Published in final edited form as:
Science. 2014 February 14; 343(6172): 747–751. doi:10.1126/science.1243518.

A genetic atlas of human admixture history

Garrett Hellenthal[1], **George B.J. Busby**[2], **Gavin Band**[3], **James F. Wilson**[4], **Cristian Capelli**[2], **Daniel Falush**[#5], and **Simon Myers**[#*,3,6]

[1]UCL Genetics Institute, University College London, Gower Street, London WC1E 6BT, UK

[2]Department of Zoology, Oxford University, South Parks Road, Oxford OX1 3PS, UK

[3]Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, Oxford OX3 7BN, UK

[4]Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, UK

[5]Max Planck Institute for Evolutionary Anthropology, DeutscherPlatz 6, 04103 Leipzig, Germany

[6]Department of Statistics, Oxford University, 1 South Parks Road, Oxford OX1 3TG, UK

[#] These authors contributed equally to this work.

Abstract

Modern genetic data combined with appropriate statistical methods have the potential to contribute substantially to our understanding of human history. We have developed an approach that exploits the genomic structure of admixed populations to date and characterize historical mixture events at fine scales. We used this to produce an atlas of worldwide human admixture history, constructed using genetic data alone and encompassing over 100 events occurring over the past 4,000 years. We identify events whose dates and participants suggest they describe genetic impacts of the Mongol Empire, Arab slave trade, Bantu expansion, first millennium CE migrations in eastern Europe, and European colonialism, as well as unrecorded events, revealing admixture to be an almost universal force shaping human populations.

Diverse historical, archaeological, anthropological and linguistic sources of information indicate that human populations have interacted throughout history, due to the rise and fall of empires, invasions, migrations, slavery, and trade. These interactions can result in sudden or gradual transfers of genetic material, creating admixed populations. However, the genetic legacy of these interactions remains unknown in most cases, and the historical record is incomplete. We have developed an approach that provides a detailed characterization of the mixture events in the ancestry of sampled populations, based on genetic data alone.

Admixed populations should have segments of DNA from all contributing source groups (Fig. 1A), whose size decreases over successive generations due to recombination, and approaches have been developed to date admixture events by inferring the size of ancestry segments (1-5). Between-population frequency differences of individual alleles may provide

[*]Correspondence to: myers@stats.ox.ac.uk.

Europe PMC Funders Author Manuscripts

information on ancestry sources (6, 7). Based on these principles we developed an integrated approach, using genome-wide patterns of ancestry to infer jointly both fine-scale information about groups involved in admixture, and its timing, allowing for the fact that migration and admixture events can occur at multiple times or involve numerous groups.

Our approach gains power and resolution by using alleles at multiple successive SNPs (haplotypes) (8). Given a focal population within a larger dataset containing many such groups, the chromosomes of individuals in this population share ancestors with those in other populations, resulting in shared "chunks" of DNA. We used CHROMOPAINTER (8) to decompose each chromosome as a series of haplotypic chunks, each inferred to be shared with an individual from one of the other groups, and colored (or painted) by this group (Fig. 1B). If the focal population is admixed, the changing colors along a chromosome noisily reflect true, but unknown, underlying ancestry (Fig. 1B), and so can be used to learn details of the source group(s) involved. To do this, we model haplotypes within each unsampled source group as being found across a weighted mixture of sampled "donor" populations (9). If a source group is genetically relatively similar to a single sampled population, then this population will dominate the inferred mixture. If there is no close proxy for the admixing group in the sample, especially likely for ancient admixture events or sparsely sampled regions, several donor populations will be needed to approximate its pattern of haplotype sharing. The focal population is then automatically a haplotypic mixture of the combined donors, because it is a mixture of the source groups. Inferring the reduced set of groups within the mixture allows us to produce a "cleaned" painting (Fig. 1B) using only these groups.

To assess the evidence for admixture and date events, informally we measure the scale at which the "cleaned" painting changes along the genome. Specifically, we produce a "coancestry curve" for each pair of donor populations, which plots genetic distance x against a measure of how often a pair of haplotype chunks, separated by distance x, come from each respective population (Fig. 1C), analogously to ROLLOFF curves (4), and averaging over uncertain, and typically computationally estimated, haplotypic phase (9). In theory, given a single admixture event, ancestry chunks inherited from each source have an exponential size distribution, resulting in an exponential decay of these coancestry curves (9). The *rate* of decay in all curves will be equal to the time in generations since admixture (Fig. 1C) (4, 9, 10), allowing estimation of this date: steeper decay corresponds to older admixture. Such a decay distinguishes true admixture from ancient spatial structure, and should only occur in recipient, but not donor, groups involved in non-reciprocal admixture events. We test for evidence to reject ($p<0.01$) a no-admixture null model, i.e. no exponential decay in (normalized) coancestry curves, via bootstrapping (9). Multiple admixture times result in a mixture of exponentials (9); so if admixture is detected, we test for evidence of multiple admixture times (e.g. two episodes of admixture, or more continuous admixture over a longer period; empirical $p<0.05$ in simulations), comparing the fit of a single exponential decay rate versus a mixture of rates.

