenbosch, N., and Korneliussen, T. et al. (2010). Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science* *329*, 75-78.
110. Jacovas, V., Couto-Silva, C., Nunes, K., Lemes, R., de Oliveira, M., Salzano, F., Bortolini, M., and Hünemeier, T. (2018). Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. *Scientific Reports* *8*, 12733.
111. Malaspinas, A., Westaway, M., Muller, C., Sousa, V., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J., and Crawford, J. et al. (2016). A genomic history of Aboriginal Australia. *Nature* *538*, 207-214.
112. 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature* *526*, 68.
113. Amorim, C., Nunes, K., Meyer, D., Comas, D., Bortolini, M., Salzano, F., and Hünemeier, T. (2017). Genetic signature of natural selection in first Americans. *Proceedings Of The National Academy Of Sciences* *114*, 2195-2199.
114. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., Sklar, P., de Bakker, P., and Daly, M. et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal Of Human Genetics* *81*, 559-575.
115. R Core Team (2018). R: A language and environment for statistical computing. R. Foundation for Statistical Computing *2018*.
116. Zerbino, D., Achuthan, P., Akanni, W., Amode, M., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., and Girón, C. et al. (2017). Ensembl 2018. *Nucleic Acids Research* *46*, D754-D761.
117. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* *8*.

118. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., and Morales, J. et al. (2016). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research* 45, D896-D901.
119. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739-1740.

Referências (Introdução Geral)

1. Carrier, D., Kapoor, A., Kimura, T., Nickels, M., Scott, E., So, J., and Trinkaus, E. (1984). The Energetic Paradox of Human Running and Hominid Evolution [and Comments and Reply]. *Current Anthropology* 25, 483-495.
2. Bramble, D., and Lieberman, D. (2004). Endurance running and the evolution of Homo. *Nature* 432, 345-352.
3. Lieberman, D., and Bramble, D. (2007). The Evolution of Marathon Running. *Sports Medicine* 37, 288-290.
4. Liebenberg, L. (2006). Persistence Hunting by Modern Hunter-Gatherers. *Current Anthropology* 47, 1017-1026.
5. Aiello, L., and Wheeler, P. (1995). The Expensive-Tissue Hypothesis: The Brain and the Digestive System in Human and Primate Evolution. *Current Anthropology* 36, 199-221.
6. Lieberman, D. E., Bramble, D. M., Raichlen, D. A., & Shea, J. J. (2009). Brains, brawn, and the evolution of human endurance running capabilities. In *The first humans—origin and early evolution of the genus homo*, ed (Springer, Dordrecht), pp. 77-92.
7. Pickering, T., and Bunn, H. (2007). The endurance running hypothesis and hunting and scavenging in savanna-woodlands. *Journal Of Human Evolution* 53, 434-438.
8. Lieberman, D., Bramble, D., Raichlen, D., and Shea, J. (2007). The evolution of endurance running and the tyranny of ethnography: A reply to Pickering and Bunn (2007). *Journal Of Human Evolution* 53, 439-442.
9. Predel, H. (2014). Marathon run: cardiovascular adaptation and cardiovascular risk. *European Heart Journal* 35, 3091-3098.
10. Grogan, R. (1981). Run, Pheidippides, Run! The story of the Battle of Marathon. *British Journal Of Sports Medicine* 15, 186-189.
11. Onywera, V., Scott, R., Boit, M., and Pitsiladis, Y. (2006). Demographic characteristics of elite Kenyan endurance runners. *Journal Of Sports Sciences* 24, 415-422.
12. Scott, R., Georgiades, E., Wilson, R., Goodwin, W., Wolde, B., and Pitsiladis, Y. (2003). Demographic Characteristics of Elite Ethiopian Endurance Runners. *Medicine & Science In Sports & Exercise* 35, 1727-1732.
13. Kenya National Bureau of Statistics (2010). Population and Household Distribution by Social Economic Characteristics. In *The 2009 Kenya Population and Housing Census*, pp. 397-398.
14. Githiora, C. (2002). Sheng: Peer language, Swahili dialect or emerging Creole?. *Journal Of African Cultural Studies* 15, 159-181.

15. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M., and Choudhury, A. et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature* *517*, 327-332.
16. Gouveia, M., Bergen, A., Borda, V., Nunes, K., Leal, T., Ogowang, M., Yeboah, E., Mensah, J., Kinyera, T., and Otim, I. et al. (2019). Genetic signatures of gene flow and malaria-driven natural selection in sub-Saharan populations of the "endemic Burkitt Lymphoma belt." *PLOS Genetics* *15*, e1008027.
17. Central Statistical Agency (2010). 2007 Population and Housing Census of Ethiopia.
18. Wilber, R., and Pitsiladis, Y. (2012). Kenyan and Ethiopian Distance Runners: What Makes Them so Good?. *International Journal Of Sports Physiology And Performance* *7*, 92-102.
19. Tucker, R., Santos-Concejero, J., and Collins, M. (2013). The genetic basis for elite running performance. *British Journal Of Sports Medicine* *47*, 545-549.
20. Scott, R., Moran, C., Wilson, R., Onywera, V., Boit, M., Goodwin, W., Gohlke, P., Payne, J., Montgomery, H., and Pitsiladis, Y. (2005). No association between Angiotensin Converting Enzyme (ACE) gene variation and endurance athlete status in Kenyans. *Comparative Biochemistry And Physiology Part A: Molecular & Integrative Physiology* *141*, 169-175.
21. Yang, N., Macarthur, D., Wolde, B., Onywera, V., Boit, M., Lau, S., Wilson, R., Scott, R., Pitsiladis, Y., and North, K. (2007). The ACTN3 R577X Polymorphism in East and West African Athletes. *Medicine & Science In Sports & Exercise* *39*, 1985-1988.
22. Timmons, J., Larsson, O., Jansson, E., Fischer, H., Gustafsson, T., Greenhaff, P., Ridder, J., Rachman, J., Peyrard-Janvid, M., and Wahlestedt, C. et al. (2005). Human muscle gene expression responses to endurance training provide a novel perspective on Duchenne muscular dystrophy. *The FASEB Journal* *19*, 750-760.