



Published in final edited form as:

Cell Rep. 2025 May 27; 44(5): 115577. doi:10.1016/j.celrep.2025.115577.

Genetic and training adaptations in the Haenyeo divers of Jeju, Korea

Diana Aguilar-Gómez^{1,2}, Jacob Bejder⁴, Jonathan Graae⁴, Yelin Ko³, Andrew Vaughn², Kendell Clement⁵, Martin Tristani-Firouzi^{6,7}, Joo-Young Lee^{8,9}, Nikolai B. Nordsborg⁴, Rasmus Nielsen², Melissa Ilardo^{5,7,10,*}

¹Ecology and Evolutionary Biology, University of California, Los Angeles, 612 Charles E. Young Drive South, Los Angeles, CA 90095, USA

²Center for Computational Biology, University of California, Berkeley, Berkeley, CA 94720, USA

³Department of Human-Centered Design, Cornell University, Ithaca, NY 14850, USA

⁴Department of Nutrition, Exercise, and Sports, University of Copenhagen, Nørre Allé 51, 2200 Copenhagen, Denmark

⁵Department of Biomedical Informatics, University of Utah, 421 Wakara Way #140, Salt Lake City, UT 84108, USA

⁶Division of Pediatric Cardiology, Department of Pediatrics, University of Utah, 81 Mario Capecchi Dr., Salt Lake City, UT, USA

⁷Nora Eccles Harrison CVRTI Cardiovascular Research Training Institute, University of Utah, Salt Lake City, UT, USA

⁸Research Institute for Human Ecology, College of Human Ecology, Seoul National University, Seoul 08826, Republic of Korea

⁹Graphene Research Center for Convergence Technology, Advanced Institute of Convergence Technology, Suwon 16229, Republic of Korea

¹⁰Lead contact

SUMMARY

Natural selection and relative isolation have shaped the genetics and physiology of unique human populations from Greenland to Tibet. Another such population is the Haenyeo, the all-female Korean divers renowned for their remarkable diving abilities in frigid waters. Apnea diving

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*Correspondence: melissa.ilardo@utah.edu.

AUTHOR CONTRIBUTIONS

M.I., J.-Y.L., R.N., and N.B.N. designed all experiments. M.I., J.-Y.L., Y.K., J.B., and J.G. conducted experiments. D.A.G., M.I., N.B.N., J.B., J.G., R.N., M.T.-F., A.V., and K.C. analyzed data. D.A.G. and M.I. wrote the manuscript, and all authors reviewed, edited, and approved the manuscript.

DECLARATION OF INTERESTS

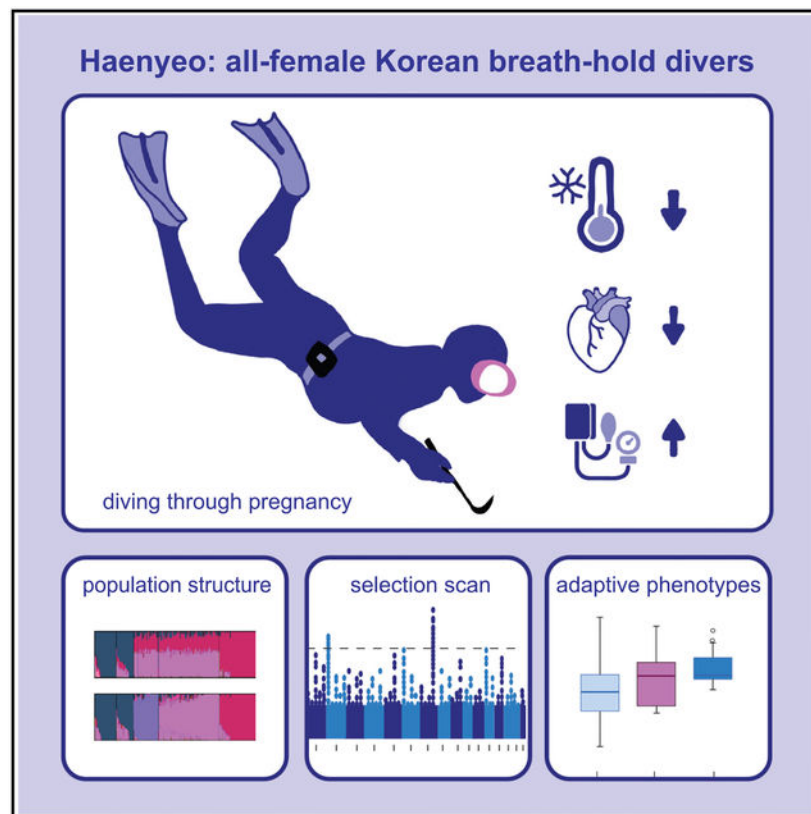
The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2025.115577>.

induces considerable physiological strain, particularly in females diving throughout pregnancy. In this study, we explore the hypothesis that breath-hold diving has shaped physiological and genetic traits in the Haenyeo. We identified pronounced bradycardia during diving, a likely training effect. We paired natural selection and genetic association analyses to investigate adaptive genetic variation that may mitigate the effects of diving on pregnancy through an associated reduction of diastolic blood pressure. Finally, we identified positively selected variation in a gene previously associated with cold water tolerance, which may contribute to reduced hypothermia susceptibility. These findings highlight the importance of traditional diving populations for understanding genetic and physiological adaptation.

Graphical Abstract



In brief

Aguilar-Gómez et al. find evidence of evolutionary adaptation with potential relevance to breath-hold diving in the Haenyeo divers of Korea. Genetic variants under selection in this population may help mitigate the physiological strain of repeated diving during pregnancy. Additionally, long-term training appears to enhance the diving response through pronounced bradycardia.

INTRODUCTION

Extraordinary human populations have inhabited nearly every environment on the planet, adopting unique lifestyles and strategies to survive. The extreme conditions faced by these

populations have imposed natural selection, introducing heritable adaptive variation. Modern humans have adapted to high altitudes,^{1,2} distinctive diets,³ and even breath-hold (or apnea) diving.^{4,5} Studying the genetics and physiology of these populations through the lens of evolution enables a deep understanding of the relationship between genetic variation and human physiology, including novel discoveries that cannot be found through genome-wide association studies (GWAS) alone. We present evidence of a novel adaptation in the second known population that may have evolved for breath-hold diving: the Haenyeo of Jeju Island, Korea.

The inhabitants of Jeju Island represent a subpopulation of Korea with a distinct language, a matriarchal family structure, and the celebrated resident community of Haenyeo divers. The practice of breath-hold diving is so integral to Jeju's culture that the shortening of words characteristic of the Jeju language is colloquially attributed to the need for divers to communicate quickly at the water's surface. Haenyeo literally translates to women of the sea.⁶ The Haenyeo dive year-round in social collectives, harvesting everything from abalone to sea urchin. Their typical diving occurs at relatively shallow depths (<10 m) and for a short duration (~30 s) but with high energy expenditure for around 4–5 h per day.⁷ Haenyeo typically dive throughout pregnancy, often continuing to dive until the day they give birth.⁸ Young women are no longer continuing this matrilineal tradition; the current Haenyeo population, with an average age of around 70 years, may represent the last generation of Haenyeo divers.⁹

During diving, all humans experience a well-characterized effect known as the mammalian dive reflex.^{10–14} This multi-system physiological response offers many biological components on which natural selection can act. The dive reflex is triggered by the combination of apnea and water submersion and induces vasoconstriction, which selectively redistributes blood to vital organs and away from non-essential tissues.^{15,16} Diving also triggers a slowing of the heart rate, or bradycardia, which reduces cardiac output and conserves oxygen.^{10,17,18} The only element of the dive response thus far demonstrated to have a link to natural selection is spleen size, which was found to be increased in the Bajau “sea nomads” of Indonesia. During a dive, splenic contraction increases the volume of circulating red blood cells and, thus, available oxygen by up to 9.6%.^{19–22} Other than spleen size, it is unknown to which extent components of the dive response are influenced by genetic variation rather than training. Genetic evidence from study of the Bajau and others suggests that variation in the gene *BDKRB2* may have been selected for its role in vascular contraction^{5,23}; however, this effect has not yet been directly measured in any traditional diving population.

The interaction between the mammalian dive reflex and pregnancy poses unique physiological challenges for female divers. Although there is limited documentation of its impact in breath-hold divers, exposure to intermittent hypoxia during early pregnancy is linked to a higher risk of hypertensive disorders of pregnancy, commonly in the form of preeclampsia.²⁴ This effect has most commonly been observed in cases of obstructive sleep apnea (OSA), where apnea triggers vasoconstriction, raising blood pressure.^{25,26} Given the parallels between OSA and breath-hold diving, it is plausible that apnea during diving may also increase preeclampsia risk.

The genetics of diving physiology in the Haenyeo and the broader Jeju population have not been studied to date. We were motivated by the discovery of natural selection on diving in the Bajau to pursue the first investigation of genetic adaptation to diving in the Haenyeo. Our participants included Haenyeo divers and age-matched controls in Jeju and Seoul. We measured baseline metrics associated with diving physiology, including spleen size. We also performed cardiovascular measurements at rest and during a series of simulated dives to identify candidate diving phenotypes that may be associated with underlying genetic variation. We then used whole genome sequencing (WGS) data to identify genetic loci that have been under selection in this population and examined whether these loci were associated with candidate phenotypes.

RESULTS

Study population

We sampled women from three populations: Haenyeo in Jeju ($n = 30$), non-Haenyeo in Jeju ($n = 30$), and non-Haenyeo in Seoul ($n = 31$). These sample sizes were selected to match those of prior studies that have detected genetic association on selected genetic variation in genetically isolated, putatively adapted populations.^{1,2,5} We chose to include participants from the Korean mainland to account for possible genetic continuity between the Haenyeo and non-Haenyeo participants on Jeju Island as well as possible genetic differences between Jeju and mainland Koreans. All Haenyeo participants in Jeju were self-identified members of a Haenyeo community in Hado, Jeju, and were all at least third-generation Haenyeo. All Haenyeo were actively working as divers, spending a minimum of 3 days per week in the water. Non-Haenyeo controls in Jeju and Seoul were all age matched to reflect the demographics of the Haenyeo, an aging population (average age at the time of the study, 65 years). Using genetic analyses, we identified three individuals as being closely related, two from Jeju and one from Seoul, and excluded these individuals from all subsequent analyses.

Spleen measurements

Previous work demonstrated that naturally selected genetic variation was associated with an increased spleen size in a traditional diving population,⁵ and subsequent work has suggested that the act of apnea training itself can correlate with an increase in spleen size.²⁷ We therefore measured spleen size in the Haenyeo to compare with control populations. We performed ultrasound measurements in two planes in order to calculate spleen volume using the same methodology employed in the Bajau⁵ that has been demonstrated to correlate well with computed tomography measurements. We found that spleen size did not differ between the Haenyeo and non-Haenyeo of Jeju; however, the spleen size of each Jeju group was significantly larger than that of controls from Seoul (two sample t test, $p = 0.0075$ and $p = 0.016$ for Haenyeo vs. Seoul and Jeju vs. Seoul, respectively). When using a linear model to test Jeju individuals compared to non-Jeju individuals, the results remained significant ($p = 0.016$, $\beta = -85.42$, $SE = 34.75$). However, when we accounted for additional covariates, including height, weight, and age, the results were no longer significant ($p = 0.052$, $\beta = -88.59$, $SE = 44.96$). This suggests that the observed spleen size difference may not be primarily correlated with genetic effects.

Hematological measurements

The large spleen adaptation in the Bajau is linked to genetic variation that is associated with changes in thyroid hormone levels. Hyperthyroidism is known to stimulate erythropoiesis,²⁸ and recent work in animal models suggests that replicating the Bajau genetic adaptation is associated with hematological variation.²⁹ Specifically, mice that underwent pharmacological inhibition of the same gene thought to be downregulated in the Bajau presented with higher red blood cell (RBC) count, hemoglobin (Hgb), and hematocrit (Hct) than control animals. On the basis of these results, we used capillary blood samples to measure baseline hematological parameters in all participants. We found no difference in Hgb or Hct between any of the groups measured.

Simulated dive measurements

We sought to determine whether natural selection on breath-hold diving is associated with changes in Haenyeo physiology. We conducted a series of simulated dives to measure diving-specific physiological responses through a protocol refined in previous studies of the human dive response.^{23,30} In this sequence, the participant maximally holds her breath while submerging her face in a pool of cold water, thus triggering the bradycardia and peripheral vasoconstriction associated with the human dive reflex. Over a series of three “dives,” with 2 min recovery between each dive, we measured the cardiovascular response to the diving stimulus in the form of heart rate and blood pressure (BP).