The curve heights (intercepts) provide complementary information to deconvolve the number and genetic composition of the ancestral sources prior to admixture (11). Fitted curves for all pairs of donor groups (Fig. 1C shows three examples) specify a pairwise

intercept matrix which, following normalization, we decompose using a series of eigenvectors. Analogous to the standard use of eigenvector decomposition in principal components analysis (PCA) in genetics to estimate relative ancestry source contributions for different *individuals* (12), the eigenvectors allow us to estimate the relative contribution to different admixing sources (e.g. source 1 vs. source 2) for each different donor *group* (9). Also as for PCA, admixture between K distinct source populations produces K-1 significant eigenvectors (13), and we test for three or more admixing sources by testing (empirically) for evidence of two or more such eigenvectors ($p<0.05$) (9). Following iterative modeling to improve results, this allows us to attempt to "reverse" the admixture process (Fig. 1D) and infer the haplotypic makeup of admixing source groups as well as admixture date(s), in our method GLOBETROTTER.

To test our approach under diverse single, complex and no-admixture scenarios, incorporating many of the complexities – such as unsampled or admixed donor groups – likely to be present in real data, we simulated admixture scenarios involving real (but hidden to our analysis) human populations (4, 9) and populations generated under a coalescent framework (14) incorporating inferred (15-18) past demographic events. Admixture was simulated between 7 and 160 generations (200-4,400 years, assuming 28 years per human generation (19)) ago, with admixture fractions 3%-50%, and genetic differentiation (F_{ST}) between the admixing groups varying from 0.018 (similar to Europe vs Central Asia) to 0.185 (similar to West Africa vs. Europe). Results are detailed online (Fig. S3-7; Tables S1, S5). All populations simulated without admixture, including those with long-term migration, showed no admixture evidence ($p>0.1$). Power to detect admixture ($p<0.01$) when present was 94%, and 95% of our 95% bootstrapped confidence intervals contained the true admixture date, including cases with two distinct incidents of admixture or multiple groups admixing simultaneously. Inferred source accuracy was very high (9), with e.g. the mixture representation predicting a haplotype composition more correlated to the true, typically unsampled, source population than to any single sampled population >80% of the time. However source accuracy was lower for admixing sources contributing only 5% of DNA, with around 40% of such scenarios yielding elevated (>25%) rates of falsely inferring multiple admixture times and/or admixing groups. Further testing demonstrated robustness of GLOBETROTTER, in simulations and real data, to haplotypic phase inference approach used, inclusion/exclusion of particular chromosomes, genetic map chosen to provide genetic distances, and the presence of population bottlenecks since admixture, while GLOBETROTTER admixture dating was improved relative to ROLLOFF (4, 9).

Nevertheless, there are a number of settings which we believe are challenging for our approach. First, although the admixing sources need not be sampled – often impossible due to genetic drift, extinction, or later admixture into the sources themselves - source inference is improved when more similar extant groups are sampled, and GLOBETROTTER may miss events where we lack any extant group that can separate sources. Second, sampling of several genetically very similar groups can mask admixture events they share. Similarly, a caveat is that where genuine, recent bidirectional gene flow has occurred, admixture fractions are difficult to define and interpret. However, date estimation is predicted to still be useful, and in real data the majority of our inferred events do not appear to be bidirectional in this manner. Third, even in theory our approach finds it challenging to distinguish distinct

continuous "pulses" of admixture and continuous migration over some timeframe (9), due to the difficulty of separating exponential mixtures (20). If the time frame were narrow, we expect to infer a single admixture time, within the range of migration dates. Where we infer two admixture dates, in particular with the same source groups, the exponential decay signal could also be consistent with more continuous migration, and so we conservatively refer to this as admixture at "multiple dates". Finally, we only attempt to analyze populations with signals consistent with at most 3 groups admixing, and infer at most two admixture times, and we can provide only less precise inference of sources for the weaker/older admixture signal in these complex cases (9).

Using GLOBETROTTER, we analyzed 1,490 individuals from 95 worldwide human groups (Table S6, Fig. S11) (9), comprising 17 newly genotyped groups (21), 53 from the Human Genome Diversity Panel (HGDP) (22) and 25 from other sources (23, 24), filtered to 474,491 autosomal SNPs. We phased the individuals using IMPUTE2 (9, 25) and used fineSTRUCTURE (8) to verify homogeneity within labeled populations, identify genetically similar and clustered groups, and to remove outlying individuals (Figs. S12-S14; Tables S10-11). Of the 95 populations, 80 showed evidence ($p<0.01$) of admixture, although nine could not be characterized by our approach (Table S8). More than half of these have evidence of multiple waves of admixture ($p<0.05$), and estimated admixture times vary from <10 generations, to >150 generations (Fig. 2). We present individual results, for each population, via an interactive map online (26). We tested consistency of our results against a previous analysis of the 53 groups within the HGDP (11), which identified 34 groups with evidence of recent admixture. We identify ($p<0.01$) admixture evidence in all 34 cases (with multiple event evidence in 15 cases), and obtain 95% admixture date CIs narrower than, but consistent with, those estimated using ROLLOFF (9, 11). For 10 of 19 HGDP groups lacking previous support for recent admixture, GLOBETROTTER also identifies no events: in the remaining populations admixture is inferred as occurring between genetically similar sources ($F_{ST}<0.02$), a challenging setting where simulations suggest our method is more powerful (9).

