Methods S1 - Bioinformatic pipeline, population genetics analyses and demographic inferences.

Further analyses and information to support the main text results. *Related to STAR Methods, Figures 2-4, Figures S1-4, Tables S1-5.*

Bioinformatics pipeline

On the 15 newly sequenced individuals, we performed quality control as described in the STAR Methods section "Bioinformatics pipeline".

All samples show typical post-mortem damage (PMD) patterns of elevated PMD towards the ends of the reads (Fig. M1_1A). VC3-2 contains libraries that were not UDG-treated and therefore shows higher PMD-patterns than the other samples from this study. The molecular sex was obtained for all samples with high confidence (Fig. M1_1B). For samples where the anthropological sex was available, the molecular sex matched the assignment (Table S1).

We estimated contamination using two complementary approaches: i) the rate of authentically matching mitochondrial reads to the consensus sequence at Neanderthal-specific sites with *ContamMix* (Fu *et al.*, 2013) and ii) the heterozygosity on X-chromosomal regions in males with *ANGSD* (Rasmussen *et al.*, 2011). Both methods result in generally low contamination estimates (<2% in most cases, <4% in all cases, Fig. M1_1C), with those obtained with *ContamMix* generally lower than those obtained with *ANGSD*. Of the two samples for which *ContamMix* inferred the highest contamination rates (LEPE48, Nea3), only one was male (LEPE48). For that sample, *ANGSD* inferred a similarly low contamination as for the remaining samples. The quality transformation by base quality recalibration shows that base qualities were slightly overestimated by the sequencing machine in most cases and successfully recalibrated (Fig. M1_2). To minimise reference biases, we analysed previously published samples (Table M1_1) and modern samples with the same pipeline as the newly sequenced.

The impact of subsequent filtering steps ("Data filtering" section in the STAR Methods) is shown in Fig. M1_3. The step aiming at correcting the allelic imbalance allowed us to improve the symmetry between the singleton and non-singleton sites in the ancient and modern genomes, with the exception of CarsPas1, which was excluded from all demographic analysis (Fig. M1_4).

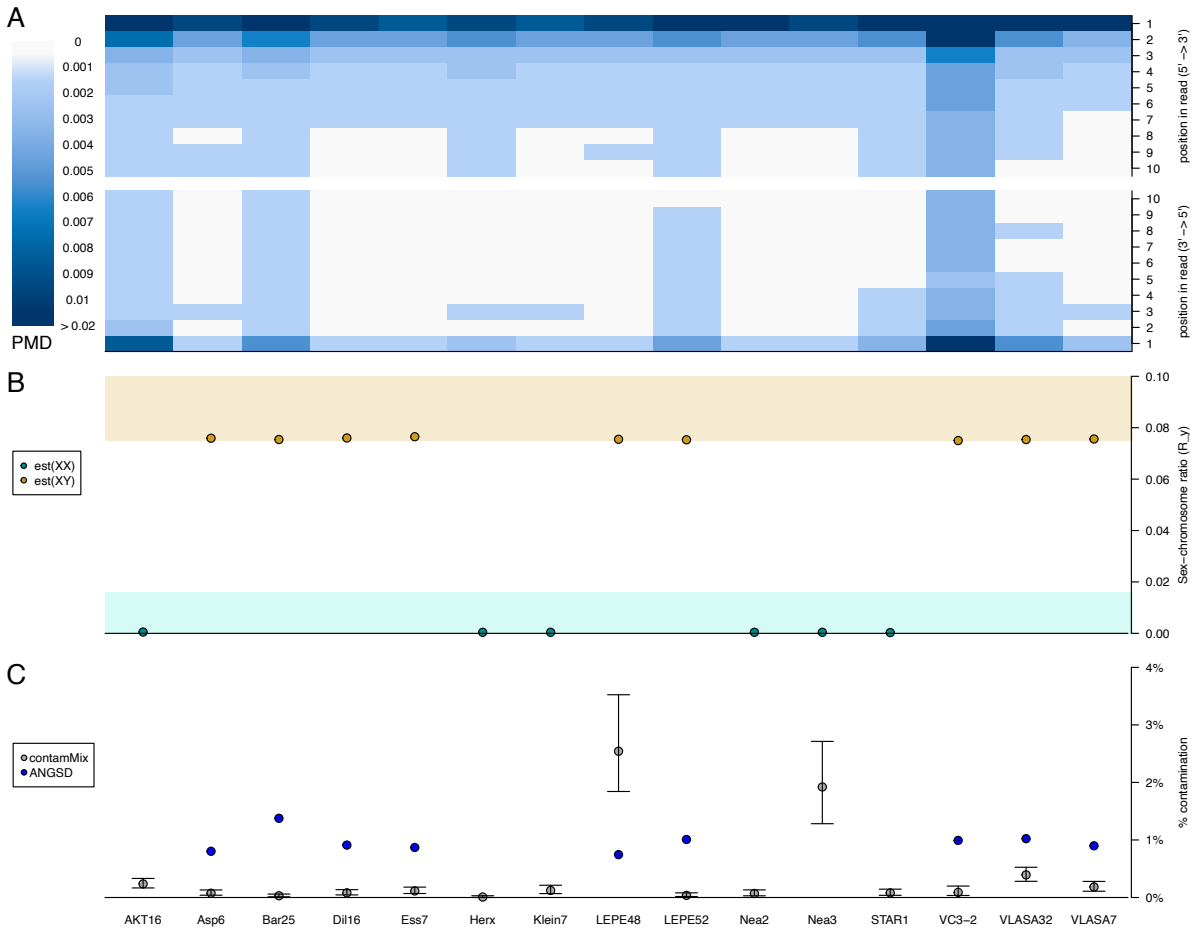

Figure M1_1 - Quality control.

(A) Estimated post mortem damage (PMD) patterns for the first 10 bp (C→ T, top) and the last 10 bp (G → A, bottom) of each read. PMD was estimated per PCR-parallel. Shown here is the average per sample weighted by read count. For VC3-2, not all libraries were treated with UDG resulting in visibly higher PMD-patterns. All samples show a pattern of elevated PMD towards the end of the reads.

(B) Inferred chromosomal sex based on the ratio of reads aligning to the X- and Y-chromosomes (R_y). A ratio of R_y < 0.016 indicates an XX genotype (green), a ratio of R_y > 0.075 indicates an XY genotype (orange). Note that error bars showing 95% confidence intervals are too small to be visible in this plot.

(C) Percentage of contamination estimated from reads aligning to the mitochondrial using *ContamMix* (grey; 1 - MapAuthentic; error bars represent 95% CI) and from the X-chromosome heterozygosity using *ANGSD* (blue; Method1, new_llh; standard errors were below 12e-12 and therefore not visible in this plot). All samples show a contamination rate below 4%.