We first quantified the extent of bradycardia over simulated dives. While participants from all groups experienced bradycardia during diving, we found that the Haenyeo demonstrated significantly greater bradycardia over the course of a simulated dive than either control group. The difference in bradycardia was particularly pronounced within Jeju populations; the Haenyeo displayed a mean difference in heart rate (Δ HR) of -18.8 bpm compared to -12.6 bpm in Jeju controls ($p = 0.015$) despite significantly shorter dive times of 27 s in the Haenyeo and 39 s in Jeju controls, respectively ($p = 0.0018$) (Figure 1).

We then examined diastolic and systolic BP. We considered these two measurements separately rather than mean arterial pressure (which is calculated as the sum of 2/3 diastolic BP and 1/3 systolic BP) in order to detect any independent changes in either variable that could be obscured by combining them. We found that the Haenyeo experience significantly higher diastolic BP on average than controls from Seoul at baseline ($p = 0.010$), during dives ($p = 0.0053$), and during the 2 min recovery period between dives ($p = 0.00051$) (Figures 2A–2C). However, diastolic BP was not significantly different between Haenyeo and Jeju controls. Isolated diastolic hypertension has not been observed previously in diving populations. The difference in diastolic BP was consistently approximately 10 mm Hg higher in Jeju residents than in Seoul controls (Figures 2D–2F). Systolic BP was not significantly different between any groups at any time point (Figures 2G–2I). Because we were most interested in natural selection on diving, we chose to focus on the measurement of diastolic BP that might best represent a physiological benefit during a dive through its effect on increasing cerebral perfusion: maximum diastolic BP reached over a simulated dive. Specifically, we examined the first simulated dive in the series. Selecting this measurement also enabled us to control for possible confounding factors through the inclusion of dive

duration as a covariate. We found that Jeju residents, regardless of whether they are Haenyeo, have significantly higher maximum diastolic BP during the simulated dive ($p = 0.014$ using a two-sample t test). Using a multiple linear regression model to account for height, weight, age, and dive duration, the diastolic BP difference between groups remained significant ($p = 0.012$, $\beta = -10.1$, $SE = 3.90$). This finding is consistent with a previous survey that indicates that Jeju has one of the highest rates of hypertension in the nation.³¹

Genetic analyses

Demographic and evolutionary history: To examine the demographic and evolutionary history of the Jeju population, we generated low-depth WGS data ($5\times$), coverage selected deliberately to be used in conjunction with genome imputation.³² As a reference panel for imputation, we used Korean mainland individuals, specifically from the Korean Genome Project (Korean1K).³³ Based on DNA sample quality, we were able to generate sequencing data from 84 individuals representing Haenyeo in Jeju ($n = 27$), non-Haenyeo in Jeju ($n = 30$), and non-Haenyeo in Seoul ($n = 27$).

For the purpose of population history analyses, we merged the data we collected with data from the Korean1K data and selected populations from the 1000 Genomes Project dataset, resulting in 5,673,091 SNPs after merging and filtering. We chose to include only female participants for consistency with the individuals in our study. We performed a principal-component analysis (PCA) including only individuals from Korea: Haenyeo (HAE), non-Haenyeo Jeju residents (JEU), Korean mainlanders in Jeju (KMJ), participants from the Korean1K project (KOA), and controls from Seoul (SEO). In this PCA, PC1 appears to separate individuals from Jeju Island and those from the Korean mainland (Figure 3). The individuals in our dataset from Seoul cluster with those from Korean1K, who were primarily recruited in Ulsan on the Korean mainland.³³ Five individuals from Jeju Island appeared to cluster tightly with mainland Koreans. In order to quantify this genetic similarity, we performed an admixture analysis using Ohana's qpas³⁴ and examined ancestry proportions of these individuals (Table S1). We relabeled any Jeju participants with more than 60% Korean ancestry and less than 15% Jeju ancestry as Korean mainlander in Jeju in subsequent analyses to more accurately reflect their genetic ancestry. Similarly, one individual from Korean1K clustered with the individuals from Jeju Island in the PCA. This individual had nearly 100% Jeju ancestry; therefore we reclassified this individual as a Jeju Islander (JEU).

The Haenyeo and Jeju controls appeared to be largely overlapping in PC1/PC2; however there was some apparent visual separation in the populations. We calculated the fixation index (F_{ST}) to quantify the genetic divergence between these populations and found that the Haenyeo and Jeju controls are not meaningfully genetically differentiated ($F_{ST} = 0.00031$) (Table 1). This suggests that all individuals from multigenerational Jeju families, regardless of whether they are Haenyeo, represent an approximately panmictic population. Given the integral role of diving for thousands of years of Jeju's history, it is therefore likely that the ancestors of all Jeju residents (Haenyeo and non-Haenyeo) are equally likely to be influenced by natural selection on diving. The ancestral separation between Jeju individuals and mainland Koreans is also reflected in F_{ST} ; the F_{ST} between Jeju Island and Mainland

Korea is 0.0034. This suggests that mainland Koreans are at least as differentiated from Jeju residents than they are from the Japanese (JPT) ($F_{ST} = 0.0030$) and northern Han Chinese (CHB) ($F_{ST} = 0.0026$). We note, however, that F_{ST} reflects genetic divergence, which depends on effective population size (N_e). The higher F_{ST} values may thus reflect a low N_e in the Jeju relative to CHB and JPT.

We performed a PCA using global populations from the 1000 Genomes Project, including four Asian populations and two outgroup populations. We observed that, in PC3 and PC4, the PCs in which Asian populations begin to separate from each other, individuals from Jeju cluster most closely with Japanese individuals rather than mainland Koreans (Figure 4). In additional admixture analyses of all populations using Ohana's qpas,³⁴ we found that, when considering 5 ancestry components ($K = 5$), two predominant ancestral components were represented in Northern East Asia: one shared by the Chinese and Mainland Koreans and a second specific to the Japanese and all Koreans (Figure 5). This component represents almost the complete ancestry of Japanese and Jeju individuals at this K , echoing the PC3/PC4 overlap between Jeju individuals and the JPT and reflecting shared ancestry between these populations. At $K = 6$, Jeju individuals are represented by their own ancestral component, and at $K = 7$, mainland Koreans are represented by another distinct component.

To contextualize any potential genetic adaptation, we examined the genetic history of Jeju. This history enables us to identify how and when adaptive variation may have arisen in the population. We started by using AdmixtureBayes to infer a population tree from the populations in our dataset. We ran AdmixtureBayes allowing for 0, 1, or 2 admixture events. When considering graphs with 0 admixture events, we saw that the Jeju Islanders (JEJ) and JPT populations were inferred to be sister populations, which is in concordance with the PCA showing significant overlap between the JEJ and JPT clusters (Figure S1A). The next most related population to both JEJ and JPT was inferred to be the mainland Koreans (KOR). When allowing for 1 admixture event, we inferred that the Southern Chinese population (CHS) is admixed between a population related to CHB and a population related to Kinh Vietnamese (KHV) (Figure S1B). These results are also in agreement with the PCA analysis showing the CHS cluster as being between the CHB and KHV clusters. The relationship between the JEJ, JPT, and KOR populations remained consistent with the 0 admixture event graph. There is some uncertainty in the exact placement of the CHB lineage as a result of the short branch connecting this population to the rest of the graph. When allowing for 2 admixture events, we found that the distribution of admixture graphs is too dispersed to draw definitive conclusions (Figure S1C). This is due to the very high number of admixture graph topologies with 2 admixture events that are compatible with the observed allele frequency covariance matrix. It is possible that using more information, such as identity by descent tracts, could help resolve more fine-scale admixture in these populations, but existing admixture graph inference methods are unable to do so.

We next wanted to estimate when the divergence between Jeju residents and mainland Koreans occurred as well as to put an upper limit on the time at which natural selection may have begun to shape Jeju genetic ancestry. We therefore estimated the N_e jointly for mainland Koreans and Jeju residents, from which we can then infer divergence time as the time when these populations stop sharing patterns of expansion or contraction. We found

that the population trajectories of Jeju and mainland Koreans noticeably diverged roughly 5,000–7,000 years ago (Figure S2). Since this time, many other East Asian populations have experienced uninterrupted growth, resulting in a roughly 10-fold increase in N_e over the past 10,000 years. However, Jeju's population has only effectively doubled in the past 10,000 years, with a dramatic reduction in population roughly 1,200–3,000 years ago, suggesting a bottleneck. We note that inferring divergence times from N_e curves involves assumptions about the demographic history of populations, such as absence of migration.

Selection analysis and association testing: To identify genes or variants that may be of adaptive significance and related to the phenotypes of interest, we used a two-step approach that previously has proven useful for detecting human adaptive genetic variation.^{2,3,5} We first scanned the genome for variants that may have been targeted by selection by comparing allele frequencies between populations. We subsequently tested for associations only on the top candidates from the selection scan. While only applicable in settings where there are closely related populations that differ phenotypically, this protocol greatly reduces the multiple testing burden of GWASs by focusing only on variants that have the potential to explain between-population phenotypic differences. This protocol has been used previously to detect variants in the *EPAS1* gene associated with altitude adaptation in Tibetans,² variation in *FADS* genes associated with multiple phenotypes in Inuit,³ and genetic variants associated with diving-related phenotypes in the Bajau.⁵

Our population genomics analyses on structure and demography revealed that the modern population of Jeju, including Haenyeo and non-Haenyeo, is descended from the same ancestral population. This indicates that modern Jeju residents are equally likely to have descended from ancestral divers, suggesting that a selection scan targeting the Jeju ancestral component is appropriate for detecting regions of positive selection in the Haenyeo and other Jeju residents. In order to detect naturally selected genetic variation, we used a method that leverages local patterns of allele frequency differences, similar to those used in previous selection studies.^{2,3,5} Specifically, we used selscan from Ohana³⁴ to identify SNPs that deviate in Jeju from the genome-wide covariance using a likelihood ratio test. This statistical test identifies variants that have experienced a stronger change in allele frequency change between populations than expected given the genome-wide patterns of allele frequency change, and it assumes a very simple Brownian motion model of allele frequency change. We therefore also later apply a more parametric test (CLUES2³⁵) that models selection and genetic drift using a Wright-Fisher model under an explicit demographic model that includes changes in population size in order to validate the selection inferences for the top candidate SNPs (this validation step is described in detail below). For the Ohana analyses, we selected mainland Koreans (KOR) as a comparison population and CHS as an outgroup. We elected to filter these results in order to annotate and further investigate the 10 strongest candidate signals of selection (Figure 6; Table S2). Among these results, we identified several SNPs in genetic regions with potential links to Haenyeo diving.

The strongest selection signal came from a peak in the region of the centromere of chromosome 8, upstream of the gene that encodes prostate, ovary, testis, and placenta expressed (POTE) ankyrin domain family member A. This peak overlaps with a signal previously identified through a GWAS in a Korean cohort, specifically a study of copy

number variations associated with hematological parameters.³⁶ Genetic variation in the region of our top SNP was significantly associated with RBC count. Hematological variation is believed to underlie the large-spleen phenotype observed in the Bajau and may provide additional oxygen capacity during diving.²⁹ We observed another selected allele in the gene that encodes sarcoglycan zeta. Variation in this gene has been shown previously to be significantly associated with pain sensitivity, as measured by the cold pressor test in a 23andMe cohort of undisclosed (likely European) ancestry.³⁷ This test, much like Haenyeo diving in the winter, involves submersion in exceptionally cold water. It may therefore be linked to the well documented Haenyeo ability to tolerate cold.^{38–43} While we did not measure thermoregulatory physiology in our study, this represents an avenue for future research.

We next wanted to test for an association between the variants putatively under selection in Jeju and diving diastolic BP. Using all individuals, including those from Jeju and Seoul, we used PLINK⁴⁴ on the top SNPs from the 10 highest-scoring genomic regions in the selection scan to measure association with untransformed maximum diastolic BP during the first simulated dive. We included the following covariates: age, height, weight, dive length, and the first 10 PCs (to control for underlying population structure). We found that only one SNP from the selection scan was significantly associated with maximum diving diastolic BP at the Bonferroni-corrected significance level (0.005) using an additive model: rs66930627 ($p = 0.0021$). When we included age¹ as an additional covariate to account for non-linear effects, as is common in BP GWASs,⁴⁵ the association remained significant ($p = 0.0014$). We also tested the association within only Jeju participants to assess whether the association reflects population-specific allele frequency differences between Jeju and Seoul and found that the association remains significant ($p = 0.0039$). We note that we were unable to include use of anti-hypertensives as an additional covariate given that these data were not collected at the time of the study.