In several instances, GLOBETROTTER clarifies or extends previous genetic analyses. For example, a previous study (27) inferred admixture in the Maya, with best source populations the Mozabites from North Africa and the Native American Surui, speculating based on historical events that this might actually represent a mixture of European, West African and Native American ancestry sources. GLOBETROTTER inferred admixture between three groups in the Maya dating to around 1670CE (9 generations ago) (28) (Figure 2A; Fig. 2D, hot pink box 1), with distinct sources from Europe (most genetically similar to the Spanish), West Africa (the Yoruba) and the Americas (the Pima, the nearest sampled group in the Americas). A different method – which aims to detect, but not date, admixture - concluded that Cambodians trace ~16% of their DNA to a group equally related to modern-day Europeans and East Asians (29). GLOBETROTTER infers a ~19% contribution from a similar source - related to modern-day Central, South and East Asians - and an ~81% contribution from a source related specifically to modern-day Han and Dai, the latter a branch of the Tai people who entered the region in historical times (30) (Fig. 2D, orange box 5). Further, this event dates to 1362CE (1194-1502CE), a period spanning the end of the

Indianized Khmer empire (802-1431CE) (30), one of the most powerful empires in Southeast Asia whose fall was hypothesized to relate to a Tai influx (30).

A comparison with the historical record becomes progressively more difficult for older episodes. Even when events are well attested their exact genetic impacts (if any) are rarely if ever known, motivating our approach. Nevertheless, we have identified nine groups of populations showing related events, incorporating almost all (19/20) with the strongest GLOBETROTTER admixture evidence (9). Results are presented as online maps (26). Some events appear to match well with particular historical occurrences such as the Bantu Expansion into Southern Africa (9). Events affecting a group of seven populations (Fig. 2D, purple box 4) correspond in time to the rapid expansion led by Genghis Khan and the subsequent Mongol empire (1206-1368CE) (31), one of the most dramatic events in human history. These populations, including the Hazara (32, 33), Uygur (34) and the Mongola themselves, were sampled from within the range of the Mongol empire and show an admixture event dating within the Mongol Period, with one source closely genetically related to the Mongola that progressively decreases in proportion westward, to 8% in the Turkish (Fig. 2D).

Seventeen populations from the Mediterranean, Near East and countries bordering the Arabian Sea (Fig. 2D, blue box 3) show signals of admixture from sub-Saharan Africa, with most recent dates in the range 890-1754CE (Fig. 2B,D). We interpret these signals, consistent with overlapping results of previous studies (4, 20) as resulting from the Arab expansion and slave trade, which originated around the 7th century (35). Our event dates are highly consistent with this, but also imply earlier sub-Saharan African gene flow into e.g. the Moroccans. The highest-contributing sub-Saharan donor is West African for all 12 Mediterranean populations, and an East or South African Bantu-speaking group for all 5 Arabian Sea populations (Fig. 2D), confirming genetically different sources for these slave trades (35).

A population group centered around Eastern Europe shows signals of complex admixture. FineSTRUCTURE did not fully separate groups from this region, suggesting “masked” shared events might be present. We therefore repainted them excluding each other as donors: we performed similar re-analyses of five additional geographic regions for the same reason (Table S16; Figs. S16-21). The easterly Russians and Chuvash both show evidence ($p<0.05$) of admixture at more than one time (Fig. 2D), at least partially predating the Mongol empire, between groups with ancestry related to Northeast Asians (e.g. the Oroqen, Mongola and Yakut) and Europeans, respectively (Table S16). Six other European populations (Fig. 2D, pink/maroon box 2) independently show evidence following the repainting for similar admixture events involving more than two groups ($p<0.02$) at approximately the same time (Fig. 3). CIs for the admixture time(s) overlap, but predate the Mongol empire, with estimates from 440-1080CE (Fig. 3). In each population, one source group has at least some ancestry related to Northeast Asians with ~2-4% of these groups’ total ancestry linking directly to East Asia. This signal might correspond to a small genetic legacy from invasions of peoples from the Asian steppes (e.g. the Huns, Magyar and Bulgars) during the first millennium CE (36). The other two source groups appear much more local. One is more North-European in the repainting - when we exclude other East

Europe PMC Funders Author Manuscripts

European groups as donors - and is largely replaced by northern Slavic-speaking groups in our original analysis (Fig. 2D; Table S12). The other source is more southerly (e.g. Greece, West Asians). This local migration could explain a recent observation of an excess of IBD sharing in Eastern Europe – including in the Greeks, in whom we infer admixture involving a group represented by Poland, at the same time -that was dated to a wide range between 1,000 and 2,000 years ago (37). We speculate that these events may correspond to the Slavic expansion across this region at a similar time, perhaps related to displacement caused by the Eurasian steppe invaders (38).

Finally, Central Asia shows a particularly complex inferred history following a reanalysis of 10 groups excluding each other as donors, with nine of ten groups showing diverse recent events (Fig. 4A). The exception are the Kalash, a genetically isolated (39) population from the Hindu Kush mountains of Pakistan (40). Distinct, ancient and partially shared admixture signals (always dated older than 90BCE) are seen in six groups (Fig. 4B), including the Kalash (Figure 2C), whose strongest signal suggests a major admixture event (990-210BCE) from a source related to present-day Western Eurasians, though we cannot identify the geographic origin precisely. This period overlaps that of Alexander the Great (356-323BCE) whose army, local tradition holds, the Kalash are descended from (40), but these ancient events predate recorded history in the region, precluding confident interpretation.