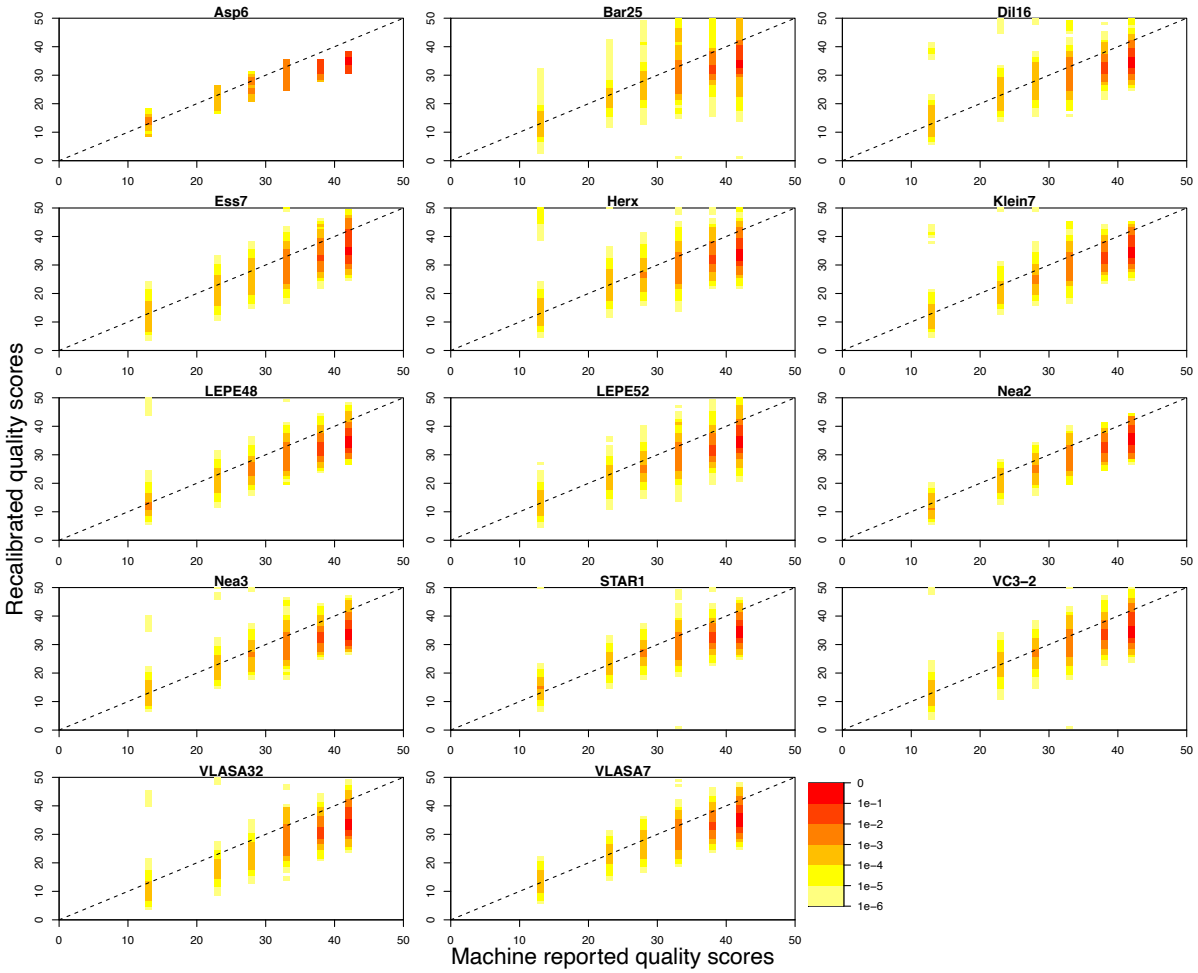

Figure M1_2 - Quality Transformation for single-end genomes. Base quality scores before and after recalibration. For all samples, base qualities were overestimated by the sequencing machines on average.

Table M1_1 - General and genetic information on the 10 ancient genomes from the literature used in our study. Palaeolithic and Mesolithic hunter-gatherers (HGs) are shown in grey, early farmers (EFs) in white. Note that the mean depth was recalculated from the pipeline described in this study.

Individual	Period (culture)	Site	Country	Age (cal. BP)	Mean Depth (X)	Genetic sex	Haplogroups mtDNA	Haplogroups Y	Phenotype (Eye; Hair)	Publication
Bichon	LUP	Grotte du Bichon	Switzerland	13,770-13,560	7.47	M	U5b1h	I2	Br; D	Jones, 2015
SF12	Meso	Stora Förvar	Sweden	9,033–8,757	51.21	F	U4a1	–	Br; D	Günther, 2018
Loschbour	LM	Heffingen	Luxembourg	8,170-7,940	22.43	M	U5b1a	I2a1b*	Bl; D	Lazaridis, 2014
KK1	Meso	Kotias Klde	Georgia	9,895-9,529	15.4	M	H13c	J	Br; D	Jones, 2015
Bon002	Neo	Boncuklu	Turkey	10,229-9,927	5.45	F	K1a	–	–	Kilinc, 2016
WC1	EN	Wezmeh Cave	Iran	9,405-9,032	12.66	M	J1d6	G2b	Bl; D	Broushaki, 2016
Bar8	EN	Barcın	Turkey	8,162-7,980	7.37	F	K1a2	–	Br; D	Hofmanova, 2016
NE1	MN (ALP)	Polgar-Ferenci-hat	Hungary	7,260-7,020	17.66	F	U5b2c	–	Br; D	Gamba, 2014
Stuttgart	Neo (LBK)	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	7,050-6,750*	15.58	F	T2c1d1	–	Br; D	Lazaridis, 2014
CarsPas1	EN	Carsington Pasture Cave, Brassington	England	5,606-5,471	10.14	M	J1c1	I2a2a1a1a	Br; D	Brace, 2019

LUP, Late Upper Palaeolithic; Meso, Mesolithic; LM, Late Mesolithic; Neo, Neolithic; EN/MN, Early/Middle Neolithic; LBK, Linearbandkeramik; ALP, Alföld Linear Pottery
Br, Brown eyes; Bl, Blue eyes; D, Dark hair (listed as "Brown or Black", "Black/Dark", "Dark" in the original publications). Samples with genetic sex determined as XX and XY are noted as F and M respectively. *approximate date based on the archaeological context

