

Title: A Neolithic Expansion but Strong Genetic Structure in the Independent History of New Guinea

Authors: Anders Bergström^{1*}, Stephen J Oppenheimer², Alexander J Mentzer³, Kathryn Auckland³, Kathryn Robson⁴, Robert Attenborough^{5,6}, Michael P Alpers^{7,8}, George Koki⁸, William Pomat⁸, Peter Siba⁸, Yali Xue¹, Manjinder S Sandhu¹, Chris Tyler-Smith^{1*}

Affiliations:

¹Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.

²School of Anthropology and Museum Ethnography, University of Oxford, Oxford, OX2 6PE, UK.

³Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.

⁴MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, OX3 9DS, UK.

⁵Biological Anthropology, Department of Archaeology & Anthropology, University of Cambridge, Cambridge, CB2 1QH, UK.

⁶School of Archaeology & Anthropology, Australian National University, Canberra, ACT 2601, Australia.

⁷International Health Research, Curtin University, Perth, 6845, Australia

⁸Papua New Guinea Institute of Medical Research, P.O. Box 60, Goroka, Papua New Guinea.

*Correspondence to: ab34@sanger.ac.uk (A.B); cts@sanger.ac.uk (C.T.-S)

Abstract:

New Guinea shows human occupation since ~50 thousand years ago (kya), independent adoption of plant cultivation ~10kya, and great cultural and linguistic diversity today. We performed genome-wide SNP genotyping on 381 individuals from 85 language groups in Papua New Guinea (PNG) and find a sharp divide originating 10-20kya between lowland and highland groups, and a lack of non-New Guinean admixture in the former. All highlanders share ancestry within the last 10kya, with major population growth in the same period, suggesting population structure was reshaped following the Neolithic lifestyle transition. However, genetic differentiation between groups in PNG 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.

One Sentence Summary: A Neolithic transition reshaped population structure in Papua New Guinea, but in comparison to other areas of the world, present-day genetic differentiation remains high.

Main Text:

The island of New Guinea contains some of the earliest archaeological evidence for modern humans outside of Africa, dating back to approximately 50kya (1). Starting ~10 kya, 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 approximately 850 languages (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 prehistory likely represents an independent instance of human genetic and cultural evolution over ~50ky. 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 (Figs. 1A, S1, tables S1, S2, (11)), and analysed 39 previously generated high-coverage whole-genome sequences (6, 12), including the PNG samples from the HGDP-CEPH panel (which we find consists of one highland and one lowland subset (fig. S2, 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, S5). The lowlands, however, harbour widespread Southeast Asian ancestry, with substantially higher levels in Austronesian speakers than in non-Austronesian speakers (mean of 38.7% vs 11.6%, $p = 1.4 \times 10^{-13}$, Wilcoxon rank sum test). The lowest levels (mean of 4.3%) are found in northern groups speaking 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 ~8kya, 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 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) (Figs. 2B, 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 south-eastern highlands (Figs. 2A, 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 PCA and D-statistics (Figs. 2A,C, S10, (11)). Thus, all highlanders seem to share the same relationship to lowland groups regardless of geographic location within the highlands. In line with this apparent genetic unity of highlanders, 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. 2E, 2G).

In contrast, highlanders as a group are not equally similar to all lowlanders (Fig. 2D), displaying 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). While 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 populations in PNG separated, we applied MSMC (18) to whole-genome sequences for six highland groups and one group from the Sepik lowlands. We used 10x Genomics linked-read whole-genome sequencing (19) to physically infer phase for eight genomes (table S4), and also analysed perfectly phased male X chromosomes (11). The results suggest that highlanders and Sepik lowlanders separated 10-20kya. All splits within the highlands seem to have occurred within the last ~10kya (Figs. 3A, B, S12). A Y-chromosomal phylogeny similarly revealed shared ancestry across groups within these timescales (fig. S13). We also find evidence of major increase in effective population sizes in most highlander groups in the last 10 ky (Figs. 3C, S14), using the SMC++ (20) and MSMC methods. Sepik lowlanders do not share this expansion, consistent with anthropological records of lower lowland population densities, likely linked to widespread malaria (21).

