



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Genomic analyses inform on migration events during the peopling of Eurasia

### Citation for published version:

Pagani, L, Lawson, DJ, Jagoda, E, Mörseburg, A, Eriksson, A, Mitt, M, Clemente, F, Hudjashov, G, Degiorgio, M, Saag, L, Wall, JD, Cardona, A, Mägi, R, Sayres, MAW, Kaewert, S, Inchley, C, Scheib, CL, Järve, M, Karmin, M, Jacobs, GS, Antao, T, Iliescu, FM, Kushniarevich, A, Ayub, Q, Tyler-smith, C, Xue, Y, Yunusbayev, B, Tambets, K, Mallick, CB, Saag, L, Pocheshkhova, E, Andriadze, G, Muller, C, Westaway, MC, Lambert, DM, Zoraqi, G, Turdikulova, S, Dalimova, D, Sabitov, Z, Sultana, GNN, Lachance, J, Tishkoff, S, Momynaliev, K, Isakova, J, Damba, LD, Gubina, M, Nymadawa, P, Evseeva, I, Atramentova, L & Utevska, O 2016, 'Genomic analyses inform on migration events during the peopling of Eurasia', *Nature*, vol. 538, no. 7624, pp. 238-242. <https://doi.org/10.1038/nature19792>

### Digital Object Identifier (DOI):

[10.1038/nature19792](https://doi.org/10.1038/nature19792)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Peer reviewed version

### Published In:

Nature

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



1 **Geographical barriers, environmental challenges, and**  
2 **complex migration events during the peopling of**  
3 **Eurasia**

4  
5 **Authors List**

6 Luca Pagani<sup>1,2\*†</sup>, Daniel John Lawson<sup>3\*</sup>, Evelyn Jagoda<sup>1,4\*</sup>, Alexander Mörseburg<sup>1\*</sup>,  
7 Anders Eriksson<sup>5,6\*</sup>, Mario Mitt<sup>7,8</sup>, Florian Clemente<sup>1,9</sup>, Georgi Hudjashov<sup>10,11,12</sup>,  
8 Michael DeGiorgio<sup>13</sup>, , Lauri Saag<sup>10</sup>, Jeffrey D. Wall<sup>14</sup>, Alexia Cardona<sup>1,15</sup>, Reedik  
9 Mägi<sup>7</sup>, Melissa A. Wilson Sayres<sup>16,17</sup>, Sarah Kaewert<sup>1</sup>, Charlotte Inchley<sup>1</sup>,  
10 Christiana L. Scheib<sup>1</sup>, Mari Järve<sup>10</sup>, Monika Karmin<sup>10,18</sup>, Guy S. Jacobs<sup>19,20</sup>, Tiago  
11 Antao<sup>21</sup>, Florin Mircea Iliescu<sup>1</sup>, Alena Kushniarevich<sup>10,22</sup>, Qasim Ayub<sup>23</sup>, Chris  
12 Tyler-Smith<sup>23</sup>, Yali Xue<sup>23</sup>, Bayazit Yunusbayev<sup>10,24</sup>, Kristiina Tambets<sup>10</sup>, Chandana  
13 Basu Mallick<sup>10</sup>, Lehti Saag<sup>18</sup>, Elvira Pocheshkhova<sup>25</sup>, George Andriadze<sup>26</sup>, Craig  
14 Muller<sup>27</sup>, Michael C. Westaway<sup>28</sup>, David Lambert<sup>28</sup>, Grigor Zoraqi<sup>29</sup>, Shahlo  
15 Turdikulova<sup>30</sup>, Dilbar Dalimova<sup>31</sup>, Zhaxylyk Sabitov<sup>32</sup>, Gazi Nurun Nahar  
16 Sultana<sup>33</sup>, Joseph Lachance<sup>34,35</sup>, Sarah Tishkoff<sup>36</sup>, Kuvat Momynaliev<sup>37</sup>, Jainagul  
17 Isakova<sup>38</sup>, Larisa D. Damba<sup>39</sup>, Marina Gubina<sup>39</sup>, Pagbajabyn Nymadawa<sup>40</sup>, Irina  
18 Evseeva<sup>41,42</sup>, Lubov Atramentova<sup>43</sup>, Olga Utevska<sup>43</sup>, François-Xavier Ricaut<sup>44</sup>,  
19 Harilanto Razafindrazaka<sup>44</sup>, Herawati Sudoyo<sup>45</sup>, Thierry Letellier<sup>44</sup>, Murray P.  
20 Cox<sup>12</sup>, Nikolay A. Barashkov<sup>46,47</sup>, Vedrana Skaro<sup>48,49</sup>, Lejla Mulahasanovic<sup>50</sup>,  
21 Dragan Primorac<sup>51,52,53,49</sup>, Hovhannes Sahakyan<sup>10,54</sup>, Maru Mormina<sup>55</sup>, Christina  
22 A. Eichstaedt<sup>1,56</sup>, Daria V. Lichman<sup>39,57</sup>, Syafiq Abdullah<sup>58</sup>, Gyaneshwer Chaubey<sup>10</sup>,  
23 Joseph T. S. Wee <sup>59</sup>, Evelin Mihailov<sup>7</sup>, Alexandra Karunas<sup>24,60</sup>, Sergei  
24 Litvinov<sup>24,60,10</sup>, Rita Khusainova<sup>24,60</sup>, Natalya Ekomasova<sup>60</sup>, Vita Akhmetova<sup>24</sup>,  
25 Irina Khidiyatova<sup>24,60</sup>, Damir Marjanovic<sup>61,62</sup>, Levon Yepiskoposyan<sup>54</sup>, Doron M.  
26 Behar<sup>10</sup>, Elena Balanovska<sup>63</sup>, Andres Metspalu<sup>7</sup>, Miroslava Derenko<sup>64</sup>, Boris  
27 Malyarchuk<sup>64</sup>, Mikhail Voevoda<sup>65,39,57</sup>, Sardana A. Fedorova<sup>47,46</sup>, Ludmila P.  
28 Osipova<sup>39,57</sup>, Marta Mirazón Lahr<sup>66</sup>, Pascale Gerbault<sup>67</sup>, Matthew Leavesley<sup>68,69</sup>,  
29 Andrea Bamberg Migliano<sup>70</sup>, Michael Petraglia<sup>71</sup>, Oleg Balanovsky<sup>72,63</sup>, Elza K.  
30 Khusnutdinova<sup>24,60</sup>, Ene Metspalu<sup>10,18</sup>, Mark G. Thomas<sup>67</sup>, Andrea Manica<sup>6</sup>,  
31 Rasmus Nielsen<sup>73</sup>, Richard Villems<sup>10,18,74\*</sup>, Eske Willerslev<sup>27\*</sup>, Toomas  
32 Kivisild<sup>1,10\*†</sup>, Mait Metspalu<sup>10,18\*†</sup>

- 1 *\*These authors contributed equally to this work.*
- 2 *†Corresponding authors: L.P. ([lp.lucapagani@gmail.com](mailto:lp.lucapagani@gmail.com)), T.K. ([tk331@cam.ac.uk](mailto:tk331@cam.ac.uk)),*
- 3 *M.M. ([mait@ebc.ee](mailto:mait@ebc.ee))*
- 4

1 **Author Affiliations**

2 1: Department of Biological Anthropology, University of Cambridge, Cambridge,  
3 United Kingdom

4 2: Department of Biological, Geological and Environmental Sciences, University of  
5 Bologna, Via Selmi 3, 40126, Bologna, Italy

6 3: Integrative Epidemiology Unit, School of Social and Community Medicine,  
7 University of Bristol, Bristol BS8 2BN, UK.

8 4: Department of Human Evolutionary Biology, Harvard University, Cambridge, MA  
9 02138, USA

10 5: Integrative Systems Biology Lab, Division of Biological and Environmental  
11 Sciences & Engineering, King Abdullah University of Science and Technology,  
12 Thuwal, Kingdom of Saudi Arabia

13 6: Department of Zoology, University of Cambridge, Cambridge, UK

14 7: Estonian Genome Center, University of Tartu, Tartu, Estonia

15 8: Department of Biotechnology, Institute of Molecular and Cell Biology, University  
16 of Tartu, Tartu, Estonia

17 9: Institut de Biologie Computationnelle, Université Montpellier 2, Montpellier,  
18 France

19 10: Estonian Biocentre, Tartu, Estonia

20 11: Department of Psychology, University of Auckland, Auckland, 1142, New  
21 Zealand;

22 12: Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey  
23 University, Palmerston North, New Zealand

24 13: Department of Biology, Pennsylvania State University, University Park, PA,  
25 16802, USA

26 14: Institute for Human Genetics, University of California, San Francisco, California  
27 94143, USA

28 15: MRC Epidemiology Unit, University of Cambridge, Institute of Metabolic  
29 Science, Box 285, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ

30 16: School of Life Sciences, Tempe, AZ, 85287 USA

31 17: Center for Evolution and Medicine, The Biodesign Institute, Tempe, AZ, 85287  
32 USA



1 18: Department of Evolutionary Biology, Institute of Molecular and Cell Biology,  
2 University of Tartu, Tartu, Estonia  
3 19: Mathematical Sciences, University of Southampton, Southampton SO17 1BJ, UK  
4 20: Institute for Complex Systems Simulation, University of Southampton,  
5 Southampton SO17 1BJ, UK  
6 21: Division of Biological Sciences, University of Montana, Missoula, MT, USA  
7 22: Institute of Genetics and Cytology, National Academy of Sciences, Minsk,  
8 Belarus  
9 23: The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United  
10 Kingdom  
11 24: Institute of Biochemistry and Genetics, Ufa Scientific Center of RAS, Ufa , Russia  
12 25: Kuban State Medical University, Krasnodar, Russia  
13 26: Scientific-Research Center of the Caucasian Ethnic Groups, St. Andrews  
14 Georgian University, Georgia  
15 27: Center for GeoGenetics, University of Copenhagen, Denmark  
16 28: Research Centre for Human Evolution, Environmental Futures Research  
17 Institute, Griffith University, Nathan, Australia  
18 29: Center of Molecular Diagnosis and Genetic Research, University Hospital of  
19 Obstetrics and Gynecology, Tirana, Albania  
20 30: Center of High Technology, Academy of Sciences, Republic of Uzbekistan  
21 31: Institute of Bioorganic Chemistry Academy of Science, Republic of Uzbekistan  
22 32: L.N. Gumilyov Eurasian National University, Astana, Kazakhstan  
23 33: Centre for Advanced Research in Sciences (CARS), DNA Sequencing Research  
24 Laboratory, University of Dhaka, Dhaka-1000, Bangladesh  
25 34: Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104-  
26 6145, USA  
27 35: School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA  
28 36: Departments of Genetics and Biology, University of Pennsylvania, Philadelphia,  
29 Pennsylvania, USA  
30 37: DNcode laboratories, Moscow, Russia  
31 38: Institute of Molecular Biology and Medicine, Bishkek, Kyrgyz Republic  
32 39: Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of  
33 Sciences, Novosibirsk, Russia

