

## HUMAN GENETICS

# A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea

Anders Bergström,<sup>1\*</sup> Stephen J. Oppenheimer,<sup>2</sup> Alexander J. Mentzer,<sup>3</sup> Kathryn Auckland,<sup>3</sup> Kathryn Robson,<sup>4</sup> Robert Attenborough,<sup>5,6</sup> Michael P. Alpers,<sup>7,8</sup> George Koki,<sup>8</sup> William Pomat,<sup>8</sup> Peter Siba,<sup>8</sup> Yali Xue,<sup>1</sup> Manjinder S. Sandhu,<sup>1,9</sup> Chris Tyler-Smith<sup>1\*</sup>

New Guinea shows human occupation since ~50 thousand years ago (ka), independent adoption of plant cultivation ~10 ka, and great cultural and linguistic diversity today. We performed genome-wide single-nucleotide polymorphism genotyping on 381 individuals from 85 language groups in Papua New Guinea and find a sharp divide originating 10 to 20 ka between lowland and highland groups and a lack of non-New Guinean admixture in the latter. All highlanders share ancestry within the last 10 thousand years, with major population growth in the same period, suggesting population structure was reshaped following the Neolithic lifestyle transition. However, genetic differentiation between groups in Papua New Guinea is much stronger than in comparable regions in Eurasia, demonstrating that such a transition does not necessarily limit the genetic and linguistic diversity of human societies.

The island of New Guinea contains some of the earliest archaeological evidence for modern humans outside of Africa, dating back to ~50 ka (1). Starting ~10 ka, systematic plant cultivation was developed in its central mountain range (2), approximately coinciding with similar, independent developments in the Near East, East Asia, and the Americas. Today, the country of Papua New Guinea (PNG) occupies the eastern half of the island and northern Island Melanesia and is the most linguistically diverse country, with ~850 languages spoken (3). About half belong to the Trans-New Guinea phylum, spoken across all of the highlands and large parts of the lowlands, and hypothesized to have spread alongside plant cultivation (4).

The Sahul continent appears to have been isolated from the rest of the world at least until

the last few thousand years (5, 6), so its pre-history likely represents an independent instance of human genetic and cultural evolution over ~50 thousand years (ky). Genetic studies are increasingly indicating that agriculture, languages, and culture in Eurasia and Africa have primarily spread through the movement of people (7–10), and it is of great interest to understand if the shift from a hunter-gatherer to a sedentary, cultivation-based lifestyle in New Guinea—which we here refer to as a Neolithic transition—followed similar patterns.

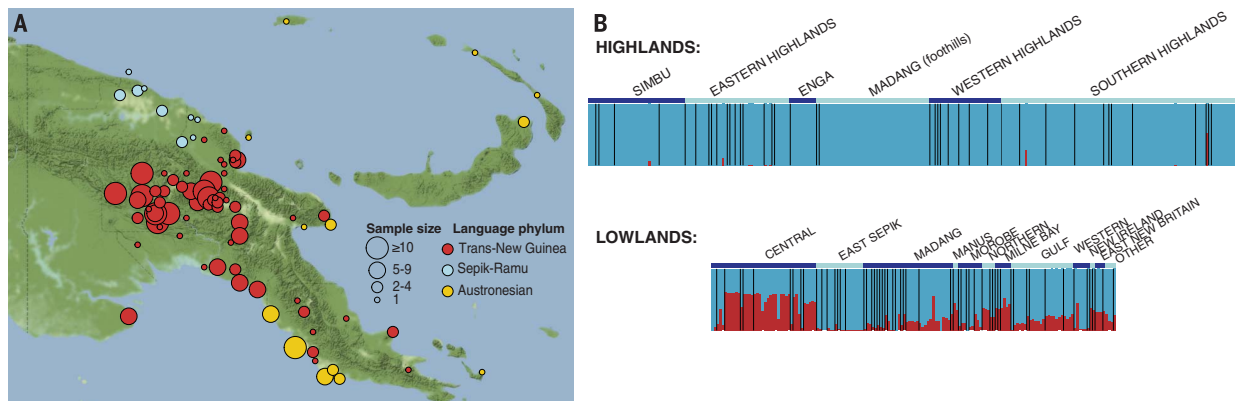
We genotyped 381 individuals from 85 language groups across PNG at 1.7 million genome-wide markers [Fig. 1A, fig. S1, and tables S1 and S2, (11)] and analyzed 39 previously generated high-coverage whole-genome sequences (6, 12), including the PNG samples from the Human Genome Diversity Project (HGDP)–CEPH panel, which we

find consist of one highland and one lowland subset [fig. S2 and table S3, (11)].

