POPULATION GENETICS

A geographic history of human genetic ancestry

Michael C. Grundler[1], Jonathan Terhorst[2], Gideon S. Bradburd[1]*

Describing the distribution of genetic variation across individuals is a fundamental goal of population genetics. We present a method that capitalizes on the rich genealogical information encoded in genomic tree sequences to infer the geographic locations of the shared ancestors of a sample of sequenced individuals. We used this method to infer the geographic history of genetic ancestry of a set of human genomes sampled from Europe, Asia, and Africa, accurately recovering major population movements on those continents. Our findings demonstrate the importance of defining the spatiotemporal context of genetic ancestry when describing human genetic variation and caution against the oversimplified interpretations of genetic data prevalent in contemporary discussions of race and ancestry.

Present-day genomes are inherited from an unbroken chain of ancestors who lived in different geographic locations at different times, creating spatial patterns of genetic relatedness (*1*). Understanding these patterns is vital both for the identification of the genomic basis of phenotypic variation (*2*) and for knowledge of the demographic history of a species (*3*). Conversely, ignoring spatial demographic history can have serious implications for genome-wide association studies or the identification of loci involved in local adaptation (*4*).

Genetic variation in humans is often summarized with discrete labels, but these can be inaccurate and misleading (*5*, *6*). Even when based on geographic history, genetic ancestry labels oversimplify a complex picture because they implicitly focus on only a single point in time. For example, based on our current understanding of human origins, all living individuals are "African" (regardless of the geography of their recent ancestors) when considering their ancestry ~200,000 years before present. Advances in the study of ancient DNA have revealed a lack of genetic continuity within geographic regions (*7*–*10*), further highlighting the shortcomings of genetic ancestry labels. The fact that these labels are generated using statistical genetics approaches gives them the veneer of authenticity, further reifying problematic notions of race and ancestry in society (*6*, *11*).

At a technical level, many of the existing methods for quantifying ancestry average over the ages of shared ancestors in the sample, effectively "flattening" the temporal component of the genealogy that connects all individuals within a species (*12*, *13*). In reality, any pair of individuals is connected by many shared ancestors from whom each has inherited some portion of their genome (*14*). This flattening has the effect of painting a static notion of ancestry rather than one that changes as it proceeds backward in time.

If we knew the identities, locations, and ages of the ancestors of a sample, we could more precisely and accurately report their geographic ancestry through time. Moreover, we could learn about their history of dispersal, identifying major population movements, demographic events, and barriers to migration. Although such detailed pedigree information is rare [but see (*15*) and (*16*)], it is nevertheless possible to learn about pedigree ancestors that are shared among individuals in a sample. This is because genetic relationships between samples, as well as the identities of the shared ancestors with whom they are related, are encoded in an interwoven collection of gene genealogies called an ancestral recombination graph (ARG) (*17*–*19*).

Recent advances in statistical and computational population genetics (*20*–*23*) have facilitated the inference of an ARG from large numbers of genomes. The ARG is a record of all coalescence and recombination events since the divergence of the sequences under study, and therefore it specifies a complete genealogy of the sample at each genomic position. This record can be represented as a tree sequence (*24*, *25*), an ordered set of trees localized to adjacent regions of the genome describing the gene genealogies of a set of samples at every genomic position. Each internal node in a local genealogy represents a haplotype within an ancestor from whom two or more sampled individuals have coinherited a portion of their genome. By estimating where and when each of these ancestors lived, we can reconstruct the geographic history of a set of modern-day individuals, documenting the path through space and time by which their genomes came to them.

Here, we present and validate a method for achieving this goal. Our method, called GAIA (geographic ancestor inference algorithm), efficiently infers the geographic locations of the shared ancestors of a modern, georeferenced sample. We applied GAIA to a tree sequence of humans sampled in Europe, Asia, the Middle East, and Africa (*22*) and reconstructed the geographic history of human ancestry over the past 2 million years.

Inferring the locations of shared ancestors

Conceptually, GAIA is like many tree-based methods in phylogeography that attempt to reason about the locations and movements of genetic ancestors based on the geographic distribution of modern-day samples and the phylogenetic relationships among them (*26*–*28*). Instead of working with a single gene tree or with a collection of independent gene trees, GAIA generates inferences using a sequence of locally correlated gene trees. It is similar to several recent existing methods in this regard. For example, Wohns *et al.* (*22*) introduced a nonparametric approach that estimates ancestor locations by successively averaging the coordinates of sample locations in a postorder traversal of a tree sequence. In addition, Osmond and Coop (*29*) described a likelihood method for locating genetic ancestors and estimating migration rates based on a model of branching Brownian motion using a sample of gene trees from a tree sequence, an approach that was recently extended by Deraje *et al.* (*30*) to the full ARG. GAIA differs from these approaches, both in its choice of optimality criterion and in its flexible representation of geographic space, which can be continuous or discrete.

GAIA works by fitting a minimum migration cost function to each genomic position in an ancestral haplotype using the generalized parsimony algorithm (Fig. 1). Because neighboring gene genealogies in a tree sequence are highly correlated, we were able to efficiently maintain the state of parsimony calculations as we iterated over the local genealogies in a tree sequence containing the ancestral haplotype. Once these cost functions were computed for all genomic positions, we averaged them and assigned the ancestral haplotype to the geographic location that minimized its average cost function. These assignments then have a straightforward interpretation: They correspond to the geographic location that minimizes the overall migration cost of an average ancestral base pair. In our implementation, migration cost is a function of geographic distance, and the overall migration cost in a local genealogy is simply the sum of all ancestor-descendant migration costs.