- 1 40: Mongolian Academy of Medical Sciences, Ulaanbaatar, Mongolia
- 2 41: Northern State Medical University, Arkhangelsk, Russia
- 3 42: Anthony Nolan, London, United Kingdom
- 4 43: V. N. Karazin Kharkiv National University, Kharkiv, Ukraine
- 5 44: Evolutionary Medicine group, Laboratoire d'Anthropologie Moléculaire et
- 6 Imagerie de Synthèse, UMR 5288, Centre National de la Recherche Scientifique,
- 7 Université de Toulouse 3, Toulouse, France
- 8 45: Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular
- 9 Biology, Jakarta, Indonesia
- 10 46: Department of Molecular Genetics, Yakut Scientific Centre of Complex Medical
- 11 Problems, Yakutsk, Russia
- 12 47: Laboratory of Molecular Biology, Institute of Natural Sciences, M.K. Ammosov
- 13 North-Eastern Federal University, Yakutsk, Russia
- 14 48: Genos, DNA laboratory, Zagreb, Croatia
- 15 49: University of Osijek, Medical School, Osijek, Croatia
- 16 50: Center for Genomics and Transcriptomics, CeGaT, GmbH, Tübingen, Germany
- 17 51: St. Catherine Speciality Hospital, Zabok, Croatia
- 18 52: Eberly College of Science, The Pennsylvania State University, University Park,
- 19 PA, USA
- 20 53: University of Split, Medical School, Split, Croatia
- 21 54: Laboratory of Ethnogenomics, Institute of Molecular Biology, National
- 22 Academy of Sciences, Republic of Armenia, 7 Hasratyan Street, 0014, Yerevan,
- 23 Armenia
- 24 55: Department of Applied Social Sciences, University of Winchester, Sparkford
- 25 Road, Winchester SO22 4NR, UK
- 26 56: Thoraxclinic at the University Hospital Heidelberg, Heidelberg, Germany
- 27 57: Novosibirsk State University, Novosibirsk, Russia.
- 28 58: RIPAS Hospital, Bandar Seri Begawan, Brunei Darussalam
- 29 59: National Cancer Centre Singapore, Singapore
- 30 60: Department of Genetics and Fundamental Medicine, Bashkir State University,
- 31 Ufa, Russia

- 1 61: *Department of Genetics and Bioengineering. Faculty of Engineering and*  
2 *Information Technologies, International Burch University, Sarajevo, Bosnia and*  
3 *Herzegovina*
- 4 62: *Institute for Anthropological Researches, Zagreb, Croatia*
- 5 63: *Research Centre for Medical Genetics, Russian Academy of Sciences, Moscow*  
6 *115478, Russia*
- 7 64: *Genetics Laboratory, Institute of Biological Problems of the North, Russian*  
8 *Academy of Sciences, Magadan, Russia*
- 9 65: *Institute of Internal Medicine, Siberian Branch of Russian Academy of Medical*  
10 *Sciences, Novosibirsk, Russia*
- 11 66: *Leverhulme Centre for Human Evolutionary Studies, Department of*  
12 *Archaeology and Anthropology, University of Cambridge, Cambridge, United*  
13 *Kingdom*
- 14 67: *Research Department of Genetics, Evolution and Environment, University*  
15 *College London, London, United Kingdom*
- 16 68: *Department of Archaeology, University of Papua New Guinea, University PO*  
17 *Box 320, NCD, Papua New Guinea*
- 18 69: *College of Arts, Society and Education, James Cook University, PO Box 6811,*  
19 *Cairns QLD 4870, Australia*
- 20 70: *Department of Anthropology, University College London, London, United*  
21 *Kingdom*
- 22 71: *School of Archaeology, University of Oxford, 75 George Street, Oxford, OX1 2BQ,*  
23 *UK*
- 24 72: *Vavilov Institute for General Genetics, Russian Academy of Sciences, Moscow,*  
25 *Russia*
- 26 73: *Department of Integrative Biology, University of California Berkeley, Berkeley*  
27 *94720, CA, USA*
- 28 74: *Estonian Academy of Sciences, 6 Kohtu Street, Tallinn 10130, Estonia*
- 29
- 30

1 **Abstract**

2 Previous human genetic studies, based on sampling small numbers of  
3 populations, have supported a recent Out-of-Africa dispersal model with minor  
4 additional input from archaic humans. Here, we present a novel dataset of 379  
5 high-coverage human genomes from 125 populations worldwide. The  
6 combination of high spatial and genomic coverage enabled us to refine current  
7 knowledge of continent-wide patterns of heterozygosity, long- and short-  
8 distance gene flow, archaic admixture, and changes in effective population size.  
9 Compared to Eurasians, the examined Papuan genomes show an excess of highly  
10 derived modern human haplotypes and deeper split times from Africans. This is  
11 compatible with an early and largely extinct expansion of modern humans Out-  
12 of-Africa. This is also indicated by the Western Asian fossil record and the recent  
13 discovery of modern human and Neanderthal admixture 100,000 years ago,  
14 which significantly predates the main Out-of-Africa expansion of modern  
15 humans. Our tests of positive and balancing selection highlight a number of new  
16 metabolism- and immunity-related loci as candidates for local adaptation.

17

## 1 **Introduction**

2

3        Previous genome-wide sequencing efforts have aimed at characterizing  
4 common variants in the human genome by targeting moderate numbers of  
5 geographically distinct populations and combining genotyping, low-coverage  
6 whole-genome and exome sequencing data<sup>1,2</sup>. High-coverage whole-genome  
7 sequence studies have so far been limited to focusing on specific populations<sup>3</sup>  
8 and geographic regions<sup>4-7</sup>, or targeted at specific diseases, e.g. cancer<sup>8</sup>.  
9 Nevertheless, the availability of high-resolution genomic data has led to the  
10 development of new methodologies for inferring population history<sup>9-13</sup> and  
11 refuelled the debate on the mutation rate in humans<sup>14</sup>. From these initial studies,  
12 the unprecedented potential of high-coverage genomic data to reveal  
13 geographically specific patterns of genetic diversity has become evident. Here,  
14 we present a new dataset of high-coverage human genomes from nearly 150  
15 populations distributed worldwide. This comprehensive population sample,  
16 which, among others, includes new samples from Siberia, Island Southeast Asia  
17 and Papua New Guinea, allows us to infer human demographic history in finer  
18 detail and to investigate signatures of natural selection. We estimate split times  
19 among populations, test how the different populations conform to the model of a  
20 single expansion out of Africa with archaic admixture (OoA), and assess patterns  
21 of neutral and adaptive variation associated with different environments.

22        **Data description.** Our worldwide panel of 483 high-coverage human  
23 genomes from 148 populations includes 379 new genomes from 125 populations  
24 (Figure 1) (Table S1.7-I). All genomes were sequenced by Complete Genomics  
25 Inc. and mapped, called and phased using the same bioinformatic pipeline,  
26 thereby minimizing platform and processing bias conflicts (Supplementary  
27 Section 1.1). We maximised the number of groups in this study by limiting the  
28 number of individuals to three for most populations. Existing SNP-chip  
29 information was used in most cases to choose unrelated individuals and to avoid  
30 cases of recent admixture between geographically distant populations. For  
31 demographic inferences, we combined previously published and new sequences  
32 to generate a geographically balanced sample (Figure 1, Diversity Set, N=447).  
33 For selection scan analysis, we focussed on well-covered geographic regions,

1 combining a subset of the Diversity Set with published sequences (Figure 1,  
2 Selection Set, N = 396, Supplementary Section 1.7).

3 **The current view on the peopling of Eurasia.** The timing and route of  
4 human movements out of Africa, as well as the degree to which migrating  
5 populations interbred with archaic humans during their expansion across  
6 Eurasia, have been the subject of considerable debate over the past two  
7 decades<sup>15</sup>. Fossil evidence demonstrates that *Homo sapiens* was present in  
8 Levant between ca. 120-70 kya<sup>16</sup>. This colonization has, however, been viewed as  
9 a failed expansion OoA<sup>17</sup>. Nevertheless craniometrical studies of African and  
10 Asian populations<sup>18</sup> and fossil data from eastern Asia<sup>15</sup>, including the very recent  
11 reports of human remains in China from before 80 kya<sup>19</sup>, admit the possibility of  
12 an early dispersal. Moreover, archaeological finds in Arabia and South Asia  
13 indicate the presence of human populations in ameliorated environments  
14 between 125 and 75 kya<sup>15</sup>. Previous genetic analyses of living populations have  
15 revealed a steady decline in genetic diversity with distance from Africa, which is  
16 consistent with a serial founder event model<sup>20-22</sup>.

17 Ancient DNA (aDNA) sequencing has further contributed to our  
18 understanding of the peopling of Eurasia and revealed admixture with at least  
19 two archaic human lineages. Neanderthals have left a genetic signature in all  
20 non-Africans from around 55 kya<sup>23</sup>, while admixture with Denisova was largely  
21 restricted to the ancestors of modern Papuan and Australian populations<sup>24</sup>. In  
22 addition aDNA from modern humans indicates population structuring and turnover,  
23 but little additional archaic admixture, in Eurasia over the last 35-45 thousand  
24 years<sup>25-27</sup>. Overall, these findings provide support for a model<sup>28,29</sup> by which the  
25 vast majority of human genetic diversity outside Africa derives from a single  
26 dispersal event that was followed by admixture with archaic humans<sup>23,29</sup>.

27

## 28 **Results**

29

30 **Population structure in Eurasia.** We used ADMIXTURE<sup>30</sup> to infer  
31 genetic structure and admixture patterns in our Diversity Set (Figure 1 for K=8  
32 and K=14, Supplementary Sections 2.1.1-2 for Ks=2-14). Western Eurasia is  
33 characterised by two predominant genetic clusters, whilst the much less

1 populous Siberia shows evidence of three differentiated clusters (Figure 1,  
2 K=14), consistent with previous reports<sup>31</sup>. Island Southeast Asia also exhibits  
3 high population structuring. Both these latter two regions have histories of small  
4 effective population densities (Figure S2.2.3-I, as inferred by MSMC<sup>10</sup>), which  
5 increase genetic drift and local differentiation.

6 We compared the haplotype similarity of our samples using fineSTRUCTURE<sup>32</sup>.  
7 This shows that our sampling strategy retains the power to identify population  
8 structure at fine resolution. We inferred 106 genetically distinct populations  
9 forming 12 major regional clusters, corresponding well to the 148 self-identified  
10 population labels. This clustering is based on an individual level measure of  
11 haplotype similarity, which is sensitive to small and recent genetic  
12 differentiation, and forms the basis for the groupings used in the scans of natural  
13 selection.

14

15 **The importance of geography.** The dense geographic coverage of our  
16 samples allowed us to investigate the importance of geographic barriers in  
17 shaping gene flow. We did so by interpolating genetic variation spatially,  
18 focussing on measures of pairwise similarity between genomes in pairs of  
19 populations (Supplementary Section 2.2.2). We considered several similarity  
20 measures (Supplementary Section 2.2.2) and report gradients of allele  
21 frequencies in Figure 2. We validated the approach using isolation by distance  
22 patterns across major gradients and migration surfaces reconstructed using  
23 EEMS<sup>33</sup>. The main features are the East-West Eurasian split near the Ural  
24 Mountains, and the Tibetan plateau, as expected. To formally link these patterns  
25 to geographic features, we quantified the effects of elevation, temperature, and  
26 precipitation on genetic gradients while controlling for pairwise geographic  
27 distances (Supplementary section 2.2.2). This analysis identifies precipitation  
28 and elevation as environmental variables that correlate most strongly with the  
29 genetic gradients estimated from allele frequencies (inset of Figure 2).