Genetic differentiation as measured by F_{ST} is much stronger in PNG than in regions of similar size in Eurasia, where 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} between eastern and western groups are 2-3%, and values between the Angan-speaking and other groups reach 4-5% (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 50ky, and the independence of New Guinea from Australia for much of this time. Present-day population structure on the mainland, marked by a very sharp highland-lowland division, does not date back to the initial peopling of Sahul, but instead appears to have formed within the last 20ky. Highland structure formed subsequently, mostly within the last 10ky, which is within the general timescale 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 the eastern highlands (figs. S9,S14), consistent with a hypothesized origin of cultivation in the former (22). Our results thus 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 like 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:

1. J. F. O'Connell, J. Allen, The process, biotic impact, and global implications of the human colonization of Sahul about 47,000 years ago. *J Archaeol Sci* **56**, 73-84 (2015).
2. T. Denham, H. Barton, The emergence of agriculture in New Guinea: a model of continuity from pre-existing foraging practices. In: Kennett DJ, Winterhalder B, editors. *Behavioral Ecology and the Transition to Agriculture*, 237–264 (2006).
3. M. P. Lewis, G. F. Simons, C. D. Fennig, *Ethnologue: Languages of the World*, Nineteenth edition. (2016).
4. A. Pawley, The chequered career of the Trans New Guinea hypothesis: recent research and its implications. In: *Papuan Pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. A. Pawley, R. Attenborough, J. Golson, R. Hide. Eds., 67–107 (2005).
5. A. Bergström, N. Nagle, Y. Chen, S. McCarthy, M. O. Pollard, Q. Ayub, S. Wilcox, L. Wilcox, R. A. van Oorschot, P. McAllister, L. Williams, Y. Xue, R. J. Mitchell, C. Tyler-Smith, Deep Roots for Aboriginal Australian Y Chromosomes. *Curr Biol* **26**, 809-813 (2016).
6. A. S. Malaspinas, M. C. Westaway, C. Muller, V. C. Sousa, O. Lao, I. Alves, A. Bergström, G. Athanasiadis, J. Y. Cheng, J. E. Crawford, T. H. Heupink, E. Macholdt, S. Peischl, S. Rasmussen, S. Schiffels, S. Subramanian, J. L. Wright, A. Albrechtsen, C. Barbieri, I. Dupanloup, A. Eriksson, A. Margaryan, I. Moltke, I. Pugach, T. S. Korneliussen, I. P. Levkivskyi, J. V. Moreno-Mayar, S. Ni, F.