In order to validate the evidence for selection on this variant, we used an entirely independent method of detecting selection that accounts for changes in population size and demography and uses information from the haplotype and allele frequency distribution within the focal population: CLUES2.³⁵ In contrast to the previous test, this method does not use allele frequency differences between populations but, instead, independent evidence of allele frequency changes obtained from the ancestral recombination graph. Using CLUES2, we found significant evidence in rs66930627 of selection favoring the C allele in the JEJ population across all 3 models considered: a 1-epoch model in which selection was constant through time ($p = 3.3 \times 10^{-5}$), a 2-epoch model in which the selection coefficient changed 1,200 years ago ($p = 5.6 \times 10^{-5}$), and a 2-epoch model in which the selection coefficient changed 7,000 years ago ($p = 1.0 \times 10^{-4}$). As measured by Akaike information criterion, an estimator of prediction error, the 2-epoch model with selective pressure beginning 1,200 years ago was chosen as the best model, thus providing evidence in support of the hypothesis that the selection on rs66930627 to decrease diastolic BP was very recent and very strong ($\hat{S}^{MLE} = 0.021$ in the more recent epoch) (Figure S3). Interestingly, this timing aligns with a recent bottleneck experienced by the Jeju population.

The putatively selected allele in Jeju, C, has an allele frequency of 33% in Jeju and 7% in mainland Korea. It is associated with a decrease in diastolic BP corresponding to roughly a 10 mm Hg decrease in diastolic BP per selected allele. The established contribution of elevated diastolic BP to preeclampsia risk in pregnancy could induce a strong selective pressure, especially when provoked through the dive response. The association between rs66930627 and a decreased diastolic BP could therefore suggest that natural selection acted to mitigate elevated diastolic BP in the Jeju population. The associated SNP is found in an intergenic region; however, it is an expression quantitative trait locus (eQTL) for a gene that encodes Fc γ receptor IIA (Fc γ RIIA, $p = 1.2 \times 10^{-4}$) (data source: GTEx Analysis Release v.8; dbGaP: phs000424.v8.p2). Fc γ RIIA is a low-affinity Fc γ receptor that binds IgG2 and may influence hypertension through its effect on the pro-inflammatory activities of vascular smooth muscle cells.^{46,47} Manipulation of the gene Fc γ RIIB, which encodes the largely homologous Fc γ receptor IIB, has been demonstrated to directly correlate with hypertension in animal models.⁴⁸ In humans, variation in Fc γ RIIB is significantly associated with preeclampsia risk.⁴⁹

The SNP rs66930627 is also an eQTL for FC receptor-like B (FCRLB; $p = 7.0 \times 10^{-5}$), which shares a common ancestor with Fc γ RIIB,⁵⁰ and heat shock 70 kDa protein 7 (HSPA7; $p = 1.2 \times 10^{-6}$). HSPA7 is also referred to as a pseudogene or long non-coding RNA and has been shown to promote the inflammatory transition of vascular smooth muscle cells. HSPA7, like FC γ RIIA and FCRLB, may therefore also play a role in inflammation-mediated hypertension.⁵¹ The functional overlap between the eQTLs of the associated SNP suggests a potential redundancy in their roles.

We were able to validate the association between rs66930627 and diastolic BP in a European ancestry cohort. In a sample of European ancestry individuals from the All of Us cohort ($n = 1,376$), filtered to age match our Korean cohort and exclude individuals on anti-hypertensives, we found a significant association between rs66930627 and diastolic BP ($p = 0.040$). In this cohort, we found that each additional C allele corresponded to a roughly 0.75 mm Hg decrease in diastolic BP, a more modest effect size than what we observed in the Korean cohort. This discrepancy underscores the challenges of validating such associations in populations with different genetic backgrounds.

DISCUSSION

We present the first evidence of naturally selected genetic variation in the Haenyeo divers of Jeju Island, Korea. Using natural selection combined with association, we found a genetic variant that is significantly associated with diastolic BP. We also identified a training effect in the Haenyeo that likely contributes to their skillful diving. Populations like the Haenyeo are integral for uncovering relationships between genes and the regulation of physiology due to their demographic and evolutionary history.

Our first finding was that Jeju represents a genetically distinct subset of the Korean population. Jeju, which is found ~80 km off the Korean peninsula, has long been relatively isolated from the Korean mainland. The earliest signs of human settlement on Jeju date back 10,000 years to a community that eventually grew into the Tamna Kingdom. For a brief

period roughly spanning the 14th century, Mongolian forces controlled the island, possibly introducing genetic ancestry distinctive to the region.⁵² Mongol control was succeeded by half a millennium of Jeoson rule, during which Jeju faced significant restrictions with a lasting cultural impact. For example, inhabitants of Jeju were prohibited from leaving the island, thus increasing Jeju's relative isolation until well into the 19th century.⁶ Owing to this unique history, the Jeju language is by some measures as different from standard Korean as Dutch is from Norwegian.⁵³ We found that this linguistic distinction is reflected in genetics; the Jeju population appears to be genetically diverged from mainland Korea. We also found that the population of Jeju is approximately panmictic, with Haenyeo divers and non-Haenyeo Jeju residents descending from the same ancestral population. For the purposes of distinguishing between genetic adaptation and training effects, Jeju residents can therefore be considered non-diving Haenyeo.

We found that the Haenyeo and other Jeju residents demonstrate significantly elevated diastolic BP compared to residents of Seoul at baseline, over the course of a simulated dive, and during recovery between dives. We found no corresponding elevation of systolic BP, although hypertension has been reported previously in Jeju.³¹ Elevated diastolic BP, particularly when exacerbated during apnea, has been linked to pregnancy risk factors, including preeclampsia.⁵⁴ We identified a genetic variant with evidence of strong, recent natural selection in Jeju that is significantly associated with a reduction of diastolic BP in this population. This association may reflect natural selection to mitigate the complications of diastolic hypertension experienced by female divers while diving through pregnancy. Pregnancy is increasingly thought to be a driver of natural selection in extreme environments and lifestyles, with particular focus on the relationship between natural selection and pregnancy in women living at high altitude.⁵⁵ Preeclampsia has a well-established relationship with long-term cardiovascular health outcomes.^{56,57} The BP mitigation correlated with the identified genetic variation may therefore also be linked to the observation that, while Jeju has the second highest prevalence of hypertension in Korea, it also has the lowest stroke mortality in the nation.³¹ Further data are needed to fully control for environmental factors that may influence regional differences in BP.

The genetic variant presumed to be under selection and associated with a decrease in diastolic BP is found in a region that is associated with changes in expression of the gene *FcγRIIA*. By binding IgG2, *FcγRIIA* may mitigate diastolic hypertension through its role in vascular inflammation, resembling the mechanism of the related *Fcγ Receptor FcγRIIB*.^{46–48} In fact, genetic variation in *FcγRIIB* is significantly associated with preeclampsia.⁴⁹ This association may be linked to the mitigation of vascular inflammation. The nature of this relationship is unclear with current data, especially as our understanding of the role of inflammation in vascular biology and secondary hypertension continues to evolve.

We were able to validate the association between rs66930627 and diastolic BP in a European ancestry subset of the All of Us cohort, albeit with a lower-magnitude effect (~0.75 mm Hg). In the general population, a 1 mm Hg rise in systolic BP corresponds to a log-linear incremental risk of cardiovascular disease, so the effect size we observed in All of Us is likely clinically significant.⁵⁸ BP is known to be a multifactorial and polygenic trait,⁴⁵ and SNPs identified in GWASs have yet to provide sufficient magnitude effect to cause

observed BP phenotypes outside of monogenic BP syndromes.⁵⁹ Additionally, the majority of genetic variants previously identified as associated with BP traits are common rather than rare variants due to the late onset nature of hypertension, which protects this variation from purifying selection.^{59,60}

We observed that the Haenyeo exhibit pronounced bradycardia in response to simulated diving, displaying a significantly greater reduction in HR over the course of a dive. Remarkably, one Haenyeo participant experienced an HR drop of over 40 bpm in a simulated dive lasting just 15 s. This feature appears to be a training effect, as it was only observed in diving Haenyeo and not in their non-diving genetic relatives in Jeju. These results are consistent with previous studies in the Haenyeo and in other trained divers, where rigorous apnea dive training has been suggested to increase the magnitude of bradycardia over a dive.^{38,61–64}

We were able to demonstrate physiological outcomes in the Haenyeo divers of Jeju that may arise from both genetic adaptations and a lifetime of dedicated training. The Haenyeo are the second known breath-hold diving population in whom medically relevant physiology is correlated with novel, adaptive genetic variation. Our findings reinforce the extraordinary biology of diving communities and the power of natural selection to illuminate novel genetic variation. Continued investigation of populations like the Haenyeo will provide opportunities for breakthroughs in our understanding of the genetic regulation of disease.

Limitations of the study

While our study illustrates the strength of leveraging an evolutionary approach to identify novel genetic variation associated with changes in physiology, it is limited by the sample size, the limited physiological data collected, and sequencing depth due to the constraints of the population and study magnitude. Generalizability of the results is limited by the fact that all participants were female. As with all studies of genetic adaptation in modern human populations, it is impossible to ever know the true selective pressure that is driving observed signals of selection, in part due to the pleiotropic nature of most genes. Furthermore, any parametric claims about selection in humans or other organisms using population genetic data rely on assumptions regarding population history and demography. Some uncertainty regarding the possible confounding effects of demography will always remain. A future study with more extensive experimental physiological, clinical, and environmental data will be necessary to more deeply unravel the genetic basis of diving adaptation and to better understand potential relationships with hypertensive disorders of pregnancy.

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to and will be fulfilled by the lead contact, Melissa Ilardo (melissa.ilardo@utah.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Newly generated, de-identified WGS data have been deposited at the NCBI Sequence Read Archive and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- All original code has been deposited at <https://github.com/aguilar-gomez/haenyeo> and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

STAR★METHODS

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human subjects research: We obtained approval to conduct these human subject studies from the IRB boards of both the University of Utah and Seoul National University. Prior to implementing the study, we conducted community engagement and outreach to gauge community interest in the project. Informed consent was obtained from all participants. All participants were consented by a native Korean speaker (who also spoke Jeju dialect) after a thorough explanation of the experimental protocol. Following data analysis, we returned to share the results with the communities to ensure comfort with the results and their presentation.

We recruited female participants from three groups: Haenyeo in Jeju ($n = 30$ individuals), non-Haenyeo in Jeju ($n = 30$ individuals), and non-Haenyeo in Seoul ($n = 31$ individuals). All Haenyeo participants in Jeju were self-identified members of a Haenyeo community in Hado, Jeju and were all at least third-generation Haenyeo. All Haenyeo were actively working as divers, spending a minimum of 3 day per week in the water. Non-Haenyeo controls in Jeju and Seoul had no Haenyeo relatives and were all age matched to reflect the demographics of the Haenyeo, an aging population. Participants were all female due to the nature of the Haenyeo tradition. Participants had an average age of 65 years old. Participant selection is further described in the results section of the manuscript.

METHOD DETAILS

We obtained and electronically recoded the following basic demographic information: age, self-identified ancestry, whether or not the individual is a diver, and how frequently they dive. We then measured the participants height and weight, took capillary blood samples, and measured spleen size at rest.

We next had each participant perform the diving protocol. We asked each participant perform a series of simulated dives through a standardized procedure that has been used in previous studies of the human dive response.²³ In this procedure, the volunteer was asked to lie in a prone position on a comfortable, elevated surface. We attached a Nexfin monitor to the participant's right hand middle finger to measure heart rate as well as systolic and diastolic blood pressure. After attaching the device, the participant performed a test 'dive' in order to acclimatize to the procedure, followed 10 minutes of rest. A 'dive' included the

following: the volunteer took a deep breath and placed their face in a bowl of water that was 10°C cooler than room temperature and was elevated to the same height as the surface on which they were resting. When the volunteer could no longer hold their breath with reasonable comfort, they raised their face from the water and breathed normally. Following the test dive, the participant rested for an additional 10 minutes to return to a physiological baseline. The participant then performed three (non-test) dives with two minutes of rest between each dive.