30 **Differentiation in Eurasia after the expansion out of Africa.** We  
31 observe the well-documented decrease in the number of heterozygous sites per  
32 genome as a function of distance from East Africa (Figure 1); a pattern consistent  
33 with a model of serial founder events during the peopling of Eurasia<sup>20,21</sup>.



1 While this pattern is relatively smooth, there are a number of discontinuities that  
2 potentially highlight geographic regions that acted as barriers during the  
3 expansion. Such discontinuities can be visualised by plotting the outgroup  $f_3$   
4 statistic<sup>13,34</sup> in the form  $f_3(X, Y; \text{Yoruba})$ , which here measures shared drift  
5 between non-African populations  $X$  and  $Y$  from Yoruba as an African outgroup  
6 (Supplementary Section 2.2.6, Figures S2.2.6-I-II). We tested all possible  
7 combinations of  $X$  and  $Y$  within our Diversity Set and 25 published aDNA  
8 genomes. While recapitulating the main groupings inferred by ADMIXTURE and  
9 fineSTRUCTURE, the outgroup  $f_3$  statistic also flags populations that have  
10 experienced additional drift. For example, the  $f_3$  values are similar for  
11 comparisons within Caucasus populations and between populations from  
12 Europe and Caucasus. The  $f_3$  values for comparisons within Europe, however, are  
13 significantly higher. These findings are consistent with a simple model of  
14 population splits within the Caucasus dating to approximately the same time as  
15 the split between European and Caucasus populations<sup>35</sup>.

16 **An excess of old haplotypes in Sahul.** Our fineSTRUCTURE analysis  
17 highlights an excess of shorter African haplotypes in Papuans, as well as  
18 Philippine Negritos, compared to all other non-African populations. This pattern  
19 remains after correcting for potential confounders such as phasing errors and  
20 sampling bias (Figure S2.2.1-VII, Supplementary Section 2.2.1). A natural  
21 interpretation from population genetics theory is that these shorter shared  
22 haplotypes reflect an older population split<sup>36</sup>.

23 We further investigated whether Sahul populations differ from other Eurasian  
24 populations by estimating population splits using MSMC<sup>10</sup>. We focussed on 23  
25 populations (Supplementary Figure 2.2.3-II), chosen to represent major genetic  
26 groups (Supplementary Section 2.2.3) and used a novel method to predict all  
27 pairwise split times (Methods, Supplementary Figure 2.2.3-III). The split of all  
28 mainland Eurasian populations from Yorubans consistently appears as a gradual  
29 process with a median time  $\sim 75$  kya (Table S2.2.3-I, Figure 3A). Importantly,  
30 Papuans are an exception to this broad picture, showing a deeper median split  
31 time from Yoruba at around 90 kya; a conclusion robust to phasing artefacts (See  
32 Methods). The Papuan-Eurasian MSMC split time of  $\sim 40$  kya is slightly older  
33 (Figure S2.2.3-III) than splits between West Eurasian and East Asian populations



1 (~30 kya). The Papuan split times from Yoruban and Eurasian are incompatible  
2 with a simple bifurcating population tree model, implying that modern Papuan  
3 individuals are admixed between different topologies. Some of their genome is  
4 an outgroup to most modern Africans and Eurasians, while the rest of their  
5 genome shares a history with Eurasia.

6 **Ancient or modern introgression in Sahul?** At least two main models  
7 could account for Sahul populations having older split dates from Africa than  
8 mainland Eurasians in our sample:

9 a) Admixture in Sahul with an archaic human population that split from modern  
10 humans either before or at the same time as did Denisova and Neanderthal. This  
11 introgressing population could potentially have diverged from the available  
12 aDNA samples more than 350 kya.

13 b) Admixture in Sahul with a modern human population (xOoA) that left Africa  
14 well after the split between modern humans and Neanderthals, but before the  
15 main expansion of modern humans in Eurasia (main OoA).

16 We performed a large number of tests to distinguish these scenarios. Because the  
17 introgressing lineage has not been observed with aDNA, standard methods are  
18 limited in their ability to distinguish between these hypotheses. Our approach  
19 therefore relies on building multiple lines of evidence using haplotype-based  
20 MSMC and fineSTRUCTURE comparisons. The two hypotheses are not mutually  
21 exclusive and we can only hope to identify the source of the strongest  
22 contribution.

23 **Single site statistics cannot identify the source of introgression.** We  
24 first tried traditional statistical approaches, most notably Patterson's *D*  
25 statistic<sup>13,23</sup>, which we applied to all possible tree relationships between our  
26 samples from Africa, Sahul and Eurasia (Figure S2.2.7-I). The best-supported  
27 topology among those tested shows a contribution to the Sahul genome from a  
28 population (xOoA) that diverged early from West Africans, Baka and Mbuti. This  
29 predates the separation of the ancestors of the modern Africans and Eurasians in  
30 our dataset (topology 3 in Figure S2.2.7-I) as previously proposed<sup>37</sup>. However,  
31 when including the documented Denisova admixture into the analysis<sup>38</sup> and  
32 allowing Denisova introgressed segments to have strongly (350 kya) diverged  
33 from the observed Denisova genome, the *D*-based test could not discriminate

1 between a putative xOoA and the Denisova genomic components  
2 (Supplementary Section S2.2.7).  
3 We also counted non-African Alleles (nAAs), i.e. derived alleles present outside  
4 Africa, but absent in Africans and also archaic humans (Altai Neanderthal and  
5 Denisova genomes) (Figure S2.2.7-II). When compared to Eurasians, both Sahul,  
6 including two admixed Australian Aborigine genomes, and Philippine Negrito  
7 samples do show an excess of nAAs. This is independent of potential  
8 demographic confounders, such as inbreeding or drift (Figure S2.2.7-III). Again,  
9 the excess of nAAs could be explained by admixture with xOoA, which had more  
10 time to accumulate such alleles. However, simulations show that, when allowing  
11 sufficient within-Denisova divergence time, archaic introgression could generate  
12 the same pattern. In this case, we fail to fully mask the derived alleles in Papuans  
13 originating from the introgressing Denisova by relying only on a single Denisova  
14 sample (Figure S2.2.7-IV). Our D-based and nAAs results and related simulations  
15 show empirically that these kind of single site statistics lack the power to  
16 discriminate between the hypothesised scenarios: either Denisova introgression  
17 or a xOoA scenario would result in an increase of non-African derived alleles in  
18 Papuans. The extent of such increase, at the genome-wide level, is a function of  
19 the admixture proportion and divergence time of the introgressing population  
20 from the main human lineage. Therefore, two admixture events with unknown  
21 proportions and time depth are equally able to explain the data and cannot be  
22 disentangled by single site statistics alone.

23 **Haplotype-based analyses indicate an early modern human**  
24 **expansion signature in Sahul.** Using a previously published method<sup>39</sup>, we  
25 located and masked putatively introgressed Denisova haplotypes from the  
26 genomes of Papuans. We also tried symmetrically phasing Papuans and  
27 Eurasians (see Methods) to evaluate the contribution of phasing errors to the  
28 observed shift in MSMC split dates. Neither modification (Figure 3A,  
29 Supplementary Section 2.2.9, Table S2.2.9-I) changed the estimated split time  
30 (based on MSMC) between Africans and Papuans, suggesting that Denisova  
31 admixture or phasing artefacts are not the main driver of this pattern (See  
32 Methods, Supplementary Section 2.2.8, Figure S.2.2.8-I, Table 2.2.8-I). We further  
33 tested the possible role of Denisova admixture by extensive coalescent

1 simulations (Figures S2.2.8-I-II). Without assuming an implausibly large  
2 contribution from a Denisova-like population, we could not simulate the large  
3 Papuan-African and Papuan-Eurasian split times inferred from the data.  
4 Assuming that MSMC dates behave linearly under admixture, the results also  
5 indicate that the hypothesised xOoA lineage may have split from most Africans  
6 as early as 120 kya. This assumption is validated in Supplementary Section 2.2.4  
7 by checking that split dates behave as a mixture in known admixture events.  
8 However, for very old divergences the linearity does not hold true as we  
9 demonstrate in Supplementary Section 2.2.8. Here we show with additional  
10 simulations that the observed shift in the African-Papuan MSMC split curve can  
11 be qualitatively reproduced when including a 4% genomic component that  
12 diverged 120 kya from the main human lineage within Papuans, but that a  
13 similar quantity of Denisova admixture does not produce any significant effect  
14 (Figure S2.2.8-III). Together with the previous simulations, this favours a small  
15 presence of xOoA lineages rather than Denisova admixture alone as the likely  
16 cause of the observed deep African-Papuan split.

17 We further tested our hypothesised xOoA model by focussing on genomic  
18 regions in Papuans that have African ancestry not found in other Eurasian  
19 populations. We reran fineSTRUCTURE on an “ancient diversity panel”, a subset  
20 of the Diversity Set with the addition of the Denisova, Altai Neanderthal and the  
21 Human Ancestral Genome sequences<sup>2</sup>, with sites that are heterozygous in  
22 archaic humans removed. FineSTRUCTURE infers chunks of the genome that  
23 have a single inferred most recent common ancestor (MRCA). An MRCA between  
24 different populations occurs either because the lineage first coalesces before two  
25 populations split, or because of a more recent introgression event. Papuan  
26 genomic chunks that have an African MRCA assignment in the sample, like the  
27 genome-wide nAAs results above, had an elevated level of non-African derived  
28 alleles compared to such chunks in Eurasians. They therefore have an older  
29 mean coalescence time with our African samples, as would be expected if  
30 Papuans contained genetic contributions from a xOoA lineage.

31 On the other hand there may also be a deep divergence between the  
32 sampled Denisova and the one introgressing into modern humans. We were  
33 hence concerned that some introgressed archaic haplotypes have an MRCA with

1 Africans due to coalescence in the ancestral population, and hence are assigned  
2 to be African. However, we can resolve the age and hence origin of these chunks  
3 by their sequence similarity with modern Africans. To account for the archaic  
4 introgression we modelled these genomic portions as a mixture of chunks  
5 assigned African or Denisova in Eurasians, as well as chunks assigned Denisova  
6 in Papuans. Chunks are modelled (see Methods) in terms of the distribution of  
7 length and mutation rate, which is characterised in terms of the density of non-  
8 African derived alleles, which are nAAs that are fixed ancestral in our Africans.

9 This approach captures lineages that coalesce before the  
10 human/Denisova split since the properties of these chunks should not depend on  
11 the population they were found in, and since Eurasians (specifically Europeans)  
12 have not experienced Denisova admixture. By this way we could disentangle the  
13 various introgressing lineages by looking at their mutation density. From the  
14 discrepancy between the distribution of Papuan chunks assigned to Africans and  
15 the fitted distribution (Figure 3B-D) we can identify the characteristics of xOoA  
16 chunks (Supplementary Section 2.2.10). Including a xOoA component was  
17 necessary to account for the number of short chunks with “moderate” mutation  
18 density, i.e. higher than Eurasian chunks assigned African but significantly lower  
19 than those assigned Denisova in either Eurasians or Papuans. Inferred xOoA  
20 chunks have 1.5 times more nAAs than that observed in chunks assigned to be  
21 Eurasian, compared to 4 times for chunks assigned to be Denisova. These  
22 proportions can be interpreted as a relative mean time to the most recent  
23 common ancestor, implying a xOoA-Africa split 1.5 times older than the main  
24 OoA, consistent with our MSMC findings (Supplementary Section 2.2.4).