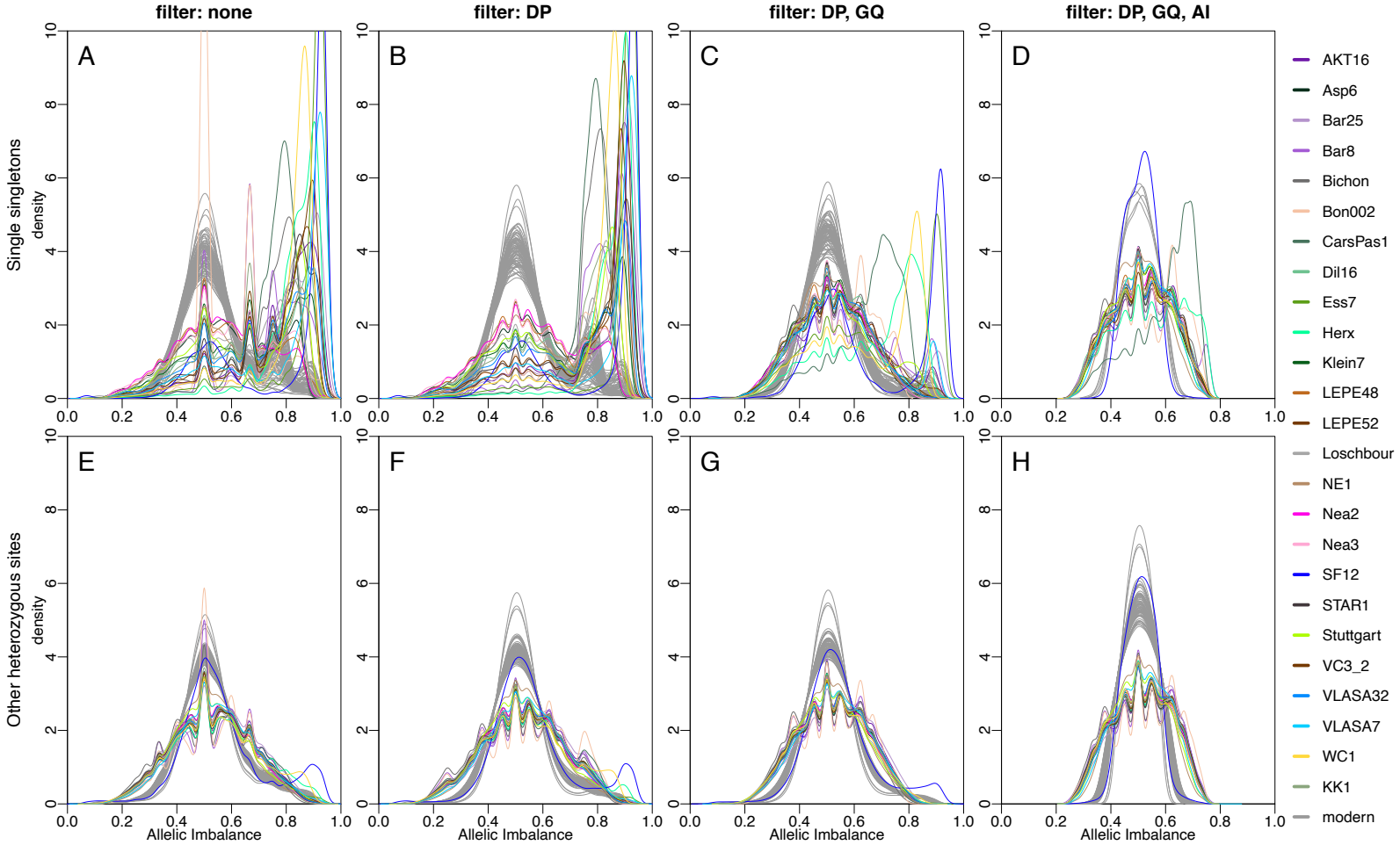

Figure M1_3 - Successive effects of filtering on allelic imbalance. Shown are the distributions of the allelic imbalance for singleton positions (top line panel) and other heterozygous positions (bottom line panel) for each sample prior to applying any filter (**A**,**E**), after removing sites with excess depth (**B**,**F**), after additionally removing calls with low genotype quality (GQ < 30, **C**,**G**) and after additionally correcting for allelic imbalance (AI > 0.1, **D**,**H**). Modern samples are shown in grey in all panels.

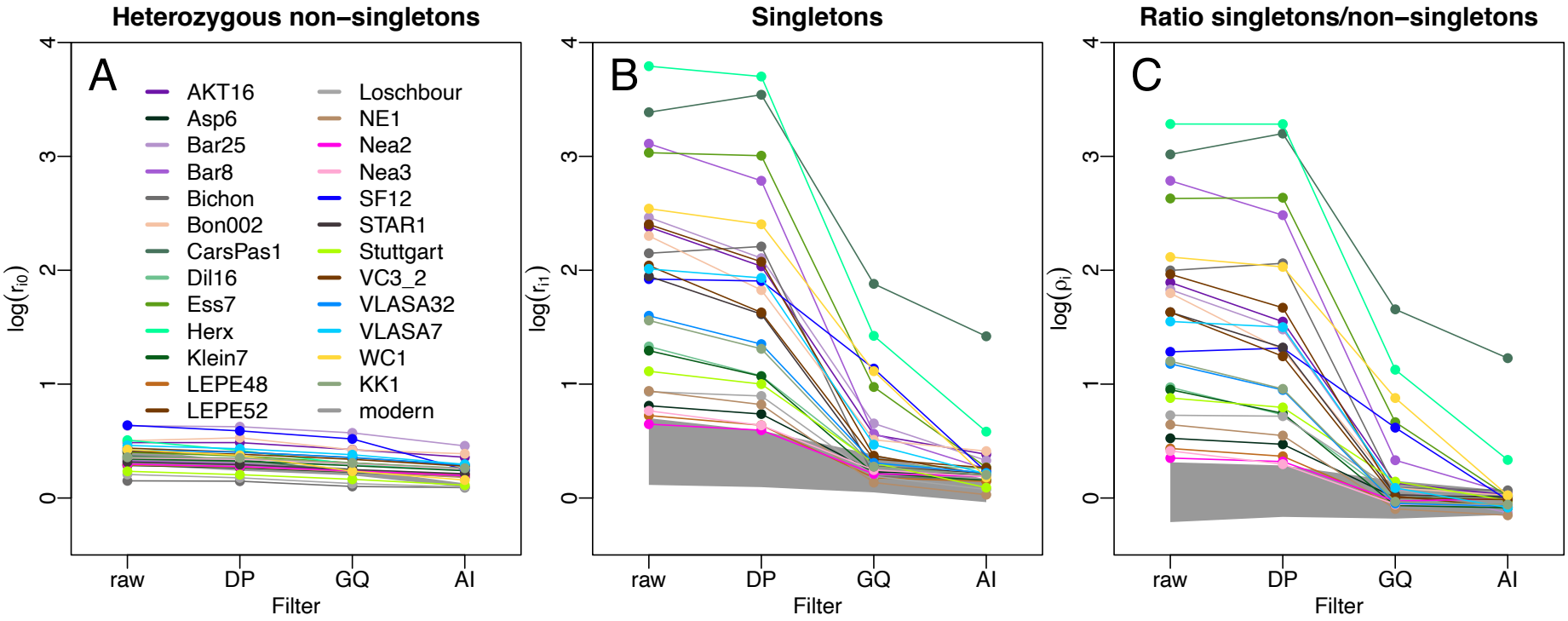

Figure M1_4 - Imbalance statistics as a function of filtering. (**A**) r_{i0} calculated for non-singleton sites, (**B**) r_{i1} calculated for singleton sites, (**C**) imbalance ratio $\rho_i = r_{i1}/r_{i0}$. The range of these statistics observed for modern samples is shown as grey shades in all panels.

Population Genetics Analyses

Genetic structure analysis

Sex-specific markers and haplogroups

Consistent with the other Mesolithic individuals from the Danube Gorges/Iron Gates published to date (González-Fortes *et al.*, 2017; Mathieson *et al.*, 2018), the two newly sequenced hunter-gatherers (HGs) from Vlasac were found to carry mitochondrial haplogroup U5 (subclade U5a2a). Their Y-chromosomal haplogroups (VLASA32: R1b1, VLASA7: I2) also fall within the known diversity as, to date, all Mesolithic individuals from this area exclusively show I or R (R1b) haplogroups (González-Fortes *et al.*, 2017; Mathieson *et al.*, 2018).

In contrast, Y-chromosomal haplogroups of the newly sequenced western EF males either belong to haplogroup G (Bar25: G2a2b2a1, Ess7: G2a2b2a1a1, LEPE52: G2a2b2a1a1c, VC3-2: G2a2a1a3~, Asp6: G2a2b2a3) or haplogroup C (Dil16 & LEPE48: C1a2b). While haplogroup G was common among early Neolithic farmers (Mathieson *et al.*, 2018), haplogroup C was less frequent in comparison (G: 56.52%, C: 17.39%, estimates based on version v42.4 of https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V42/V42.4/SHARE/public.dir/v42.4.1240K.anno).

Mitochondrial haplogroup frequencies observed in the 13 newly sequenced Neolithic individuals were consistent with those previously reported from a Neolithic context: we found five individuals with haplogroup K (Nea2: K1a, LEPE48: K1a1, Nea3: K1a2c, AKT16: K1a3, Herx: K1a4a1i), one individual each with haplogroups N1a1a1 (Bar25), J1c6 (Dil16), W1-119 (Klein7), T2e2 (STAR1) and HV-16311 (VC3-2). Furthermore, U5 haplogroups, common among HGs (Mathieson *et al.*, 2018) and increasing in frequencies during the middle and late Neolithic periods, were found in some early Neolithic individuals from Germany and Austria (Asp6: U5a1c1, Ess7: U5b2c1). Additionally, haplogroup H3 was inferred for an individual from the early-Middle Neolithic in the Danube Gorges (LEPE52); this haplogroup is rare or even absent in early Neolithic individuals but is found more frequently in the Middle Neolithic in Germany (Lipson *et al.*, 2017; Mittnik *et al.*, 2019) as well as the Iberian Peninsula (Mathieson *et al.*, 2015; Olalde *et al.*, 2018, 2019). While it was previously suggested that haplogroup H3 was associated with a glacial Iberian refugium and has spread throughout Europe from there (Brotherton *et al.*, 2013), our data indicates that H3 was also already present in Neolithic Serbia.

Multi-Dimensional Scaling (MDS) on pairwise average nucleotide divergence

Contrasting to the analyses performed only on neutrally evolving sites (Fig. 2A; *Neutral* dataset), an MDS analysis done on the whole genome including sites potentially affected by selection (Fig. M1_5A; *102samples* dataset) reveals a slightly different picture as it suggests strongest affinities of European Neolithic early farmers (EFs) with modern individuals from Southern Europe other than Sardinians (Crete, Greece, Italy, Albania, Spain), which is at odds with previous analyses based on the projections of ancient individuals on Principal Components computed from modern individuals only (Skoglund *et al.*, 2012). It also suggests some genetic continuity since at least Neolithic times as the EF from Iran (WC1) and the Caucasus HG (KK1) show strongest genetic affinities with modern Iranians and individuals from the Northern Caucasus. Note that the MDS performed on ancient genomes only (Fig. M1_5B; *Ancient* dataset) shows more differences between WC1 and KK1 than when moderns are also included.