QUANTIFICATION AND STATISTICAL ANALYSIS

Physiological analyses: We tested between-group differences in heart rate, spleen volume, and blood pressure measurements using an unpaired t-test. For heart rate, our sample sizes were as follows: Jeju controls: $n = 26$, Jeju Haenyeo: $n = 29$, Seoul controls: $n = 31$. For spleen volume, our sample sizes were as follows: Jeju controls: $n = 26$, Jeju Haenyeo: $n = 29$, Seoul controls: $n = 31$. For all blood pressure measurements, our sample sizes were as follows: Jeju controls: $n = 25$, Jeju Haenyeo: $n = 29$, Seoul controls: $n = 31$. In Figures 1 and 2, we plotted the distribution of values with error bars where the center represents the mean and the error bars indicate one standard deviation. Statistical significance in the unpaired t test, as plotted, was defined as *ns* = not significant ($p > 0.05$), * = $p \leq 0.05$, and ** = $p \leq 0.01$ (See Figures 1 and 2).

The measurement we selected for genetic association testing represented the maximum diastolic blood pressure during the first (non-test) dive. This dive represented the dive in which the participants were maximally normalized to each other because of the preceding 10 minute rest period and because they had already experienced a test dive to acclimatize them to the experimental setup. To further interrogate diastolic blood pressure during diving, we used a multiple linear regression model to account for covariates that could affect this measurement. We included dive duration as a covariate in our multiple linear regression model to account for any differences arising from dive time. Further details on these analyses can be found in RESULTS, Simulated dive measurements.

Genetic analyses

Preprocessing reads: We first assessed the quality of the reads using FastQC v0.11.8, which generates summary statistics and visualizations to evaluate and identify potential biases or artifacts in the sequencing data.⁶⁶ We found G repeats as overrepresented sequences and k-mer content. We did not detect adapter content; the base and sequence qualities were high with no warnings. We cleaned reads using PRINSEQ-lite 0.20.4, which filters and trims sequences based on quality scores, removes duplicate and low-complexity sequences, and eliminates reads containing ambiguous bases (Ns) to improve downstream analysis reliability.⁶⁷ We filtered low-complexity reads using the dust method with a threshold of 7. We also included a custom parameter to delete repeats of 50 Gs in a row.

We mapped the reads to the human reference genome GRCh38 using *bwa-mem*,⁶⁸ a Burrows-Wheeler Aligner algorithm optimized for high-throughput sequencing data. It performs local alignment, automatically detects split reads indicative of structural variants, and efficiently handles mismatches and indels, producing SAM format output for further

processing. We used samtools⁶⁹ to filter the bam files removing: unmapped reads, with mate unmapped, not primary alignment, reads that fail quality checks, and PCR or optical duplicates (-F 1804).

Imputation: Our average sequencing depth of 5x was selected deliberately for cost-effectiveness in a design based on imputation-based genotype calling. As a reference panel for the imputation, we used individuals from the Korean Genome Project (Korean1K).³³ From this resource, we obtained a vcf file with data from 400 randomly chosen individuals. We refer to this dataset as Korean1K (KOA). We performed imputation using *GLIMPSE v1.1.1*.³² Using *bcftools view*,⁷⁰ we only kept biallelic SNPs from the reference panel. We then calculated genotype likelihoods for each individual at each position. We did this by first using *bcftools mpileup* to annotate the number of reads in each position (-a 'FORMAT/DP'), excluding indels (-I) and recalculating BAQ (-E), and then by *bcftools call* to calculate the likelihoods, using multiallelic caller (-m), inserting sited missed in *mpileup* but present in the reference panel (-i), and keeping all possible alternative alleles (-a). We next split the genome into pieces with a window size of 2Mb and buffer size of 200,000 (GLIMPSE_chunks). Then, we imputed each chromosome per individual (GLIMPSE_phase) by 2Mb piece. We ligated all pieces together (GLIMPSE_ligate) and phased the positions by sampling haplotypes (GLIMPSE_sample). Finally, we merged our files to generate a file for each chromosome with all the individuals using *bcftools merge*.

Relatedness: We calculated relatedness using ngsToolsV2,⁷¹ a revision of ngsRelateV1 that now allows for estimation of relatedness from low depth sequencing data even in the presence of endogamy. This program uses identity by descent (IBD) to quantify the genetic relationship between two individuals. K0 is the proportion of the genome where two individuals do not share any segments by IBD, and we therefore used 0.75 as a threshold for K0 to determine relatedness between any given pair of samples. We found five pairs of individuals with K0>0.75. The 10 individuals of these pairs were non-overlapping, so we excluded one individual from each pair and recalculated relatedness. The excluded individuals included two from Jeju, two from the Korea1K project and one from Seoul. We recalculated the relatedness for the individuals and no remaining pairs had K0>0.75.

Processing samples and SNP filtering: We downloaded data from individuals representing select global populations of the 1000 Genomes Phase 3 Project.⁶⁵ All participants that make up our study data are female to match the demographics of the Haenyeo population. Therefore, we selected only female individuals from the 1000 genomes project for consistency. These individuals represented several East Asian populations; Japanese in Tokyo (JPT), Han Chinese In Beijing (CHB), Han Chinese South (CHS), and Kinh Vietnamese (KHV); as well as participants from Europe and Africa as outgroups: British in England and Scotland (GBR) and Yoruba from Nigeria (YRI).

We applied the strict 1000 genomes accessibility mask. We also filtered for biallelic sites, excluded sites with the flag 'MAF=0' to remove invariant sites, removed positions with the flag that the reference allele had been switched ('INFO=REF_SWITCH'), and applied an excess of heterozygosity filter with *bcftools* (ExHet<=.000001). These filters resulted in a set of 25,899,170 SNPs. We then merged the Korean data with the downloaded 1000

genomes samples and kept only positions present in both datasets, which resulted in a further reduction to 7,955,800 SNPs. We applied a 1% minor allele frequencies filter across individuals from all populations ($n = 468$ samples), resulting in 5,673,091 SNPs that were used in the downstream analyses.

Population structure analysis: We performed Principal Component Analysis (PCA) using plink,⁴⁴ which computes eigenvectors and eigenvalues from genotype data to capture population structure and genetic variation. First, we performed a PCA with only Korean individuals. PC1 separated Jeju Island from the mainland. The individuals from Seoul in our dataset cluster with those from the Korean 1K project. One individual from the Korean 1K project clustered with the Jeju residents. After quantifying this individual's ancestry and finding it consistent with Jeju, we re-labeled this individual as non-Haenyeo Jeju (JEU). We also identified five individuals from Jeju that appear to have predominantly mainland heritage, and we labeled them as Korean Mainlander in Jeju (KMJ). The remaining PCs do not further separate the Korean populations.

We also performed a PCA using all the Korean samples and other outgroup populations from the 1000 Genomes Project. We downloaded only female individuals from the following populations: Japanese in Tokyo (JPT), Han Chinese In Beijing (CHB), Han Chinese South (CHS), Kinh Vietnamese (KVV), British in England and Scotland (GBR) and Yoruba from Nigeria (YRI). PC1 and PC2 separate populations by continent. We observed three clusters: British (GBR), Yoruba (YRI), and Asian samples (HAE, JEU, SEO, KOA, JPT, CHB, CHS, KHV). PC3 and PC4 separate the Asian populations: PC3 separates populations in latitude, and PC4 separates them longitudinally. While Jeju Island individuals cluster with the Japanese in PC3/PC4, PC5 separates Japanese from Jeju Island individuals.

In order to estimate admixture proportions, we sampled 3% of sites to use as input for Ohana.³⁴ We ran *qpas*, which estimates individual's ancestry proportions assuming k ancestral components using maximum likelihood. We varied k from 4 to 7, with a maximum of 450 iterations, and used 0.08 as the minimum likelihood delta between iterations. From both the PCA and the structure analysis it became apparent that five of the individuals from Jeju (individuals 39,44,47,48, and 60) clustered and shared ancestry components with mainland Korean participants and therefore were relabeled as Korean mainlanders in Jeju (KMJ). At $k=7$, we noticed that one of the outgroups (YRI) was split into two groups, and we consequently used a supervised version of the algorithm, where individuals from the outgroups (YRI and GBR) were forced to have 100% ancestry from their respective outgroups. We used pong to visualize the structure plots.⁷⁸ We used nemeco from Ohana to model the allele frequencies across ancestral components and obtain a matrix of the variances and covariances between them. We then used this matrix to estimate a tree of the ancestry components.

Demographic history and population divergence: In order to understand population divergence in our data, we calculated the Weir and Cockerman fixation index (F_{ST}) using plink2 (`-fst "CATPHENO" 'method=wc'`).^{72,73} F_{ST} values representing genetic differentiation between pairs of relevant populations can be found in Table 1.

We used two different methods for inferring a population tree and admixture events between populations: Treemix,⁷⁴ which uses a Gaussian approximation for allele frequency change and a greedy search, and requires the user to specify the number of admixture events, and Admixture Bayes,⁷⁵ which assumes a similar model but uses Bayesian inference and Markov Chain Monte Carlo to explore the space of possible admixture graphs. The latter method infers a credible set of admixture graphs and the number of admixture events from the data using a prior that penalizes having too many admixture events. For these analyses, we merged the individuals from Seoul and the Korean1K project into a single group labeled KOR. Similarly, we merged the Haenyeo and Jeju controls into a single population labeled JEJ and representing all Jeju residents. In both analyses, we used British individuals (GBR) as an outgroup. The graphs generated by both programs are consistent with each other and with our F_{ST} , PCA, and admixture analyses.

To understand the effective population size (N_e) history of Korean populations, we used the method implemented in Relate^{76,77} for estimating trajectories of N_e . To provide an initial estimate of N_e for Ancestral recombination Graph (ARG) estimation in Relate, we first calculated the average number of pairwise differences (π) in KOR and JEJ populations using vcfTools -windows-pi, setting the window size to 250,000,000, to get an estimate of π per chromosome. Then, we estimated effective population size using $\widehat{N_e} = \pi/4\mu$. Using a mutation rate, $\mu=1.25e-8$, the average N_e per chromosome was $N_{eJEJ}=5204$, and $N_{eKOR}=5194$. We then estimated the ARG using Relate.^{76,77} We used $N_e=5,000$ as an initial estimate of population size in Relate to jointly infer population size over time for the JEJ and KOR populations. We chose a generation time of 27 years, consistent with previous estimates of generation times in East Asian populations⁷⁹ We used Relate^{76,77} to infer ancestral recombination graphs (ARGs). We set the mutation rate to 1.25×10^{-8} and the effective population size of haplotypes (N) to 10,000. We then estimated population size over time for both the JEJ and KOR populations.

Selection scan: We performed a selection scan in order to identify regions of the Jeju population genome that have been under selection. We hypothesized that the dangers of breath hold diving, especially during pregnancy, combined with the integral history of this practice in Jeju, will have affected genetic variation in the population in ways that we can identify using selection analyses. For these analyses, we used the same 5,673,091 SNPs as those used in the demographic analyses. We focused on participants from only the following populations: Haenyeo (HAE), non-Haenyeo Jeju residents (JEU), Korean Mainlanders in Jeju (KMJ), Korean1K (KOA), Seoul (SEO), and Han Chinese South (CHS). Based on the structure analysis, two distinct Korean ancestral components emerged, one for HAE and JEU (representing Jeju heritage) and another for KOA, SEO, and KMJ (representing mainland heritage). We therefore decided to categorize the populations as follows for subsequent analyses: Jeju residents (JEJ), consisting of HAE and JEU, and Korean mainland individuals (KOR), comprising KOA, KMJ, and SEO. We chose to restrict our analysis to these three populations rather than using all global populations used in the demographic analyses to avoid identifying as outliers SNPs that are fixed in 1000 genomes and unique to the Korean populations, an artifact of using many outgroup populations instead of restricting the number of comparison populations to two.

We considered the results from $K = 3$ for downstream analysis as this is the number of ancestry components for which each of the three populations (Jeju residents, mainland Koreans, and Han Chinese) obtain their own distinct component. We considered mainland Koreans and Han Chinese from the South (CHS) as comparison populations for the selection scan. To estimate admixture proportions, we sampled 3% of sites to use as input for Ohana.³⁴ We ran qpas, varying K from 2 to 3, with a maximum of 450 iterations, and used 0.08 as the minimum likelihood delta between iterations. We used nemeco from Ohana to model the allele frequencies across ancestral components and obtain a matrix of the variances and covariances between them. We then used this matrix to estimate a tree of the ancestry components.

We employed the selscan program in Ohana³⁴ to identify SNPs that exhibited significant allele frequency changes in Jeju residents compared to that expected by the genome-wide covariance structure, using a likelihood ratio test. In selscan, this is done by first estimating a covariance matrix of allele frequencies among populations for the genome-wide data. For the specific SNP tested, a scalar variable is then multiplied on the variance component associated with population of interest (in this case the Jeju population), and the value of this scalar parameter is estimated using maximum likelihood. This leads to the establishment of a likelihood ratio test comparing a model in which the allele frequency change is as expected from the genome-wide pattern and one which allows increased allele frequency change for the SNP being tested. In our case, large likelihood ratios indicate support for the hypothesis of increased change in allele frequency in a SNP the Jeju population compared to that expected from the genome-wide pattern.