We first examined the impact of external gene flow to PNG, particularly that derived from Holocene migrations from Southeast Asia (13). Highlanders show no excess shared ancestry with Asians relative to Aboriginal Australians [*D*-statistics (14),  $z > 2$  (11), and ADMIXTURE (15)], except for four individuals who likely reflect recent admixture via the lowlands (Fig. 1B). We also find no mitochondrial or Y chromosomes of recent non-Sahul origin in any highlander (figs. S4 and S5). The lowlands, however, harbor widespread Southeast Asian ancestry, with substantially higher levels in Austronesian speakers than in non-Austronesian speakers (mean of 38.7 versus 11.6%,  $P = 1.4 \times 10^{-13}$ , Wilcoxon rank sum test). The lowest levels (mean of 4.3%) are found in northern groups that speak Sepik-Ramu phylum languages. Our results thus demonstrate a variable Southeast Asian genetic impact on different parts of PNG and independence of highlander ancestry from non-Sahul sources.

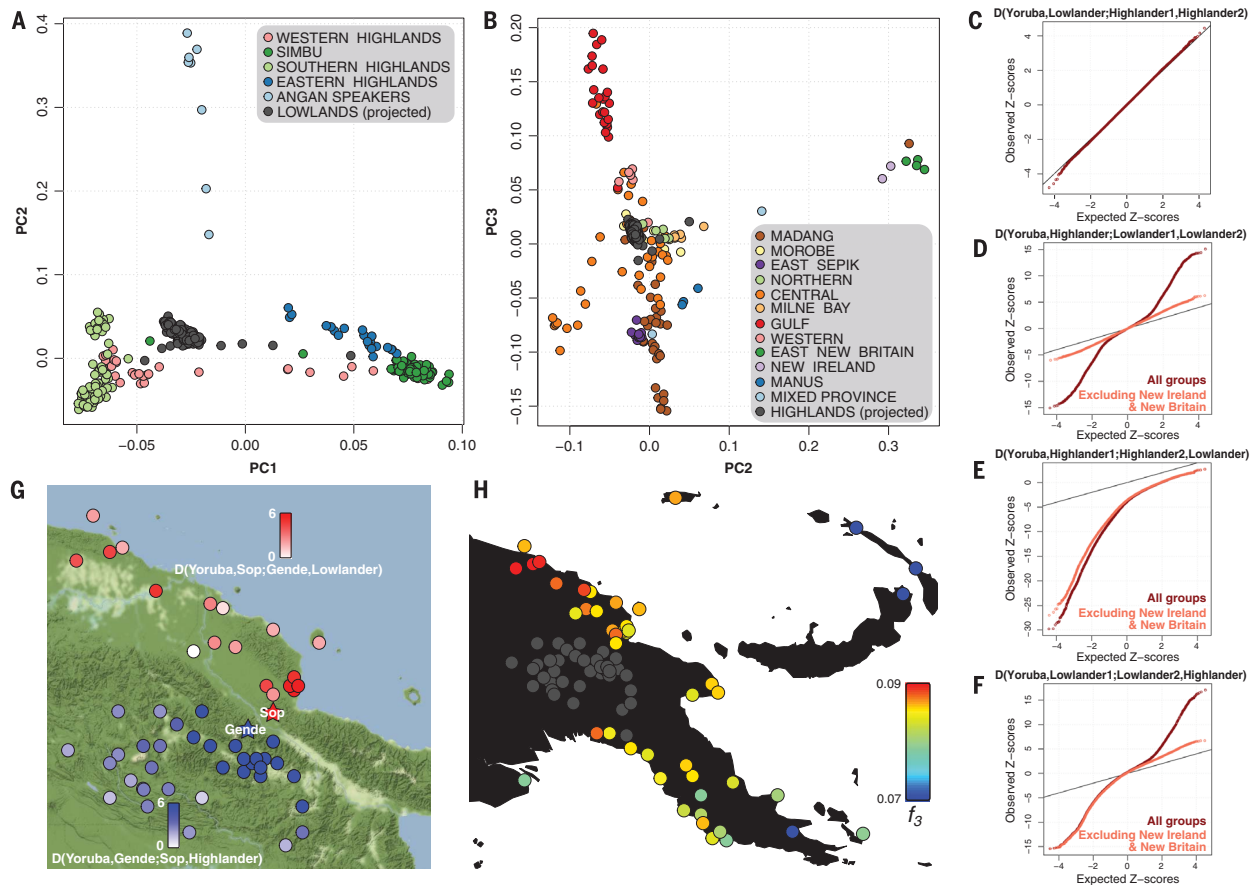
Papuans diverged genetically from Aboriginal Australians long before rising sea levels separated New Guinea and Australia ~8 ka, and different groups across Australia display a uniform relationship to Papuans (6). When accounting for Southeast Asian admixture using admixture graphs and *D*-statistics (14), we similarly find that all genotyped Papuan individuals share a