Uniform Manifold Approximation and Projection (UMAP)

We performed a UMAP analysis on 90 European and SW Asian individuals (65 modern and 25 ancient samples) genotyped at "neutral" sites without missing data for any individual (2035 sites from the *Neutral* dataset). No clear cluster seems to emerge in this UMAP analysis (Fig. M1_6). Modern samples and EFs appear indeed completely mixed and scattered over the whole panel. Only European HGs (shown with blue and grey triangles) seem to stand out in a relatively separate cluster (even though three modern individuals strangely cluster with these HGs). This dimension reduction analysis thus seems more difficult to interpret than the MDS analysis in Fig. 2, which is consistent with previous palaeogenetic studies. In this case, it does not lead to the definition of additional sub-groups by using potentially more dimensions than in the MDS analysis. This lack of resolution might be due to the fact that UMAP clustering exaggerates the genetic variation within the most represented groups (i.e. moderns and EFs) focusing more on local than on global distances (Diaz-Papkovich *et al.*, 2019, 2020).

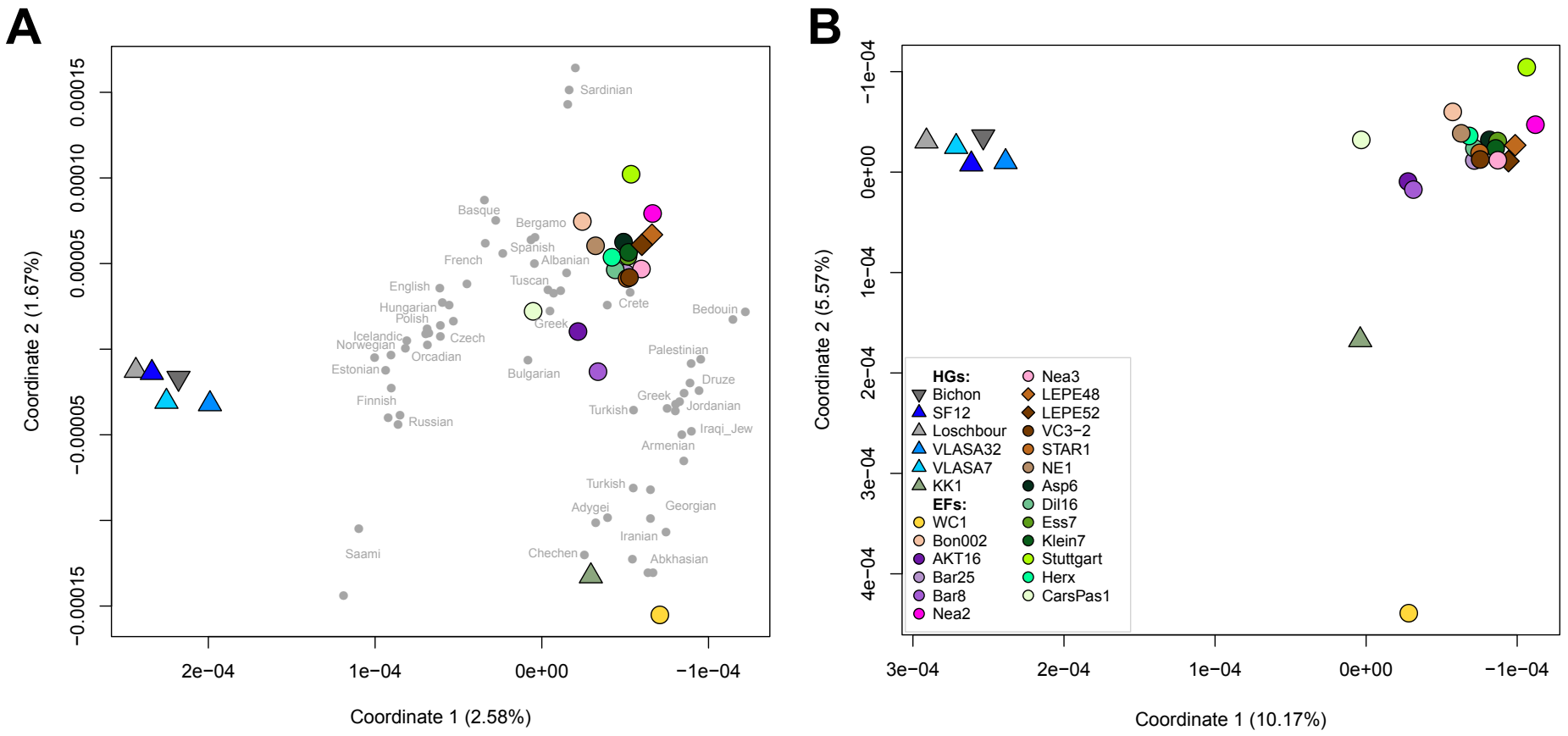

Figure M1_5 - 2D-MDS performed on the average nucleotide divergence (π_{XY}) matrix computed over the whole genome of (**A**) all Europeans and SW Asians (n = 90; *102samples* dataset), ancient and modern (shown with small circles); (**B**) ancient individuals only (n = 25; *Ancient* dataset). Distance matrices available in Table S2.