We used the same 5,673,091 SNPs as in the previous analyses, but eliminated apparent fixed differences between the Korean1K data and the 1000 Genomes data set, assuming that these SNPs were suffering from allele flipping errors (see e.g.,⁸⁰). We verified that a selection of these SNPs were allele flips in the 1000 Genomes project by inspecting the original reads from this project using samtools pileup. A total of 40 SNPs were excluded for this reason, leaving a total of 5,673,051 (See Table S3 for excluded SNPs).

To select outliers from selscan, we identified SNPs with likelihood ratios in the 99th percentile. We kept only outliers with at least one other SNP within in a 200 kb window (± 100 kb from the SNP) that also was in the 99th percentile. To select the top SNP from each selection signal, we removed the outliers with a neighboring SNP in a 200kb window (± 100 kb of the outlier) that had a higher likelihood ratio. Among our top candidate SNPs, multiple hits were located within a wide peak in chromosome 8. This peak was located around the centromere of chromosome 8 and spanned more than 5Mb, much larger than the 200kb maximum peak size assumed by our filters. Despite multiple apparent hits, the SNPs in this region represent a single, wide peak caused by lower levels of recombination at the centromere of metacentric (and submetacentric) chromosomes⁸¹ See Figure 6. We therefore kept only the top SNP within that region as the representative SNP of the signal from that peak, the strongest signal of selection in our results. We identified the next 9 strongest signals of selection not found within the peak for a total of 10 SNPs. We performed gene annotation using the GRCh38 annotation (NCBI Accession: PRJNA31257 ID: 3125). We

filtered the gff file to extract the gene coordinates and annotated a SNP if it fell within the gene, including intronic regions.

In order to calibrate the p-values obtained from our *seiscan*, we used an inflation factor of $\lambda=1.94$. The likelihood ratio test *seiscan* employs is bounded at zero, implying that the expected distribution of the log-likelihood ratio (LLR) statistic ($-2 * \text{LLR}$) is a 50:50 mixture of a point mass at zero and a χ^2_1 distribution, i.e. we would expect 50% of the p-values to be equal to one under the null hypothesis. For the genomic control, we therefore excluded the lowest half of all LLRs, all of which had a value of zero. Worth noting is that, in fact, 73.9% of all SNPs we examined (4,190,414 out of 5,675,051) had a likelihood ratio of zero. This additional zero inflation is due to the discrete nature of the SNP allele frequency data, and we do not do further correction for this. For the remaining 50% of SNPs (2,836,525 SNPs), we compared their LLRs to the expected χ^2_1 distribution to determine the inflation factor. Commonly, the inflation factor, λ , is calculated using ratio of the medians of the empirical and theoretical distributions. However, due to the excess of zero values even after excluding the lowest 50%, it would not be appropriate to use the median for genomic control. Instead, we use 99th percentile to correct, i.e. we calculated the inflation factor as

$$\lambda = \frac{P_{99}^{\text{observed}}}{P_{99}^{\chi^2_1}}$$

We divided the LLRs by this inflation factor and recalculated the p-values to obtain the calibrated p-values (Figure S4).

Genetic association: We ran an association test using PLINK⁴⁴ which performed a linear regression analysis on the top 10 SNPs from the selection scan and untransformed maximum diastolic blood pressure (BP) during the first simulated dive. The analysis included the following covariates: age; height; weight; dive time; and the first 10 principal components (PCs), to control for any underlying population structure. The linear regression model was adjusted for these covariates, and we used an additive genetic model for SNP analysis. We checked for inflation in the association values with a qq-plot genome-wide and found that values were not inflated ($\lambda = 1.01$, Figure S5), suggesting that the first 10PCS properly control for underlying population structure. We therefore continued with association testing using the untransformed maximum diastolic pressure values. Only one SNP from the selection scan, the 4th strongest signal, was significantly associated with maximum diving diastolic BP at the Bonferroni-corrected significance level (0.005) using an additive model: rs66930627 ($p = 0.0021$). When we included age² as an additional covariate to account for non-linear effects, as is common in BP GWAS,⁴⁵ the association remained significant ($p = 0.0014$). We also tested the association within only Jeju participants, to assess whether the association reflects population-specific allele frequency differences between Jeju and Seoul and found that the association remains significant ($p = 0.0039$). We note that we were unable to include use of anti-hypertensives as an additional covariate given that these data were not collected at the time of the study.

The associated SNP, rs66930627, is found in an intergenic region, however it has a FORGEdb score of 7 indicating it likely has a regulatory function. FORGEdb scores range between 0 and 10, with 10 indicating highest likelihood of affecting regulation, and are calculated combining multiple sources of information regarding regulation, including data for transcription factor (TF) binding and chromatin accessibility. Specifically, we used the GTEx portal⁸² to identify a strong expression quantitative trait locus (eQTL) with this SNP and the gene that encodes Fc γ receptor IIA (Fc γ RIIA, $p = 0.00012$) in skeletal muscle (Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Fc γ RIIA is a low-affinity Fc γ receptor that binds IgG2 and may influence hypertension through its effect on the pro-inflammatory activities of vascular smooth muscle cells (VSMCs).⁴⁶ Another low-affinity Fc γ R (Fc γ RIIB) has been demonstrated to correlate with hypertension directly,⁴⁸ and has been suggested that it influences vascular contraction. Fc γ receptors, therefore, may play a crucial role in regulating diastolic blood pressure via their effect on vasculature. Using GTEx, we also found a strong eQTL with rs66930627 and the pseudogene or lncRNA Heat shock 70kDa protein 7 (HSPA7) ($p = 0.0000012$) in whole blood. Knockdown of HSPA7 has been shown to inhibit the expression of inflammatory mediators, thus highlighting its potential role in vascular inflammation.

Selection validation: In order to validate that natural selection has acted on the variant we identified as under selection (using the *selscan* genome-wide selection scan) and associated with our phenotype of interest, we used an alternative strategy for detecting selection. Specifically, we tested selection over time using CLUES2.³⁵ Using our previous results from Relate (see demographic history and population divergence), we extracted the subtrees corresponding to the JEJ population using Relate's SubTreesForSubpopulation feature. We then used the SampleBranchLengths function with the inferred historic JEJ population size to extract 200 samples of the gene tree at the SNP rs66930627 in Newick format. These samples were then converted to the CLUES2 input format. We then ran CLUES2 with the default 450 discretization points and a time cutoff of 450 generations (12150 years) on our sample of trees for 3 distinct models: constant selection through time, a selection breakpoint of 45 generations (1215 years), and a selection breakpoint of 259 generations (6993 years). These generation-to-year conversions are based on using a generation time of 27 years.

All of Us association validation: We were also able to validate the association between diastolic blood pressure in All of Us ($p = 0.040$) with the following filters: WGS data, "Are you currently prescribed medications and/or receiving treatment for high blood pressure (hypertension)?" (no, to reduce the confounding effect of antihypertensives) Ages 55–75 (to match the ages in the Korean cohort), and Race: White (to produce an approximately single-ancestry cohort: European). This represented the largest single-ancestry cohort currently available through All of Us with the aforementioned filters. We also removed samples flagged as related, resulting in a final cohort size of $n = 1376$ individuals. We used a multiple linear regression to calculate association using the first 10 PCS, age, height, and weight as covariates, as in the Korean cohort (Table S4). Systolic blood pressure was not significantly associated at the same locus.

ADDITIONAL RESOURCES

Newly generated, de-identified whole genome sequencing data have been deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) as PRJNA1235291. They are publicly available as of the date of publication. All original code has been deposited at <https://github.com/aguilar-gomez/haenyeo> and is publicly available as of the date of publication.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank first and foremost the participants in our study, including the Haenyeo community from Hado, Jeju controls from Jeju City, and Seoul controls, for their participation and support. We acknowledge All of Us participants for their contributions and thank the National Institutes of Health's [All of Us Research Program](#) for making available the participant data examined in this study. We would also like to acknowledge DNALINK, Inc. and its staff. This project was supported by the Office of Naval Research (N00014-20-1-2556), the National Institute of Health (NIGMS R35GM153400 and NHGRI R00HG011658), the National Science Foundation (Graduate Research Fellowship 2146752), the UC MEXUS-CONACYT Doctoral Fellowship, and Fulbright-García Robles. The Genotype-Tissue Expression (GTEx) Project was supported by the common fund of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. For this project, we used data from the GTEx Analysis Release v.8 (dbGaP: phs000424.v8.p2).

REFERENCES

1. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, et al. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science* 329, 72–75. 10.1126/science.1189406. [PubMed: 20466884]
2. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329, 75–78. [PubMed: 20595611]
3. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, Korneliussen TS, Gerbault P, Skotte L, Linneberg A, et al. (2015). Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 349, 1343–1347. [PubMed: 26383953]
4. Rees JS, Castellano S, and Andrés AM (2020). The genomics of human local adaptation. *Trends Genet.* 36, 415–428. [PubMed: 32396835]
5. Ilardo MA, Moltke I, Korneliussen TS, Cheng J, Stern AJ, Racimo F, de Barros Damgaard P, Sikora M, Seguin-Orlando A, Rasmussen S, et al. (2018). Physiological and Genetic Adaptations to Diving in Sea Nomads. *Cell* 173, 569–580.e15. 10.1016/j.cell.2018.03.054. [PubMed: 29677510]
6. Yang C, Yang S, and O'Grady W (2019). *Jejueo: The Language of Korea's Jeju Island* (University of Hawaii Press).
7. Ferretti G, and Costa M (2003). Diversity in and adaptation to breathe-hold diving in humans. *Comp. Biochem. Physiol. Mol. Integr. Physiol* 136, 205–213.
8. Hong SK, and Rahn H (1967). The diving women of Korea and Japan. *Sci. Am* 216, 34–43. [PubMed: 6042533]
9. Lee J-Y, and Lee HH (2014). Korean women divers 'Haenyeo': bathing suits and acclimatization to cold. *J. Hum. Environ. Syst* 17, 001–011.
10. Foster GE, and Sheel AW (2005). The human diving response, its function, and its control. *Scand. J. Med. Sci. Sports* 15, 3–12. [PubMed: 15679566]
11. Asmussen E, and Kristiansson NG (1968). The "diving bradycardia" in exercising man. *Acta Physiol. Scand* 73, 527–535. [PubMed: 5708178]

