PNAS
www.pnas.org

Supplementary Information for

Association of *EGLN1* gene with high aerobic capacity of Peruvian Quechua at high altitude

Tom D. Brutsaert[1*], Melissa Kiyamu[2], Gianpietro Elias Revollendo[1], Jenna L. Isherwood[3], Frank S. Lee[4], Maria Rivera-Ch.[2], Fabiola Leon-Velarde[2], Sudipta Ghosh[5], Abigail W. Bigham[3].

Corresponding author

Tom D Brutsaert
Email: tdbrutsa@syr.edu

This PDF file includes:

Supplementary Text, Materials and Methods
Figures S1 to S3
Tables S1 to S9
Dataset S1

www.pnas.org/cgi/doi/10.1073/pnas.1906171116

Supplementary Information Text

MATERIALS AND METHODS:

Study design and sub-group structure of the genetic sample.

The overall sampling structure comprised four study sub-groups defined by ancestry (Quechua vs non-Hispanic White), by altitude of birth (sea level vs. above 3,000 meters), and in the case of one group, by migration status (highland migrants to sea-level). These groups included: (1) Quechua-High Altitude Residents (HAR, n=195) who were recruited and tested in Cerro de Pasco, Peru at 4,338 m above sea-level; (2) Quechua-Migrants (M, n=111) who were born and raised above 3,000 m but migrated permanently to sea-level (Lima) at some point during their lives; (3) Quechua-Born at Sea level (BSL, n=123), who were recruited and tested in Lima, Peru at sea-level; and (4) Syracuse non-Hispanic whites (n=94) who were recruited from undergraduate and graduate courses at Syracuse University and tested in Syracuse, NY, USA, at ~140 m. All participants were classified as Quechua-Andean or non-Hispanic white based on self-identified ancestry that was later verified through PCA and population structure analysis performed on genome-wide SNP data generated using the Affymetrix (Santa Clara, CA) Axiom Biobanking Array. Roughly equal numbers of healthy males and non-pregnant, non-lactating females were recruited between the ages of 18-35. For the Quechua-HAR sample, individuals involved in mining activities or those who presented with chronic mountain sickness were excluded from the study. For the Quechua-M, the mean age at migration to sea-level was about 15 years, but there was a wide range with the earliest migrants arriving to Lima at several months of age and the latest migrants arriving as adults. All Quechua-M had been at sea-level for at least two months prior to study participation. Many of the Syracuse sample were recruited from the Exercise Science Department at Syracuse University, and these participants were generally involved in sport and aerobically fit. In contrast, many of the Andean participants described themselves as not physically active or involved in recreational sport. All participants provided written informed consent in their native language for study procedures approved by the *Syracuse University Office of Research Integrity and Protections*, and the *Research Ethics Committee of the Universidad Peruana Cayetano Heredia (UPCH), Lima, Peru.* The study also was approved by *The University of Michigan Institutional Review Board.*

General subject characteristics.

Standard anthropometry was used to obtain height, weight, and skin-fold measurements. Body fat percentage (% Body Fat) was calculated from biceps, triceps, subscapular, and suprailiac skinfolds using the equations of Durnin and Wormersley (1). Hemoglobin concentration [Hb] was measured on venous blood obtained by forearm venipuncture using a Hemocue blood hemoglobin analyzer (Angelhom, Sweden). Many additional physiological traits were measured on study participants, but only the exercise testing is described in detail as other traits were not relevant to the analyses presented in this paper.

Exercise Testing.

Exercise testing was conducted in Cerro de Pasco, Peru, (for Quechua-HAR), at UPCH in Lima, Peru (for Quechua-M and Quechua-BSL), or at Syracuse University for Syracuse participants. Our goal was to measure the VO_2max in hypoxia. Thus, Quechua-HAR were tested under ambient conditions in their native high-altitude environment (4,338 m), while the three sea-level resident groups (Quechua-M, Quechua-BSL, and Syracuse) were tested under simulated altitude conditions at sea-level (normobaric hypoxia). To simulate the altitude of Cerro de Pasco, Peru the fractional concentration of O_2 (F_iO_2) was lowered to approximately 0.126 using a Hypoxico Hypoxic generator (New York, NY). The generator delivered a constant flow of air to a large reservoir bag, which was then delivered to the participant via the in-port of a 1-way low resistance Hans Rudolph breathing valve.

VO_2max was measured on a mechanically braked Monarch (Langley, WA) 818e research ergometer (Peru participants) or on a Lode (Groningen, Netherlands) Excaliber X (Syracuse participants). After a 3-minute resting period, participants started cycling at 60 watts (W) external work. They remained at this work load for nine minutes as a warm-up/stabilizing period, and thereafter resistance was incremented by 30 W every three minutes until subject volitional fatigue. Participants were given verbal encouragement, and VO_2max was defined as the highest level of oxygen consumption averaged over the final minute of the test concomitant with a respiratory exchange ratio (RER) greater than 1.10 and a maximal heart rate within 10% of the age-predicted maximum.

