Evolutionary history and adaptation of a human pygmy population of Flores Island, Indonesia

Serena Tucci1,2,3, Samuel H. Vohr4, Rajiv C. McCoy1,2, Benjamin Vernooy5, Matthew R. Robinson6,7, Chiara Barbieri8,9, Brad J. Nelson10, Wenqing Fu10, Peter M. Visscher14,15, Joshua M. Akey1,2

Flores Island, Indonesia, was inhabited by the small-bodied hominin species *Homo floresiensis*, which has an unknown evolutionary relationship to modern humans. This island is also home to an extant human pygmy population. Here we describe genome-scale single-nucleotide polymorphism data and whole-genome sequences from a contemporary human pygmy population living on Flores near the cave where *H. floresiensis* was found. The genomes of Flores pygmys reveal a complex history of admixture with Denisovans and Neanderthals but no evidence for gene flow with other archaic hominins. Modern individuals bear the signatures of recent positive selection encompassing the FADS (fatty acid desaturase) gene cluster, likely related to diet, and polygenic selection acting on standing variation that contributed to their short-stature phenotype. Thus, multiple independent instances of hominin insular dwarfism occurred on Flores.

To infer population relationships, we integrated our sequencing data with SNP array data from 2507 individuals spanning 225 worldwide populations, as well as sequencing data from Melanesia (13). A principal component analysis (PCA) places the Flores population in close proximity to a cluster of East Asian and ISEA samples, with a notable affinity toward Oceanic populations (Fig. 1B and table S4). Population structure inferred by ADMIXTURE shows that most of the ancestry in the Flores pygmys can be explained by an East Asian–related component and by a smaller New Guinean–related component, shared with Oceanic populations (Fig. 1C and fig. S4). The New Guinean component accounts for 23.2% of Flores ancestry (z-score > 85) (fig. S5) (13).

These results, along with multiple sequential Markovian coalescent inferences of effective population size and divergence times, analyses of inbreeding, and mitochondrial DNA and Y chromosome variation, indicate that the Flores pygmys likely trace their ancestry back to the ancestors of Near Oceanic populations and experienced a recent admixture event with populations of East Asian ancestry (figs. S6 to S11 and tables S4 and S5) (13).

We used a PCA projection method to describe the relationship between the Flores pygmys and archaic hominins (13). Flores individuals exhibit affinity to both Neanderthals and Denisovans, suggesting that they harbor ancestry from both archaic hominins (Fig. 2A). Using F3 ratios, we estimated that the Flores pygmys harbor, on average, 0.8% Denisovan ancestry (z-score > 4) (13), which is higher than the amount for other ISEA populations but lower than that for Oceanic populations. Consistent with previous observations (7), the amount of inferred Denisovan ancestry is positively correlated with proportion of New Guinean ancestry (Pearson’s correlation = 0.97, P < 10^-16) (fig. S12) (13).

We identified regions inherited from archaic hominins by applying the S* statistical framework (fig. S13) (13) to 9 Flores, 27 Melanesians, 103 East Asian, and 91 European genomes. On average, we retrieved 53.5 Mb of significant S* sequence in the Flores sample (false discovery rate ≤ 5%) (Fig. 2B). Of this, 47.5 and 4.2 Mb were assigned with high confidence to the Neanderthal and Denisovan groups, respectively, whereas 1.8 Mb were classified as “ambiguous” (for which Neanderthal or Denisovan status cannot be confidently distinguished) (Fig. 2B, inset, and fig. S14). The average amount of Neanderthal sequence per individual (47.5 Mb) among Flores pygmys is intermediate between that among East Asians (54.5 Mb) and Melanesians (40.2 Mb), whereas the average amount of Denisovan sequence was less than that identified in Melanesians (32 Mb) (fig. S14). These data suggest that the source of Denisovan ancestry was localized east of Wallace’s Line and that such ancestry was diluted in Flores by subsequent admixture with Asian populations carrying less (or no) Denisovan ancestry (14).

The S* statistic does not require information from an archaic reference genome and thus can
potentially identify sequences from unknown hominin lineages. We searched for signatures of admixture with an unknown archaic hominin source by analyzing significant \( S^* \) sequences that did not match the Neanderthal or Denisovan genomes (hereafter called “unknown” sequences) (fig. S14). We found no evidence that unknown sequences in Flores are enriched for older or more divergent lineages (figs. S15 and S16) \(^{(13)} \), as would be expected if they contained lineages inherited from a more deeply divergent hominin group, such as \( H. floresiensis \) or \( H. erectus \). Although it is difficult to exclude very low levels of admixture from such groups given current methodological limitations, our data are inconsistent with substantial levels of ancestry from a deeply divergent hominin lineage.