Our results demonstrate that it is possible to elucidate the effect of ancient and modern migration events, and to provide fine-scale details of the sources involved, complexity of events, and timing of mixing of groups, using genetic information alone. Where independent information exists from alternative historical or archaeological sources, our approach provides results consistent with known facts, and determines the amount of genetic material exchanged. In other cases, novel mixture events we infer are plausible, and often involve geographically nearby sources, supporting their validity. Admixture events within the past several thousand years affect most human populations, and this needs to be taken into account in inferences aiming to look at the more distant history of our species. Future improvements in whole-genome sequencing, greater sample sizes, and incorporation of ancient DNA – together with additional methodological extensions, are likely to allow better understanding of ancient events where little or no historical record exists, to identify many additional events, to infer sex biases, and to provide more precise event characterization than currently possible. We believe our approach will extend naturally to these settings, as well as to other species.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful for The John Fell Fund-University of Oxford, the NIH, the Wellcome Trust (SM, grant 098387/Z/12/Z), the BBSRC, the Royal Society/Wellcome Trust (GH, grant 098386/Z/12/Z), and The Istituto Italiano di Antropologia for funding. J.F.W. is a director, stockholder and employee of ScotlandsDNA (and formerly of EthnoAncestry). We thank S. Karachanak, D. Toncheva, P. Anagnostou, F. Cali, F. Brisighelli, V. Romano, G. LeFranc, C. Buresi, J. Ben Chibani, A. Haj-Khelil, S. Denden, R. Ploski, T. Hervig, T. Moen, P. Krajewski and R. Herrera for providing samples for our genotyping. We thank the blood donors and the staff of the

Unità Operativa Complessa di Medicina Trasfusionale, Azienda Ospedaliera Umberto I, Siracusa (Italy). Data analyzed in this study may be downloaded via http://admixturemap.paintmychromosomes.com/.

References and Notes

1. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003; 164:1567. [PubMed: 12930761]
2. Baird SJE. Fisher's markers of admixture. Heredity. 2006; 97:81–83. [PubMed: 16773121]
3. Pool JE, Nielsen R. Inference of Historical Changes in Migration Rate From the Lengths of Migrant Tracts. Genetics. 2009; 181:711–719. [PubMed: 19087958]
4. Moorjani P, et al. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. PLoS Genet. 2011; 7:e1001373. [PubMed: 21533020]
5. Pugach I, Matveyev R, Wollstein A, Kayser M, Stoneking M. Dating the age of admixture via wavelet transform analysis of genome-wide data. Genome Biol. 2011; 12:R19. [PubMed: 21352535]
6. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 2000; 155:945–959. [PubMed: 10835412]
7. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009; 19:1655–1664. [PubMed: 19648217]
8. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of Population Structure using Dense Haplotype Data. PLoS Genet. 2012; 8:e1002453. [PubMed: 22291602]
9. Information on materials and methods is available online.
10. Cavalli Sforza, LL.; Bodmer, W. The Genetics of Human Populations. W.H. Freeman; San Francisco, USA: 1971.
11. Loh P-R, et al. Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. Genetics. 2013; 193:1233–1254. [PubMed: 23410830]
12. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 2006; 38:904–909. [PubMed: 16862161]
13. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]
14. Chen GK, Marjoram P, Wall JD. Fast and flexible simulation of DNA sequence data. Genome Res. 2009; 19:136–142. [PubMed: 19029539]
15. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. Bayesian inference of ancient human demography from individual genome sequences. Nat. Genet. 2011; 43:1031–1034. [PubMed: 21926973]
16. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011; 475:493–496. [PubMed: 21753753]
17. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLoS Genet. 2009; 5:e1000695. [PubMed: 19851460]
18. Keinan A, Mullikin JC, Patterson N, Reich D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nat. Genet. 2007; 39:1251–1255. [PubMed: 17828266]
19. Fenner JN. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am. J. Phys. Anthropol. 2005; 128:415–423. [PubMed: 15795887]
20. Henn BM, et al. Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. PLoS Genet. 2012; 8:e1002397. [PubMed: 22253600]
21. Busby GBJ, et al. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. Proc. Biol. Sci. 2012; 279:884–892. [PubMed: 21865258]
22. Li JZ, et al. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science. 2008; 319:1100–1104. [PubMed: 18292342]
23. Behar DM, et al. The genome-wide structure of the Jewish people. Nature. 2010; 466:238–242. [PubMed: 20531471]