A minimum migration optimality criterion may miss migration events because ancestors in distinct geographic locations can sometimes be assigned to a single geographic location to lower the overall migration cost given the sampling configuration. This can create a simplified picture of ancestry when the reality is more complex (*31*). Additionally, variance in the coalescent process itself can interact with the sampling configuration such that ancestors in the same location are inferred to be in different locations, potentially resulting in mistaken

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [2]Department of Statistics, University of Michigan, Ann Arbor, MI, USA.
*Corresponding author. Email: bradburd@umich.edu

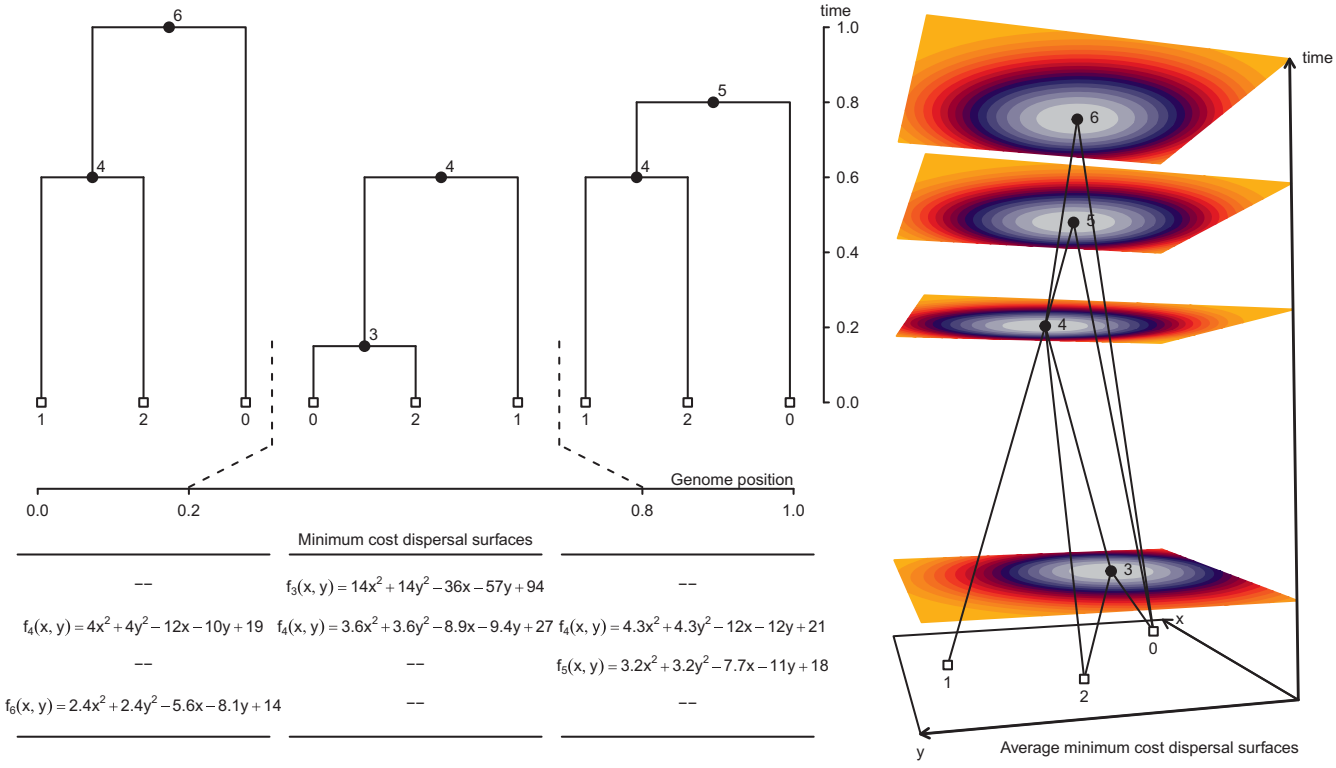

Fig. 1. Conceptual overview of GAIA. For each local tree, we used the dynamic programming method of Sankoff and Rousseau (*38*) to fit a minimum cost dispersal surface to the genealogical relationships of the georeferenced sample nodes. In this example, we used squared Euclidean distance as the cost function, and $f_u(x, y)$ returns the smallest sum of squared dispersal distances between all ancestor-descendant node pairs that can be obtained when node u is at location (x, y). Using the genomic spans of local trees as weights, we then took a weighted average of local surfaces to assign each node a single average minimum cost dispersal surface. Here, node 4 appears in all three local trees, and its final fitted surface is the weighted average of the three local surfaces. By contrast, nodes 3, 5, and 6 appear in a single local tree, and their final fitted surfaces are identical to the surface in the local tree in which they appear. The perspective plot in the rightmost panel displays the ancestral recombination graph encoding the local trees along with the final fitted surface for each node. Nonsample nodes are positioned at the minimum cost point on the surface. Warmer colors denote higher costs.

inferences of migration. Because GAIA is principally an exploratory tool, we did not attempt exhaustive exploration of these biases, but users of GAIA should bear them in mind when interpreting results.