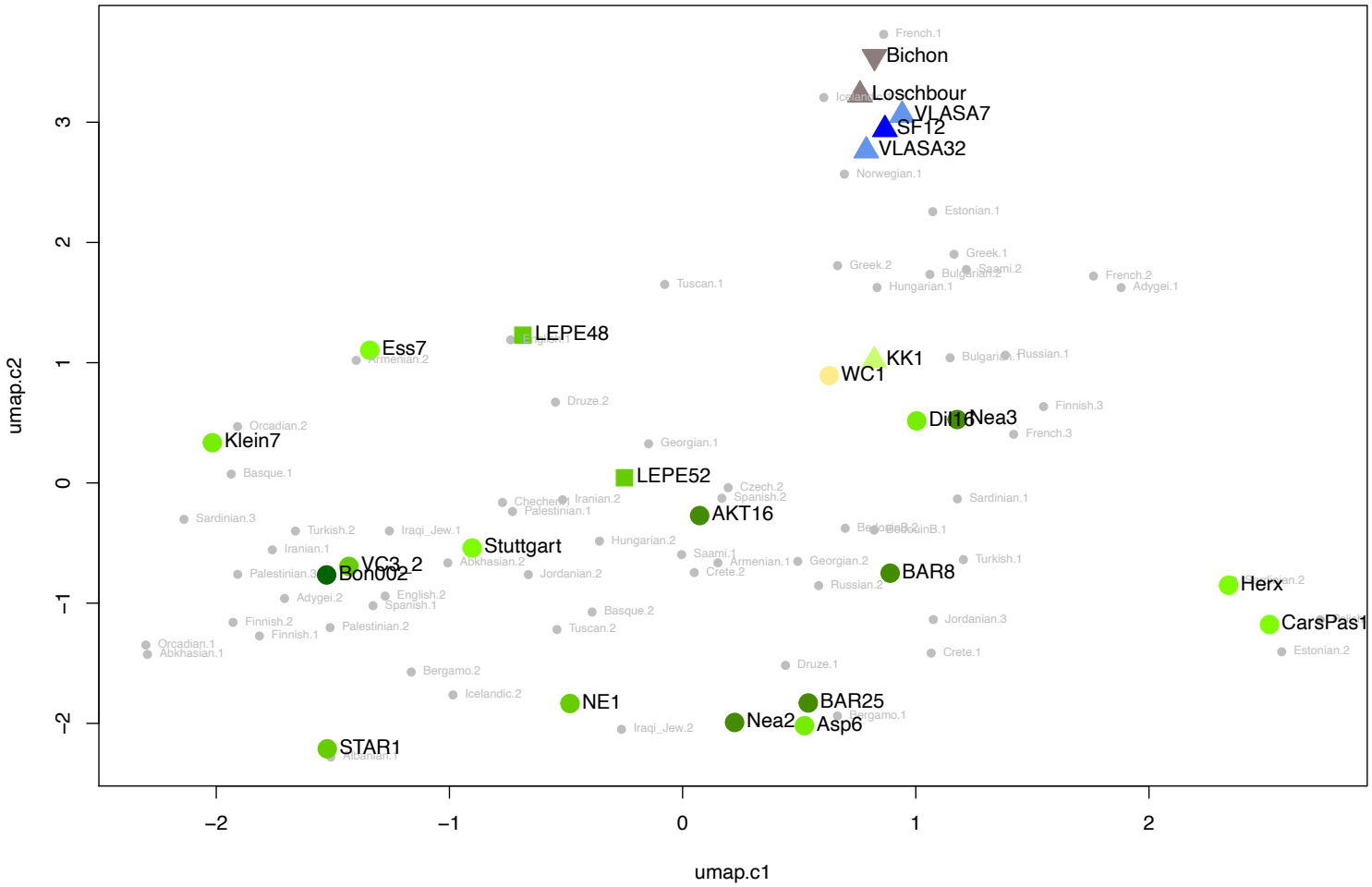

Figure M1_6 - UMAP analysis performed on 65 modern and 25 ancient individuals from Europe and SW Asia. Modern samples are shown in light grey and ancient samples with larger coloured symbols (HGs with triangles and EFs with circles).

Runs of Homozygosity (ROHs)

We detected intermediate (2-10 Mb) and long (>10 Mb) segments of homozygosity-by-descent or Runs of Homozygosity (ROHs) from the imputed genomes of the European and SW Asian modern and ancient individuals (Fig. 2D, Fig. M1_7). The number and total genome size made up of the intermediate ROHs (Fig. M1_7A) are high in ancient genomes, especially those of HGs in line with previous results (Ceballos *et al.*, 2021; Kılınç *et al.*, 2016; Ringbauer *et al.*, 2021) and of a few EFs (Bon002, WC1 and LEPE52; STAR1 and AKT16 to a lesser extent), as well as of some modern individuals. These ROHs are expected to be negatively correlated with the population size, so that a larger number of intermediate ROHs indicate remote inbreeding due to background relatedness and potentially smaller effective population sizes. We indeed found a negative correlation between the cumulative length of the intermediate ROHs and the expected neutral heterozygosity observed in ancient genomes (Fig. M1_7C; Spearman's rho test = -0.5, p-value < 0.01).

For the long ROHs indicative of recent inbreeding between close relatives (potentially second cousins or closer; Gazal *et al.*, 2014), some modern individuals present the larger amount and some ancient genomes to a lesser degree (WC1 and LEPE52 are inferred to be the most consanguineous ancient individuals, followed by Stuttgart, Bichon and Loschbour). The three other HGs were not found to have long ROHs nor Bon002 and AKT16. For modern individuals, most of the consanguineous individuals have both intermediate and long ROHs but not necessarily (e.g. Turkish-2 and Russian-1 exhibit long ROHs but only a few intermediate ROHs, respectively 3 and 0 Mb in intermediate ROHs).

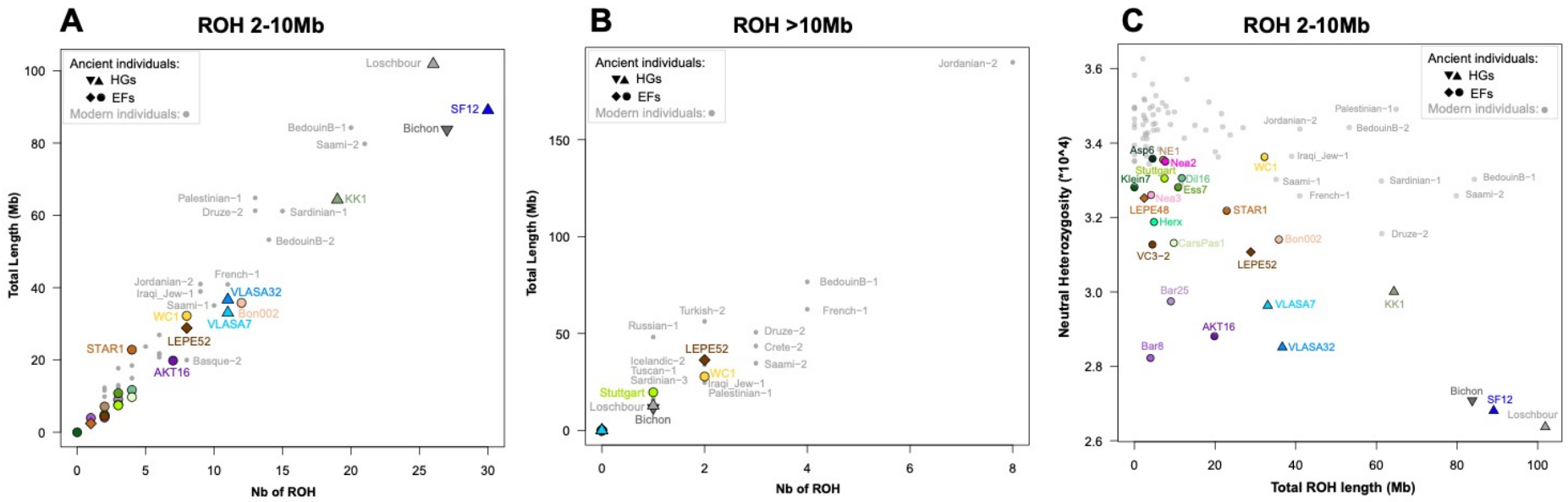

Figure M1_7 - ROHs found in the imputed European and SW Asian genomes (n = 90). Number and cumulated length of (**A**) intermediate ROHs (size between 2 and 10Mb) and (**B**) long ROHs (>10Mb); (**C**) Neutral expected heterozygosity as a function of the cumulative length of intermediate ROHs.

Neanderthal introgression

As a result of interbreeding between Neanderthals and modern humans some time after the out-of-Africa migration, all non-African individuals derive a small portion of their genome from Neanderthals. We estimated the Neanderthal ancestry proportions in the ancient individuals used in *fastsimcoal2* demographic modelling (Fig. M1_8) as described in the STAR Methods. Besides elevated values in Bon002, Bichon, and Bar8, proportions are approximately 4%. This may be expected as all analysed individuals share a common ancestor ~25kya and are linked by a number of admixture events. The high ancestry proportions inferred in Bon002, Bichon, and Bar8 are likely artefacts of lower quality genomes. In line with previous results (Coll Macià *et al.*, 2021; Petr *et al.*, 2019), we find no significant difference in Neanderthal ancestry levels between ancient HGs and EFs (permutation test p-value = 0.711).

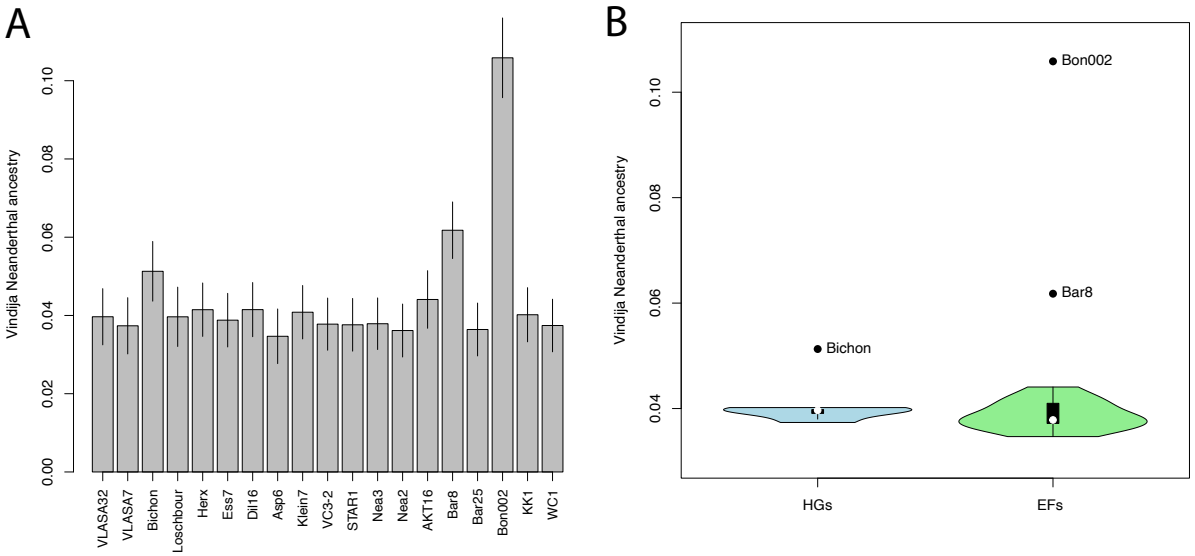

Figure M1_8 - Vindija Neanderthal ancestry proportions (**A**) inferred for the 19 ancient individuals used in the *fastsimcoal2* demographic modelling. (**B**) Comparison between hunter-gatherers (HGs) and early farmers (EFs).