Europe PMC Funders Author Manuscripts

24. Henn BM, et al. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc. Natl. Acad. Sci. 2011; 108:5154–5162. [PubMed: 21383195]
25. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]
26. An online atlas of worldwide human genetic admixture history (available at www.admixturemap.paintmychromosomes.com).
27. Patterson N, et al. Ancient Admixture in Human History. Genetics. 2012; 192:1065–1093. [PubMed: 22960212]
28. We convert g inferred generations to the admixture year:1950-(g+1)x28.
29. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet. 2012; 8:e1002967. [PubMed: 23166502]
30. Chandler, D. A history of Cambodia. 4th ed.. Westview Press; USA: 2007.
31. Atwood P., C. Encyclopedia of Mongolia and the Mongol Empire. Facts on File, Inc; New York, USA: 2004.
32. Bacon EE. The Inquiry into the History of the Hazara Mongols of Afghanistan. Southwest. J. Anthr. 1951; 7:230–247.
33. Zerjal T, et al. The genetic legacy of the Mongols. Am. J. Hum. Genet. 2003; 72:717–721. [PubMed: 12592608]
34. Xu S, Jin L. A genome-wide analysis of admixture in Uyghurs and a high-density admixture map for disease-gene discovery. Am. J. Hum. Genet. 2008; 83:322–336. [PubMed: 18760393]
35. Gordon, M. Slavery in the Arab world. 1st ed.. New Amsterdam Books; New York, USA: 1989.
36. Heather, P. Empires and Barbarians: migration, development and the birth of Europe. Macmillan; London, UK: 2009.
37. Ralph P, Coop G. The Geography of Recent Genetic Ancestry across Europe. PLoS Biol. 2013; 11:e1001555. [PubMed: 23667324]
38. Beckwith, CI. Empires of the Silk Road: A History of Central Eurasia from the Bronze Age to the Present. Princeton University Press; Princeton, US: 2006.
39. Rosenberg NA, et al. Genetic structure of human populations. Science. 2002; 298:2381–2385. [PubMed: 12493913]
40. Lines, M. The Kalasha people of North-western Pakistan. Emjay Books International; Peshawar, Pakistan: 1999.