To validate GAIA's performance, we simulated genetic data under different spatial models using SLiM (*32*). GAIA performs well under a variety of demographic histories and over a range of dispersal kernels (both magnitude and shape), achieving greater accuracy and faster computation times than related nonparametric methods (figs. S3 to S6), and it is able to accurately reconstruct pairwise distributions of ancestor distances over a range of temporal scales (fig. S13). We also demonstrate that we can use the reconstructed geographic distances between nodes in the tree sequence to estimate the parent-offspring dispersal distance for both Gaussian and non-Gaussian dispersal kernels (figs. S1 and S2).

Tracking human ancestors through space and time

We inferred the geographic locations of ancestors of a contemporary sample of 2140 georeferenced human genomes from the Human Genome Diversity Project (*33*, *34*) using a dated tree sequence inferred for chromosome 18 by Wohns *et al.* (*22*) (Fig. 2). To avoid the complexity introduced by large-scale postcolonial migrations, we focused on ancestry of the subset of individuals sampled from the continents of Europe, Asia, and Africa, consisting of 1070 contemporary individuals. The tree sequence for these individuals consists of 28,154 local genealogies containing 114,606 ancestral nodes and spanning ~80,000 generations of human history. An equal area discrete global grid (*35*) (cell spacing ~800 km) intersected with Earth's landmass provided a set of habitable areas; individual sample locations were assigned to the nearest grid cell. Although a variety of complex migration cost matrices can be envisioned, we opted for the simplest model that only allows migration between neighboring grid cells and assigns a unit cost to each migration event. We used GAIA to locate ancestral nodes to the grid cells with the lowest migration cost. For many ancestral nodes, especially older nodes, multiple grid cells may be optimal or near optimal (fig. S7). Because our summaries ignore near-optimal solutions, they do not explore the full range of uncertainty in ancestral locations and should be viewed, not as precise statements on where ancestors lived and how they moved, but rather as summaries of major trends.

Our inferred geographic chronology of the ancestors of the sample largely reconstructs major population movements in human prehistory, including the out-of-Africa expansion and the peopling of Eurasia (Fig. 3). Our estimates place some genetic ancestors of the sample in Asia and the Middle East well before the earliest fossil evidence of human dispersal out of Africa. The oldest nodes in the tree sequence are estimated to have occurred ~2 million years before present, and a small minority of these are inferred to be located in Asia and the Middle East. A similar pattern was observed by Wohns *et al.* (*22*) in their analysis of chromosome 20. Geographic estimates of ancient ancestry must be viewed with caution because phylogenetic signal is highly attenuated at deep timescales. In the extreme case of complete loss of phylogenetic signal, estimates from GAIA will be no better than random