25 We went on to estimate the proportion of xOoA in Papuan chunks  
26 assigned as both Eurasian (0.1%, 95% CI 0-2.6) and Papuan (4%, 95% CI 2.9-4.5)  
27 (Supplementary Section 2.2.10), by using the estimated mutation density in  
28 xOoA. To do this we used the same mixture model as above (additionally  
29 considering Eurasian chunks assigned to be Eurasian) to obtain a xOoA-free  
30 prediction. When this predicted too few mutations, we assumed that the  
31 difference is due to the xOoA admixture. Adding up the contributions from all  
32 assignments of chunks leads to a genome-wide estimate of 1.9% xOoA (95% CI  
33 1.5-3.3) in Papuans.

1           Our results consistently point towards a predominantly modern human  
2 source for the abundance of alleles found in Papuans that are absent in Africans  
3 and are derived according to the ancestral human sequence. It follows that the  
4 genome of modern Papuans is best described as consisting of two human  
5 components. The predominant component is an early split from the major  
6 migration out of Africa that colonized Eurasia while the lesser component is  
7 derived from an earlier, otherwise extinct, dispersal.

8

## 9   **Adaptation outside Africa**

10

11           Humans faced a number of ecological challenges as they encountered new  
12 environments outside Africa. To study the nature and extent of any resultant  
13 adaptation, we explored the distribution of functional variants among  
14 populations, performed tests of purifying, balancing and positive selection and,  
15 finally, identified loci that showed the highest allelic differentiation among  
16 groups (Supplementary Section 3). It is important to emphasise that our  
17 sampling strategy may be underpowered to detect certain types of selection.  
18 Despite this, strong signals are present in the data.

19           **Relationship to other findings.** The results of our positive selection  
20 tests corroborated the identification of a number of selective sweeps that are  
21 well supported by functional evidence (Table S3.3.4-I), suggesting that,  
22 regardless of our sample pooling strategy, our dataset is able to detect region-  
23 specific signals of haplotype homozygosity and allelic differentiation. Our tests  
24 for purifying selection are also consistent with previous studies<sup>2,40,41</sup>, in terms of  
25 both the lack of differential purifying selection between Africans and non-  
26 Africans, as well as the distribution of alleles across frequency classes and  
27 populations (Supplementary Section 3.1, Figure S3.1-I,II; Table S3.1-IV,VI).

28           **Novel findings.** Our results show novel signals of purifying, balancing  
29 and positive selection. With regard to purifying selection, we report evidence for  
30 significant differences in the strength of selection in systematically defined  
31 phenotype-related sets of genes. We infer more purifying selection in Africans in  
32 genes involved in pigmentation (bootstrapping p value for  $R_{X/Y}$ -scores < 0.05 )  
33 (Figure S3.5-II) and immune response against viruses ( $p < 0.05$ ), whilst more

1 purifying selection was indicated on olfactory receptor genes in Asians ( $p < 0.01$   
2 in the Southeast Asia Island population,  $p < 0.05$  in the Southeast Asia Mainland,  
3 South American and Northeast Siberia populations) (Table S3.1.1-II). A genome-  
4 wide scan for ancient balancing selection in populations grouped into 12  
5 geographical regions according to their genetic clustering (Supplementary  
6 Section 3.2) revealed a significant enrichment (false discovery rate  $q$ -value  $<$   
7 0.01) for antigen processing/presentation, antigen binding, and MHC and  
8 membrane component genes (Tables S3.3.2-I-III). The HLA (*HLA-C*)-associated  
9 gene (*BTNL2*) was the top candidate in eight of 12 geographic regions (Table  
10 S3.3.1-I).

11 Our positive selection scans and variant-based analyses (Supplementary  
12 Sections 3.2 and 3.2) revealed many novel signals, especially in the less-studied  
13 populations, a subset of which is highlighted in Table 1. Benefiting from the  
14 availability of high resolution sequencing information, we were also able to  
15 identify new potentially causal variants in both novel and previously-detected  
16 positive selection signals.

17 Given the geographic distribution of our samples, we were particularly  
18 interested in assessing whether genes associated with phenotypes highly-  
19 correlated to local environmental features, such as temperature, UV exposure,  
20 diet, and pathogen load, are systematically overrepresented in the signals of  
21 positive selection in the sampled populations (Supplementary Section 3.4; Tables  
22 S3.5-I-VI). All categories reported as enriched have chi-square  $p$ -values less than  
23 0.01. We observed that genomic regions containing pigmentation-related genes  
24 were overrepresented in some of our positive selection tests in West Eurasian  
25 populations (Table S3.5-I), while those containing genes relating to  
26 thermoregulation were enriched, albeit for different genes, in Africans and  
27 Central Siberians (Table S3.5-II). Unlike Khrameeva and colleagues<sup>42</sup>, we do not  
28 observe an enrichment of fatty acid metabolism (or specifically lipid catabolism)  
29 genes in the positive selection tests for our European samples. We do, however,  
30 observe enrichment of such genes in Island Southeast Asian and Central Siberian  
31 populations (Table S3.5-IV, Figure S3.5-IV).

32 With regard to immunity, we found enrichment of bacterial immunity genomic  
33 windows in Island Southeast Asians (Table S3.5-V), which was lost after the

1 exclusion of Philippine Negritos from the tests, suggesting that the observation  
2 partially reflects elevated selection in these hunter-gatherer groups.  
3 Furthermore, both western Asian and the South Asian groups showed significant  
4 enrichment in innate immune response annotations based on Tajima's *D* statistic  
5 (Table S3.5-VI, Figure S3.5-V), which was the only category that showed any  
6 enrichment by that test. This is consistent with selection represented by these  
7 signatures being older than those detected by the haplotype homozygosity tests.  
8 The fact that most innate immunity signals are shared between at least two  
9 populations supports this interpretation.

10

## 11 **Discussion**

12

13 **A valuable resource.** The collection of worldwide high-coverage  
14 genomes presented here has allowed us to: (i) provide a finer resolution  
15 description of human genetic diversity; (ii) identify the genetic trace of a so-far  
16 unidentified component in Sahul populations; and (iii) increase the number of  
17 candidate genome regions that have been subjected to distinct selective  
18 pressures on physiological processes. The latter is key to unravelling our  
19 adaptation history. The data and inferences presented here provide the  
20 groundwork to refine hypotheses about human evolution that are essential to  
21 the understanding of modern patterns of genetic diversity, disease vulnerability  
22 and distribution.

23 **Methodological difficulties.** Existing methods based on single-site  
24 analyses seemed unable to resolve our hypotheses about Sahul and could not be  
25 used to distinguish between a small fraction of ancient admixture and a larger  
26 fraction of more recent admixture. The power of these approaches in practice  
27 depends on appropriate ancient samples being available. The behaviour of  
28 haplotype-based inference approaches are relatively poorly characterised and  
29 there is no formal inferential framework available to address our hypotheses.  
30 However, haplotypes preserve more information on our evolution as they can  
31 persist for long periods in finite populations<sup>43</sup> at lengths that are detectable with  
32 sequence variation data (Supplementary Section 2.2.13). They allow us to  
33 calibrate drift by considering the rate of non-African alleles accumulated in



1 segments of known length, providing us with a way to estimate the age of splits  
2 from Africa.

3 A further confounder is that detecting Denisova and Neanderthal introgression  
4 mostly relies on matching to the aDNA data available, which may be a poor proxy  
5 for the actual introgressed DNA. Other possible confounders could involve a  
6 shorter generation time in Papuan and Philippine Negrito populations<sup>44</sup>,  
7 different recombination processes, or alternative demographic histories that  
8 have not been investigated here. We therefore strongly encourage the  
9 development of new model-based approaches that can explain the haplotype  
10 patterns described here.

11 **Evidence for an earlier exit out of Africa?** Our estimate of the split  
12 between African and Eurasians is in broad agreement with previous reports  
13 based on mtDNA and Y chromosome<sup>45-47</sup> and full genome sequencing data<sup>5,10</sup>,  
14 and is consistent with a major OoA expansion (likely through the Levant<sup>5</sup> and/or  
15 Arabia<sup>15</sup>) after that date. Other methods rescaled to the lower mutation rate used  
16 here<sup>14</sup> suggest slightly older dates for that split<sup>28,48</sup>. A recent IBS tract sharing-  
17 based method<sup>11</sup>, when similarly rescaled, yields a remarkably similar split time  
18 of ~80 kya.

19 Our analyses, however, provide clear evidence that the Sahul populations  
20 sampled here, and possibly other populations from the region that were not  
21 included in our study design, possess an additional genetic signal of  
22 introgression from an uncharacterised hominin. We used a series of tests to try  
23 to identify whether this hominin came from a) an archaic lineage or b) an earlier  
24 out-of-Africa, modern human branch. Current single-site approaches could not  
25 distinguish these hypotheses, but our haplotype-based approaches all point  
26 towards a small amount of admixture (at least 2%) from an earlier modern  
27 human dispersal out-of-Africa around 120 kya (Figure 4) whose genetic  
28 signature has not been identified in any other extant population. We also show  
29 (see Methods) that this is not at odds with evidence that show that Sahul shares  
30 Y chromosome and mtDNA lineages with Eurasians, as there is a high probability  
31 that older Y and mtDNA lineages would be lost as a result of random genetic  
32 drift, as was also argued by Groucutt and others colleagues<sup>15,49</sup>.



1 The inferred xOoA split time (~120 kya) corresponds with fossil and  
2 archaeological evidence for an early expansion of *Homo sapiens* from Africa<sup>15,19</sup>.  
3 Furthermore, Kuhlwilm and colleagues<sup>50</sup> recently identified modern human  
4 admixture into the Altai Neanderthal before 100 kya. This is consistent with  
5 modern human presence outside of Africa well before the main OoA expansion  
6 after 75 kya. Further studies will confirm if the xOoA we propose here and the  
7 early modern humans that admixed with ancestors of Altai Neanderthals were  
8 part of the same early expansion out of Africa. Similarly, we are agnostic to the  
9 geographic extent of such an early event. Indeed, archaeological evidence for  
10 modern human colonization of Sahul is no earlier than ca. 60-50 kya<sup>51</sup>, and  
11 perhaps as late as ca. 47 kya<sup>52</sup>. The preponderance of genomic evidence, in fact,  
12 indicates that early human expansions did not leave detectable genetic traces in  
13 most contemporary Eurasian populations, perhaps as a consequence of  
14 substantial population replacements, as indicated by aDNA from Oase,  
15 Romania<sup>53</sup>. Climatic changes over the last 120 thousand years, including glacial  
16 advances and significant fluctuations of wet and dry environmental cycles, likely  
17 influenced population structure across Eurasia<sup>54</sup>, perhaps leading to lineage  
18 extinctions and regional extirpations. The unexpected genetic traces of xOoA in  
19 Papuans, shown here for the first time, suggest that unravelling the evolutionary  
20 history of our own species will require the recovery of aDNA from additional  
21 fossils, and further archaeological investigations in under-explored regions of  
22 Eurasia.

23

#### 24 **Data availability**

25 The newly sequenced genomes were deposited in the ENA archive under  
26 accession number ENAXXXX and are also freely available through the Estonian  
27 Biocentre website ([www.ebc.ee/free\\_data](http://www.ebc.ee/free_data)).