<sup>1</sup>Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK. <sup>2</sup>School of Anthropology and Museum Ethnography, University of Oxford, Oxford OX2 6PE, UK. <sup>3</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. <sup>4</sup>MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK. <sup>5</sup>Biological Anthropology, Department of Archaeology and Anthropology, University of Cambridge, Cambridge CB2 1QH, UK. <sup>6</sup>School of Archaeology and Anthropology, Australian National University, Canberra, ACT 2601, Australia. <sup>7</sup>International Health Research, Curtin University, Perth, WA 6845, Australia. <sup>8</sup>Papua New Guinea Institute of Medical Research, Post Office Box 60, Goroka, Papua New Guinea. <sup>9</sup>Department of Medicine, University of Cambridge, Cambridge CB2 0QQ, UK. \*Corresponding author. Email: ab34@sanger.ac.uk (A.B.); cts@sanger.ac.uk (C.T.S.)



**Fig. 1. PNG samples.** (A) Each language group is represented by a circle; the area indicates the number of genotyped individuals, and the color indicates the top-level language phylum. 39 individuals are not included because either the specific language is unknown or the two parents are from different language groups. Also see fig. S1. (B) Papuan (blue) and Southeast Asian (red) ancestry

proportions as estimated by ADMIXTURE [number of ancestry components ( $K$ ) = 2, with 504 East Asian individuals from the 1000 Genomes Project; also see fig. S3]. Individuals are grouped by province and then language group (separated by black bars). Ancestry proportions correlate strongly [correlation coefficient ( $r$ ) = 0.988] with those estimated using  $f_4$ -ratios (11).



**Fig. 2. PNG population structure.** (A) When projected onto principal components (PCs) constructed with only highlander genotypes, all lowlanders (excepting a few outliers) group uniformly. Also see fig. S10. (B) When projected onto PCs constructed with only lowlander genotypes, all highlanders (excepting a few outliers) group uniformly. Also see fig. S11. (C to F) Quantile-quantile plots comparing z scores from  $D$ -statistics relating highlanders and lowlanders to those expected under a normal distribution (11). The African Yoruba population is used as an outgroup. (C) Lowlanders are equally similar to different highlander groups. (D) Highlanders have stronger affinity to some lowlander groups than to others. (E) Highlanders are more similar to each other

than to lowlanders. (F) Lowlanders are not always more similar to each other than to highlanders. (G) The z scores (capped at 6) of two different  $D$ -statistics measuring (i) if the highland Gende speakers are more similar to the lowland Sop speakers, living just 40 km away, or to other highlanders (blue indicating more highlander similarity) and (ii) if Sop speakers are more similar to Gende speakers or to other lowlanders (red indicating more lowlander similarity). (H) Genetic affinity of highlanders (treated as a single group, in gray) to different lowland groups measured by the outgroup  $f_3$ -statistic  $f_3$  (Highlanders, X; Aboriginal Australian) (red indicating higher affinity). (C) to (H) were calculated after masking lowlander genomes for Southeast Asian ancestry.

uniform relationship to Aboriginal Australians (fig. S6), revealing a lack of genetic continuity across Sahul.

The strongest genetic separation within PNG appears to be that between the mainland and the Bismarck archipelago islands (New Britain and New Ireland) (Fig. 2B and fig. S7), consistent with previous studies (16). Highlanders fall into three clusters: one western, one eastern, and one corresponding to a small set of Angan language groups from the southeastern highlands (Fig. 2A and fig. S8), the last showing evidence of genetic isolation (fig. S9).