- Racimo, M. Sikora, Y. Xue, F. A. Aghakhanian, N. Brucato, S. Brunak, P. F. Campos, W. Clark, S. Ellingvag, G. Fourmile, P. Gerbault, D. Injie, G. Koki, M. Leavesley, B. Logan, A. Lynch, E. A. Matisoo-Smith, P. J. McAllister, A. J. Mentzer, M. Metspalu, A. B. Migliano, L. Murgha, M. E. Phipps, W. Pomat, D. Reynolds, F. X. Ricaut, P. Siba, M. G. Thomas, T. Wales, C. M. Wall, S. J. Oppenheimer, C. Tyler-Smith, R. Durbin, J. Dortch, A. Manica, M. H. Schierup, R. A. Foley, M. M. Lahr, C. Bowern, J. D. Wall, T. Mailund, M. Stoneking, R. Nielsen, M. S. Sandhu, L. Excoffier, D. M. Lambert, E. Willerslev, A genomic history of Aboriginal Australia. *Nature* **538**, 207-214 (2016).
7. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Banffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szecsenyi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211 (2015).
 8. M. E. Allentoft, M. Sikora, K. G. Sjogren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlstrom, L. Vinner, A. S. Malaspinas, A. Margaryan, T. Higham, D. Chivall, N. Lynnerup, L. Harvig, J. Baron, P. Della Casa, P. Dabrowski, P. R. Duffy, A. V. Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolar, A. Kriiska, I. Lasak, C. Longhi, G. McGlynn, A. Merkevicus, I. Merkyte, M. Metspalu, R. Mkrtychyan, V. Moiseyev, L. Paja, G. Palfi, D. Pokutta, L. Pospieszny, T. D. Price, L. Saag, M. Sablin, N. Shishlina, V. Smrcka, V. I. Soenov, V. Szeverenyi, G. Toth, S. V. Trifanova, L. Varul, M. Vicze, L. Yepiskoposyan, V. Zhitenev, L. Orlando, T. Sicheritz-Ponten, S. Brunak, R. Nielsen, K. Kristiansen, E. Willerslev, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172 (2015).
 9. C. de Filippo, K. Bostoen, M. Stoneking, B. Pakendorf, Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proc Biol Sci* **279**, 3256-3263 (2012).
 10. I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gamarra, K. Sirak, S. Connell, K. Stewardson, E. Harney, Q. Fu, G. Gonzalez-Fortes, E. R. Jones, S. A. Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Eshed, A. S. Mizrahi, C. Meiklejohn, F. Gerritsen, L. Bejenaru, M. Bluhner, A. Campbell, G. Cavalleri, D. Comas, P. Froguel, E. Gilbert, S. M. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. A. Merriwether, S. O'Reilly, M. B. Richards, O. Semino, M. Shamoon-Pour, G. Stefanescu, M. Stumvoll, A. Tonjes, A. Torroni, J. F. Wilson, L. Yengo, N. A. Hovhannisyan, N. Patterson, R. Pinhasi, D. Reich, Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419-424 (2016).
 11. Materials and methods are available as supplementary materials at the Science website.
 12. M. Raghavan, M. Steinrucken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, A. Albrechtsen, C. Valdiosera, M. C. Avila-Arcos, A. S. Malaspinas, A. Eriksson, I. Moltke, M. Metspalu, J. R. Homburger, J. Wall, O. E. Cornejo, J. V. Moreno-Mayar, T. S. Korneliusen, T. Pierre, M. Rasmussen, P. F. Campos, B. Damgaard Pde, M. E. Allentoft, J. Lindo, E. Metspalu, R. Rodriguez-Varela, J. Mansilla, C. Henrickson, A. Seguin-Orlando, H. Malmstrom, T. Stafford, Jr., S. S. Shringarpure, A. Moreno-Estrada, M. Karmin, K. Tambets, A. Bergström, Y. Xue, V. Warmuth, A. D. Friend, J. Singarayer, P. Valdes, F. Balloux, I. Lebreiro, J. L. Vera, H. Rangel-Villalobos, D. Pettener, D. Luiselli, L. G. Davis, E. Heyer, C. P. Zollikofer, M. S. Ponce de Leon, C. I. Smith, V. Grimes, K. A. Pike, M. Deal, B. T. Fuller, B. Arriaza, V. Standen, M. F. Luz, F. Ricaut, N. Guidon, L. Osipova, M. I. Voevoda, O. L. Posukh, O. Balanovsky, M. Lavryashina, Y. Bogunov, E. Khusnutdinova, M. Gubina, E. Balanovska, S. Fedorova, S. Litvinov, B. Malyarchuk, M. Derenko, M. J. Mosher, D. Archer, J. Cybulski, B. Petzelt, J. Mitchell, R. Worl, P. J. Norman, P. Parham, B. M. Kemp, T. Kivisild, C. Tyler-Smith, M. S. Sandhu, M. Crawford, R. Villems, D. G. Smith, M. R. Waters, T. Goebel, J. R. Johnson, R. S. Malhi, M. Jakobsson, D. J. Meltzer, A. Manica, R. Durbin, C. D. Bustamante, Y. S. Song, R. Nielsen, E. Willerslev, Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).
 13. A. T. Duggan, M. Stoneking, Recent developments in the genetic history of East Asia and Oceania. *Curr Opin Genet Dev* **29**, 9-14 (2014).
 14. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
 15. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664 (2009).