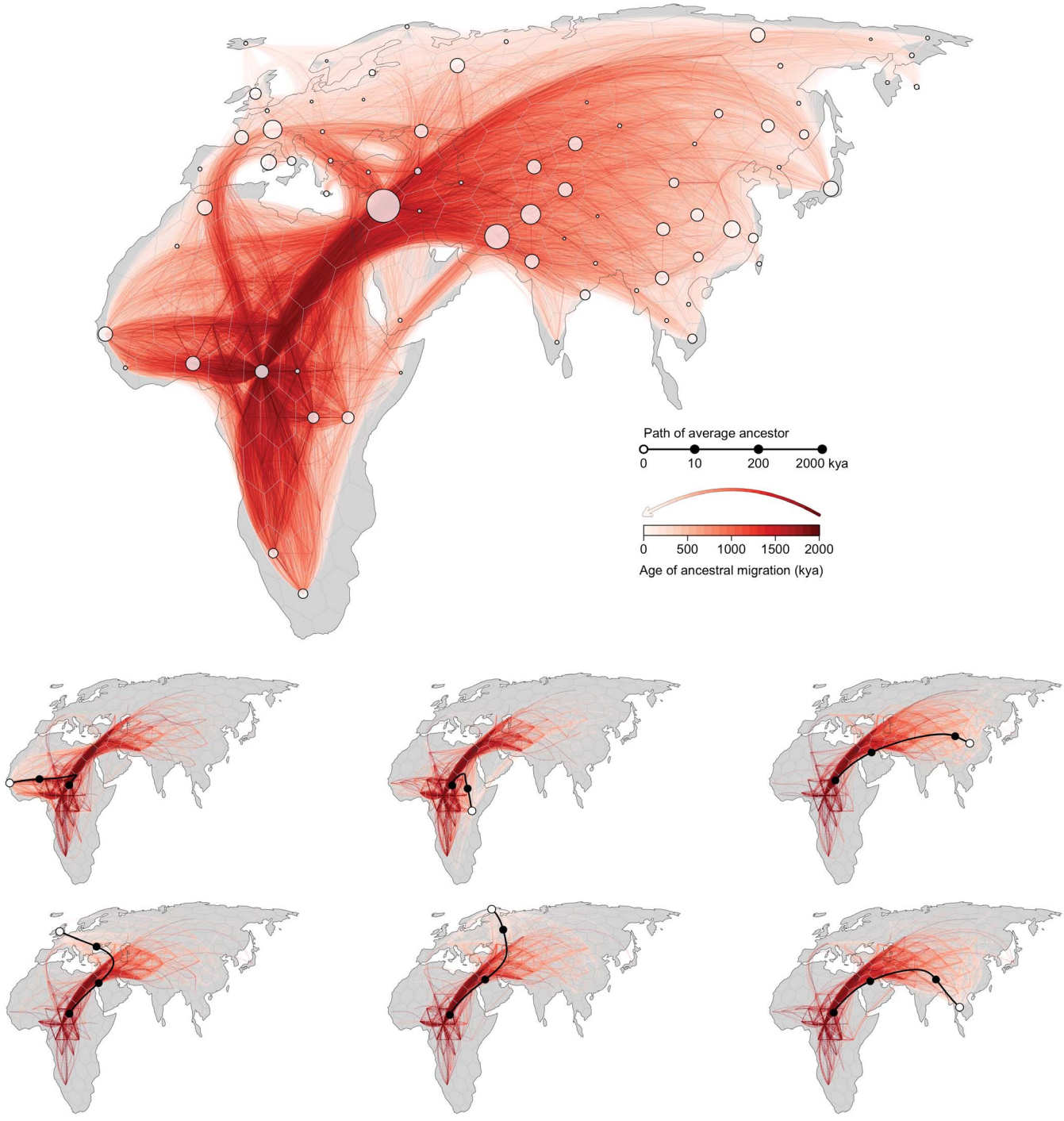

Fig. 2. A georeferenced ARG. Red lines trace the inferred historical migrations of genetic ancestors of the sample (white points), with darker shading used to indicate movements that took place in the more distant past. Six contemporary samples are highlighted in the bottom panels. Black lines trace the average position of their genetic ancestors back through time, and red lines denote the subset of edges in the ARG that are ancestral to the samples.

guessing and will tend to be pulled toward a majority rule geographic centroid location. Although this is not a concern for the dataset as a whole (where the geographic centroid occurs in central Asia), it may contribute to GAIA's placement of the oldest genetic ancestors in central Africa (fig. S12). These concerns notwithstanding, from ~2 million to 200,000 years before present, the average positions of genetic ancestors to the geographic subsets of samples from Europe, Asia, the Middle East, and Africa are all inferred to be in Africa. This coincides with the period when most genome positions in these geographic subsets trace their descent to an ancestor in Africa. Between 200,000 years ago and the present, the average location of ancestors to the geographic subsets of samples from Europe, Asia, and the Middle East diverge from one another and begin to move toward the average of position of samples from those regions, whereas the average position of ancestors to the African subset of samples remains in Africa (Fig. 4).

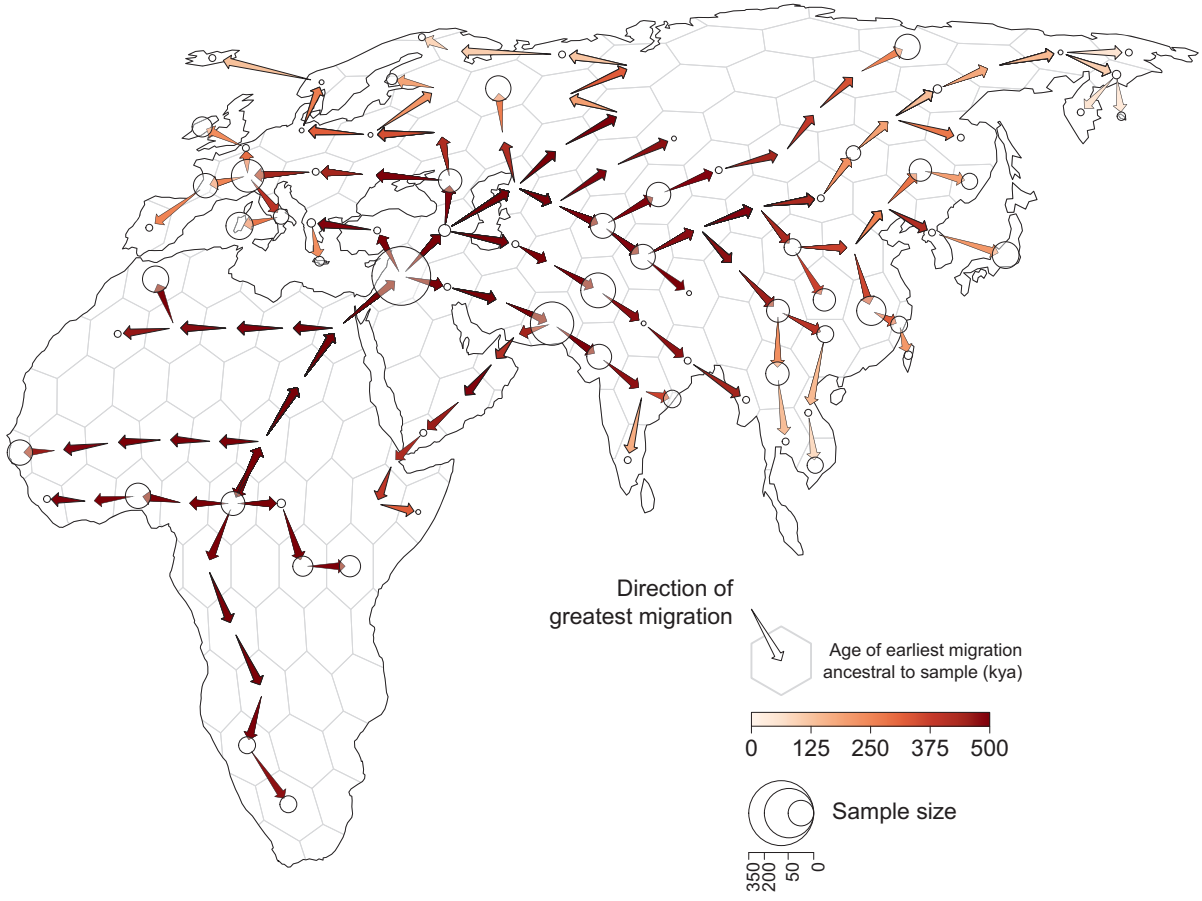

Fig. 3. Geographic chronology of human genetic ancestry. Arrows point in the direction of the greatest migration of the shared genetic ancestors of the sample and are colored according to the age of the earliest migration. Points show the distribution of sampled modern-day genomes. Point size is proportional to the number of samples from each locality.