To compare highlanders and lowlanders, we masked lowlander genomes for Southeast Asian ancestry tracts, achieving a misclassification rate of <0.5% (11). We find no differences in the affinity of lowlanders to different highlander groups using principal components analysis and  $D$ -statistics [Fig. 2, A and C, and fig. S10; (11)]. Thus, all highlanders, regardless of geographic location,

seem to form a clade relative to lowlanders. In line with this, there is not a single  $D$ -statistic (at  $z > 3$ ) in which a highlander group is more similar to any lowlander group than to any other highlander group (Fig. 2, E and G).

By contrast, highlanders as a group are not equally similar to all lowlanders (Fig. 2D), as they display slightly higher affinity to groups from the Sepik River region (Fig. 2H). This is surprising linguistically, as the local Sepik-Ramu languages are unrelated to the Trans-New Guinea languages of the highlands. There is, however, archaeological evidence for Holocene cultural contact between the two regions (17). Whereas highlanders are all similar amongst themselves, the same is not true of lowlanders (Fig. 2F)—both southern and northern lowlanders are more similar to highlanders than they are to each other.

To investigate when present-day groups in PNG separated, we applied MSMC (multiple sequentially Markovian coalescent) (18) to whole-

genome sequences for six highland groups and one Sepik lowlands group. We used 10x Genomics linked-read whole-genome sequencing (19) to physically phase eight genomes (table S4) and also analyzed perfectly phased male X chromosomes (11). The results suggest that highlanders and Sepik lowlanders separated 10 to 20 ka. All splits within the highlands seem to have occurred within the last ~10 ky (Fig. 3, A and B, and fig. S12). A Y-chromosomal phylogeny similarly revealed shared ancestry across groups within these time scales (fig. S13). We also find evidence of a major increase in effective population sizes in most highlander groups in the last 10 ky (Fig. 3C and fig. S14), using SMC++ (20) and MSMC. Sepik lowlanders do not share this increase, consistent with anthropological records of lower lowland population densities, likely linked to widespread malaria (21).

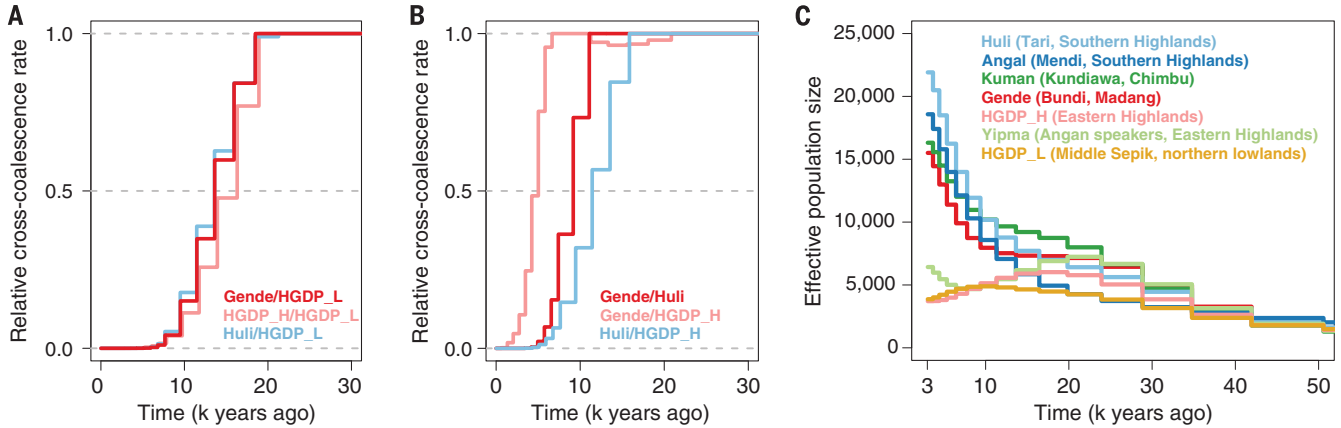
Genetic differentiation is much stronger in PNG than in regions of similar size in Eurasia, where fixation index ( $F_{ST}$ ) values between major

populations within Europe or East Asia are generally 1% or less (Fig. 4). Within the highlands, a sampled area about the size of Denmark,  $F_{ST}$  values between eastern and western groups are 2 to 3%, and values between the Angan-speaking and other groups reach 4 to 5% (as high as between European and South Asian populations). Within each of the eastern and western highland

clusters, values are below 2%, but many are above 1%. Levels of  $F_{ST}$  in the lowlands are also high, suggesting that cultural-linguistic factors, rather than terrain, drive the differentiation. Between the highland, northern lowland, and southern lowland regions, differentiation is even higher. Structure is stronger for the Y chromosome than for the mitochondrial genome, suggesting lower