16. J. S. Friedlaender, F. R. Friedlaender, F. A. Reed, K. K. Kidd, J. R. Kidd, G. K. Chambers, R. A. Lea, J. H. Loo, G. Koki, J. A. Hodgson, D. A. Merriwether, J. L. Weber, The genetic structure of Pacific Islanders. *PLoS Genet* **4**, e19 (2008).
17. P. Swadling, P. Wiessner, A. Tumu, Prehistoric stone artefacts from Enga and the implication of links between the highlands, lowlands and islands for early agriculture in Papua New Guinea. *Le Journal de la Société des Océanistes*, 271-292 (2008).
18. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple genome sequences. *Nat Genet* **46**, 919-925 (2014).
19. G. X. Zheng, B. T. Lau, M. Schnall-Levin, M. Jarosz, J. M. Bell, C. M. Hindson, S. Kyriazopoulou-Panagiotopoulou, D. A. Masquelier, L. Merrill, J. M. Terry, P. A. Mudivarti, P. W. Wyatt, R. Bharadwaj, A. J. Makarewicz, Y. Li, P. Belgrader, A. D. Price, A. J. Lowe, P. Marks, G. M. Vurens, P. Hardenbol, L. Montesclaros, M. Luo, L. Greenfield, A. Wong, D. E. Birch, S. W. Short, K. P. Bjornson, P. Patel, E. S. Hopmans, C. Wood, S. Kaur, G. K. Lockwood, D. Stafford, J. P. Delaney, I. Wu, H. S. Ordonez, S. M. Grimes, S. Greer, J. Y. Lee, K. Belhocine, K. M. Giorda, W. H. Heaton, G. P. McDermott, Z. W. Bent, F. Meschi, N. O. Kondov, R. Wilson, J. A. Bernate, S. Gauby, A. Kindwall, C. Bermejo, A. N. Fehr, A. Chan, S. Saxonov, K. D. Ness, B. J. Hindson, H. P. Ji, Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat Biotechnol* **34**, 303-311 (2016).
20. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet* **49**, 303-309 (2017).
21. I. D. Riley, Population-Change and Distribution in Papua-New-Guinea - an Epidemiological Approach. *J Hum Evol* **12**, 125-132 (1983).
22. D. K. Feil, *The Evolution of Highland Papua New Guinea Societies*. (Cambridge University Press, 1987).
23. P. Skoglund, H. Malmstrom, A. Omrak, M. Raghavan, C. Valdiosera, T. Gunther, P. Hall, K. Tambets, J. Parik, K. G. Sjogren, J. Apel, E. Willerslev, J. Stora, A. Gotherstrom, M. Jakobsson, Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747-750 (2014).
24. T. N. Huffman, Archaeology and Ethnohistory of the African Iron-Age. *Annu Rev Anthropol* **11**, 133-150 (1982).
25. G. D. Poznik, Y. Xue, F. L. Mendez, T. F. Willems, A. Massaia, M. A. Wilson Sayres, Q. Ayub, S. A. McCarthy, A. Narechania, S. Kashin, Y. Chen, R. Banerjee, J. L. Rodriguez-Flores, M. Cerezo, H. Shao, M. Gymrek, A. Malhotra, S. Louzada, R. Desalle, G. R. Ritchie, E. Cerveira, T. W. Fitzgerald, E. Garrison, A. Marcketta, D. Mittelman, M. Romanovitch, C. Zhang, X. Zheng-Bradley, G. R. Abecasis, S. A. McCarroll, P. Flicek, P. A. Underhill, L. Coin, D. R. Zerbino, F. Yang, C. Lee, L. Clarke, A. Auton, Y. Erlich, R. E. Handsaker, C. Genomes Project, C. D. Bustamante, C. Tyler-Smith, Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *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, Jonathan Friedlaender for background information on the HGDP-CEPH samples and the Wellcome Trust Sanger Institute genotyping and sequencing facilities for generating data, especially Michael Quail, David Jackson and Steven Leonard for generating 10x Genomics data. A.B., Y.X., M.S.S. and C.T.-S. were supported by the Wellcome Trust (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 Adrian VS 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).