We analyzed copy number variation (CNV) in the Flores pygmies, along with a panel of diverse modern and archaic human genomes (figs. S17 to S24 and tables S6 and S7) \(^{(13)} \). We found 1865 biallelic CNVs in Flores individuals, as well as a common \( [\text{allele frequency (AF) = 50%}] \), large segmental duplication block (>220 kilo-base pairs; chromosome 16p12.2) that to date has been observed only in Denisovan and Oceanic individuals \( (\text{AF} = 82.6\%) \) (Fig. 2C and fig. S18). Previous work suggests this duplication introgressed from Denisovans into the ancestors of present-day populations in Oceania ~40 ka ago \(^{(15)} \).

To test hypotheses of recent adaptation, we used the population branch statistic (PBS) to scan for alleles exhibiting strong population-specific structure \(^{(13)} \). We identified 2 genomic windows in the top 0.001 percentile \( (\text{PBS} > 1.04) \) and 10 additional windows in the top 0.01 percentile \( (\text{PBS} > 0.78) \) of mean genome-wide PBS scores (fig. S25 and table S8). One of the top two regions encompasses the human leukocyte antigen gene complex, a well-known substrate of diversifying selection with a critical role in adaptive immunity \(^{(13)} \). The strongest PBS signal, however, extends over a ~74-kb region of chromosome 11 that includes \( \text{FADS1} \) and \( \text{FADS2} \) (Fig. 3A). These genes encode fatty acid desaturase (FADS) enzymes that catalyze synthesis of long-chain polyunsaturated fatty acids (LC-PUFAs) from plant-based medium chain (MC)-PUFA precursors. Notably, the Flores sample is nearly fixed for an ancestral haplotype \( (\text{tagged by the C allele of SNP rs174547}) \) in a pattern consistent with a recent selective sweep. In the larger Omni2.5-genotyped sample \( (n = 21 \text{ unrelated individuals}) \), we confirmed a 95% frequency of the ancestral \( (\text{C}) \) allele of rs174547.

Other Southeast Asian populations also carry high frequencies of the ancestral allele (Fig. 3B, inset), consistent with positive selection in their common ancestor, with drift and additional selection in ISEA populations subsequent to their divergence \(^{(13)} \).

Supporting functional differences, previous data demonstrate that SNPs defining this haplotype are strongly associated with circulating levels of fatty acids \( (16) \) (table S9) and a wide variety of blood phenotypes \( (17) \) (table S10). Furthermore, these variants are known expression quantitative trait loci \( (\text{eQTLs}) \) of both \( \text{FADS1} \) and \( \text{FADS2} \) \(^{(13, 18, 19)} \). Specifically, the selected \( (\text{C}) \) allele of rs174547 is associated with up-regulation of \( \text{FADS2} \) and down-regulation of \( \text{FADS1} \) \(^{(20)} \) (Fig. 3C), in turn predicting reduced efficiency of conversion from MC- to LC-PUFA.

Fig. 1. Sampling location and genomic variation of the Flores pygmies. (A) Location of Flores pygmy village and populations integrated into the analysis. The inset shows a subset of 85 populations from East Asia (EA), ISEA, and Oceania used for the PCA. The Rampasasa village (red square) is close to the Liang Bua cave, where \( H. floresiensis \) fossils were excavated. (B) PCA performed on 85 populations \(^{(13)} \). (C) ADMIXTURE results for \( K = 6 \) clusters are shown for 96 worldwide populations \(^{(13)} \).
fixed in an Inuit population in Greenland (23), potentially in response to climate and a marine diet rich in omega-3 fatty acids. While mirroring our findings from Flores, the adaptive Greenlandic Inuit haplotype extends over a broader downstream region encompassing FADS3, potentially reflecting distinct selection events or population-specific patterns of recombination (fig. S26). Although the global history of this locus remains to be clarified, current evidence points to a critical role of FADS as an evolutionary “toggle switch” in response to changing diet.

Body size reduction is one of the best-known responses to island environments and is common among mammalian taxa. Examples on Flores include H. floresiensis and the pygmy proboscidean Stegodon. We leveraged our data to test hypotheses about the genetic architecture and evolution of short stature in our Flores sample. If the short stature phenotype on Flores was a consequence of polygenic selection acting on common variants, we would expect to see higher frequencies of alleles associated with reduced height in other populations (24). We thus performed mixed-linear model association analysis for height in 456,426 individuals of European ancestry to identify height-associated loci, unbiased of population stratification (13).

We find that European height-associated loci are significantly more differentiated between Flores and Melanesian individuals. The Denisovan D duplication is present only in Denisovan, Oceanic, and Flores individuals. A panel of worldwide populations is shown, along with the Denisovan and Neanderthal genomes (13). A copy number greater than two (four and three for light and dark blue, respectively) in region D (far right) indicates presence of the duplication.