28

29

#### 30 **Acknowledgements**

31 Support was provided by: Australian Research Council (M.W.; D.L.); Danish  
32 National Research Foundation; the Lundbeck Foundation and KU2016 (E.W.);  
33 ERC Starting Investigator grant (FP7 - 261213) (T.K.); Estonian Institutional

1 Research grant IUT24-1 (T.K.); Estonian Research Council grant PUT766 (G.C.;  
2 M.K.); Estonian Science Foundation grant 8973 (M.M.); EU European Regional  
3 Development Fund through the Centre of Excellence in Genomics to Estonian  
4 Biocentre; Estonian Institutional Research grant IUT24-1; (L.S.; M.J.; A.K.; B.Y.;  
5 K.T.; C.B.M.; Le.S.; H.Sa.; S.L.; D.M.B.; E.M.; R.V.; G.H.; M.K.; G.C.; M.M.); French  
6 Ministry of Foreign and European Affairs and French ANR grant number ANR-  
7 14-CE31-0013-01 (F.-X.R.); Gates Cambridge Trust Funding (E.J.); ICG SB RAS  
8 (No. VI.58.1.1) (D.V.L.); Leverhulme Programme grant no. RP2011-R-045 (A.B.M.,  
9 P.G. & M.G.T.); Ministry of Education and Science of Russia; Project 6.656.2014/K  
10 (S.A.F.); NEFREX grant funded by the European Union (People Marie Curie  
11 Actions; International Research Staff Exchange Scheme; call FP7-PEOPLE-2012-  
12 IRSES-number 318979) (G.H.); NIH grants 5DP1ES022577 05, 1R01DK104339-  
13 01, and 1R01GM113657-01 (S.Tis); Russian Foundation for Basic Research  
14 (grant N 14-06-00180a) (M.G.); Russian Foundation for Basic Research; grant  
15 16-04-00890 (O.B.; E.B); Russian Science Foundation grant 14-14-00827 (O.B.);  
16 The Russian Foundation for Basic Research (11-04-00725-a); The Russian  
17 Humanitarian Scientific Foundation (13-11-02014) and the Program of the Basic  
18 Research of the RAS Presidium "Biological diversity". (E.K.K.); Wellcome Trust  
19 and Royal Society grant WT104125AIA (D.J.L); Wellcome Trust grant 098051  
20 (Q.A.; C.T.-S.; Y.X.); Wellcome Trust Senior Research Fellowship grant  
21 100719/Z/12/Z (M.G.T); Young Explorers Grant from the National Geographic  
22 Society (8900-11) (C.A.E.); ERC Consolidator Grant 647787 'LocalAdaptatio'  
23 (AM); Program of the RAS Presidium "Basic research for the development of the  
24 Russian Arctic" (B.M.).

25

1 **Figure and Table Legends**

2

3 **Table 1** Subset of novel positive selection Findings in our 12 macro-regional  
4 groups defined using fineSTRUCTURE.

5

6 **Figure 1** Panel A: Map of samples location highlighting Diversity/Selection Set;  
7 Panel B: ADMIXTURE plot (K=8 and 14) which relates general visual inspection  
8 of genetic structure to studied populations and their region of origin; Panel C:  
9 Sample level heterozygosity is plotted against distance from Addis Ababa. The  
10 trend line represents only non-African samples. The inset shows the waypoints  
11 used to arrive at the distance in kilometres for each sample.

12

13 **Figure 2** Spatial visualisation of genetic barriers inferred from genome-wide  
14 genetic distances, quantified as the magnitude of the gradient of spatially  
15 interpolated allele frequencies (value denoted by colour bar; grey areas have  
16 been land during the last glacial maximum but are currently under water). Here  
17 we used a novel spatial kernel smoothing method based on the matrix of  
18 pairwise average heterozygosity. **Inset:** partial correlation between magnitude  
19 of genetic gradients and combinations of different geographic factors, elevation  
20 (E), temperature (T) and precipitation (P), for genetic gradients from  
21 fineSTRUCTURE (red) and allele frequencies (blue). This analysis (see  
22 Supplementary Section 2.2.2 for details) shows that despite the large number of  
23 prehistoric movements across Eurasia, genetic differences within this region  
24 have been strongly shaped by physical barriers such as mountain ranges,  
25 deserts, and open water (such as the Wallace line).

26

27 **Figure 3** Panel A: MSMC split times plot. The Yoruba-Eurasia split curve shows  
28 the mean of all Eurasian genomes against one Yoruba genome. The grey area  
29 represents top and bottom 5% of runs. We chose a Koinanbe genome as  
30 representative of the Sahul populations. Panels B-D: Decomposition of the  
31 ChromoPainter inferred African chunks in Papuans. Panel **B:** Semi-parametric  
32 decomposition of the joint distribution of chunk lengths and non-African derived  
33 allele rate per SNP, showing the relative proportion of chunks in K=20  
34 components of the distribution, ordered by non-African derived allele rate,  
35 relative to the overall proportion of chunks in each component. The four datasets  
36 produced by considering (African/Denisova) chunks in (Europeans/Papuans)  
37 are shown with our inferred "extra Out-of-Africa xOoA" component. Panel **C:** The  
38 reconstruction of African chunks in Papuans using a mixture of the other data  
39 (red) and adding the xOoA component (black). Panel **D:** The properties of the  
40 components in terms of non-African derived allele rate, on which the  
41 components are ordered, and length.

42

43 **Figure 4** A subway map figure illustrating, as suggested by the novel results  
44 presented here, the model of an early, extinct Out-of-Africa (xOoA) entering the  
45 genome of Sahul populations at their arrival in the region. Given the overall small  
46 genomic contribution of this event to the genome of modern Sahul, we could not  
47 determine whether the documented Denisova admixture (question marks) and  
48 putative multiple Neanderthal admixtures took place along this extinct OoA.

## 1 **Methods**

2       **Data Preparation:** In the final dataset, we retained only one second  
3 (Australians, to make use of all the available samples)- and five third-degree  
4 relatives pairs (Table S1.7-1). All genomes were annotated against the Ensembl  
5 GRCh37 database and compared to dbSNP Human Build 141 and Phase 1 of the  
6 1000 Genomes Project dataset<sup>2</sup> (Supplementary Sections 1.1-6). We found  
7 10,212,117 new SNPs, 401,911 of which were exonic. As expected from our  
8 sampling scheme, existing lists of variable sites have been extended mostly by  
9 the Siberian, South-East Asian and South Asian genomes, which contribute  
10 89,836 (22.4%), 63,964 (15.9%) and 40,758 (10.1%) of the new exonic variants  
11 detected in this study.

12 Compared to the genome-wide average, we see fewer heterozygous sites on  
13 chromosomes 1 and 2, and an excess on chromosomes 16, 19 and 21. This  
14 pattern is independent of simple potential confounders, such as rough estimates  
15 of recombination activity and gene density (Supplementary Section 1.8), and  
16 mirrors the inter-chromosomal differences in divergence from chimpanzee<sup>55</sup>,  
17 suggesting large-scale differences in mutation rates among chromosomes. We  
18 confirmed this general pattern using 1000Genomes Project data (Supplementary  
19 Section 1.8).

20  
21       **Geographic gradient analyses.** We used a Gaussian kernel smoothing  
22 (based on the shortest distance on land to each sample) to interpolate genetic  
23 patterns across space. Averaging over all markers, we obtained an expression for  
24 the mean square gradient of allele frequencies in terms of the matrix of genetic  
25 distance between pairs of samples (Supplementary Section 2.2.2). This provides  
26 a simple way to identify spatial regions that contribute strongly to genetic  
27 differences between samples, and can be used, in principle, for any measure of  
28 genetic difference (for fineSTRUCTURE data, we used negative shared haplotype  
29 length as a measure of differentiation).

30  
31 To quantify the link between the magnitude of genetic gradients (from  
32 fineSTRUCTURE and allele frequency data) and geographic factors, we fitted a  
33 generalised linear model to the sum of genetic magnitude gradients on the

1 shortest paths between samples to elevation, minimum quarterly temperature,  
2 and annual precipitation summed in the same way, controlling for path length  
3 and spatial random effects (Supplementary Section 2.2.2), and calculated partial  
4 correlations between genetic gradient magnitudes and geographic factors.

5

6 **Finestructure Analysis.** FineSTRUCTURE<sup>32</sup> was run as described in  
7 Supplementary Section S2.2.1. Within the 106 genetically distinct genetic groups,  
8 labels were typically genetically homogeneous - 113 of the 148 population labels  
9 (76%) were assigned to only one 'genetic cluster'. Similarly, genetic clusters  
10 were typically specific to a label, with 66 of the 106 'genetic clusters' (62%)  
11 containing only one population label.

12 Correction for phasing errors: To check whether phasing errors could produce  
13 the shorter Papuan chunks, we focussed on regions of the genome that had an  
14 extended (>500Kb) run of homozygosity. We ran ChromoPainter for each  
15 individual on only these regions, meaning each individual was only painted  
16 where it had been perfectly phased. This did not change the qualitative features  
17 (Supplementary Section 2.2.1).

18 Removal of similar samples: Papuans are genetically distinct from other  
19 populations due to tens of thousands of years of isolation. We wanted to check  
20 whether African chunk lengths were biased by the inclusion of a large number of  
21 relatively homogeneous Eurasians with few Papuans. To do this we repeated the  
22 N=447 painting allowing only donors from dissimilar populations, including only  
23 individuals who donated <2% of a genome in the main painting. This did not  
24 change the qualitative chunk length features (Supplementary Section 2.2.1).

25 Inclusion of ancient samples: We ran our smaller individual panel with (N=109)  
26 and without (N=106) ancient samples (Denisova, Neanderthal and ancestral  
27 human). This did not change the qualitative chunk length features  
28 (Supplementary Section 2.2.1).

29

30 **MSMC, Denisova masking, simulations of alternative scenarios and**  
31 **assessment of phasing robustness.** Genetic split times were initially calculated  
32 following the standard MSMC procedure<sup>10</sup>, and subsequently modified as  
33 follows. To estimate the effect of archaic admixture, putative Denisova

1 haplotypes were identified in Papuans using a previously published method<sup>39</sup>  
2 and masked from all the analysed genomes. Particularly, whether a putative  
3 archaic haplotype was found in heterozygous or homozygous state within the  
4 chosen Papuan genome, the “affected” locus was inserted into the MSMC mask  
5 files and, hence, removed from the analysis.

6 We note that a fraction of the Denisova and Neanderthal contributions to the  
7 Papuan genomes may be indistinguishable, due to the shared evolutionary  
8 history of these two archaic populations. As a result, some of the removed  
9 “Denisova” haplotypes may have actually entered the genome of Papuans  
10 through Neanderthal. Regardless of this, our exercise successfully shows that  
11 the MSMC split time estimates are not affected by the documented presence of  
12 archaic genomic component (whether coming entirely from Denisova or partially  
13 shared with Neanderthal).

14 We further excluded the role of Denisova admixture in explaining the deeper  
15 African-Papuan MSMC split times through coalescent simulations (using ms to  
16 generate 30 chromosomes of 5 Mbp each, and simulating each scenario 30  
17 times). These showed that the addition of 4% Denisova lineages to the Papuan  
18 genomes does not change the MSMC results, while the addition of 4% xOoA  
19 lineages recreates the qualitative shift observed in the empirical data.

20 Phasing artefacts were also taken into account as putative confounders of the  
21 MSMC split time estimates. We re-run MSMC after re-phasing one Estonian, one  
22 Papuan and 20 West African and Pygmies genomes in a single experiment. By  
23 this way we ruled out potential artefacts stemming from the excess of Eurasian  
24 over Sahul samples during the phasing process. Both the archaic and phasing  
25 corrections yielded the same split time as of the standard MSMC runs.