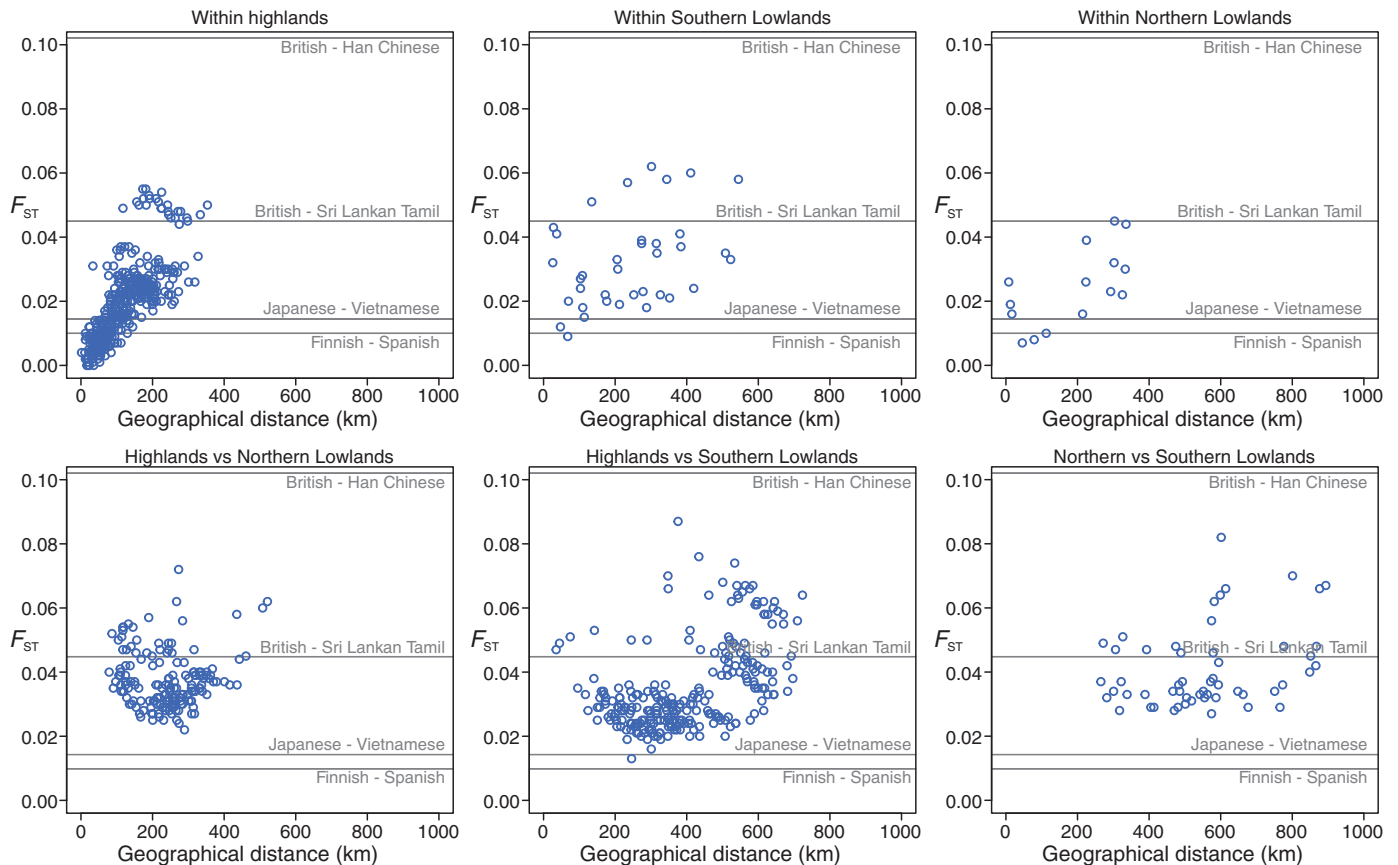
male effective population sizes and/or more female movement between groups (fig. S15).