**Fig. 2. Archaic hominin ancestry in the Flores pygmies.** (A) PCA to investigate genetic similarities of present-day humans and archaic hominins. Mean values for the top two principal components were plotted for each population. (B) Amounts of total archaic introgressed sequences in 9 Flores pygmies, 27 Melanesians, 103 East Asians, and 91 Europeans. The inset shows amounts of Neanderthal, Denisovan, and ambiguous sequences in Flores and Melanesian individuals. (C) The Denisovan D duplication is present only in Denisovan, Oceanic, and Flores individuals. A panel of worldwide populations is shown, along with the Denisovan and Neanderthal genomes (13). A copy number greater than two (four and three for light and dark blue, respectively) in region D (far right) indicates presence of the duplication.
and the Flores population is 1 SD lower than the average of neighboring populations (Fig. 4A), we predict the Flores population to be ~2 cm ($\sqrt{0.085 \times 6} = 1.75$ cm) shorter in stature than populations in East Asia and Oceania and ~5 cm shorter if the genetic predictor explained 70% of the phenotypic variance. Collectively, these data provide evidence that polygenic selection acting on standing genetic variation was an important determinant of short stature in this Flores pygmy population.

High-coverage genomes provide insights into the history of demographic changes and adaptation in a pygmy population on Flores Island (Indonesia). Although these individuals possess ancestry from both Neanderthals and Denisovans, we found no evidence of admixture with a deeply diverged hominin group. This observation, combined with evidence that their short-stature phenotype resulted from polygenic selection acting on standing variation, suggests that insulin dwarfism arose independently in two separate hominin lineages on Flores Island.

REFERENCES AND NOTES

13. See supplementary materials.


ACKNOWLEDGMENTS

We thank members of the Akey, Green, and Barbujani laboratories, I. Mathieson, J. G. Schraiber, A. Manica, and S. Browning for helpful feedback related to this work. We acknowledge P. Kusuma and other members of the Eijkman Institute for Molecular Biology (EIMB) for providing logistical support in facilitating sample collection. Funding: This work was supported in part by an NIH grant (R01GM110068) to J.M.A., Searle Scholars Program and Gordon and Betty Moore Foundation grants.
Fig. 4. Polygenic selection for reduced stature in the Flores pygmies. (A) Comparison of the genome-wide genetic predictor of height from observed genotypes (purple) versus the null model (blue). The Flores panel is significantly enriched for height-reducing alleles ($P < 0.001$) in a multivariate chi-square test of population genetic differentiation from the expectation under the null model. (B) Frequency differences between the Flores population and neighboring 1000 Genomes Project populations for 4000 genome-wide loci of the strongest association, with height plotted against effect size for the height-increasing allele. The regression slope shown in the figure between the height-increasing effect size and the frequency difference was $-0.71$ ($t = -3.18$, $P = 0.002$), reflecting height-increasing alleles being at lower frequency in the Flores population. (C) Association between height in the Flores population and a genetic predictor of height.

Transfer agreement. Individuals interested in obtaining the materials should contact the Eijkman Institute.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/361/6401/511/suppl/DC1
Materials and Methods
Figs. S1 to S26
Tables S1 to S10
References (26–126)
22 December 2017; accepted 22 June 2018
10.1126/science.aar8486
Evolutionary history and adaptation of a human pygmy population of Flores Island, Indonesia


Science 361 (6401), 511-516.
DOI: 10.1126/science.aar8486

The genetics of human short stature
Flores Island in Indonesia has a long history of hominin occupation, including by the extinct Homo floresiensis and a more recent settlement by modern humans. Furthermore, Flores has an extant population of pygmy humans, and H. floresiensis exhibited a diminutive adult size relative to other hominins. Tucci et al. examined genetic variation among 32 individuals, including 10 sequenced genomes, from a population of pygmies living close to the cave where H. floresiensis remains were discovered. These individuals exhibit signatures of polygenic selection explaining the short stature and have genomic content from both Neanderthals and Denisovans, but no additional archaic lineages. Thus, restricted height is under selection at this location and has evolved independently at least twice in hominins.

Science, this issue p. 511

ARTICLE TOOLS http://science.sciencemag.org/content/361/6401/511
SUPPLEMENTARY MATERIALS http://science.sciencemag.org/content/suppl/2018/08/01/361.6401.511.DC1
RELATED CONTENT http://science.sciencemag.org/content/sci/361/6401/439.full
REFERENCES This article cites 124 articles, 30 of which you can access for free
http://science.sciencemag.org/content/361/6401/511#BIBL
PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title Science is a registered trademark of AAAS.