Europe PMC Funders Author Manuscripts

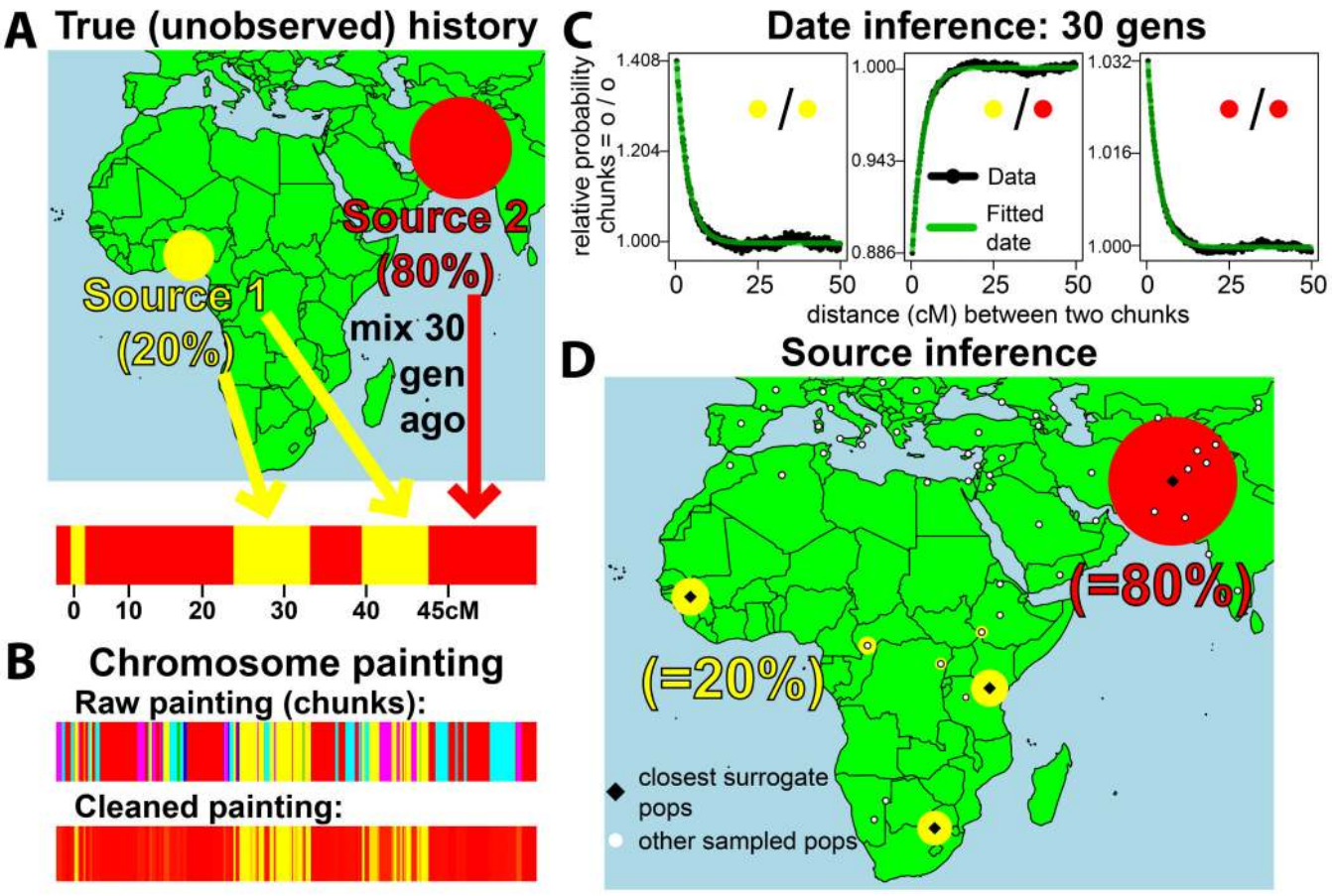

Fig. 1. Ancestry painting and admixture analysis of simulated admixture

(A) We illustrate a simulated event 30 generations ago between Brahui (80%, red) and Yoruba (20%, yellow), resulting in admixed individuals having DNA segments from each source (bottom). The true sources are then treated as unsampled (B) CHROMOPAINTER's painting of the same region (yellow=Africa, green=America, red=Central-South-Asia, blue=East-Asia, cyan=Europe, pink=Near-East, black=Oceania), showing haplotypic segments ("chunks") shared with these groups. Our model fitting narrows the donor set largely to Central-South Asia and Africa, generating a "cleaned" painting. (C) Coancestry curves (black line) show relative probability of jointly copying two chunks from red (Balochi; F_{ST}=0.003 with Brahui) and/or yellow (Mandenka; F_{ST}=0.009 with Yoruba) donors, at varying genetic distances. The curves closely fit an exponential decay (green line) with rate 30 generations (95% CI: 27-33). The positive slope for the Balochi – Mandenka curve (middle) implies these donors represent different admixing sources. (D) GLOBETROTTER's source inference, with black diamonds indicating sampled populations with greatest similarity (F_{ST}≤0.001 over minimum) to true sources, white circles other sampled populations. Red and yellow circles, with areas summing to 20% and 80%, respectively, show inferred haplotypic make-up of the two admixing sources.

Europe PMC Funders Author Manuscripts

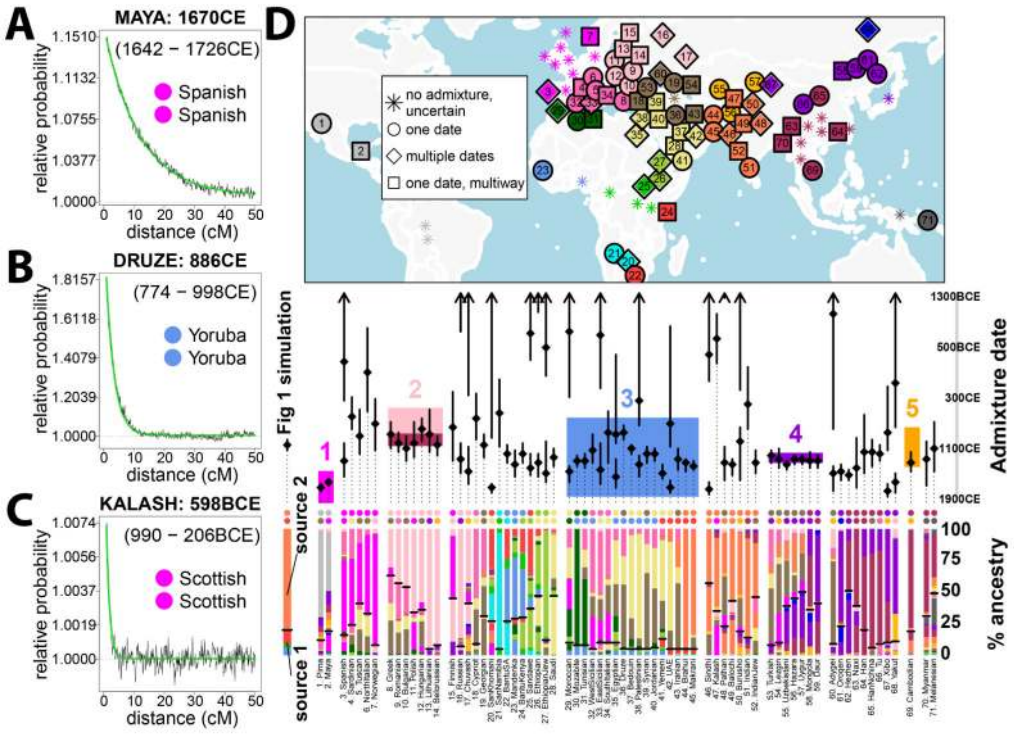

Fig. 2. Overview of inferred admixture for 95 human populations

(A) Coancestry curve for the Maya for Spanish donor group (inferred as closest to minor admixing source), with green fitted line showing inferred exponential decay curve and a corresponding recent admixture date (with 95% CI). (B, C) As A, but showing the Druze and Kalash respectively, with different indicated donors (donors indicated are proxies for minor admixing source, inferred as closest to Yoruba and Germany/Austria, respectively) and with successively older admixture. (D) On the map (locations approximate in densely sampled regions), shapes (see legend) indicate inference: no admixture, a single admixture event, or more complex admixture; colors indicate fineSTRUCTURE clustering into 18 clades (Table S11, Figs. S12-S13). Inferred date(s), 95% CIs are directly below the map, with two inferred admixing sources (dots and vertical bars) shown below each date (see example for simulation of Fig. 1 at left). For multiple admixture times, these two sources correspond to the more recent event; for multiple groups, they reflect the strongest admixture "direction". Colored dots above each bar indicate clades best representing the major (top) and minor (bottom) sources. The bar is split at the inferred admixture fraction (horizontal line, fractions < 5% shown as 5%). Each bar section indicates inferred donor group haplotypic make-up, colored as the map, for one source. Shaded boxes on the inferred admixture times denote events referred to in the text, specifically 1. European colonization of the Americas (1492CE-present; hot pink), 2. Slavic (500-900CE; pink), Turkic (500-1100CE; maroon) migrations, 3. Arab slave trade (650-1900CE; cyan), 4. Mongol Empire (1206-1368CE; purple), and 5. Khmer Empire (802-1431CE; orange).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