12. Arnold RW (1985). Extremes in human breath hold, facial immersion bradycardia. *Undersea Biomed. Res* 12, 183–190. [PubMed: 4071845]
13. Campbell LB, Gooden BA, and Horowitz JD (1969). Cardiovascular responses to partial and total immersion in man. *J. Physiol* 202, 239–250. [PubMed: 5770894]
14. Gooden BA (1994). Mechanism of the human diving response. *Integr. Physiol. Behav. Sci* 29, 6–16. [PubMed: 8018553]
15. Lin YC, Shida KK, and Hong SK (1983). Effects of hypercapnia, hypoxia, and rebreathing on circulatory response to apnea. *J. Appl. Physiol* 54, 172–177. [PubMed: 6826401]
16. Zapol WM, Liggins GC, Schneider RC, Qvist J, Snider MT, Creasy RK, and Hochachka PW (1979). Regional blood flow during simulated diving in the conscious Weddell seal. *J. Appl. Physiol* 47, 968–973. [PubMed: 511722]
17. Elsner R, Franklin DL, Van Citters RL, and Kenney DW (1966). Cardiovascular defense against asphyxia. *Science* 153, 941–949. [PubMed: 5917556]
18. Lindholm P, Sundblad P, and Linnarsson D (1999). Oxygen-conserving effects of apnea in exercising men. *J. Appl. Physiol* 87, 2122–2127. 10.1152/jappl.1999.87.6.2122. [PubMed: 10601158]
19. Bakovic D, Eterović D, Saratlija-Novaković Z, Palada I, Valic Z, Bilopavlović N, and Dujčić Z (2005). Effect of human splenic contraction on variation in circulating blood cell counts. *Clin. Exp. Pharmacol. Physiol* 32, 944–951. 10.1111/j.1440-1681.2005.04289.x. [PubMed: 16405451]
20. Bakovic D, Valic Z, Eterović D, Vukovic I, Obad A, Marinović-Terzić I, and Dujčić Z (2003). Spleen volume and blood flow response to repeated breath-hold apneas. *J. Appl. Physiol* 95, 1460–1466. 10.1152/japplphysiol.00221.2003. [PubMed: 12819225]
21. Schagatay E, Andersson JP, Hallen M, and Palsson B (2001). Selected contribution: role of spleen emptying in prolonging apneas in humans. *J. Appl. Physiol* 90, 1623–1629, discussion 1606. 10.1152/jappl.2001.90.4.1623. [PubMed: 11247970]
22. Stewart IB, and McKenzie DC (2002). The human spleen during physiological stress. *Sports Med.* 32, 361–369. [PubMed: 11980500]
23. Baranova TI, Berlov DN, Glotov OS, Korf EA, Minigalin AD, Mitrofanova AV, Ahmetov II, and Glotov AS (2016). Genetic determination of the vascular reactions in humans in response to the diving reflex. *Am. J. Physiol. Heart Circ. Physiol* 312, H622–H631. [PubMed: 27923785]
24. Jaimchariyatam N, Na-Rungsri K, Tungsanga S, Lertmaharit S, Lohsoonthorn V, and Totientchai S (2019). Obstructive sleep apnea as a risk factor for preeclampsia–eclampsia. *Sleep Breath.* 23, 687–693. [PubMed: 30484015]
25. Pankow W, Nabe B, Lies A, Becker H, Köhler U, Kohl FV, and Lohmann FW (1997). Influence of sleep apnea on 24-hour blood pressure. *Chest* 112, 1253–1258. [PubMed: 9367465]
26. Kario K (2009). Obstructive sleep apnea syndrome and hypertension: ambulatory blood pressure. *Hypertens. Res* 32, 428–432. [PubMed: 19494815]
27. Bouten J, Caen K, Stautemas J, Lefevere F, Derave W, Lootens L, Van Eenoo P, Bourgois JG, and Boone J (2019). Eight weeks of static apnea training increases spleen volume but not acute spleen contraction. *Respir. Physiol. Neurobiol* 266, 144–149. 10.1016/j.resp.2019.04.002. [PubMed: 31009754]
28. Das KC, Mukherjee M, Sarkar TK, Dash RJ, and Rastogi GK (1975). Erythropoiesis and erythropoietin in hypo- and hyperthyroidism. *J. Clin. Endocrinol. Metab* 40, 211–220. 10.1210/jcem-40-2-211. [PubMed: 1117976]
29. Ilardo M, Dos Santos MCF, Grote Beverborg N, Rajan M, Said MA, Verweij N, Van Der Harst P, Van Der Meer P, and Leibold EA (2021). An Erythropoietin-Independent Mechanism of Erythrocytic Precursor Proliferation Underlies Hypoxia Tolerance in Sea Nomads. *Front. Physiol* 12, 760851. 10.3389/fphys.2021.760851. [PubMed: 35177992]
30. Whayne TF Jr., and Killip T 3rd. (1967). Simulated diving in man: comparison of facial stimuli and response in arrhythmia. *J. Appl. Physiol* 22, 800–807. [PubMed: 6023194]
31. Kim JY, Kang K, Kang J, Koo J, Kim DH, Kim BJ, Kim WJ, Kim EG, Kim JG, Kim JM, et al. (2019). Executive Summary of Stroke Statistics in Korea 2018: A Report from the Epidemiology Research Council of the Korean Stroke Society. *J. Stroke* 21, 42–59. 10.5853/jos.2018.03125. [PubMed: 30558400]

32. Rubinacci S, Ribeiro DM, Hofmeister RJ, and Delaneau O (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet* 53, 120–126. 10.1038/s41588-020-00756-0. [PubMed: 33414550]
33. Jeon S, Bhak Y, Choi Y, Jeon Y, Kim S, Jang J, Jang J, Blazyte A, Kim C, Kim Y, et al. (2020). Korean Genome Project: 1094 Korean personal genomes with clinical information. *Sci. Adv* 6, eaaz7835. 10.1126/sciadv.aaz7835. [PubMed: 32766443]
34. Cheng JY, Mailund T, and Nielsen R (2017). Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics* 33, 2148–2155. [PubMed: 28334108]
35. Vaughn A, and Nielsen R (2023). Fast and accurate estimation of selection coefficients and allele histories from ancient and modern DNA. Preprint at bioRxiv. 10.1101/2023.12.16.572012.
36. Kim K-K, Cho Y-S, Cho N-H, Shin C, and Kim J-W (2010). A Genome-wide Association Study of Copy Number Variation in Hematological Parameters in the Korean Population. *Genom. Informatics* 8, 122–130.
37. Fontanillas P, Kless A, Bothmer J, and Tung JY; 23andMe Research Team (2022). Genome-wide association study of pain sensitivity assessed by questionnaire and the cold pressor test. *Pain* 163, 1763–1776. 10.1097/j.pain.0000000000002568. [PubMed: 34924555]
38. Lee JY, Lee HH, Kim S, Jang YJ, Baek YJ, and Kang KY (2016). Diving bradycardia of elderly Korean women divers, haenyeo, in cold seawater: a field report. *Ind. Health* 54, 183–190. 10.2486/indhealth.2015-0043. [PubMed: 26632118]
39. Lee JY, Park J, Koh E, and Cha S (2017). Finger cold-induced vasodilation of older Korean female divers, haenyeo: effects of chronic cold exposure and aging. *Int. J. Biometeorol* 61, 1299–1307. 10.1007/s00484-017-1308-2. [PubMed: 28160059]
40. Park J, Kim S, Kim DH, Cha S, and Lee JY (2018). Whole-body cold tolerance in older Korean female divers “haenyeo” during cold air exposure: effects of repetitive cold exposure and aging. *Int. J. Biometeorol* 62, 543–551. 10.1007/s00484-017-1463-5. [PubMed: 29098422]
41. Paik KS, Kang BS, Han DS, Rennie DW, and Hong SK (1972). Vascular responses of Korean ama to hand immersion in cold water. *J. Appl. Physiol* 32, 446–450. 10.1152/jappl.1972.32.4.446. [PubMed: 5026490]
42. Hong SK (1973). Pattern of cold adaptation in women divers of Korea (ama). *Fed. Proc* 32, 1614–1622. [PubMed: 4705209]
43. Hong SK, Lee CK, Kim JK, Song SH, and Kennie DW (1969). Peripheral blood flow and heat flux of Korean women divers. *Fed. Proc* 28, 1143–1148. [PubMed: 5783513]
44. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, and Sham PC (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet* 81, 559–575. [PubMed: 17701901]
45. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, et al. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet* 50, 1412–1425. [PubMed: 30224653]
46. Ryu J, Lee CW, Shin JA, Park CS, Kim JJ, Park SJ, and Han KH (2007). FcγRIIa mediates C-reactive protein-induced inflammatory responses of human vascular smooth muscle cells by activating NADPH oxidase 4. *Cardiovasc. Res* 75, 555–565. 10.1016/j.cardiores.2007.04.027. [PubMed: 17531211]
47. Tanigaki K, Sundgren N, Khera A, Vongpatanasin W, Mineo C, and Shaul PW (2015). Fcγ receptors and ligands and cardiovascular disease. *Circ. Res* 116, 368–384. [PubMed: 25593280]
48. Song X, Zou X, Ge W, Hou C, Cao Z, Zhao H, Zhang T, Jin L, Fu Y, Kong W, et al. (2021). Blocking FcγRIIb in Smooth Muscle Cells Reduces Hypertension. *Circ. Res* 129, 308–325. 10.1161/CIRCRESAHA.120.318447. [PubMed: 33980031]
49. Wang Y, Sugita N, Kikuchi A, Iwanaga R, Hirano E, Shimada Y, Sasahara J, Tanaka K, and Yoshie H (2012). FcγRIIb-nt645+ 25A/G gene polymorphism and periodontitis in Japanese women with preeclampsia. *Int. J. Immunogenet* 39, 492–500. [PubMed: 22594540]
50. Davis RS (2007). Fc receptor-like molecules. *Annu. Rev. Immunol* 25, 525–560. 10.1146/annurev.immunol.25.022106.141541. [PubMed: 17201682]

51. Ann S. j., Bang H, Lee CJ, Oh J, Park S, Kang SM, Choi JK, and Lee SH (2021). LncRNA HSPA7 in human atherosclerotic plaques sponges miR-223 and promotes the proinflammatory vascular smooth muscle cell transition. *Exp. Mol. Med* 53, 1842–1849. [PubMed: 34857901]
52. Cho S, Kim MY, and Lee SD (2023). Biogeographic origin and genetic characteristics of the peopling of Jeju Island based on lineage markers. *Genes Genomics* 45, 307–318. 10.1007/s13258-022-01363-5. [PubMed: 36607592]
53. Yang C, O’Grady W, Yang S, Hilton NH, Kang SG, and Kim SY (2020). Revising the language map of Korea. In *Handbook of the Changing World Language Map* (Springer Nature), pp. 215–229.
54. Nørgaard SK, Vestgaard MJ, Jørgensen IL, Ásbjörnsdóttir B, Ring-holm L, McIntyre HD, Damm P, and Mathiesen ER (2018). Diastolic blood pressure is a potentially modifiable risk factor for preeclampsia in women with pre-existing diabetes. *Diabetes Res. Clin. Pract* 138, 229–237. [PubMed: 29475019]
55. Moore LG, Shriver M, Bemis L, Hickler B, Wilson M, Brutsaert T, Parra E, and Vargas E (2004). Maternal adaptation to high-altitude pregnancy: an experiment of nature—a review. *Placenta* 25, S60–S71. [PubMed: 15033310]
56. Choi E-S, Jung YM, Kim D, Cho SE, Park ES, Park CW, Park JS, Jun JK, and Lee SM (2024). Long-term cardiovascular outcome in women with preeclampsia in Korea: a large population-based cohort study and meta-analysis. *Sci. Rep* 14, 7480. [PubMed: 38553468]
57. Wu P, Haththotuwa R, Kwok CS, Babu A, Kotronias RA, Rushton C, Zaman A, Fryer AA, Kadam U, Chew-Graham CA, and Mamas MA (2017). Preeclampsia and future cardiovascular health: a systematic review and meta-analysis. *Circ. Cardiovasc. Qual. Outcomes* 10, e003497. [PubMed: 28228456]
58. Collaboration PS (2002). Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 360, 1903–1913. [PubMed: 12493255]
59. Padmanabhan S, and Dominiczak AF (2021). Genomics of hypertension: the road to precision medicine. *Nat. Rev. Cardiol* 18, 235–250. [PubMed: 33219353]
60. Zeng J, de Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, Yap CX, Xue A, Sidorenko J, McRae AF, et al. (2018). Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet* 50, 746–753. [PubMed: 29662166]
61. Bonneau A, Friemel F, and Lapierre D (1989). Electrocardiographic Aspects of Skin Diving. *Eur. J. Appl. Physiol* 58, 487–493. 10.1007/Bf02330702.
62. Ferrigno M, Grassi B, Ferretti G, Costa M, Marconi C, Cerretelli P, and Lundgren C (1991). Electrocardiogram during Deep Breath-Hold Dives by Elite Divers. *Undersea Biomed. Res* 18, 81–91. [PubMed: 2042264]
63. Hong SK, Moore TO, Seto G, Park HK, Hiatt WR, and Bernauer EM (1970). Lung Volumes and Apneic Bradycardia in Divers. *J. Appl. Physiol* 29, 172–176. 10.1152/jappl.1970.29.2.172. [PubMed: 5428891]
64. Schagatay E, van Kampen M, Emanuelsson S, and Holm B (2000). Effects of physical and apnea training on apneic time and the diving response in humans. *Eur. J. Appl. Physiol* 82, 161–169. 10.1007/s004210050668. [PubMed: 10929209]
65. 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, and Abecasis GR (2015). A global reference for human genetic variation. *Nature* 526, 68–74. 10.1038/nature15393. [PubMed: 26432245]
66. Andrews S (2010). *Babraham Bioinformatics* (Cambridge, United Kingdom: Babraham Institute).
67. Schmieder R, and Edwards R (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. 10.1093/bioinformatics/btr026. [PubMed: 21278185]
68. Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. 10.48550/arXiv.1303.3997.
69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. 10.1093/bioinformatics/btp352. [PubMed: 19505943]