26

27 **Emulation of all pairwise MSMC split times.** We confirmed that none of  
28 the other populations behaved as an outlier from those identified in the N=22  
29 full pairwise analysis by estimating the MSMC split times between all pairs. We  
30 chose 9 representative populations (including Papuan, Yoruba and Baka) from  
31 the 22, and compared each of the 447 diversity panel genomes to them. We  
32 learn a model for each individual  $l$  not in our panel,

33  $\hat{t}_{lj} = \sum_{k=1}^9 \alpha_{lk} t_{kj}$  for  $j \in (1..9)$ ,

1 where the positive mixture weights  $\alpha_k$  sum to 1 and are otherwise learned from  
 2 the  $j \in (1..9)$  observations which we have data under quadratic loss. We can  
 3 then predict the unobserved values

$$\hat{t}_{li} = \sum_{k=1}^9 \alpha_k t_{ki}.$$

4 Examination of this matrix (Supplementary Section S2.2.3, Table S2.2.3-III)  
 5 implies no other populations are expected to have unusual MSMC split times  
 6 from Africa.

7

8 **Mixture model for African haplotypes in Papuans.** *Obtaining*

9 *haplotypes from painting:* We define as African or Archaic chunk in Eurasians or  
 10 Papuans a genomic locus spanning at least 1000bp, and showing SNPs that were  
 11 assigned by chromopainter a 50% chance of copying from either an African or  
 12 Archaic genome, respectively. For each chunk we then calculated the number of  
 13 non-African mutations, defined as sites found in derived state in a given chunk  
 14 and in ancestral state in all of the African genomes included in the present study.

15 *Modelling:* We used a non-parametric model for the joint distribution of length  
 16 and non-African derived allele mutation rate of chunks. We fit K (=20)  
 17 components to the joint distribution. Each component has a characteristic length  
 18  $l_k$ , variability  $\sigma_k$  and mutation rate  $\mu_k$ . A chunk of length  $l_i$  with  $X_i$  such  
 19 mutations from component  $I_i = k$  has the following distribution:

$$l_i | \{l_k, \sigma_k^2, I_i = k\} \sim \text{log-Normal}(l_k, \sigma_k^2)$$

$$X_i | \{l_k, \mu_k, I_i = k\} \sim \text{Binomial}(l_k, \mu_k)$$

20 This model for chunk lengths is motivated by the extreme age of the split times  
 21 we seek to model. Recent splits would lead to an exponential distribution of  
 22 haplotype lengths. However, due to haplotype fixation caused by finite  
 23 population size, very old splits have finite (non-zero) haplotype lengths.  
 24 Additionally, the data are left-censored since we cannot reliably detect chunks  
 25 that are very short. We note that whilst this makes a single component a  
 26 reasonable fit to the data, as K increases the specific choice becomes less  
 27 important.

28 We then impose the prior  $p(I_i = k) = 1/K$  and use the Expectation-  
 29 Maximization algorithm to estimate the mixture proportions  $\pi_{ik} = \mathbb{E}(I_{ik} | l_i, X_i)$



1 along with the maximum likelihood parameter estimates  $\{l_k, \sigma_k^2, \mu_k\}$ . We do this  
 2 for the four combinations of African (AFR) and Denisova (DEN) chunks found in  
 3 Papuans (PNG) or Europeans (EUR), in order to learn the parameters.  
 4 Supplementary Section S2.2.10 describes this in more detail. We then describe  
 5 the distribution of chunks for each class  $c$  of chunk in terms of the expected  
 6 proportion of chunks found in each component,

$$7 \quad \pi_{ck} = \frac{\pi'_{ck}}{\sum_{k=1}^K \pi'_{ck}} \text{ where } \pi'_{ck} = \sum_{i=1}^{N_c} \pi_{cik},$$

8 where  $N_c$  is the number of chunks of class  $c$ .  $\pi_c$  is a vector of the proportions  
 9 from each of the  $K$  components.

10

11 Single-out-of-Africa model: We fit African chunks in Papuans as a mixture of the  
 12 others in a second layer of mixture modelling:

$$\pi_{PNG.AFR} = \sum_{c \in \{PNG.DEN, EUR.AFR, EUR.DEN\}} \alpha_c \pi_c,$$

13 where  $\alpha_c$  sum to 1. This is straightforward to fit.

14

15 xOoA model: We jointly estimate an additional component  $\pi_{xOoA}$  and the  
 16 mixture contributions  $\beta_c$  under the mixture

$$\pi_{PNG.AFR} = \sum_{c \in \{PNG.DEN, EUR.AFR, EUR.DEN, xOoA\}} \beta_c \pi_c.$$

17 This is non-trivial to fit. We use a penalisation scheme to simultaneously ensure  
 18 we a) obtain a valid mixture for  $\beta_c$ , b) give a prediction  $x_k$  that is also a valid  
 19 mixture, c) leave little signal in the residuals, and d) obtain a good fit. Cross-  
 20 validation is used to obtain the optimal penalisation parameters ( $A$  and  $B$ ) with  
 21 the loss function:

$$\text{loss} = \sum_{k=1}^K e_k^2 + AP_A + BP_B,$$

22 where  $e_k$  are the residuals in each component,  $P_A = |(\sum_c \beta_c) - 1| +$   
 23  $|(\sum_k x_k) - 1|$  (for a valid mixture) and  $P_B = s.d(e_k)$  (for requirement c, good  
 24 solutions will have similar residuals across components). The loss is minimised  
 25 via standard optimization techniques. Supplementary Section S2.2.10 details  
 26 how initial values are found and explores the robustness of the solution to



1 changes in A and B - the results do not change qualitatively for reasonable  
2 choices of these parameters, and the mixtures are valid to within numerical  
3 error.

4 Genome-wide xOoA estimation: We used the estimated xOoA derived allele  
5 mutation rate estimate  $\theta_{xOoA}$  to estimate the xOoA contribution in haplotypes  
6 classed as Eurasian or Papuan by ChromoPainter. First we obtained estimates of  
7  $\pi_{PNG.EUR}$  and  $\pi_{PNG.PNG}$  using the single out-of-Africa model above, additionally  
8 allowing a EUR.EUR contribution. We then estimate  $\alpha_{xOoA}$  using the observed  
9 mutation rate  $\theta_{obs}$  and that predicted under the mixture model  $\theta_{mix}$  by rearranging  
10 the mixture:

$$\theta_{obs} = \alpha_{xOoA}\theta_{xOoA} + (1 - \alpha_{xOoA})\theta_{mix}$$

11 Estimates less than zero are set to 0. The genome wide estimate is obtained by  
12 weighting each  $\theta$  by the proportion of the genome that was painted with that donor.  
13 Neanderthal and Denisova chunks were assumed to be proxied by PNG.DEN (0% xOoA  
14 by assumption); African chunks by PNG.AFR; Papuan and Australian by PNG.PNG and all  
15 other chunks by PNG.EUR. We obtain confidence intervals by bootstrap resampling of  
16 haplotypes for each donor/recipient pair.

17

18 **Y chromosome and mtDNA haplogroup analysis.** The presence of an  
19 extinct xOoA trace in the genome of modern Papuans may seem at odds with  
20 analyses of mtDNA and Y chromosome phylogenies, which point to a single,  
21 recent origin for all non-African lineages (mtDNA L3, which gives rise to all  
22 mtDNA lineages outside Africa has been dated at  $\sim 70$  kya,<sup>45,46</sup>). However,  
23 uniparental markers inform on a small fraction of our genetic history, and a  
24 single origin for all non-African lineages does not exclude multiple waves OoA  
25 from a shared common ancestor. We show analytically (Supplementary Section  
26 2.2.12) that, if the xOoA entered the Papuan genome  $>40$  kya, their mtDNA and Y  
27 lineages could have been lost by genetic drift even assuming an initial xOoA  
28 mixing component of up to 35%. Similar findings have been reported recently<sup>15</sup>.

29

1   **References**

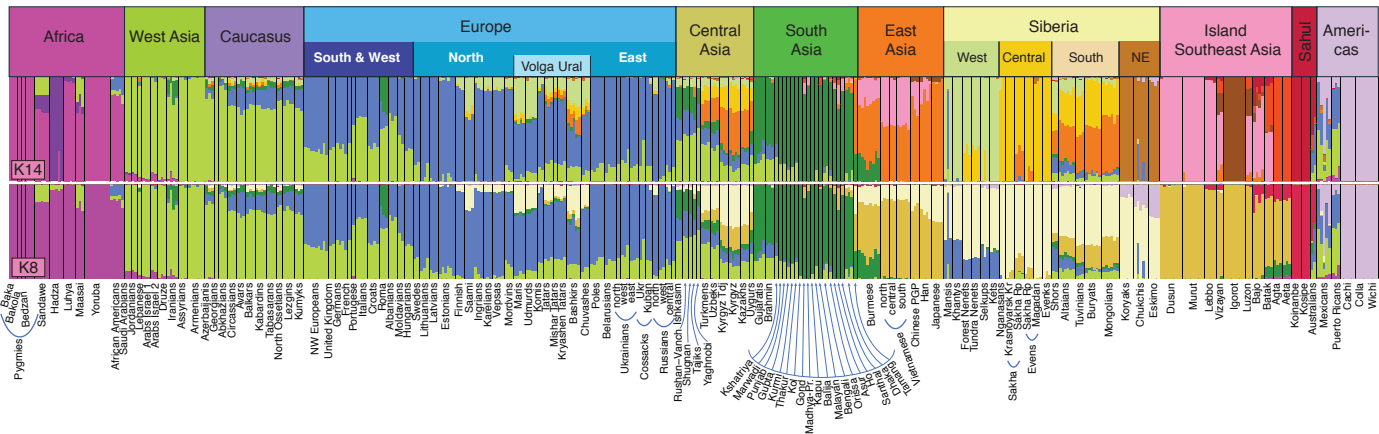
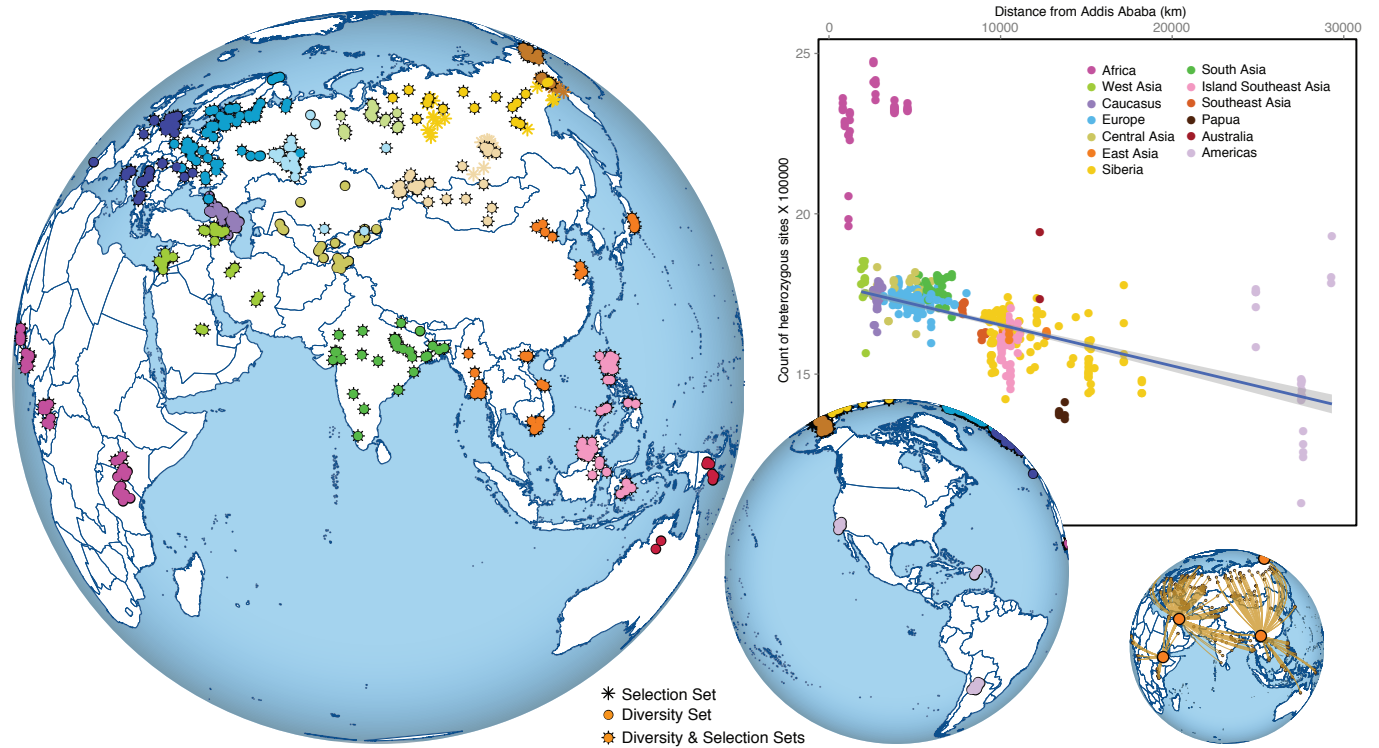
2