During VO_2 testing, expired ventilation (VE, $l \cdot min^{-1}$) as well as the fractional concentrations of O_2 and CO_2 in expired air were processed by a Parvo Medics True Max metabolic measuring system (Sandy, Utah) to produce 1-minute interval calculations of VO_2. For altitude simulation testing a precise measure of the F_iO_2 was made just prior to the beginning of VO_2 measurement as this value typically fluctuated by several tenths of a percent. This final value was used in the calculation of VO_2. Gas analyzers were calibrated with standard gases before each exercise test. A heated pneumotach was used to measure the VE. It was calibrated prior to each test with a 3-liter calibration syringe. Heart rate (HR) was continuously monitored via telemetry (Polar Electric, Oy, Sweden) interfaced with the metabolic measuring system.

Determination of a final sample for genetic analysis.

A total of n=748 participants were recruited for exercise testing, with Quechua-HAR=298, Quechua-M=150, Quechua-BSL=152, and Syracuse=148. Given the large role of participant motivation during maximal exercise testing, participants were only retained in the final sample if they met objective criteria for a true VO_2max i.e., a respiratory exchange ratio greater than or equal to 1.1 and a maximal heart rate within 10% of the age-predicted maximum. Applying these criteria resulted in the exclusion of about 20% of all participants to yield a sample of n=570. Having only 80% of participants reach a true VO_2max when testing non-athlete populations is typical, and the success rate was similar between Quechua and Syracuse participants. Of the remaining "true VO_2max" participants, 47 additional samples were excluded to yield a final genetic sample of n=523. These included 19 Quechua participants who were excluded after later genetic analysis revealed that they were third degree relatives or higher with another participant (we retained the relative with the higher genotyping call rate) (2); eight samples who were excluded as they failed best practices guidelines for the Affymetrix Axiom Biobanking Array (six Quechua and two Syracuse); three Syracuse samples who did not provide DNA; and 17 Syracuse samples did not have available DNA at the time of genotyping.

Variant Genotyping.

Microarray genotype data were generated using the Affymetrix (Santa Clara, CA) Axiom Biobanking Array featuring approximately 610,000 markers. All DNA available at the time of genotyping was assayed. The Biobanking Array contains 29 markers in and around (50KB upstream and downstream) *EGLN1.* Of the 29 *EGLN1* markers analyzed, 14 were monomorphic in our sample and five had minor allele frequencies (MAFs) less than 1% and thus were removed from association analysis (SI Appendix, Table S8). Four of the remaining *EGLN1* markers did not meet Affy best practices for QC and were also removed from association analysis. In addition, we manually genotyped two *EGLN1* SNPs (rs479200, rs480902) using PCR and restriction enzyme digestion and tetra-primer amplification refractory mutation system-PCR (ARMS-PCR), respectively, in our final sample of 523 participants. This left a final selection of 8 *EGLN1* SNPs for genetic analysis. The two manually genotyped SNPs were selected based on the analysis of whole genome sequence data obtained from 10 high altitude-adapted Andeans (3). These SNPs exhibited substantial differences in MAF compared to Mexican controls from the 1,000 Genomes Phase 3 (Table 2) and resided in transcriptionally active regions according to ENCODE data obtained from the UCSC Genome Browser. SNPs were phased using FastPHASE (4). Haploview was used for r^2 calculations of LD (5). In our replication cohort of 67 Peruvian Quechua high-

altitude residents, the most significant SNP identified in our association models (rs1769793) was manually genotyped using PCR and restriction enzyme digestion. Primers and restriction enzymes for these three SNPs are provided in SI Appendix, Table S9.

Genome-wide association analysis.

For GWAS, we tested 215,512 autosomal genomic variants for associations with VO_2Max. Each of these variants passed initial QC filtering of the full array data (628,679 variants) including genotyping rate > 95%, minor allele frequency (MAF) > 0.05, and Affy best practices QC. SNP associations with VO_2max were performed using standard linear regression in Plink version 1.9 (https://www.cog-genomics.org/plink2/). Genome-wide significance was assessed by applying the false discovery rate of Benjamini and Hochberg as well as genomic control. Sex, group, age, and height were included as covariates. In the Andes, a particular concern is admixture (population stratification), which can lead to spurious genotype-phenotype association i.e., false positives (6). We controlled for stratification by introducing into all statistical models the first principle component (PC) of a principal component analysis (PCA) performed using genotype data from the Affymetrix Axiom Biobanking array. Peruvian Quechua and Syracuse data were combined with publicly available genotype data from three HapMap populations including 60 Yorubans from Ibadan, Nigeria, 90 East Asians including 45 Han Chinese from Beijing and 45 Japanese from Tokyo, and 60 CEPH Europeans with northern and western European ancestry. SNPs with call rates greater than 95% and with r^2 values below 0.8 were retained in the analysis (n = 383,916). PCA was performed in Plink version 1.9 (7). The first three PCA plots are provided in SI Appendix, Fig. S1.