Geographic history of human ancestry

We used the georeferenced tree sequence to define a spatiotemporally explicit ancestry coefficient, which we then tracked across space and time to understand and quantify the genetic and geographic history of our sample. Only a subset of any given contemporary individual's genome, $A_i(t)$ in individual i, is found in the ancestral nodes or branches in the tree sequence at a given point in time t (once the local genealogy in a particular genomic region has coalesced, that portion of the genome is no longer represented in deeper sections of the tree sequence; therefore, the farther back in time we look, the less of any modern-day individual's genome can be found). We define this spatiotemporally explicit ancestry coefficient, which we call $z_{ik}(t)$, as $A_{ik}(t)/A_i(t)$: the proportion of individual i's genome that exists in its ancestors in the tree sequence at time t and that is inherited from ancestors living within some prescribed geographic region k. This ancestry coefficient can be broadly understood as the proportion of a sample's genome inherited from individuals living within a specified geographic region at a specific time.

Unlike ancestry labels, $z_{ik}(t)$ is explicitly associated with both a point in time and a region of space. Therefore, we can use it to understand how a sample's ancestry has changed across space and time. At the present moment, the distribution of z_{ik} simply reflects the geography of modern-day sampling. By 100,000 years ago, we find that only 2% of modern-day samples are inheriting their genomes from ancestors in Europe (Fig. 4); almost all ancestors contributing genomic material to the modern-day sample are inferred to have lived in Africa, Asia, and the Middle East. The spatiotemporal ancestry coefficients show similar trajectories for modern-day samples from Asia. During this same period, the proportion of sample ancestors found in Africa increases almost monotonically backward in time, from 16% at the present (again, reflecting the geography of modern-day sampling) to ~95% by 1 million years ago. The proportion of the genomes of the sample inherited from ancestors inferred to be in Africa plateaus at ~95%, with ~5% of genome continuing to be inherited from ancestors in the Middle East along the eastern Mediterranean (fig. S8). Randomly downsampling the data so that each population is represented by only a single individual does not change this result (fig. S9). However, because attenuation of phylogenetic signal at deep timescales can cause misplacement of genetic ancestors, this result should be viewed with caution.

Large-scale movement in human ancestry

To study large-scale movements of human ancestry, we introduce a new statistical summary of the georeferenced tree sequence: ancestry flux. Formally, if $A_i(t_l, t_r)$ is the amount of individual i's genome that inherits from ancestors alive between $[t_l, t_r)$, and $A_{ijk}(t_l, t_r)$ is the same amount that inherits from ancestors who moved from j to k, we define ancestry flux as $\phi_{ijk}(t_l, t_r) = A_{ijk}(t_l, t_r)/A_i(t_l, t_r)$: the proportion of i's genome that exists in its ancestors in the tree sequence during the period $[t_l, t_r)$ and that is inherited from ancestors who moved from j to k during that same period. This coefficient can be broadly understood as the proportion of a sample's genome inherited from ancestors who moved between specified geographic regions during a particular period. However,