- 3   **1**     Altshuler, D. M. *et al.* Integrating common and rare genetic variation in  
4           diverse human populations. *Nature* 467, 52-58, doi:10.1038/nature09298  
5           (2010).
- 6   **2**     The 1000 Genomes Project Consortium. An integrated map of genetic  
7           variation from 1,092 human genomes. *Nature* 491, 56-65,  
8           doi:10.1038/nature11632 (2012).
- 9   **3**     Drmanac, R. *et al.* Human genome sequencing using unchained base reads  
10          on self-assembling DNA nanoarrays. *Science* 327, 78-81,  
11          doi:10.1126/science.1181498 (2010).
- 12   **4**     Lachance, J. *et al.* Evolutionary history and adaptation from high-coverage  
13          whole-genome sequences of diverse African hunter-gatherers. *Cell* 150,  
14          457-469, doi:10.1016/j.cell.2012.07.009 (2012).
- 15   **5**     Pagani, L. *et al.* Tracing the Route of Modern Humans out of Africa by Using  
16          225 Human Genome Sequences from Ethiopians and Egyptians. *American*  
17          *journal of human genetics* 96, 986-991, doi:10.1016/j.ajhg.2015.04.019  
18          (2015).
- 19   **6**     Clemente, F. J. *et al.* A Selective Sweep on a Deleterious Mutation in CPT1A  
20          in Arctic Populations. *American journal of human genetics* 95, 584-589,  
21          doi:10.1016/j.ajhg.2014.09.016 (2014).
- 22   **7**     Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the  
23          Icelandic population. *Nat Genet* 47, 435-444, doi:10.1038/ng.3247 (2015).
- 24   **8**     Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project.  
25          *Nat Genet* 45, 1113-1120, doi:10.1038/ng.2764 (2013).
- 26   **9**     Li, H. & Durbin, R. Inference of human population history from individual  
27          whole-genome sequences. *Nature* 475, 493-496, doi:10.1038/nature10231  
28          (2011).
- 29   **10**    Schiffels, S. & Durbin, R. Inferring human population size and separation  
30          history from multiple genome sequences. *Nat Genet* 46, 919-925,  
31          doi:10.1038/ng.3015 (2014).
- 32   **11**    Harris, K. & Nielsen, R. Inferring demographic history from a spectrum of  
33          shared haplotype lengths. *PLoS genetics* 9, e1003521,  
34          doi:10.1371/journal.pgen.1003521 (2013).
- 35   **12**    Sheehan, S., Harris, K. & Song, Y. S. Estimating variable effective population  
36          sizes from multiple genomes: a sequentially markov conditional sampling  
37          distribution approach. *Genetics* 194, 647-662,  
38          doi:10.1534/genetics.112.149096 (2013).
- 39   **13**    Patterson, N. *et al.* Ancient admixture in human history. *Genetics* 192,  
40          1065-1093, doi:10.1534/genetics.112.145037 (2012).
- 41   **14**    Sally, A. & Durbin, R. Revising the human mutation rate: implications for  
42          understanding human evolution. *Nat Rev Genet* 13, 745-753,  
43          doi:10.1038/nrg3295 (2012).
- 44   **15**    Groucutt, H. S. *et al.* Rethinking the dispersal of Homo sapiens out of Africa.  
45          *Evol Anthropol* 24, 149-164, doi:10.1002/evan.21455 (2015).

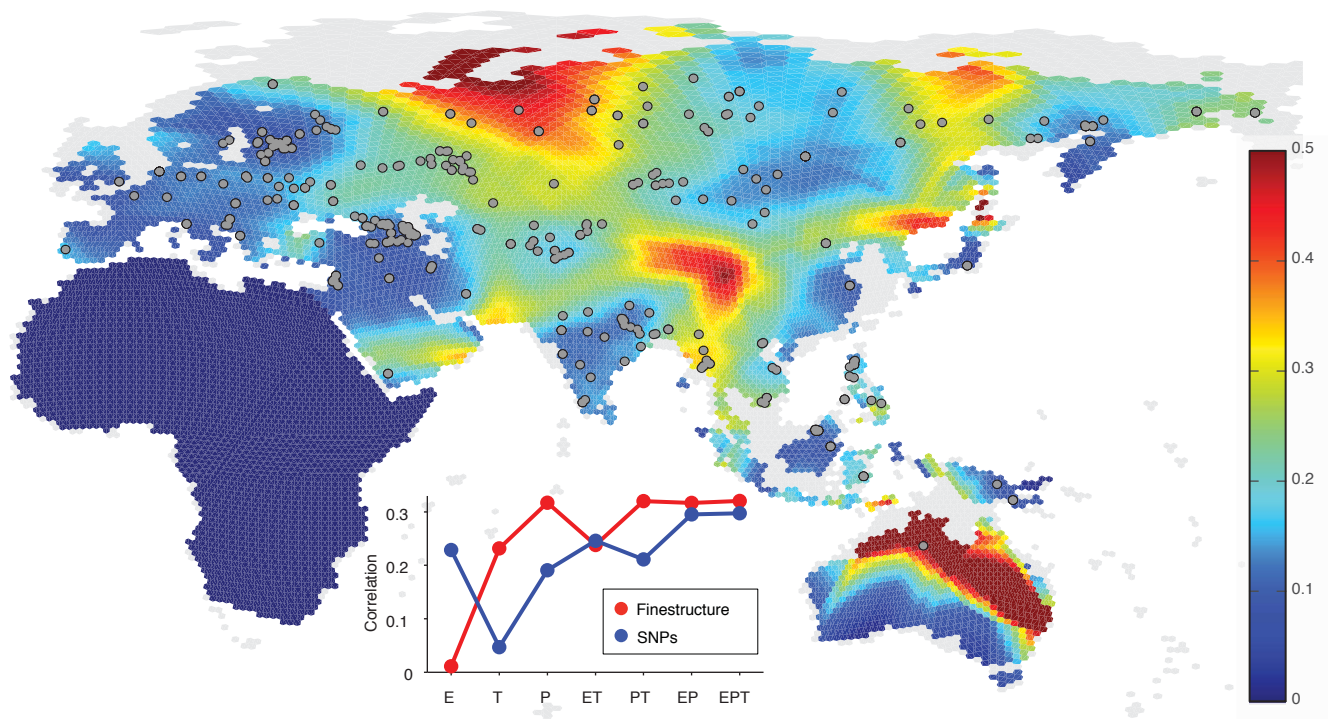
- 1 16 Grove, M. *et al.* Climatic variability, plasticity, and dispersal: A case study  
2 from Lake Tana, Ethiopia. *Journal of human evolution* 87, 32-47,  
3 doi:10.1016/j.jhevol.2015.07.007 (2015).
- 4 17 Mellars, P., Gori, K. C., Carr, M., Soares, P. A. & Richards, M. B. Genetic and  
5 archaeological perspectives on the initial modern human colonization of  
6 southern Asia. *Proceedings of the National Academy of Sciences of the*  
7 *United States of America* 110, 10699-10704, doi:Doi  
8 10.1073/Pnas.1306043110 (2013).
- 9 18 Reyes-Centeno, H. *et al.* Genomic and cranial phenotype data support  
10 multiple modern human dispersals from Africa and a southern route into  
11 Asia. *Proceedings of the National Academy of Sciences of the United States*  
12 *of America* 111, 7248-7253, doi:Doi 10.1073/Pnas.1323666111 (2014).
- 13 19 Liu, W. *et al.* The earliest unequivocally modern humans in southern China.  
14 *Nature* 526, 696-699, doi:10.1038/nature15696 (2015).
- 15 20 Prugnolle, F., Manica, A. & Balloux, F. Geography predicts neutral genetic  
16 diversity of human populations. *Current Biology* 15, R159-R160 (2005).
- 17 21 Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide  
18 patterns of variation. *Science* 319, 1100-1104, doi:DOI  
19 10.1126/science.1153717 (2008).
- 20 22 Ramachandran, S. *et al.* Support from the relationship of genetic and  
21 geographic distance in human populations for a serial founder effect  
22 originating in Africa. *Proceedings of the National Academy of Sciences of*  
23 *the United States of America* 102, 15942-15947 (2005).
- 24 23 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* 328,  
25 710-722, doi:10.1126/science.1188021 (2010).
- 26 24 Reich, D. *et al.* Denisova admixture and the first modern human dispersals  
27 into Southeast Asia and Oceania. *American journal of human genetics* 89,  
28 516-528, doi:10.1016/j.ajhg.2011.09.005 (2011).
- 29 25 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from  
30 western Siberia. *Nature* 514, 445-449, doi:10.1038/nature13810 (2014).
- 31 26 Fu, Q. *et al.* A revised timescale for human evolution based on ancient  
32 mitochondrial genomes. *Current Biology* 23, 553-559,  
33 doi:10.1016/j.cub.2013.02.044 (2013).
- 34 27 Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans  
35 dating back at least 36,200 years. *Science* 346, 1113-1118,  
36 doi:10.1126/science.aaa0114 (2014).
- 37 28 Gravel, S. *et al.* Demographic history and rare allele sharing among human  
38 populations. *Proceedings of the National Academy of Sciences of the United*  
39 *States of America* 108, 11983-11988, doi:10.1073/pnas.1019276108 (2011).
- 40 29 Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic  
41 Denisovan Individual. *Science* 338, 222-226, doi:Doi  
42 10.1126/Science.1224344 (2012).
- 43 30 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of  
44 ancestry in unrelated individuals. *Genome research* 19, 1655-1664, doi:DOI  
45 10.1101/gr.094052.109 (2009).