List of Supplementary materials:

Materials and Methods

Figs. S1 to S15

Tables S1 to S4

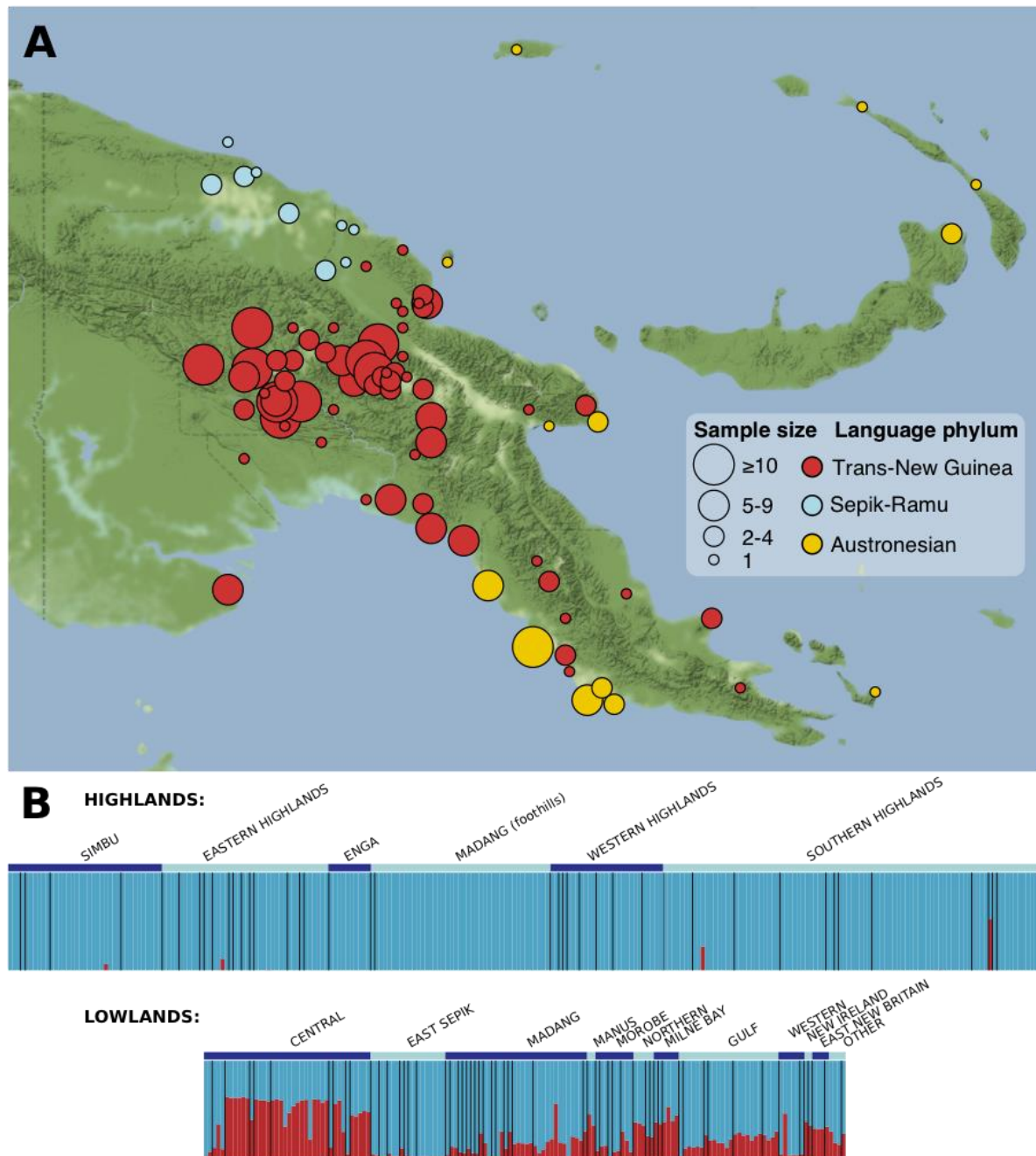


Fig. 1. PNG samples. (A) Each language group is represented by a circle, the area of which indicates the number of genotyped individuals and the colour the top-level language phylum. 39 individuals are not included as the specific language is unknown or the two parents are from different language groups. Also see fig. S1. (B) Papuan (blue) and Southeast Asian (red) ancestry proportions as estimated by ADMIXTURE (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 ($r = 0.988$) with those estimated using f_4 -ratios (11).