A priori ANCOVA.

SNP association testing was performed using the GLM procedure of IBM SPSS Statistics Software, version 23.0. VO_2max is a multifactorial complex phenotype. Before testing for association with *EGLN1* SNPs we identified the major covariates and covariate interactions that explained the majority of the variance in the absolute VO_2max ($l \cdot min^{-1}$). These included sex (i.e., VO_2max is generally higher in males than females), age (i.e., VO_2max decreases with age), weight (i.e., absolute VO_2max is larger in larger individuals), and the sub-group designation. [Hb] was not associated with VO_2max (controlling for sex), and thus not included as a covariate in association models. An alternative method to control for body-weight is to express VO_2max as a ratio standard i.e., the relative VO_2max ($ml \cdot min^{-1} \cdot kg^{-1}$). However, ratio standards are problematic as they can introduce bias and so we controlled for size/weight statistically, an approach that is favored by most methodologists (8). The study-group designation was used as a covariate to control for differences between study groups related to acclimatization status, developmental exposure to high altitude, and differences in the level of physical activity between groups. In the Andes, a particular concern is admixture (population stratification), which can lead to spurious genotype-phenotype association i.e., false positives (6). We controlled for stratification by introducing into all statistical models the first five principle components (PCs) of a PCA performed using genotype data (n= 383,916 markers) from the Affymetrix Axiom Biobanking array in Plink version 1.9 (7). We then observed the change (if any) on the SNP model coefficients.

In model building, we used the conventional p-value cutoff of $p<0.05$ for covariate and interaction effects. Once a final model was determined, including interactions, a specific *EGLN1* SNP marker was introduced into the model. For SNP associations, we applied a Bonferroni correction for multiple testing with a p-value cutoff of $p<0.004$ ($\alpha= 0.05$, 12 tests). If the main SNP effect was significant at $p<0.004$, we also examined interactions with other factors in the model and retained the interaction effect using the conventional p-value cutoff of $p<0.05$. Finally, after detecting a significant *EGLN1* SNP association, we corrected for the possibility of population stratification by introducing the first five principle components from the PCA into the model and observing the change (if any) on the model coefficients.

1000 Genomes comparison populations.

Comparison data from the 1000 Genomes Phase 3 were used to construct Fig. 1b and d, and supplemental Fig 3 (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/). The Comparison populations are GBR, British in England and Scotland; ACB, African Caribbeans in Barbados; ASW, Americans of African ancestry in southwestern USA; BEB, Bengali from Bangladesh; CDX, Chinese Dai in Xishuangbanna, China; CEU, Utah residents with Northern and Western European ancestry; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese; CLM, Colombians from Medellin, Colombia; ESN, Esan in Nigeria; FIN, Finnish in Finland; GWD, Gambian; GIH, Gujarati Indian from Houston, Texas, USA; IBS, Iberian population in Spain; ITU, Indian Telugu from the UK; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; MXL, Mexican ancestry from Los Angeles, USA; PEL, Peruvians from Lima, Peru; PJL, Punjabi from Lahore, Pakistan; PUR, Puerto Ricans from Puerto Rico; STU, Sri Lankan Tamil; TSI, Toscani in Italy; YRI, Yoruba in Ibadan, Nigeria.

The replication sample.

We replicated the association results for rs1769793 using data from a previously recruited Quechua cohort from Cerro de Pasco, Peru. These data were collected in 2001-2 from 67 male and female Quechua by the same investigators using the same exercise testing equipment and protocol (9-11). In general, the participants from the earlier study were more physically active and more were sampled from nearby surrounding rural areas rather than directly from the city of Cerro de Pasco.