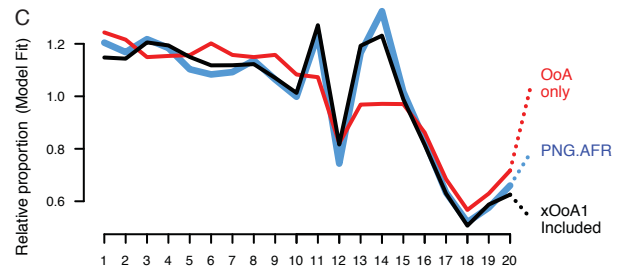
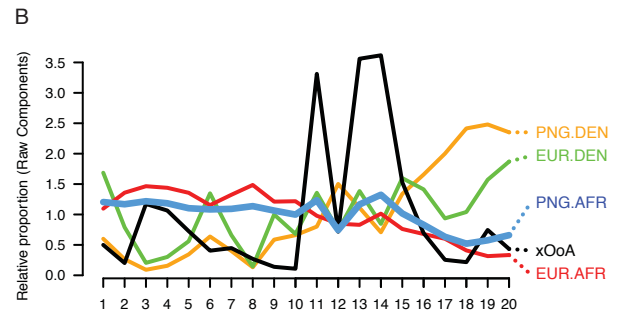
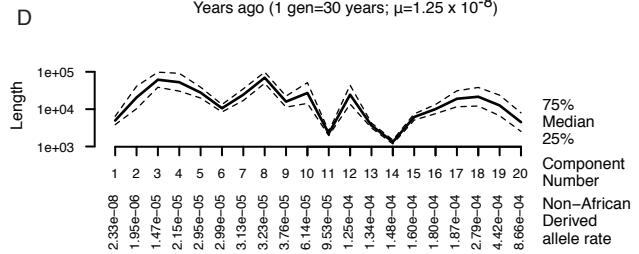
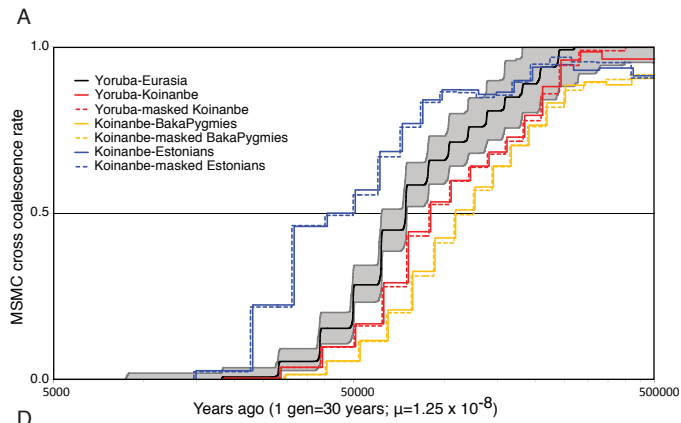
- 1 31 Cardona, A. *et al.* Genome-wide analysis of cold adaptation in indigenous  
2 siberian populations. *PloS one* 9, e98076,  
3 doi:10.1371/journal.pone.0098076 (2014).
- 4 32 Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population  
5 structure using dense haplotype data. *PLoS genetics* 8, e1002453,  
6 doi:10.1371/journal.pgen.1002453 (2012).
- 7 33 Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population  
8 structure with estimated effective migration surfaces. *Nat Genet* 48, 94-  
9 100, doi:10.1038/ng.3464 (2016).
- 10 34 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual  
11 ancestry of Native Americans. *Nature* 505, 87-91, doi:10.1038/nature12736  
12 (2014).
- 13 35 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern  
14 Eurasians. *Nature communications* 6, 8912, doi:10.1038/ncomms9912  
15 (2015).
- 16 36 Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science*  
17 343, 747-751, doi:10.1126/science.1243518 (2014).
- 18 37 Rasmussen, M. *et al.* An Aboriginal Australian Genome Reveals Separate  
19 Human Dispersals into Asia. *Science* 333, 94-98, doi:Doi  
20 10.1126/Science.1211177 (2011).
- 21 38 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova  
22 Cave in Siberia. *Nature* 468, 1053-1060, doi:10.1038/nature09710 (2010).
- 23 39 Wall, J. D. *et al.* Higher levels of neanderthal ancestry in East Asians than in  
24 Europeans. *Genetics* 194, 199-209, doi:10.1534/genetics.112.148213 (2013).
- 25 40 Hughes, A. L. *et al.* Widespread purifying selection at polymorphic sites in  
26 human protein-coding loci. *Proceedings of the National Academy of  
27 Sciences of the United States of America* 100, 15754-15757,  
28 doi:10.1073/pnas.2536718100 (2003).
- 29 41 Lohmueller, K. E. *et al.* Proportionally more deleterious genetic variation in  
30 European than in African populations. *Nature* 451, 994-997 (2008).
- 31 42 Khrameeva, E. E. *et al.* Neanderthal ancestry drives evolution of lipid  
32 catabolism in contemporary Europeans. *Nature communications* 5, 3584,  
33 doi:10.1038/ncomms4584 (2014).
- 34 43 Chapman, N. H. & Thompson, E. A. A model for the length of tracts of  
35 identity by descent in finite random mating populations. *Theoretical  
36 population biology* 64, 141-150 (2003).
- 37 44 Migliano, A. B., Vinicius, L. & Lahr, M. M. Life history trade-offs explain the  
38 evolution of human pygmies. *Proceedings of the National Academy of  
39 Sciences of the United States of America* 104, 20216-20219,  
40 doi:10.1073/pnas.0708024105 (2007).
- 41 45 Soares, P. *et al.* The Archaeogenetics of Europe. *Current Biology* 20, R174-  
42 R183 (2010).
- 43 46 Behar, D. M. *et al.* A "Copernican" Reassessment of the Human  
44 Mitochondrial DNA Tree from its Root. *American journal of human genetics*  
45 90, 675-684, doi:Doi 10.1016/J.Ajhg.2012.03.002 (2012).

- 1 47 Karmin, M. *et al.* A recent bottleneck of Y chromosome diversity coincides  
2 with a global change in culture. *Genome research* 25, 459-466,  
3 doi:10.1101/gr.186684.114 (2015).
- 4 48 Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G. & Siepel, A. Bayesian  
5 inference of ancient human demography from individual genome  
6 sequences. *Nat Genet* 43, 1031-1034, doi:10.1038/ng.937 (2011).
- 7 49 Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major  
8 Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe.  
9 *Current biology : CB*, doi:10.1016/j.cub.2016.01.037 (2016).
- 10 50 Kuhlwilm, M. *et al.* Ancient gene flow from early modern humans into  
11 Eastern Neanderthals. *Nature* 530, 429-433, doi:10.1038/nature16544  
12 (2016).
- 13 51 Clarkson, C. *et al.* The archaeology, chronology and stratigraphy of  
14 Madjedbebe (Malakunanja II): A site in northern Australia with early  
15 occupation. *Journal of human evolution* 83, 46-64,  
16 doi:10.1016/j.jhevol.2015.03.014 (2015).
- 17 52 O'Connell, J. F. & Allen, J. The process, biotic impact, and global  
18 implications of the human colonization of Sahul about 47,000 years ago.  
19 *Journal of Archaeological Science* 56, 73 - 84,  
20 doi:<http://dx.doi.org/10.1016/j.jas.2015.02.020> (2015).
- 21 53 Fu, Q. *et al.* An early modern human from Romania with a recent  
22 Neanderthal ancestor. *Nature* 524, 216-219, doi:10.1038/nature14558  
23 (2015).
- 24 54 Stewart, J. R. & Stringer, C. B. Human evolution out of Africa: the role of  
25 refugia and climate change. *Science* 335, 1317-1321,  
26 doi:10.1126/science.1215627 (2012).
- 27 55 Mikkelsen, T. *et al.* Initial sequence of the chimpanzee genome and  
28 comparison with the human genome. *Nature* 437, 69-87 (2005).  
29  
30

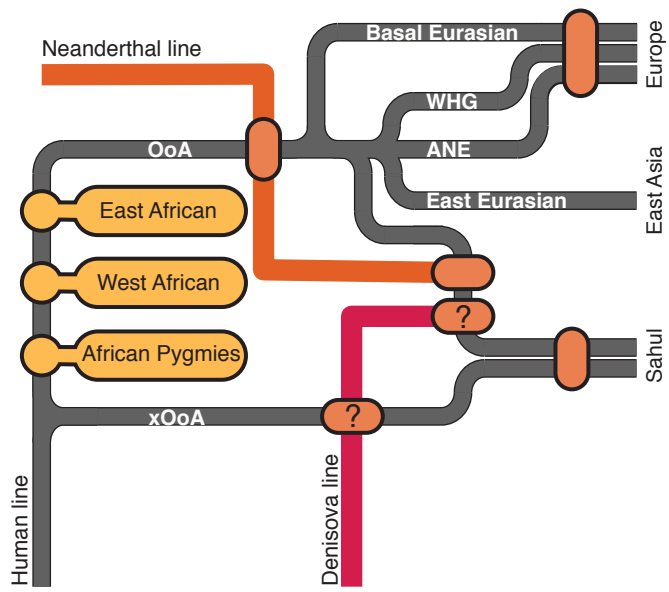












**Table 1 Eurasian subset of variants highlighted by positive selection tests**

Gene	SNP	Variant Type	Test	Population	Phenotype
<i>FADS2</i>	rs2524296	intronic	di	Wsi	Fatty acid desaturation
<i>ZNF646</i>	rs749670	missense	dDAF,DIND	CSi	Lipid metabolism, bile synthesis
<i>PPARA</i>	rs6008197	missense	iHS,nSL,TD,DIND	SoA	Lipid metabolism
<i>GANC</i>	rs8024732	missense	iHS,DIND	SoA	Carbohydrate metabolism
<i>PKDREJ</i>	rs6519993	missense	iHS,nSL,TD,DIND	SoA	Sperm-Receptor, kidney disease
<i>CSMD1</i>	rs7816731	non-coding	di	Wsi	Blood pressure
<i>LYPD3</i>	rs117823872	non-coding	di	Wsi	Wound healing
<i>POU2F3</i>	rs882856	missense	dDAF	WEu	Wound healing
<i>B9D1</i>	rs4924987	missense	dDAF	EEu	Ciliogenesis
<i>PCDH15</i>	rs4935502	missense	dDAF	CSi	Ciliogenesis
<i>TMEM216</i>	rs10897158	missense	dDAF	Wsi	Ciliogenesis
<i>PLCB2</i>	rs936212	missense	dDAF	NSi	Ciliogenesis
<i>MYO18B</i>	rs2236005	missense	dDAF	Sel	Motor activity
<i>FLNB</i>	rs12632456	missense	dDAF	Sel	Motor activity
<i>TTN</i>	rs10497520	missense	dDAF	MIe	Motor activity

Note the abbreviations of the population group names are according to Table S2.2

*iHS,nSL, or TD, indicates that the variant is a from a top 1% window by that test for the indicated population. DIND indicates that the variant is significantly (>5SD) above the neutral background by the DIND test (See Supplementary Section 3)*

*di indicates that the variant was in the top 12 of the most highly divergent SNVs by the di score in each of the twelve population groups (See Supplementary Section 3)*

*dDAF indicates that the variant was in the top 20 most highly differentiated SNPs in its class in a given comparison (See Supplementary Section 3)*