Our results confirm the independent evolution of Sahul for most of the last 50 ky and the independence of New Guinea from Australia for much of this time. Present-day mainland population structure, marked by a very sharp highland-lowland division, does not date back



**Fig. 3. Time depth of population separation and growth in PNG.** (A) Cross-coalescence curves between highlanders and a northern lowlands Middle Sepik group suggests a split time between 10 and 20 ka. (B) Cross-coalescence curves between highland groups suggest split times within the last ~10 ky (Huli

representing the western cluster, Gende and HGDP\_H the eastern, figs. S8 and S10). These were inferred using MSMC on genomes physically phased by performing linked-read sequencing. Also see fig. S12. (C) Effective population-size histories inferred using SMC++ on five genomes per group. Also see fig. S14.



**Fig. 4. Genetic differentiation in PNG.** Geographical distance between groups plotted against  $F_{ST}$ , after masking lowlander genomes for Southeast Asian ancestry. Gray lines indicate  $F_{ST}$  between selected 1000 Genomes Project populations.

to the initial peopling of Sahul but instead appears to have formed within the last 20 ky. Highland structure formed subsequently, mostly within the last 10 ky, which is within the general time scale of the spread of cultivation (2) and the Trans–New Guinea languages (4). We thus propose that an expansion of cultivating groups across the highlands could explain our observations, including the uniform relationship of highlanders to lowlanders and the recent increase in population sizes. Our data also suggest higher diversity in the western than in the eastern highlands (figs. S9 and S14), consistent with a hypothesized origin of cultivation in the former (22). Thus, our results suggest that, as in many other parts of the world, the spread of cultivation in PNG was associated with an expansion of peoples and a reshaping of population structure.

The strong genetic differentiation within PNG, however, sets it apart from other parts of the world that also underwent Neolithic lifestyle transitions. Ancient DNA studies in Europe and the Near East have documented a gradual, but dramatic, decrease in differentiation, showing that the genetic homogeneity of present-day west Eurasia emerged in the last few thousand years (10, 23).  $F_{ST}$  values in PNG fall between those of hunter-gatherers and present-day populations of west Eurasia, suggesting that a transition to cultivation alone does not necessarily lead to genetic homogenization.

A key difference might be that PNG had no Bronze Age, which in west Eurasia was driven by an expansion of herders and led to massive

population replacement, admixture, and cultural and linguistic change (7, 8), or Iron Age such as that linked to the expansion of Bantu-speaking farmers in Africa (24). Such cultural events have resulted in rapid Y-chromosome lineage expansions due to increased male reproductive variance (25), but we consistently find no evidence for this in PNG (fig. S13). Thus, in PNG, we may be seeing the genetic, linguistic, and cultural diversity that sedentary human societies can achieve in the absence of massive technology-driven expansions.