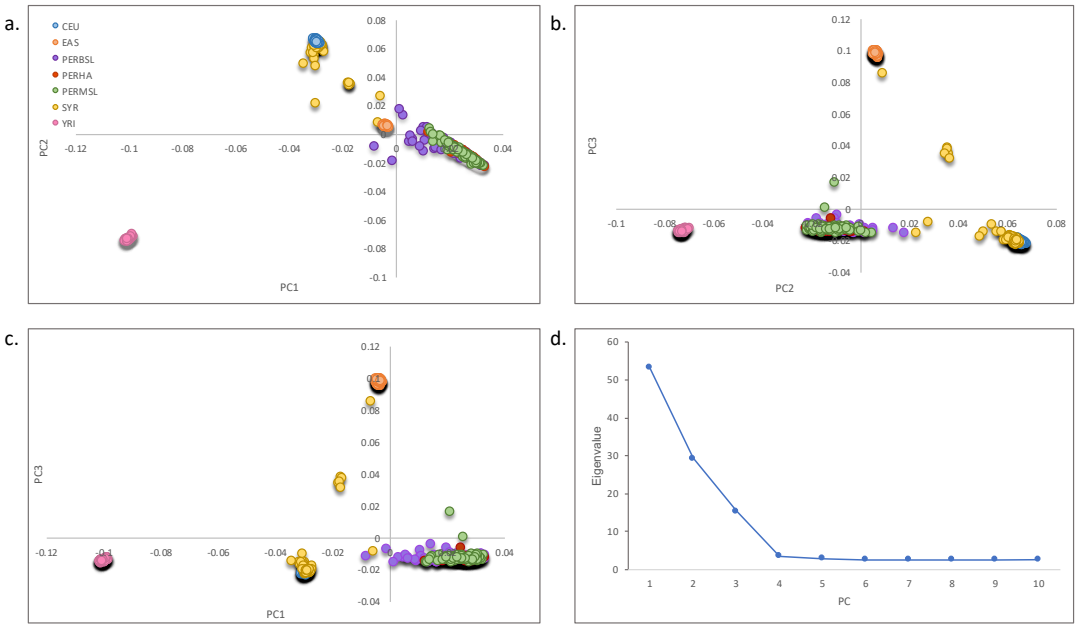

Fig. S1. The first three principal components (PCs) calculated for 523 Peruvian Quechua (PERBSL, PERHA, and PERMSL) and Syracuse (SYR) participants recruited as part of this effort as well as CEPH Europeans (CEU), East Asians including Han Chinese from Beijing and Japanese from Tokyo (EAS), and Yorubans from Ibadan Nigeria (YRI). (a) PC1 versus PC2.(b) PC2 versus PC3. (c) PC1 versus PC3. PCA was calculated in Plink v.1.9 using 383,916 autosomal SNPS genotyped on the Affymetrix biobanking array. SNPs with call rates < 95%, r^2>0.8, those that failed Affy best practices QC, and monomorphic SNPs were excluded from the analysis. Population labels are shown in (a) apply to panels (b) and (c). (d) Scree Plot depicting the Eigenvalues for PCs one through ten.

.

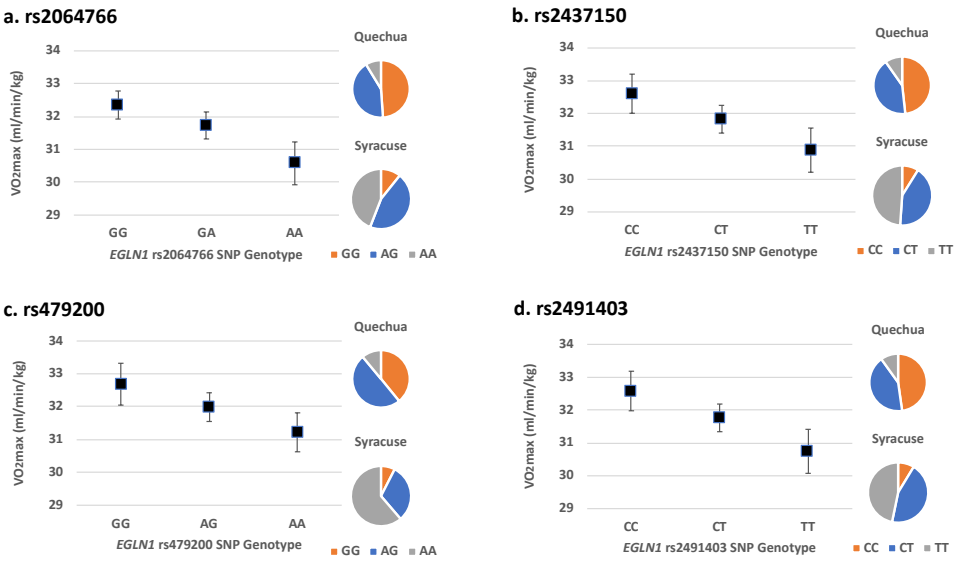

Fig. S2. Marginal mean values of VO_2max ($ml \cdot min^{-1} \cdot kg^{-1}$) and genotype frequencies for Quechua and Syracuse samples for four additional *EGLN1* SNP markers that were associated with VO_2max (a-d). The high VO_2max genotype (TT) is given in orange, the heterozygote genotype (CT) in blue, and the low VO_2max genotype (CC) in grey. The four SNPs shown in panels a-d show similar patterns of association as that described for rs1769793 in Fig. 1. That is, allele frequency differences between Quechua and Syracuse samples were similar to that for rs1769793, and the high frequency genotype in Quechua was always the high VO_2max genotypic category. Error bars are standard error of the mean (SEM).