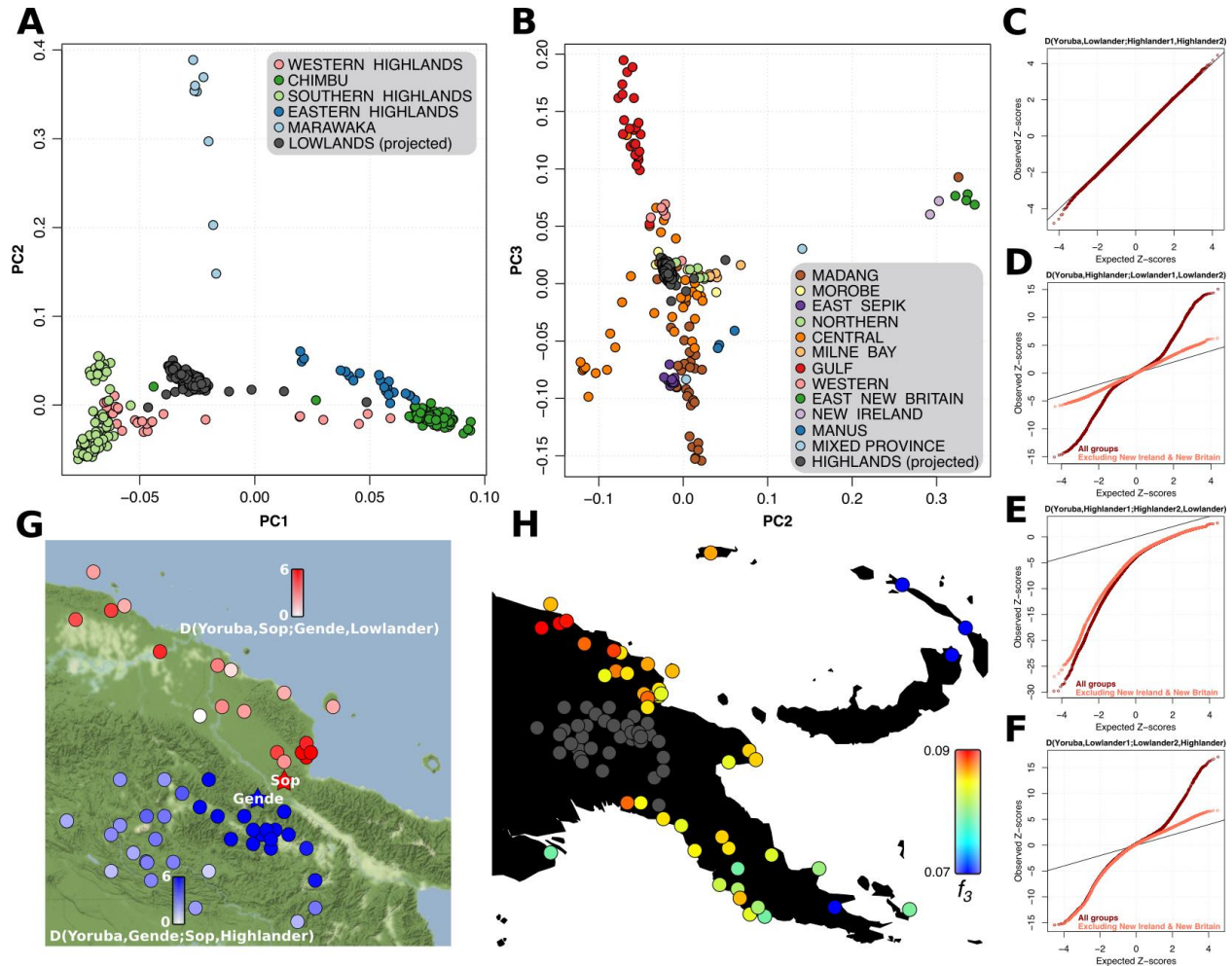


Fig. 2. PNG Population structure. (A) When projected onto principal components constructed with only highlander genotypes, all lowlanders (excepting a few outliers) group uniformly. Also see fig. S10. (B) When projected onto principal components constructed with only lowlander genotypes, all highlanders (excepting a few outliers) group uniformly. Also see fig. S11. (C-F) Quantile-quantile plots comparing Z-scores from D-statistics relating highlanders and lowlanders to those expected under a normal distribution (11). (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) Z-scores of two

different D-statistics, the first measuring if the highland Gende speakers are more similar to the lowland Sop speakers, living just 40 km away, or to other highlanders (blue meaning more highlander similarity), and the second if Sop are more similar to Gende or to other lowlanders (red meaning more lowlander similarity). Z-scores are