Phenotypic analysis

Pigmentation

With the *HIrisPlex-S* webtool (Chaitanya *et al.*, 2018; Walsh *et al.*, 2013), we managed to predict pigmentation phenotypes for the 15 newly sequenced and the previously published ancient individuals (except for the eye colour in VLASA32 for which no allele at the SNP rs12913832 could be retrieved) (Fig. M1_9, Table S3).

For some individuals, one or several SNPs in the *MC1R* and *TUBB3* genes associated with hair colour and skin pigmentation (Latreille *et al.*, 2009; Yamaguchi *et al.*, 2012) were missing in the *Ancient* dataset. As described in the STAR Methods, we considered the genotypes for these SNPs to be either the allele found directly in the BAM files and the non-effect allele (0 in *HIrisPlex-S* input file) or the found allele and the effect allele (1 in *HIrisPlex-S* input file). This second approach resulted in some predictions where red hair was the most likely outcome (Dil16, Ess7, Herx, Klein7). Given the overall low frequency of the derived alleles associated with these pigmentation phenotypes in European populations (Cavalli-Sforza *et al.*, 2004; Harding *et al.*, 2000) as well as in ancient data (Mathieson *et al.*, 2015), a red haired phenotype is highly unlikely and we consider it as an artefact (Fig. M1_9).

We found that the vast majority of the newly sequenced EFs most likely had an intermediate to light skin complexion, while the two Serbian Mesolithic HGs were inferred to have a darker skin tone in comparison (Fig. M1_9) similar to the HG from Bichon and the Iranian EF (WC1). A dark (brown to black) hair colour was inferred for all newly sequenced individuals except two EFs: LEPE52 and VC3-2, for which a light brown phenotype was more likely. Eye colour variation was similarly low, with the majority of newly sequenced individuals showing highest probabilities for brown eyes, except for two EFs associated with the Starčevo culture (STAR1 and VC3-2) who were likely blue-eyed. Light/blue-eyed phenotypes were otherwise only predicted for the HGs from Loschbour and Scandinavia, in accordance with the original publications (Günther *et al.*, 2018; Lazaridis *et al.*, 2014). Thus, the highest phenotypic variation within the newly sequenced samples seems to originate from Serbian individuals.

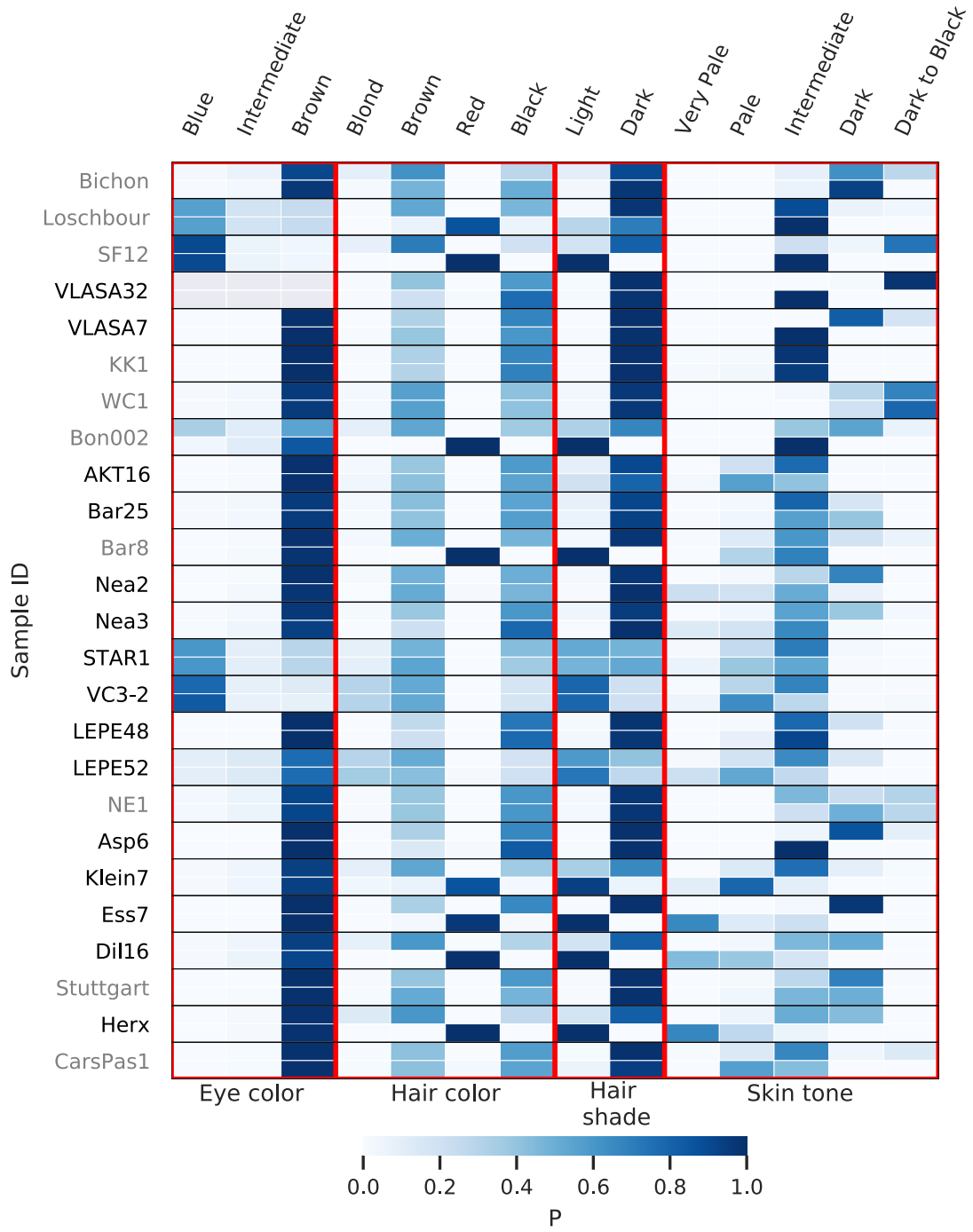

Figure M1_9 - Probabilities for eye, hair and skin pigmentation phenotypes inferred using the *HIrisPlex-S* webtool for newly sequenced (black) and previously published (grey) individuals. For each individual, the results of both runs performed to account for missing genotypes are shown (top line: BAM allele + non-effect allele, bottom line: BAM allele + effect allele). In all cases where red hair was the most likely outcome, one or several SNPs associated with hair colour genotypes could not be called properly. Individuals are grouped by subsistence and date, with the six hunter-gatherers (Bichon, Loschbour, SF12, VLASA32, VLASA7, KK1) on top.

Additional phenotypic variation

Lactase persistence

All individuals newly sequenced in this study had the ancestral G allele at rs4988235, the variant in the *MCM6* gene of which the derived allele is highly associated with lactase persistence into adulthood in Eurasia (Enattah *et al.*, 2002). Since no derived alleles were found, it is unlikely that any of the individuals was able to digest lactose, consistent with an increase in frequency of the lactase-persistence allele at a later stage (Burger *et al.*, 2020).