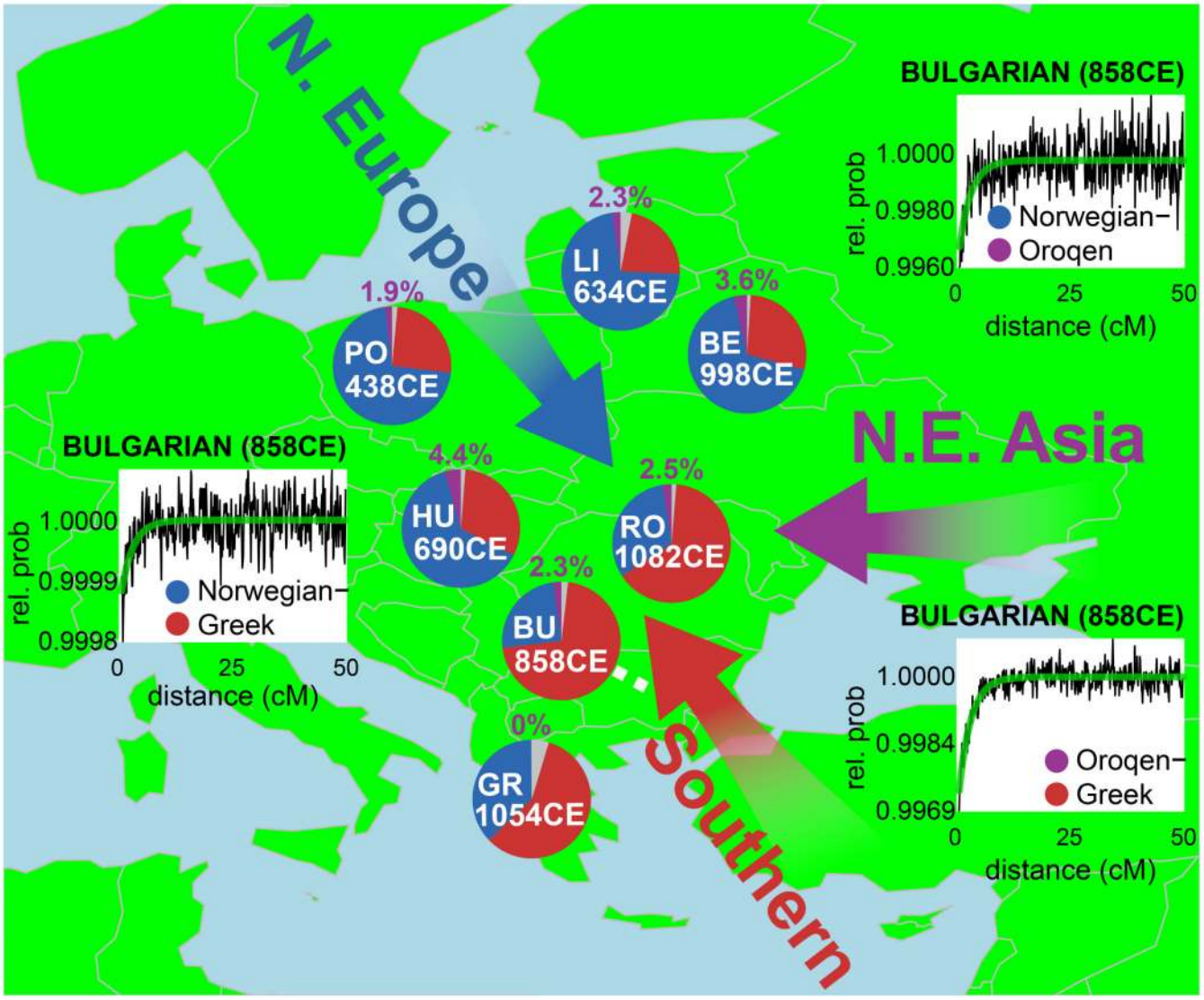

Fig. 3. Multi-way admixture in Eastern Europe

Mixing percentages (pie graphs) and dates (white text) inferred using the strongest admixture "direction" for 6 eastern European groups: Belarus (BE), Bulgaria (BU), Hungary (HU), Lithuania (LI), Poland (PO), Romania (RO), analyzed when disallowing copying from nearby groups, and Greece (GR), analyzed using the full set of 94 donors. Mixing percentages indicate percentages for three geographic regions: "N. Europe" (Northwest Europe and East Europe from clades of Table S11; blue), "Southern" (South Europe and West Asia; red) and "N.E.Asia" (Northeast Asia and Yakut; purple, also given above each pie), plus "other" (grey). All groups except Greece show evidence ($p<0.05$) of multi-way admixture involving sources along the approximate directions show by the arrows. Coancestry curves (black lines) for Bulgaria, fitted with an exponential decay curve (green lines), exemplify this multi-way signal. Each pairing of the three donor groups, each a proxy for the admixture source from a different region (Norway: N/E Europe, Oroqen: NE Asia, Greece: S. Europe and W. Asia), exhibits negative correlation (a dip) in ancestry weights at

short genetic distances, implying at least three identifiably distinct ancestral sources mixing (approximately) simultaneously (9).