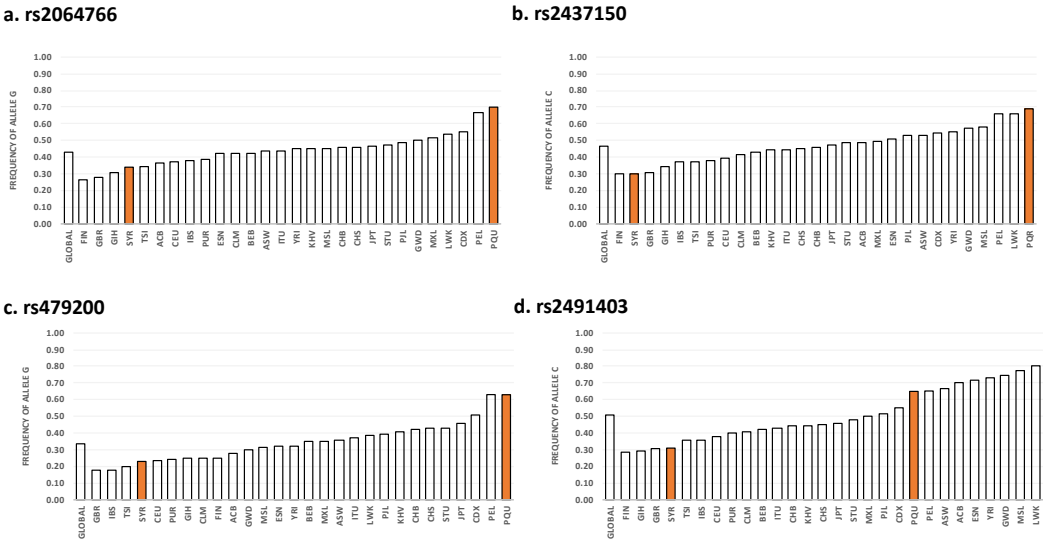

Fig. S3. Worldwide frequency of the putative adaptive allele at four additional *EGLN1* SNP loci (panels a-d) in Peruvian Quechua (PQU) and Syracuse (SYR) samples compared to allele frequency data from the 1000-genomes browser. At three of these loci (panels a, b, and c) Quechua had the highest frequency worldwide for the putative adaptive allele. SNPs represented in panels **a-d** correspond to the same panels in Supplemental Fig. 3. Allele frequency data come from the 1000-genomes browser. Global is the global mean frequency of the putative adaptive allele. Comparison populations are GBR, British in England and Scotland; ACB, African Caribbeans in Barbados; ASW, Americans of African ancestry in southwestern USA; BEB, Bengali from Bangladesh; CDX, Chinese Dai in Xishuangbanna, China; CEU, Utah residents with Northern and Western European ancestry; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese; CLM, Colombians from Medellin, Colombia; ESN, Esan in Nigeria; FIN, Finnish in Finland; GWD, Gambian; GIH, Gujarati Indian from Houston, Texas, USA; IBS, Iberian population in Spain; ITU, Indian Telugu from the UK; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; MXL, Mexican ancestry from Los Angeles, USA; PEL, Peruvians from Lima, Peru; PJL, Punjabi from Lahore, Pakistan; PUR, Puerto Ricans from Puerto Rico; STU, Sri Lankan Tamil; TSI, Toscani in Italy; YRI, Yoruba in Ibadan, Nigeria.
<insert page break here>

Table S1. Sample characteristics by sub-group.

	Born at high altitude						**Born at sea level**					
	Quechua-HAR **n=195 (m=106/f=89)**			***Quechua-M*** **n=111 (m=58/f=53)**			***Quechua-BSL*** **n=123 (m=62/f=61)**			***Syracuse*** **n=94 (m=48/f=46)**		
	Mean		S.E.	Mean		S.E.	Mean		S.E.	Mean		S.E.
Age, yrs	24.0	±	0.3 [B,D]	25.5	±	0.5 [A,D]	24.5	±	0.4 [D]	21.8	±	0.4 [A,B,C]
Mean age at migration to sea level	N/A			15.6	±	0.6	N/A			N/A		
Weight, kg	59.0	±	0.6 [C,D]	61.8	±	0.9 [C,D]	66.6	±	1.0 [A,B]	68.8	±	1.1 [A,B]
Ht, cm	158.2	±	0.6 [C,D]	158.5	±	0.7 [C,D]	161.5	±	0.8 [A,B,D]	172.2	±	0.9 [A,B,C]
Hemoglobin, g/dl	17.9	±	0.1 [B,C,D]	13.9	±	0.1 [A]	13.9	±	0.1 [A]	13.7	±	0.1 [A]
% Body Fat	23.9	±	0.6 [C]	26.0	±	0.7 [D]	27.5	±	0.6 [A,D]	23.1	±	0.8 [B,C]
VO_2max, l/min	2.02	±	0.04 [C,D]	1.92	±	0.04 [D]	1.86	±	0.04 [A,D]	2.42	±	0.06[A,B,C]
VO_2max, ml/kg/min	34.2	±	0.5 [B,C]	31.0	±	0.6 [A,C,D]	28.0	±	0.5 [B,D]	35.0	±	0.6 [B,C]

[A]Significantly different from Quechua-HAR, P<0.05.
[B]Significantly different from Quechua-M, P<0.05.
[C]Significantly different from Quechua-BSL, P<0.05.
[D]Significantly different from Syracuse, P<0.05.