capped at 6. **(H)** Genetic affinity of highlanders (treated as a single group, marked in grey) to different lowland groups measured by the outgroup f_3 statistic $f_3(\text{Highlanders}, X; \text{Aboriginal Australian})$ (red meaning higher affinity). C-H were calculated after masking lowlander genomes for Southeast Asian ancestry.

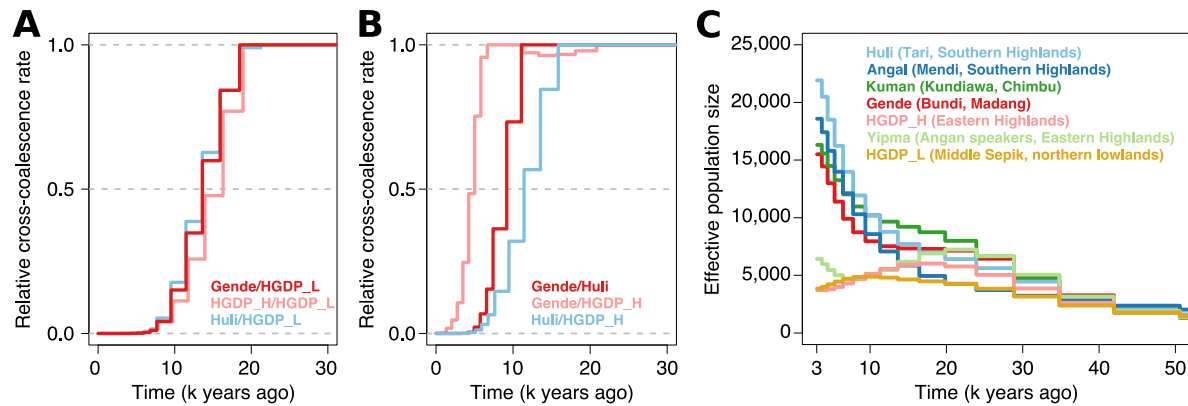


Fig. 3. Time depth of population separation and growth in PNG. (A) Cross-coalescence curves between highlanders and a northern lowlands Middle Sepik group suggests a split time between 10 and 20kya. (B) Cross-coalescence curves between highland groups suggest split times within the last ~10ky (Huli representing the western cluster, Gende and HGDP_H the eastern, figs. S8, S10). These were inferred using MSMC on genomes physically phased using linked-read sequencing. Also see fig. S12. (C) Effective population size histories inferred using SMC++ on five genomes per group. Also see fig. S14.