EDAR/ABCC11

We analysed two SNPs associated with incisor shape/hair thickness (rs3827760, *EDAR*; derived alleles are associated with thick straight hair and shovel shaped incisors (Park *et al.*, 2012)) and earwax type/body odour (rs17822931, *ABCC11*; derived alleles at rs17822931 are associated with dry earwax and reduced body odour (Shang *et al.*, 2013)), both having high derived allele frequencies in modern East-Asian populations (Fujimoto *et al.*, 2008; Ohashi *et al.*, 2010; Xue *et al.*, 2009). All ancient individuals newly sequenced in our study only had ancestral alleles for both SNPs.

Fatty acids synthesis

We further investigated seven SNPs located in the *FADS1/2* gene complex, reported to have been selected for in various populations (Buckley *et al.*, 2017; Fumagalli *et al.*, 2015). Variation in *FADS1/2* is known to influence the ability to synthesise omega-3 and omega-6 long-chain polyunsaturated fatty acids (LC-PUFAs) from precursor molecules (Ameur *et al.*, 2012; Buckley *et al.*, 2017). Since LC-PUFAs are also found in food, especially from animal sources, the dietary shift during the Neolithic transition is hypothesised to have created a selective pressure that lowered diversity at *FADS* associated loci.

Individuals were grouped by subsistence into HGs and EFs (excluding WC1, KK1 and the Lepenski Vir individuals, see Table M1_2). A two-sided binomial test was used to compare counts of derived alleles in the ancient individuals with frequencies estimated in the CEU population from phase3 of the 1000 genomes project (1000 Genomes Project Consortium *et al.*, 2015), chosen as a proxy for Central European modern populations.

For rs74771917, while the derived T allele is almost lost in CEU (0.030), it is found in significantly larger numbers in farmers ($3/26 = 0.115$, $p<0.043$). No differences could be found between CEU and EFs for any of the other SNPs. For HGs, the allele counts suggest lower

frequencies for the SNPs rs174546_C (f_{CEU} = 0.641; HGs: 1/10 = 0.100, p<0.001), while higher frequencies in comparison to CEU are inferred for rs174570_T (f_{CEU} = 0.162; HGs: 3/4 = 0.750, p<0.015) and rs174594_A (f_{CEU} = 0.616, HGs: 10/10 = 1.000, p<0.009). For rs174594_A, counts suggest a higher frequency in HGs compared to EFs (EFs: 17/28 = 0.607, HGs: 10/10 = 1.000, p<0.008). A similar result was found for rs97384_C (EFs: 12/26 = 0.462, HGs: 7/8 = 0.875, p<0.028).

These results are consistent with selection affecting variation in *FADS1/2* already during the early Neolithic (Mathieson *et al.*, 2015), rather than starting later in the Bronze Age as proposed more recently (Buckley *et al.*, 2017). This is because, with the exception of one SNP, frequencies in EFs are close to those in modern populations, but multiple SNPs differ in frequencies between HGs and modern populations. While it is possible that migration and admixture during and after the Late Neolithic led to a shift in frequencies at *FADS* related loci, our observations are also consistent with selective pressure associated with the Neolithic diet having shaped allele frequencies in early Neolithic populations to approach the levels observed in modern populations today.

Table M1_2 - Derived allele frequencies for seven SNPs located in the FADS1/2 gene complex for modern Europeans (CEU) and counts for the derived and total number of alleles for the ancient samples, grouped by subsistence (n_{HGs} = 5, n_{EFs} = 16).

SNP	Derived Allele	CEU derived frequency	No. of derived alleles	
			EFs	HGs
rs174546	C	0.641	12/24	1/10
rs174570	T	0.162	8/28	3/4
rs174594	A	0.616	17/28	10/10
rs97384	C	0.606	12/26	7/8
rs74771917	T	0.030	3/26	1/8
rs174455	A	0.647	13/20	5/10
rs174465	T	0.702	7/12	4/8

Polygenic Score for Height

Height is a classical polygenic trait, which is highly heritable, but also strongly influenced by environmental variables (McEvoy and Visscher, 2009). Hundreds of variants are known to influence height in humans, with the majority alleles having only a minor effect (Marouli *et al.*, 2017). To measure the distribution of height-related alleles, we computed generalised polygenic scores (PS) for standing height using a set of 670 SNPs (Chan *et al.*, 2015) for 20 ancient individuals (Fig. M1_10).

The highest scores were found for the two individuals from Vlasac, Serbia (VLASA32: 20.033, VLASA7: 20.147), while the lowest score was obtained for the individual from Herxheim, Germany (Herx: 18.086). We found a strong positive correlation of individual height PS with the average age of the samples (Pearson's r = 0.77577, p<6e-05). The correlation remained strong when considering only EFs between 8,300 and 7,000 BP, both with (Pearson's r = 0.6537, p<0.0082) and without the Lepenski Vir individuals (Pearson's r: 0.6565, p<0.0148), suggesting that decreasing standing height was under selection during the Neolithic expansion along the Danubian corridor.

These findings are consistent with previous studies reporting selection for decreasing predicted height in early Neolithic populations from Southern and Central Europe (Mathieson *et al.*, 2015) as well as modern populations from Sardinia (Zoledziewska *et al.*, 2015), but differ from a morphometric analysis of prehistoric skeletons (Rosenstock *et al.*, 2019). Considering the high amount of ancestry attributed to EFs in modern Sardinians (Lazaridis *et al.*, 2014), these signals may be related. Although we find significant differences between mean scores for Mesolithic HGs and Neolithic EFs including Lepenski Vir (Student t-test, t = 2.8257, p<0.0112), these differences could be explained by the older date (except for Loschbour and WC1) of the Mesolithic HGs in comparison. The decrease in height PS values could be a continuation of a trend of decreasing stature between the Upper Palaeolithic and the Mesolithic (Cox *et al.*, 2019), a result supported both by genetics and physical measurements; however, (Cox *et al.*, 2019) did not find any differences between Mesolithic and Neolithic individuals.