Table S2. ANCOVA model showing association of *EGLN1* SNP (rs1769793) with dependent variable VO_2max in hypoxia.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	R^2
Corrected Model	125.6	22	5.71	70.9	<0.00001	0.762
Intercept	0.5	1	0.49	6.1	0.014	0.012
age	0.3	1	0.31	3.8	0.051	0.008
Weight	10.1	1	10.07	125.0	<0.00001	0.204
Sex	36.9	1	36.87	457.8	<0.00001	0.484
Study sub-group	5.2	3	1.73	21.4	<0.00001	0.116
PC1	0.1	1	0.14	1.7	0.196	0.003
PC2	0.1	1	0.08	1.0	0.329	0.002
PC3	0.1	1	0.12	1.5	0.224	0.003
PC4	0.1	1	0.15	1.8	0.176	0.004
PC5	0.0	1	0.00	0.0	0.849	0
Sex*study sub-group	1.2	3	0.39	4.9	0.196	0.029
***EGLN1* SNP rs176973**	**1.4**	**2**	**0.68**	**8.5**	**0.000241**	**0.034**
rs1769793*group interaction	**1.2**	**6**	**0.20**	**2.4**	**0.024**	**0.029**
Error	39.3	488	0.08			
Total	2273.6	511				
Corrected Total	164.9	510				
Model R-Squared = .762						

*PCs are principle components used to assess admixture in the sample and control for stratification, as described in the text.
Inclusion/exclusion of PCs had no qualitative effect on SNP associations, but are presented in the complete model for theoretical reasons.

Table S3. ANCOVA model showing association of *EGLN*1 SNP (rs2064766) with dependent variable VO_2max in hypoxia

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	R^2
Corrected Model	125.9	16	7.87	93.8	P<0.001	0.748
Intercept	0.7	1	0.74	8.9	0.003	0.017
age	0.3	1	0.34	4.1	0.044	0.008
Weight	10.5	1	10.46	124.7	P<0.001	0.198
Sex	38.1	1	38.14	454.5	P<0.001	0.474
Study sub-group	5.4	3	1.79	21.4	P<0.001	0.113
PC1	0.3	1	0.29	3.5	0.063	0.007
PC2	0.2	1	0.19	2.3	0.133	0.004
PC3	0.3	1	0.34	4.1	0.044	0.008
PC4	0.3	1	0.32	3.8	0.051	0.008
PC5	0.0	1	0.02	0.3	0.605	0.001
Sex*study sub-group	1.1	3	0.38	4.5	0.004	0.026
***EGLN1* SNP rs2064766**	**0.7**	**2**	**0.36**	**4.3**	**0.014**	**0.017**
Error	42.4	505	0.08			
Total	2333.4	522				
Corrected Total	168.3	521				
Model R-Squared = .748						

*PCs are principle components used to assess admixture in the sample and control for stratification, as described in the text.
Inclusion/exclusion of PCs had no qualitative effect on SNP associations, but are presented in the complete model for theoretical reasons.

Table S4. ANCOVA model showing association of *EGLN1* SNP (rs2437150) with dependent variable VO_2max in hypoxia

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	R^2
Corrected Model	126.6	22	5.76	70.9	P<0.001	0.759
Intercept	0.4	1	0.37	4.6	0.033	0.009
age	0.2	1	0.24	3.0	0.084	0.006
Weight	9.6	1	9.57	117.8	P<0.001	0.192
Sex	39.1	1	39.15	481.8	P<0.001	0.493
Study sub-group	3.5	3	1.18	14.5	P<0.001	0.081
PC1*	0.1	1	0.08	0.9	0.338	0.002
PC2	0.0	1	0.03	0.3	0.575	0.001
PC3	0.1	1	0.10	1.2	0.279	0.002
PC4	0.1	1	0.12	1.5	0.228	0.003
PC5	0.0	1	0.03	0.4	0.551	0.001
Sex*study sub-group	0.9	3	0.31	3.8	0.01	0.023
***EGLN1* SNP rs2437150**	**0.5**	**2**	**0.23**	**2.8**	**0.059**	**0.011**
rs2437150*group interaction	**1.2**	**6**	**0.19**	**2.4**	**0.029**	**0.028**
Error	40.3	496	0.08			
Total	2314.7	519				
Corrected Total	166.9	518				
Model R-Squared = .759						

*PCs are principle components used to assess admixture in the sample and control for stratification, as described in the text.
Inclusion/exclusion of PCs had no qualitative effect on SNP associations, but are presented in the complete model for theoretical reasons.