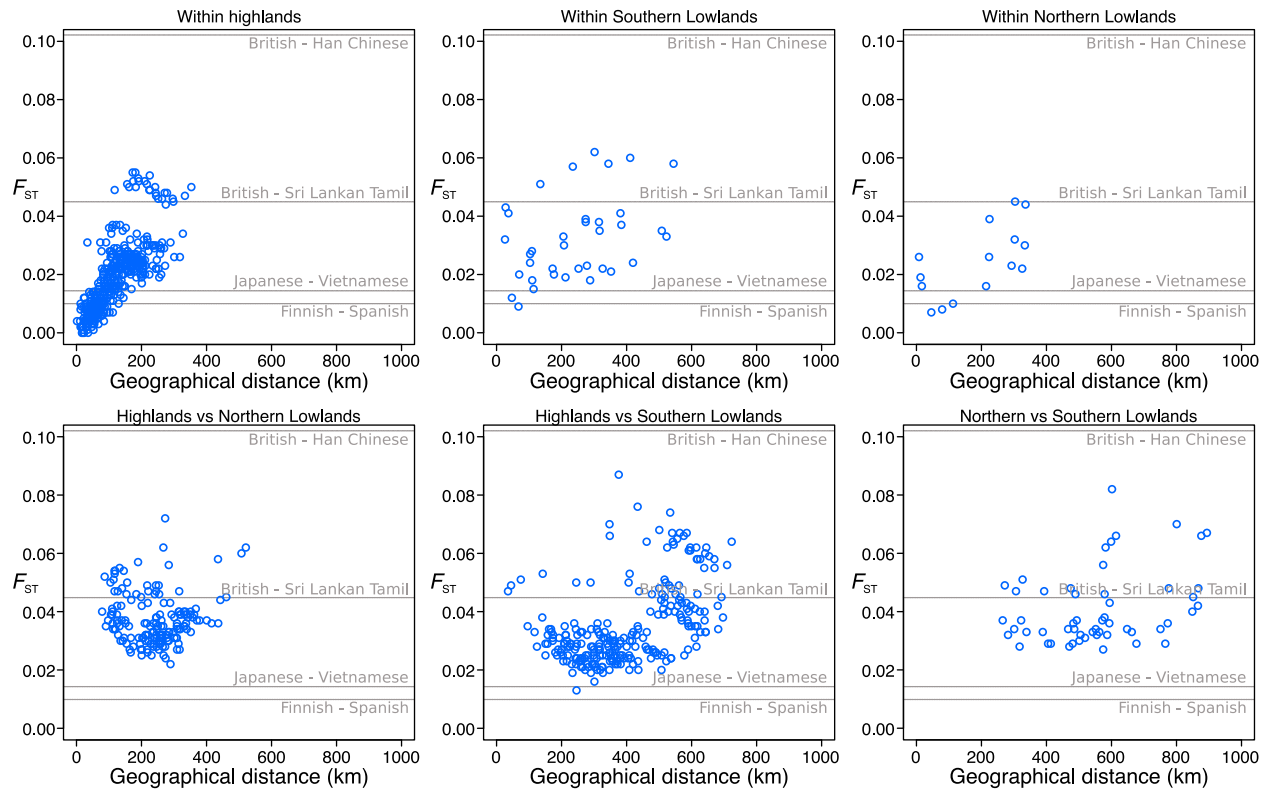


Fig. 4. Genetic differentiation in PNG. Geographical distance between groups plotted against F_{ST} , calculated after masking lowlander genomes for Southeast Asian ancestry. Grey lines indicate F_{ST} between selected 1000 Genomes Project populations.