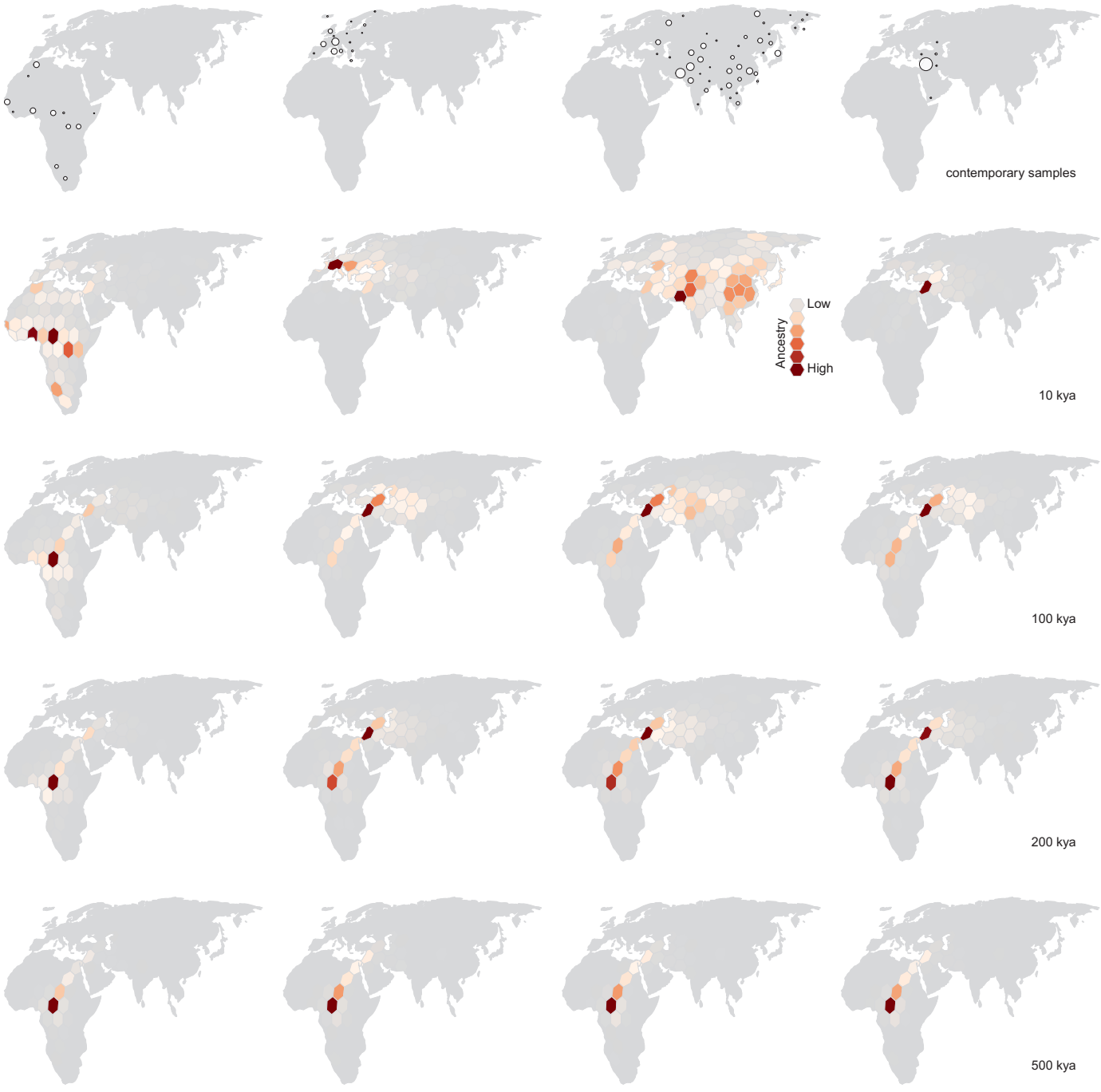

Fig. 4. Inferred spatiotemporal ancestry coefficients through time. The fraction of genomic positions in each geographic subset of samples (top row) that trace ancestry to different geographic regions at different times in the past is represented by shades of red, with darker shading indicating greater ancestry (i.e., a larger fraction). Point size is proportional to the number of sampled genomes from each locality.

because of the nature of the coalescent process, we expect that the movement of genetic ancestry quantified in this fashion will predate the movement of individuals carrying that ancestry. Additionally, estimates of ancestry flux are subject to error in ancestor location estimates and error incurred by interpolating migration routes between estimated locations (fig. S14), so care is warranted when interpreting the directions and timings associated with the coefficients.

We discretized Europe, Asia, and Africa into approximately equal-area hexagons (each ~800 km^2) and quantified ancestry flux between them in 2500-year intervals between the present and 0.5 million years ago. We found consistent ancestry flux out of Africa into the Middle East during this time, with a peak occurring between 100,000 and 150,000 years ago (Fig. 5). Nearly all ancestry flux out of Africa is estimated to have occurred through a northern route across the Sinai Peninsula rather than a southern route across the strait at Bab-el-Mandeb. Approximately 30% of genomic positions in the modern-day sample trace ancestry to a northerly migration out of Africa, but only 0.1% trace ancestry to a southerly migration