Europe PMC Funders Author Manuscripts

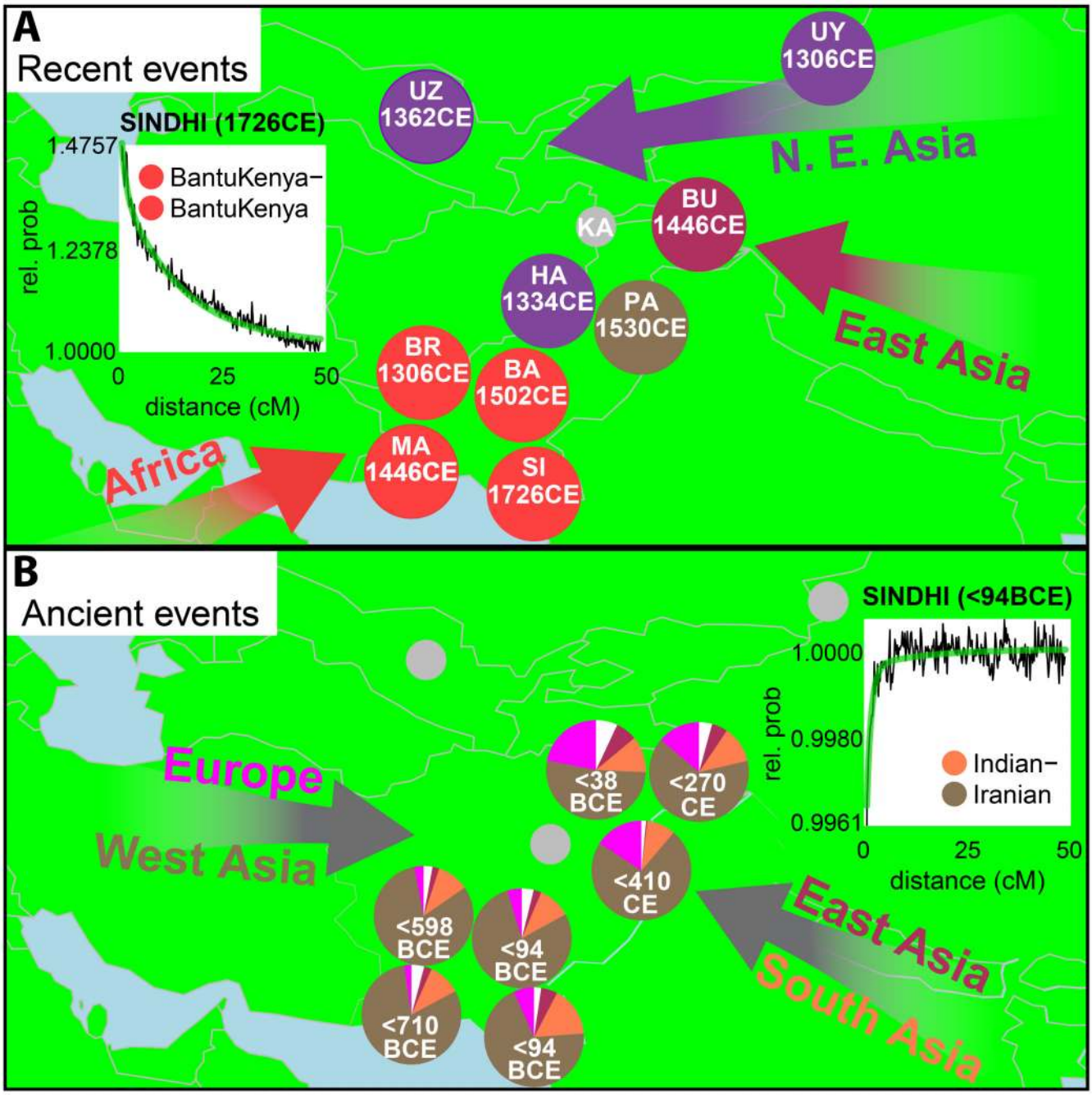

Fig. 4. Ancient and modern admixture in Central Asia

(A) Dates (white text) and minority contributing sources for recent inferred events in 9 populations (circles), analyzed disallowing copying from nearby groups, showing contributions from Northeast Asia (purple) in the Hazara (HA), Uygur (UY) and Uzbekistani (UZ); East Asia (maroon) in Burusho (BU); West Asia (brown) in Pathan (PA); Africa (red) in Balochi (BA), Brahui (BR), Makrani (MA) and Sindhi (SI). Kalash (KA, grey) have no inferred recent event. (B) Inferred mixing percentages (pie graphs) and dates (white text gives upper CI bound) for additional, possibly shared, ancient events in 7 groups.

Pie graphs show inferred donor make-up of each group after removing the recent event contribution from (A), if any, with colors referring to donors from “East Asia” (Southeast Asia from clades of Table S11; maroon), “Europe” (Northwest, East and South Europe; hot pink), Central “South Asia” (orange), “West Asia” (brown) and “other” (white). Arrows indicate “directions” of ancient admixture, with donor regions splitting into two pairs representing different sources. Coancestry curves (black lines) for Sindhi are superimposed for two different donor pairs representing proxies for admixing groups with ancestry indicated by the filled circles, indicating highly different exponential decay rates fit as a mixture of 7 and 94 generations (green lines).

Europe PMC Funders Author Manuscripts