Table S5. ANCOVA model showing association of *EGLN1* SNP (rs2491403) with dependent variable VO_2max in hypoxia

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	R^2
Corrected Model	126.3	22	5.75	68.7	P<0.001	0.754
Intercept	0.6	1	0.56	6.7	0.01	0.013
age	0.3	1	0.31	3.7	0.055	0.007
Weight	9.7	1	9.71	116.2	P<0.001	0.191
Sex	37.7	1	37.73	451.3	P<0.001	0.478
Study sub-group	4.0	3	1.32	15.8	P<0.001	0.088
PC1*	0.2	1	0.18	2.2	0.142	0.004
PC2	0.1	1	0.10	1.2	0.278	0.002
PC3	0.2	1	0.20	2.4	0.123	0.005
PC4	0.2	1	0.24	2.9	0.09	0.006
PC5	0.0	1	0.03	0.3	0.56	0.001
Sex*study sub-group	1.0	3	0.34	4.1	0.007	0.024
***EGLN1* SNP rs2491403**	**0.6**	**2**	**0.28**	**3.3**	**0.037**	**0.013**
rs2491403*group interaction	**1.1**	**6**	**0.18**	**2.2**	**0.045**	**0.026**
Error	41.2	493	0.08			
Total	2307.4	516				
Corrected Total	167.6	515				
Model R-Squared = .754						

*PCs are principle components used to assess admixture in the sample and control for stratification, as described in the text.
Inclusion/exclusion of PCs had no qualitative effect on SNP associations, but are presented in the complete model for theoretical reasons.

Table S6. ANCOVA model showing association of *EGLN1* SNP (rs479200) with dependent variable VO_2max in hypoxia

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	R^2
Corrected Model	126.1	22	5.73	70.3	P<0.001	0.757
Intercept	0.6	1	0.55	6.8	0.01	0.013
age	0.3	1	0.33	4.0	0.045	0.008
Weight	10.3	1	10.25	125.8	P<0.001	0.202
Sex	37.6	1	37.60	461.2	P<0.001	0.481
Study sub-group	4.9	3	1.62	19.9	P<0.001	0.107
PC1*	0.2	1	0.17	2.0	0.155	0.004
PC2	0.1	1	0.10	1.2	0.28	0.002
PC3	0.2	1	0.17	2.1	0.147	0.004
PC4	0.2	1	0.22	2.7	0.099	0.005
PC5	0.0	1	0.02	0.2	0.618	0.001
Sex*study sub-group	1.1	3	0.36	4.4	0.005	0.026
***EGLN1* SNP rs479200**	**0.5**	**2**	**0.24**	**2.9**	**0.056**	**0.012**
rs479200*group interaction	**1.3**	**6**	**0.21**	**2.6**	**0.019**	**0.03**
Error	40.5	497	0.08			
Total	2323.1	520				
Corrected Total	166.6	519				
Model R-Squared = .757						

*PCs are principle components used to assess admixture in the sample and control for stratification, as described in the text.
Inclusion/exclusion of PCs had no qualitative effect on SNP associations, but are presented in the complete model for theoretical reasons.

Table S7. Replication sample, ANCOVA model showing association of *EGLN1* SNP (rs1769793) with dependent variable VO_2max in hypoxia

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	R^2
Corrected Model	31.1	5	6.21	68.0	P<0.001	0.848
Intercept	1.7	1	1.66	18.2	P<0.001	0.230
age	1.0	1	0.98	10.7	0.002	0.150
Weight	0.7	1	0.72	7.9	P<0.001	0.115
Sex	14.6	1	14.64	160.6	0.007	0.725
***EGLN1* SNP rs1769793**	**0.66**	**2**	**0.33**	**3.6**	**0.033**	**0.106**
Error	5.6	61	0.09			
Total	389.4	67				
Corrected Total	36.6	66				

Model R-Squared = .848

Table S8. *EGLN1* markers assayed using the Affymetrix (Santa Clara, CA) Axiom Biobanking Array.