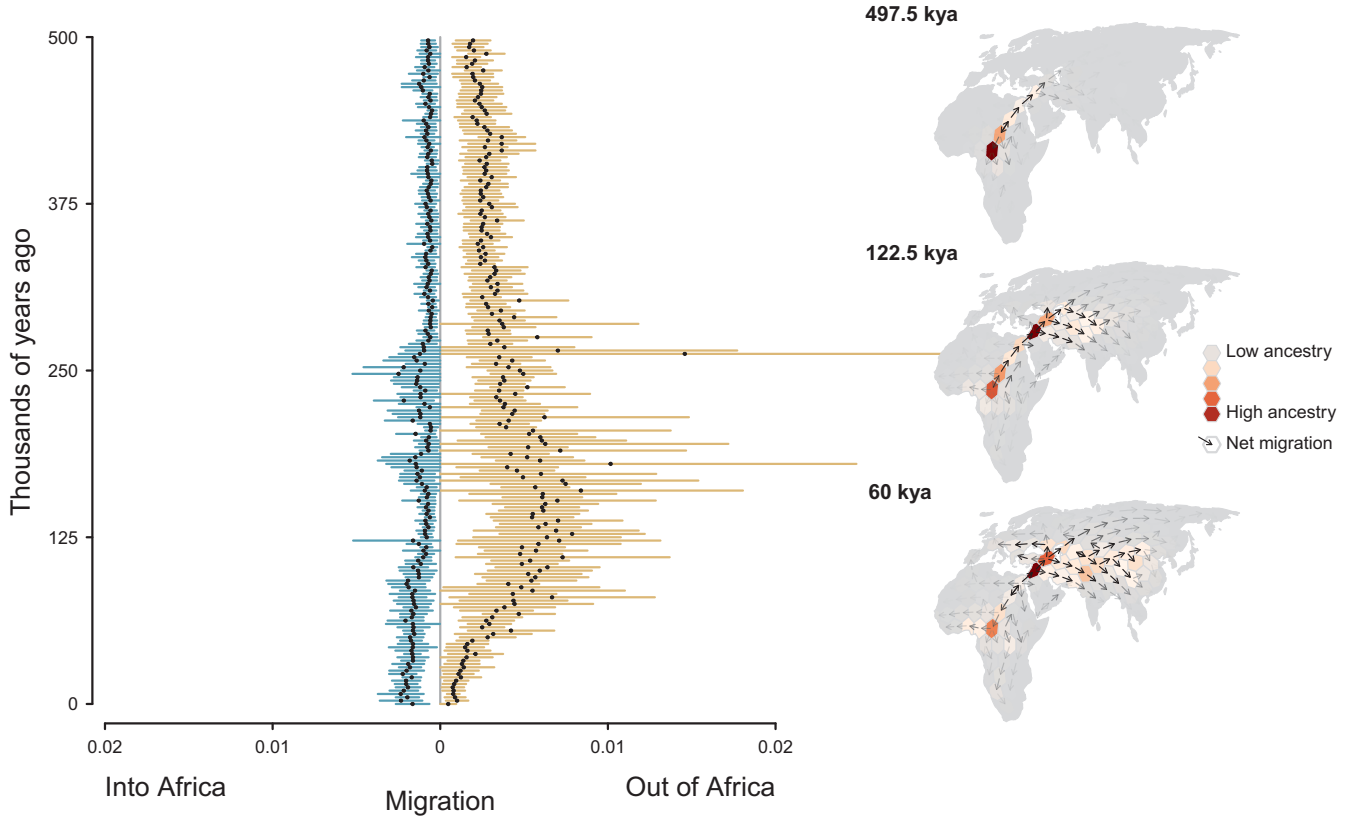

Fig. 5. Inferred flux in spatiotemporal ancestry coefficients through time. Each point is the mean estimate (computed from random subsets of the data and bracketed by the SD) of the amount of historical migration of shared genetic ancestors of the modern-day sample into or out of Africa over time. Inset maps highlight three different time intervals. Cells are colored by the fraction of genomic positions that find ancestry in them at that time, and arrows depict the direction of greatest migration, with opacity scaled by the magnitude of migration.

route out of Africa. However, this result should be viewed with caution, because simulations indicate that GAIA would detect such a pattern even if the true route had been through the south (fig. S11).

Discussion

We have described a simple heuristic approach for reconstructing the geographic genetic history of a sample, which can be used to explore the timing and location of demographic events. Patterns of diversity and relatedness between individuals are created by a complex interplay between evolutionary forces acting over the history of a population. By uncovering these patterns, our method enables researchers to go beyond the standard practice of grouping individuals into a small number of static and ill-defined "populations." Although this practice may be useful in some limited scenarios (e.g., for conservation and management), our results illustrate that it discards an immense amount of additional information. Furthermore, when applied to humans, population assignment has the potential to lead to greater harm and confusion by conflating race, ethnicity, geography, and ancestry (*5*). A spatiotemporal summary of ancestors offers an intuitive and informative path toward understanding ancestry, demography, and major population movements through time.

Of course, the geographic histories inferred by GAIA are not perfectly accurate; individual dispersal decisions are the result of countless factors that we cannot capture using a simple heuristic model. For example, GAIA assumes that lineages disperse independently of one another and that the dispersal process is independent of the coalescent process in a reconstructed gene tree, and both of these assumptions are certainly violated in empirical datasets. Moreover, errors in the tree sequences inferred by upstream tools such as RELATE (*20*) and TSINFER (*21*) have the potential to propagate into the output of GAIA. Additionally, the sampling process itself may affect our results insofar as the accuracy of geographic inference of ancestor locations depends on the distribution of sampled individuals relative to that of their ancestors. Finally, we caution that GAIA sheds light on the geographic and temporal distribution of shared genetic ancestors. Even if we had perfect information on the geographic locations of all ancestors, it would not necessarily inform our understanding of the distribution of the census population alive at any point in the past. As with all population genetic approaches, historical individuals who contributed no ancestry to the modern-day sample remain inscrutable. We therefore present GAIA as an exploratory data analysis tool rather than a formal approach for explicitly comparing different models of geographic history.

Despite these limitations, the ability to infer the locations of ancestors in the tree sequence opens exciting avenues for research. Simply summarizing the georeferenced tree sequence can yield valuable insights about the history of a population, including identifying barriers to dispersal, shifts in dispersal regimes through time, and the geography of ecological dynamics (at least with respect to genetic ancestry). A straightforward direction for extending GAIA would be to explicitly incorporate some of these elements. Because the transitions between different geographic states occur on an arbitrary network, we can easily incorporate different geographies, such as dispersal routes that connect far-flung regions that may be of interest when species occasionally experience human-mediated dispersal (*36*). More broadly, the ability to study the geography of genealogies heralds an exciting growth in the ability of

the field of population genetics to shed light on population ecological processes governing the movement, distribution, and density of individuals across space and time.