70. Danecek P, and McCarthy SA (2017). BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* 33, 2037–2039. 10.1093/bioinformatics/btx100. [PubMed: 28205675]
71. Hanghoj K, Moltke I, Andersen PA, Manica A, and Korneliussen TS (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience* 8, giz034. 10.1093/gigascience/giz034. [PubMed: 31042285]
72. Purcell S & Chang C PLINK2 (v1. 90b6. 9) Available online: www.cog-genomics.org/plink/2.0/.
73. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, and Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7. [PubMed: 25722852]
74. Pickrell JK, and Pritchard JK (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8, e1002967. [PubMed: 23166502]
75. Nielsen SV, Vaughn AH, Leppälä K, Landis MJ, Mailund T, and Nielsen R (2023). Bayesian inference of admixture graphs on Native American and Arctic populations. *PLoS Genet.* 19, e1010410. [PubMed: 36780565]
76. Speidel L, Cassidy L, Davies RW, Hellenthal G, Skoglund P, and Myers SR (2021). Inferring population histories for ancient genomes using genome-wide genealogies. *Mol. Biol. Evol* 38, 3497–3511. [PubMed: 34129037]
77. Speidel L, Forest M, Shi S, and Myers SR (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet* 51, 1321–1329. [PubMed: 31477933]
78. Behr AA, Liu KZ, Liu-Fang G, Nakka P, and Ramachandran S (2016). pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* 32, 2817–2823. 10.1093/bioinformatics/btw327. [PubMed: 27283948]
79. Wang RJ, Al-Saffar SI, Rogers J, and Hahn MW (2023). Human generation times across the past 250,000 years. *Sci. Adv* 9, eabm7047. [PubMed: 36608127]
80. Clarke GM, and Cardon LR (2010). Aspects of Observing and Claiming Allele Flips in Association Studies. *Genet. Epidemiol* 34, 266–274. 10.1002/gepi.20458. [PubMed: 20013941]
81. Nachman MW (2002). Variation in recombination rate across the genome: evidence and implications. *Curr. Opin. Genet. Dev* 12, 657–663. 10.1016/S0959-437x(02)00358-1. [PubMed: 12433578]
82. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, et al. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet* 45, 580–585. [PubMed: 23715323]

Highlights

- Evidence of selection that may increase the safety of diving during pregnancy
- Regular diving increases the magnitude of bradycardia in response to dive stimulus
- The Haenyeo may represent the second known population evolved for diving

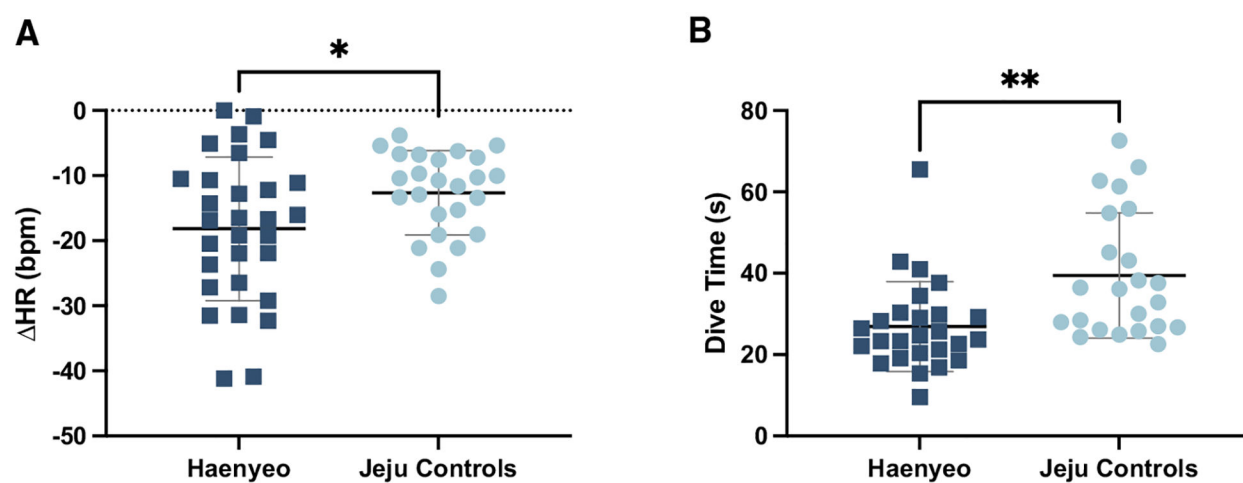


Figure 1. Haenyeo display a significant decrease in HR compared to Jeju controls during a simulated dive despite significantly shorter dive times

(A and B) Due to the shared ancestry within the population of Jeju, this is likely a training effect rather than determined by genetic variation. Data are represented as mean $\pm$ SD. Statistical significance was determined using an unpaired t test. * $p \leq 0.05$, ** $p \leq 0.01$.

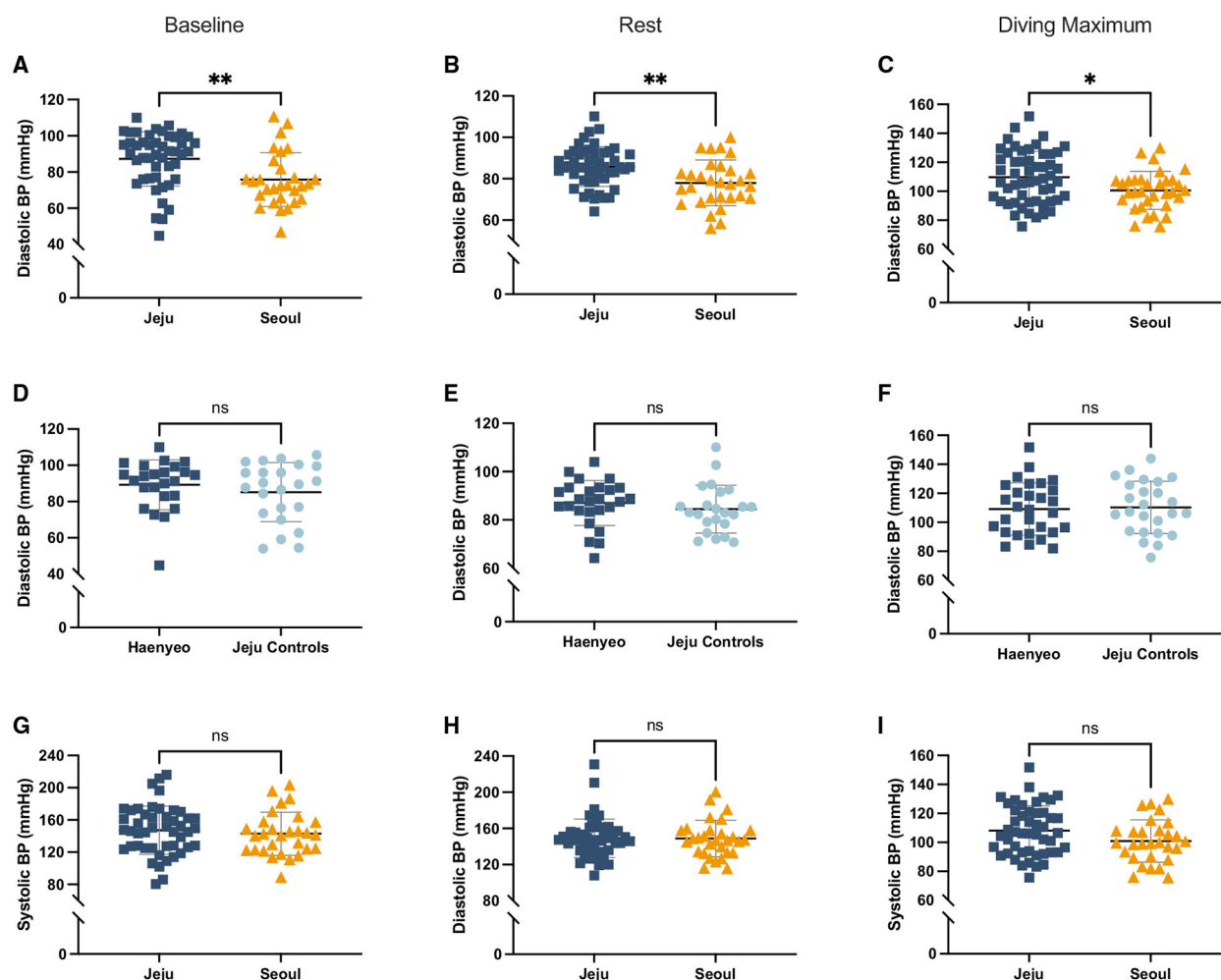


Figure 2. Diastolic BP is significantly higher in Jeju residents than in Seoul controls at all measured time points

(A–C) Diastolic BP is higher in Jeju participants than in Seoul participants at baseline (A), during a simulated dive (B), and during the 2 min recovery between dives (C).

(D–F) Diastolic BP is not significantly different between Jeju controls and Haenyeo at any of the same time points. These results suggest underlying genetic variation that contributes to elevated diastolic BP.

(G–I) There is no difference in Systolic BP between Jeju and Seoul at any time point.

Data are represented as mean $\pm$ SD. Statistical significance was determined using an unpaired t test. ns, not significant ($p > 0.05$); * $p \leq 0.05$, ** $p \leq 0.01$.

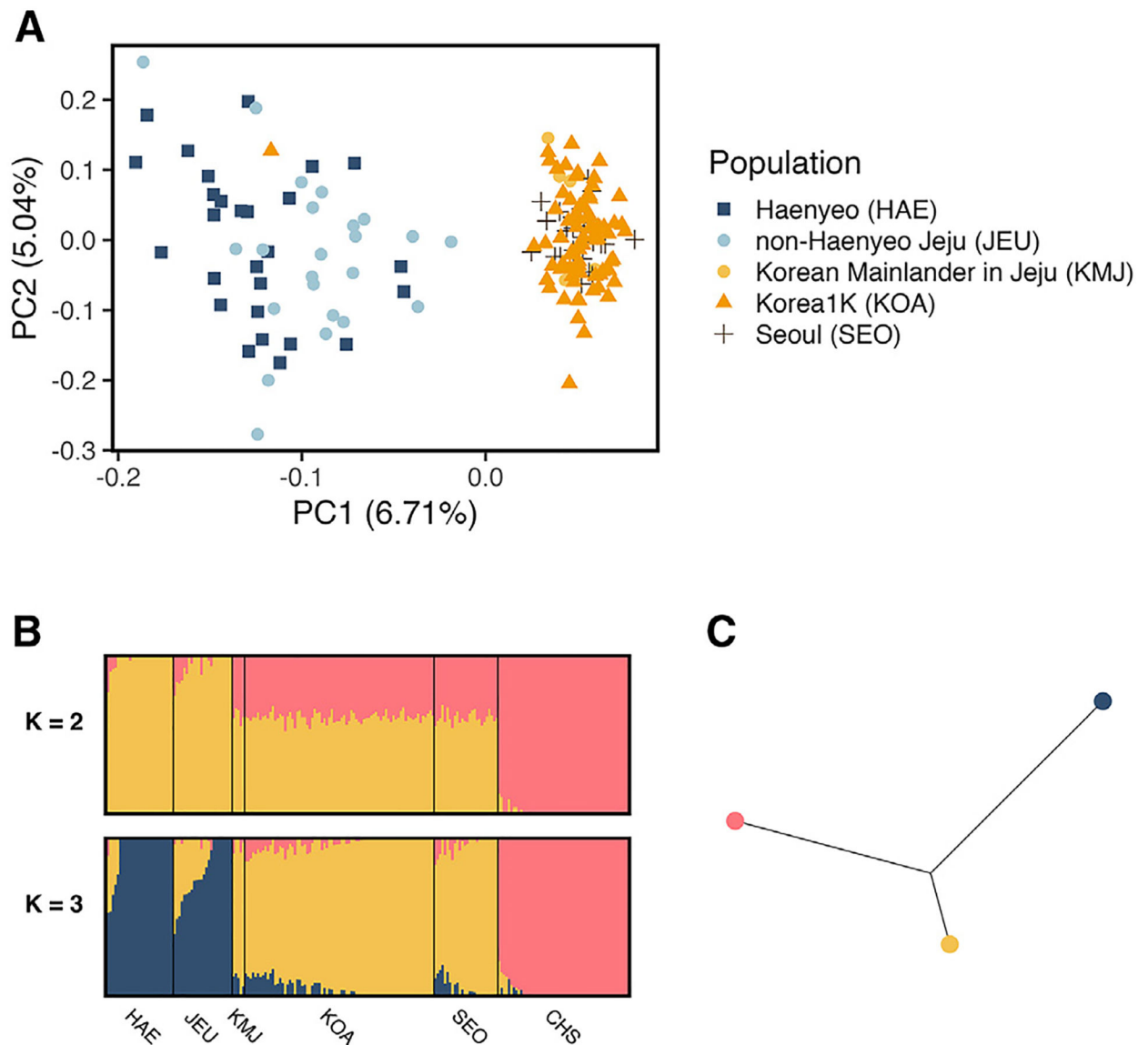


Figure 3. Genetic relationships between the Haenyeo, Jeju Island controls, and mainland Koreans

(A) In a PCA of only Korean individuals, the Haenyeo cluster with other JEUs, and all Jeju residents separate from mainland Koreans. F_{ST} values indicate that the Jeju Island populations are approximately panmictic, sharing an ancestral population, and distinct from the Korean mainland populations.