#### REFERENCES AND NOTES

- J. F. O'Connell, J. Allen, *J. Archaeol. Sci.* **56**, 73–84 (2015).
- T. Denham, H. Barton, in *Behavioral Ecology and the Transition to Agriculture*, D. J. Kennett, B. Winterhalder, Eds. (Univ. of California Press, 2006), pp. 237–264.
- M. P. Lewis, G. F. Simons, C. D. Fennig, Eds., *Ethnologue: Languages of the World* (SIL International, ed. 19, Dallas, 2016).
- A. Pawley, in *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-speaking Peoples*, A. Pawley, R. Attenborough, J. Golson, R. Hide, Eds. (Australian National Univ., 2005), pp. 67–107.
- A. Bergström *et al.*, *Curr. Biol.* **26**, 809–813 (2016).
- A. S. Malaspina *et al.*, *Nature* **538**, 207–214 (2016).
- W. Haak *et al.*, *Nature* **522**, 207–211 (2015).
- M. E. Allentoft *et al.*, *Nature* **522**, 167–172 (2015).
- C. de Filippo, K. Bostoen, M. Stoneking, B. Pakendorf, *Proc. Biol. Sci.* **279**, 3256–3263 (2012).
- I. Lazaridis *et al.*, *Nature* **536**, 419–424 (2016).
- Materials and methods are available as supplementary materials.
- M. Raghavan *et al.*, *Science* **349**, aab3884 (2015).
- A. T. Duggan, M. Stoneking, *Curr. Opin. Genet. Dev.* **29**, 9–14 (2014).
- N. Patterson *et al.*, *Genetics* **192**, 1065–1093 (2012).
- D. H. Alexander, J. Novembre, K. Lange, *Genome Res.* **19**, 1655–1664 (2009).
- J. S. Friedlaender *et al.*, *PLOS Genet.* **4**, e19 (2008).
- P. Swadlow, P. Wiessner, A. Tumu, *J. Soc. Ocean.* **126–127**, 271–292 (2008).
- S. Schiffels, R. Durbin, *Nat. Genet.* **46**, 919–925 (2014).
- G. X. Zheng *et al.*, *Nat. Biotechnol.* **34**, 303–311 (2016).
- J. Terhorst, J. A. Kamm, Y. S. Song, *Nat. Genet.* **49**, 303–309 (2017).
- I. D. Riley, *J. Hum. Evol.* **12**, 125–132 (1983).
- D. K. Feil, *The Evolution of Highland Papua New Guinea Societies* (Cambridge Univ. Press, 1987).
- P. Skoglund *et al.*, *Science* **344**, 747–750 (2014).
- T. N. Huffman, *Annu. Rev. Anthropol.* **11**, 133–150 (1982).
- G. D. Poznik *et al.*, *Nat. Genet.* **48**, 593–599 (2016).

#### ACKNOWLEDGMENTS

We thank all sample donors who contributed to this study; T. Parks, A. V. Hill, J. B. Clegg, D. Higgs, D. J. Weatherall, O. Bunari, A. Spencer, J. Barker, R. Spark, and P. Sill for assistance in sample collection and discussion; J. Friedlaender for background information on the HGDP-CEPH samples; and the Wellcome Trust Sanger Institute genotyping and sequencing facilities for generating data, especially M. Quail, D. Jackson, and S. Leonard for generating 10x Genomics data. A.B., Y.X., M.S.S., and C.T.-S. were supported by the Wellcome Trust (grant 098051). A.J.M. was supported by a Wellcome Trust Clinical Research Training grant (106289/Z/14/Z). S.J.O., A.J.M., and K.A. were supported by a Wellcome Trust Core Award (090532/Z/09/Z), and K.A. was supported by a European Research Council Advanced Grant to A. V. S. Hill (294557). The array genotypes are available for population history studies (EGA accession EGAS00001001587). The 10x Genomics sequencing data are available as two sets, the HGDP-CEPH samples with no restrictions (ENA accession ERP015796) and the others for population history studies (EGA accession EGAS00001001853).

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/357/6356/1160/suppl/DC1  
Materials and Methods  
Figs. S1 to S15  
Tables S1 to S4  
References (26–57)

6 April 2017; accepted 22 August 2017  
10.1126/science.aan3842

## A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea

Anders Bergström, Stephen J. Oppenheimer, Alexander J. Mentzer, Kathryn Auckland, Kathryn Robson, Robert Attenborough, Michael P. Alpers, George Koki, William Pomat, Peter Siba, Yali Xue, Manjinder S. Sandhu and Chris Tyler-Smith

*Science* **357** (6356), 1160-1163.  
DOI: 10.1126/science.aan3842

### Genetic history of Papua New Guinea peoples

Papua New Guinea was likely a stepping stone for human migration from Asia to Australia. Bergström *et al.* analyzed genome-wide autosomal data from several peoples in Papua New Guinea and determined population structure, divergence, and temporal size changes on the island. A sharp genetic divide is evident between the highlands and lowlands that appears to have occurred 10,000 to 20,000 years ago, concurrent with the spread of crop cultivation and the trans-New Guinea language family.

*Science*, this issue p. 1160

#### ARTICLE TOOLS

<http://science.sciencemag.org/content/357/6356/1160>

#### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2017/09/14/357.6356.1160.DC1>

#### REFERENCES

This article cites 47 articles, 6 of which you can access for free  
<http://science.sciencemag.org/content/357/6356/1160#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. The title *Science* is a registered trademark of AAAS.

Copyright © 2017, American Association for the Advancement of Science