rsID	CHR	HG19	Alleles	Minor allele[a]	MAF Quechua	MAF Syracuse	MAF Combined	*n* Quechua	*n* Syracuse	Notes
rs115887846	1	231460282	A/C	C	0.09	0.45	0.32	448	40	Failed Affy best practices
rs2064766	1	231468953	A/G	A	0.30	0.67	0.63	858	186	
rs116427277	1	231471477	C/G	C	0.00	0.00	0.00	858	188	Monomorphic
rs146418424	1	231471540	A/G	G	0.00	0.00	0.00	858	188	Monomorphic
rs75215445	1	231471673	C/G	C	0.00	0.00	0.00	858	188	Monomorphic
rs200258755	1	231472291	C/G	G	0.00	0.00	0.00	858	188	Monomorphic
rs111580847	1	231472519	T/C	C	0.00	0.00	0.00	858	188	Monomorphic
rs145329289	1	231472520	A/C	A	0.00	0.00	0.00	858	188	Monomorphic
rs144793875	1	231472537	C/G	G	0.00	0.01	0.00	858	188	MAF<0.01
rs73116365	1	231472916	A/G	A	0.00	0.00	0.00	858	188	Monomorphic
rs145173936	1	231473415	A/G	G	0.00	0.00	0.00	858	188	Monomorphic
1.231473476	1	231473476	D/I	I	0.02	0.00	0.02	832	180	Failed Affy best practices
rs143523668	1	231474181	C/G	G	0.00	0.00	0.00	856	188	Monomorphic
rs117192083	1	231474294	A/G	G	0.00	0.00	0.00	858	188	MAF<0.01
rs142478304	1	231488403	A/G	G	0.00	0.00	0.00	858	188	MAF<0.01
rs2437150	1	231488524	T/C	T	0.31	0.70	0.66	858	180	
rs78209580	1	231488541	A/G	A	0.02	0.19	0.11	822	156	Failed Affy best practices
rs62617126	1	231488952	A/T	T	0.00	0.00	0.00	858	188	Monomorphic
rs192060983	1	231489081	A/G	G	0.00	0.00	0.00	858	188	Monomorphic
rs147839743	1	231502179	A/G	A	0.00	0.00	0.00	858	188	Monomorphic
1.231506357	1	231506357	D/I	D	0.00	0.00	0.00	858	188	Monomorphic
rs2491403	1	231511185	T/C	T	0.31	0.69	0.66	848	184	
rs2749713	1	231537921	T/C	C	0.31	0.71	0.67	828	178	

rs2491419	1	231554649	T/C	C	0.02	0.08	0.06	798	174	Failed Affy best practices
rs75538505	1	231557060	T/C	C	0.00	0.00	0.00	858	188	Monomorphic
rs61750991	1	231557164	C/G	G	0.00	0.03	0.02	854	188	MAF<0.01
rs74892794	1	231583056	T/C	T	0.00	0.02	0.01	858	188	MAF<0.01
rs1769793	1	231601099	T/C	C	0.45	0.77	0.84	844	178	
rs12030600	1	231605379	A/G	A	0.14	0.08	0.18	858	188	

[a]Minor Allele defined in Peruvian Quechua
Chr = Chromosome
MAF = Minor Allele Frequency

Table S9. PCR Primers and Restriction Enzymes

SNP	Forward (5′ to 3′)	Reverse (5′ to 3′)	Inner Forward (5′ to 3′)[a]	Inner Reverse (5′ to 3′)[a]	Restriction Enzyme
rs1769793	caacctaaatgcccgctgac	tcaacaaaagccacactcaca	NA	NA	BsmAI
rs479200	ctcccgaactctgaatgtcctt	agaaatcggatggaaaggtgg t	NA	NA	HpyCH4IV
rs480902	ctggggtaatttcactggagtt gtg	gctctgggatacaatgatgaac aat	cagcacttctggtctacattaatgtgt atg	tctccaagtgatctccagtgact aat	NA
rs12097901	gtgcatggcgcagtaacgg	gaatgctgcttctcagcctag	NA	NA	BsmAI
rs186996510	gtgcatggcgcagtaacgg	gaatgctgcttctcagcctag	NA	NA	BsrBI

[a]Inner Forward and Inner Reverse primers were used for Tetra- amplification refractory mutation system (ARMS) PCR. Tetra-ARMS PCR uses four primers and does not require restriction enzyme digestion for sample genotyping.

Dataset S1. GWAS results for filtered Affymetrix Biobanking array SNPs tested for association with VO_2Max. See Excel sheet.

References

1. Durnin JV & Womersley J (1974) Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. *Br J Nutr* 32(1):77-97.
2. Manichaikul A*, et al.* (2012) Analysis of family- and population-based samples in cohort genome-wide association studies. *Hum Genet* 131(2):275-287.
3. Zhou D*, et al.* (2013) Whole-genome sequencing uncovers the genetic basis of chronic mountain sickness in Andean highlanders. *Am J Hum Genet* 93(3):452-462.
4. Scheet P & Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* 78(4):629-644.
5. Barrett JC, Fry B, Maller J, & Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263-265.
6. Kittles RA*, et al.* (2002) CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? *Hum Genet* 110(6):553-560.
7. Purcell S*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559-575.
8. Nevill AM, Ramsbottom R, & Williams C (1992) Scaling physiological measurements for individuals of different body size. *Eur J Appl Physiol* 65(2):110-117.
9. Bigham AW*, et al.* (2008) Angiotensin-converting enzyme genotype and arterial oxygen saturation at high altitude in Peruvian Quechua. *High Alt Med Biol* 9(2):167-178.
10. Brutsaert T*, et al.* (2003) Spanish genetic admixture is associated with larger VO2max decrement from sea level to 4.338 m in Peruvan Quechua. *J. Appl. Physiol.* 95:519-528.
11. Brutsaert TD*, et al.* (2005) Ancestry explains the blunted ventilatory response to sustained hypoxia and lower exercise ventilation of Quechua altitude natives. *Am J Physiol Regul Integr Comp Physiol.*