REFERENCES AND NOTES

1. G. S. Bradburd, P. L. Ralph, *Annu. Rev. Ecol. Evol. Syst.* **50**, 427–449 (2019).
2. C. J. Battey, P. L. Ralph, A. D. Kern, *Genetics* **215**, 193–214 (2020).
3. P. Ralph, G. Coop, *PLOS Biol.* **11**, e1001555 (2013).
4. A. A. Zaidi, I. Mathieson, *eLife* **9**, e61548 (2020).
5. G. Coop, Genetic similarity and genetic ancestry groups. arXiv: 2207.11595 [q-bio.PE] (2022).
6. A. C. F. Lewis *et al*., *Perspect. Biol. Med.* **66**, 225–248 (2023).
7. W. Haak *et al*., *Nature* **522**, 207–211 (2015).
8. M. E. Allentoft *et al*., *Nature* **522**, 167–172 (2015).
9. F. Racimo *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8989–9000 (2020).
10. T. M. Mattila *et al*., *Commun. Biol.* **6**, 793 (2023).
11. J. Carlson, B. M. Henn, D. R. Al-Hindi, S. Ramachandran, *Nature* **610**, 444–447 (2022).
12. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, *PLOS Genet.* **8**, e1002453 (2012).
13. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* **155**, 945–959 (2000).
14. I. Mathieson, A. Scally, *PLOS Genet.* **16**, e1008624 (2020).
15. S. M. Aguillon *et al*., *PLOS Genet.* **13**, e1006911 (2017).
16. L. Anderson-Trocmé *et al*., *Science* **380**, 849–855 (2023).
17. R. C. Griffiths, P. Marjoram, in *Progress in Population Genetics and Human Evolution* (Springer, 1997), pp. 257–270.
18. A. L. Lewanski, M. C. Grundler, G. S. Bradburd, *PLOS Genet.* **20**, e1011110 (2024).
19. Y. Wong *et al*., *Genetics* **228**, iyae100 (2024).
20. L. Speidel, M. Forest, S. Shi, S. R. Myers, *Nat. Genet.* **51**, 1321–1329 (2019).
21. J. Kelleher *et al*., *Nat. Genet.* **51**, 1330–1338 (2019).
22. A. W. Wohns *et al*., *Science* **375**, eabi8264 (2022).
23. Y. Deng, R. Nielsen, Y. S. Song, Robust and accurate Bayesian inference of genome-wide genealogies for large samples. bioRxiv [Preprint] (2024); https://doi.org/10.1101/2024.03.16.585351.
24. J. Kelleher, A. M. Etheridge, G. McVean, *PLOS Comput. Biol.* **12**, e1004842 (2016).
25. J. Kelleher, K. R. Thornton, J. Ashander, P. L. Ralph, *PLOS Comput. Biol.* **14**, e1006581 (2018).
26. J. Avise, *Phylogeography: The History and Formation of Species*, (Harvard Univ. Press, 2000), **vol. 447**.
27. L. L. Knowles, W. P. Maddison, *Mol. Ecol.* **11**, 2623–2635 (2002).
28. L. L. Knowles, *Annu. Rev. Ecol. Evol. Syst.* **40**, 593–612 (2009).
29. M. Osmond, G. Coop, *eLife* **13**, e72177 (2024).
30. P. Deraje, J. Kitchens, G. Coop, M. M. Osmond, Inferring the geographic history of recombinant lineages using the full ancestral recombination graph. bioRxiv [Preprint] (2024); https://doi.org/10.1101/2024.04.10.588900.
31. E. M. L. Scerri, L. Chikhi, M. G. Thomas, *Nat. Ecol. Evol.* **3**, 1370–1372 (2019).
32. B. C. Haller, P. W. Messer, *Am. Nat.* **201**, E127–E139 (2023).
33. H. M. Cann *et al*., *Science* **296**, 261–262 (2002).
34. J. Z. Li *et al*., *Science* **319**, 1100–1104 (2008).
35. R. Barnes, K. Sahr, "dggridR: Discrete Global Grids" (r package version 3.0.0., 2023); https://cran.r-project.org/package=dggridR.
36. S. H. Zhan *et al*., Towards pandemic-scale ancestral recombination graphs of SARS-CoV-2. bioRxiv [Preprint] (2023); https://doi.org/10.1101/2023.06.08.544212.
37. M. C. Grundler, J. Terhorst, G. S. Bradburd, Data for: A geographic history of human genetic diversity, Dryad (2024); https://doi.org/10.5061/dryad.p5hqbzkwz.
38. D. Sankoff, P. Rousseau, *Math. Program.* **9**, 240–246 (1975).

ACKNOWLEDGMENTS

We thank Y. Brandvain, J. Carlson, G. Coop, P. Deraje, D. Edge, J. Kitchens, M. Weber, and members of the Bradburd lab for comments and discussion that helped improve the research presented here. **Funding:** This work was supported by the National Science Foundation (award number DMS- 2052653 to J.T.) and the National Institute of General Medical Sciences of the National Institutes of Health under award numbers R35GM151145 (J.T.) and R35GM137919 (G.S.B.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. **Author contributions:** Conceptualization: M.C.G., J.T., G.S.B.; Data curation: M.C.G.; Formal analysis: M.C.G.; Funding acquisition: J.T., G.S.B.; Methodology: M.C.G., J.T., G.S.B.; Software: M.C.G.; Supervision: G.S.B.; Validation: M.C.G.; Writing – original draft: M.C.G., G.S.B.; Writing – review & editing: M.C.G., J.T., G.S.B. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All publicly available datasets used in this paper are available from their original publications. GAIA is available at https://github.com/blueraleigh/gaia under the MIT license. All code and data files used to perform analyses in this paper are available on GitHub at https://github.com/blueraleigh/gaia-paper and on Dryad (*37*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adp4642
Materials and Methods
Supplementary Text
Figs. S1 to S16
References (*39*–*43*)
MDAR Reproducibility Checklist

Submitted 26 March 2024; accepted 24 January 2025
10.1126/science.adp4642