(B) Admixture graph at $K = 2$ and $K = 3$ for the Haenyeo, mainland Koreans, and Han Chinese highlights the genetic difference between Jeju individuals (Haenyeo Jeju controls) and mainland Koreans, with Jeju individuals receiving their own, distinct component.

(C) Admixture tree estimate for the components in (B) at $K = 3$. The Jeju branch of the tree estimate is notably the longest of the three.

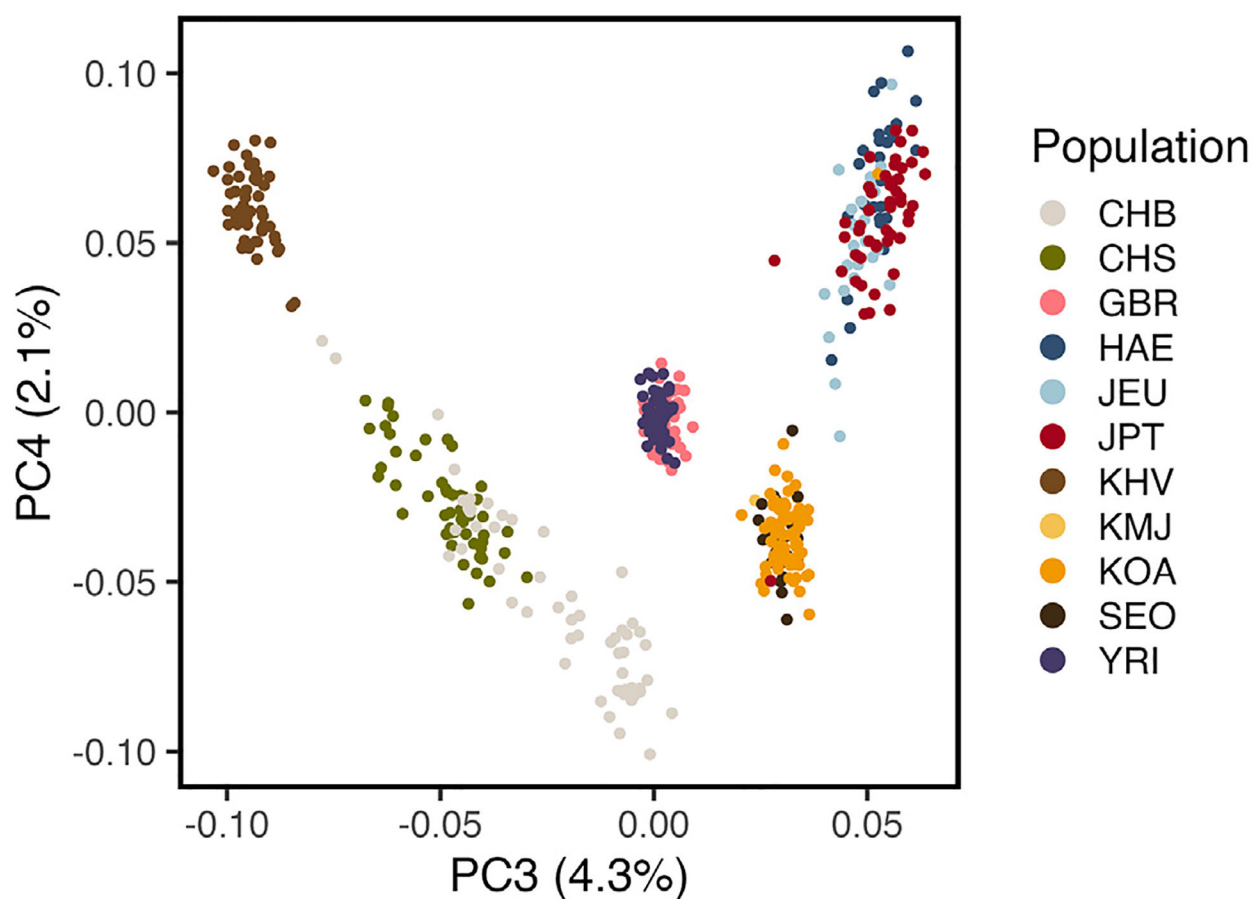


Figure 4. Global PCA of participants from this study as well as individuals from the 1000 Genomes Project and Korean1K

In these PCs, residents of Jeju Island cluster with JPT individuals rather than with mainland Koreans.

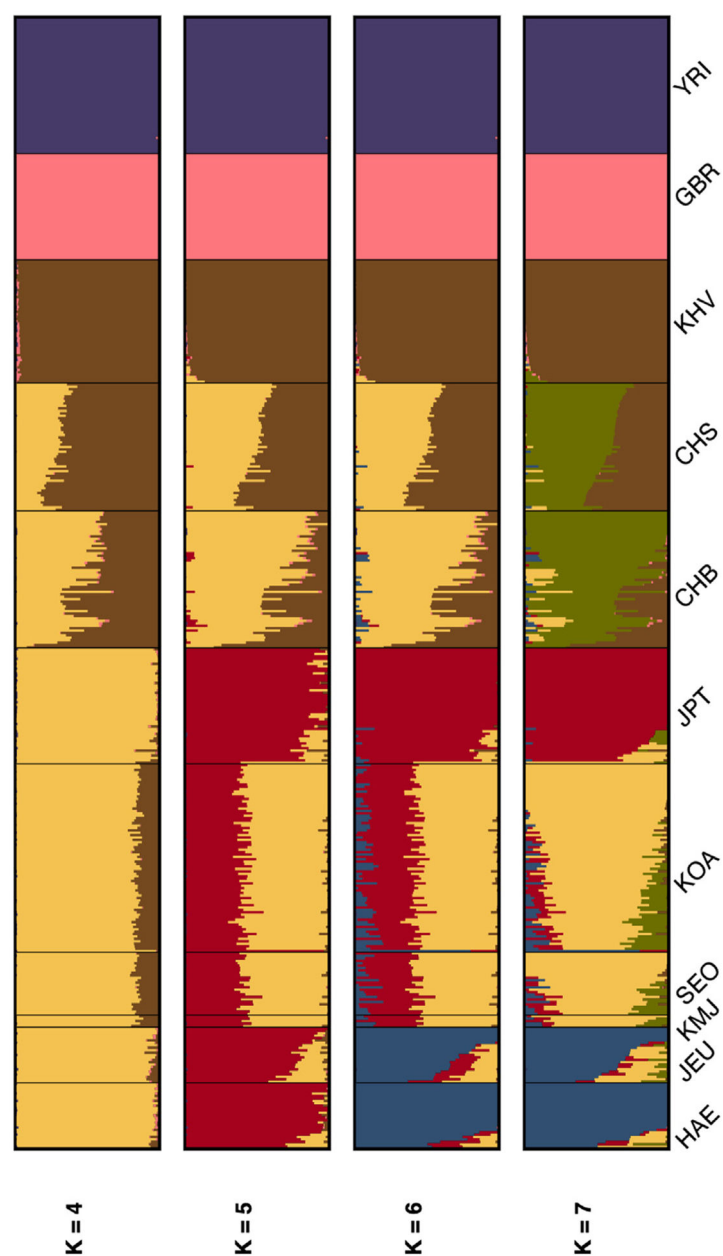


Figure 5. Admixture graph for the Haenyeo and other global populations
 The Haenyeo receive their own, distinct ancestral component at $K = 6$.

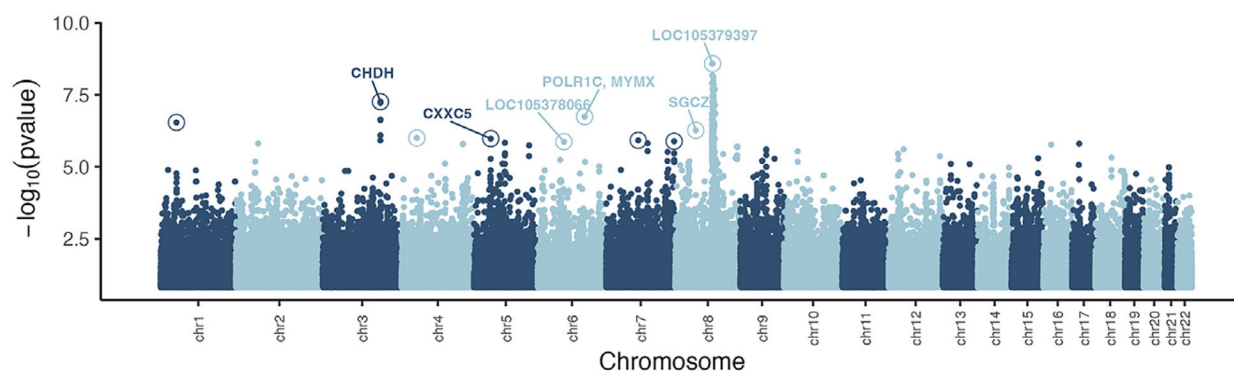


Figure 6. Manhattan plot of selection scan results, with the top 10 SNPs found in genomic regions under selection circled

SNPs are further annotated with gene name if the SNP was found within a gene.

Table 1.
Population divergence quantified by F_{ST} values of relevant populations

POP1	POP2	Weir and Cockerman F_{ST}
CHB	CHS	0.00096
JEJ	KOR	0.0034
HAE	JEU	0.00031
CHB	KOR	0.0026
JPT	KOR	0.0030
CHS	KOR	0.0046
JPT	JEJ	0.0047

Jeju resident populations, Haenyeo and Jeju controls (HAE and JEU, respectively), are not meaningfully genetically diverged. Mainland Koreans (KOR) appear more diverged from Jeju residents (JEJ) than from northern Han Chinese (CHB) and Japanese (JPT); however, this may reflect low effective population size (N_e) in Jeju relative to CHB and JPT rather than true differences in divergence. Both Korean populations, Jeju and mainland, had relatively low divergence with Japanese and Chinese populations.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Haenyeo (HAE), Jeju (JEU), Seoul (SEO) genomes	This paper	SRA: PRJNA1235291
Korean Genome Project: Korean1K (KOA) genomes	Jeon et al. ³³	https://www.science.org/doi/10.1126/sciadv.aaz7835
1000 Genomes Project: Japanese in Tokyo (JPT), Han Chinese in Beijing (CHB), Southern Chinese population (CHS), Kinh Vietnamese (KHV), British in England and Scotland (GBR) and Yoruba from Nigeria (YRI)	Genomes Project et al. ⁶⁵	http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20181203_biallelic_SNV/
All of Us Genomes	All of Us Research Program	https://www.researchallofus.org/
Software and algorithms		
FastQC v0.11.8	Andrews ⁶⁶	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
PRINSEQ-lite 0.20.4	Schmieder and Edwards ⁶⁷	https://prinseq.sourceforge.net/
bwa-mem	Li ⁶⁸	https://github.com/lh3/bwa
samtools	Li et al. ⁶⁹	https://www.htslib.org/
GLIMPSE v1.1.1	Rubinacci et al. ³²	https://odelaneau.github.io/GLIMPSE/glimpse1/
bcftools	Danecek and McCarthy ⁷⁰	https://github.com/samtools/bcftools
ngsToolsV2	Hanghoj et al. ⁷¹	https://github.com/mfumagalli/ngsTools
plink1.9	Purcell et al. ⁴⁴	https://www.cog-genomics.org/plink/1.9/
plink2	Chang et al. ^{72,73}	https://www.cog-genomics.org/plink/2.0/
Ohana	Cheng et al. ³⁴	https://github.com/jade-cheng/ohana
Treemix	Pickrell and Pritchard ⁷⁴	https://bitbucket.org/nygcresearch/treemix/wiki/Home
Admixture Bayes	Nielsen et al. ⁷⁵	https://github.com/avaughn271/AdmixtureBayes
Relate	Speidel et al. ^{76,77}	https://myersgroup.github.io/relate/
CLUES2	Vaughn and Nielsen ³⁵	https://github.com/avaughn271/CLUES2
pong	Behr et al. ⁷⁸	https://github.com/ramachandran-lab/pong
custom scripts and whole pipeline	This paper	https://github.com/aguilar-gomez/haenyeo/