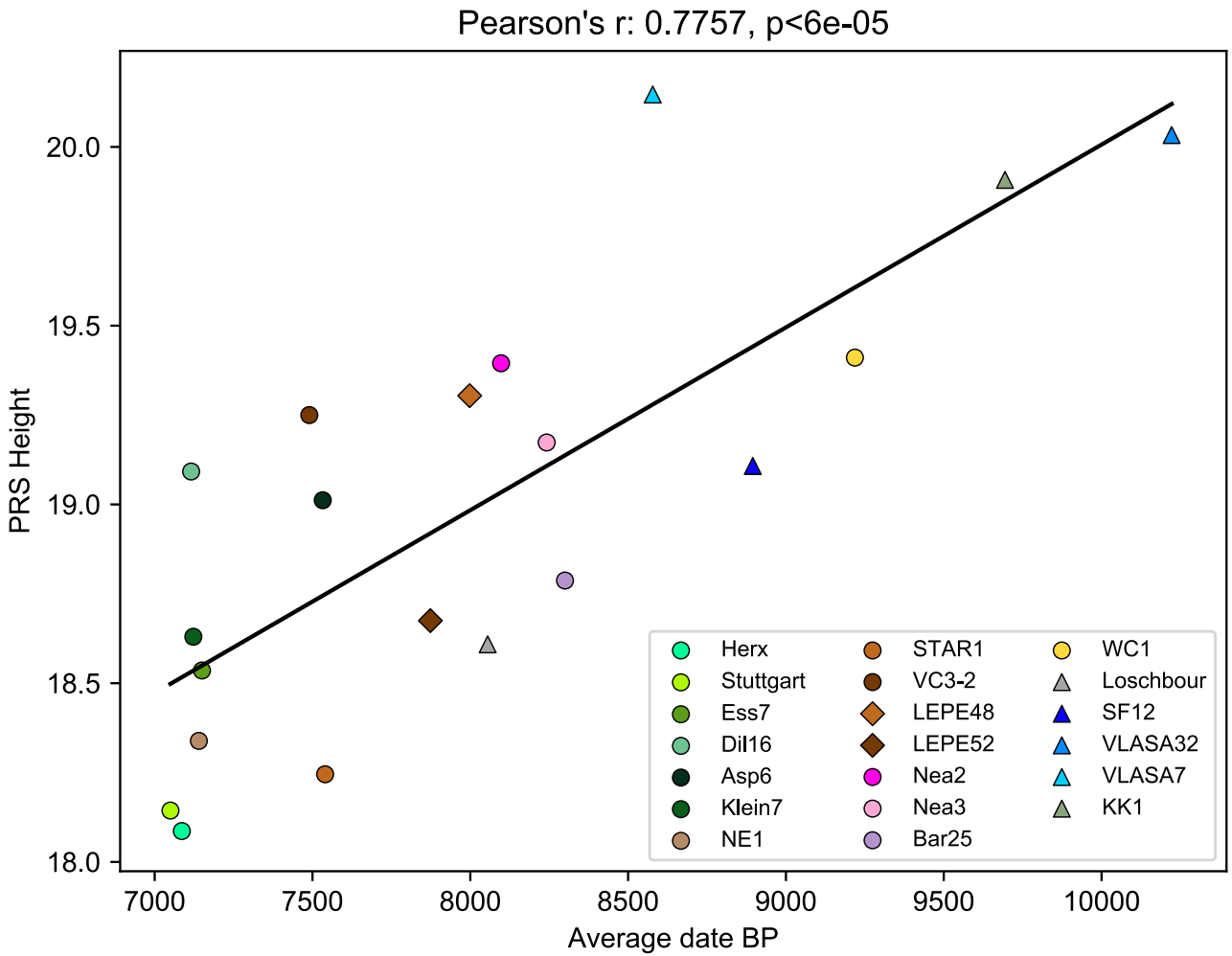

Figure M1_10 - Scatter plot of the average age in BP *vs* height PS values for each individual (n = 20; Bichon, Bon002, Bar8, CarsPas1, AKT16 being excluded based on genotype-filtering thresholds). The best-fit straight line illustrates the Pearson correlation ($r^2 = 0.6018$) between height PS and sample age, indicating a decline in PS values over time. HGs are represented with triangles, EFs with circles or diamonds.

Joint Distribution of Fitness Effects (DFE) analysis

Selective constraints may differ between Neolithic and modern populations, due to changes in cultural and environmental contexts over time. To test for such differences in purifying selection, we fit a model of the joint distribution of fitness effects (DFE) of deleterious mutations to our data using the *dadi* software (Gutenkunst *et al.*, 2009; Huang *et al.*, 2021). To facilitate modelling selection, we first fit a simple demographic model to our synonymous mutation data in which the ancestors of modern Europeans diverged from the ancestors of our Neolithic samples (Fig. M1_11A). In our joint DFE model, the fitness effects of all mutations were allowed to change after this divergence, under a joint bivariate lognormal distribution (Fig. M1_11E). When we fit this model to our nonsynonymous mutation data, our best-fit estimate of the correlation between selection coefficients in ancient versus modern populations was 0.997. We thus conclude that selective constraint has changed little in European populations between the Neolithic era and today.

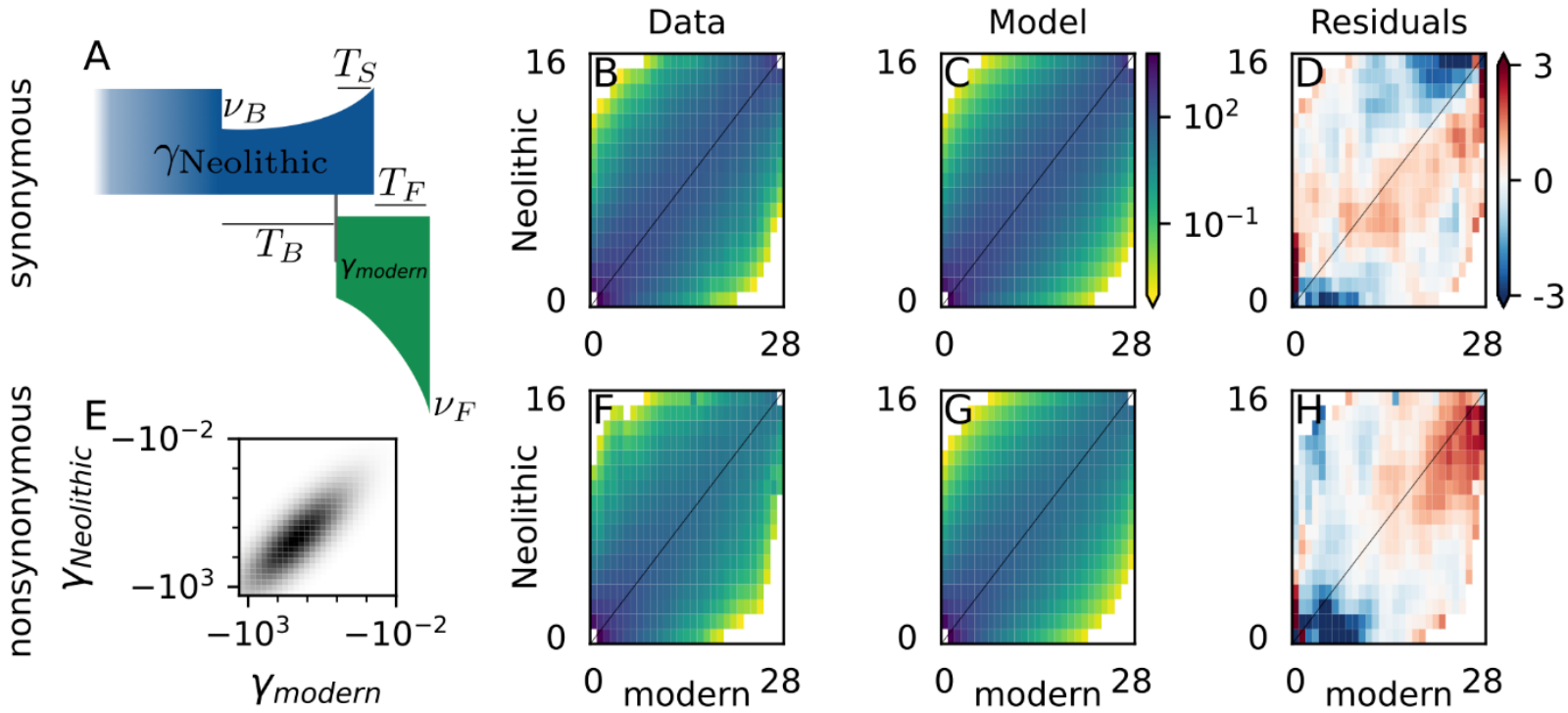

Figure M1_11 - Joint Distribution of Fitness Effects (DFE) analysis.

(**A)** The demographic model that was fit to the synonymous data. In this model, the population from which the Neolithic genomes were sampled went through a bottleneck to relative size n_B and then grew exponentially. The population from which the modern genomes were sampled diverged from that Neolithic population after T_B time units and grew at the same rate to reach its final size n_F. The population from which Neolithic genomes were sampled diverged for T_S time units until genomes were sampled T_F time units ago.

(**B-D)** The synonymous data and best-fit model allele frequency spectra, and the residuals between them.

(**E)** Illustration of the bivariate lognormal DFE that was fit to the nonsynonymous data (shown with correlation 0.8).

(**F-H)** The nonsynonymous data and best-fit model spectra, and the Anscombe residuals between them. These residuals represent the difference of model minus data, scaled to account for heteroscedasticity across the frequency spectrum

Demographic Analyses

MSMC2 analysis

Inference of human population size

The effective population size was estimated per sample (Fig. M1_12) and per population (Fig. S2). The *MSMC2* results were similar for all non-African populations before 35 kya and included a deep bottleneck at roughly 50-60 kya, a pattern that has been previously reported (Malaspinas *et al.*, 2016). As expected, we were not able to get good recent estimates from a single diploid individual, although a clear difference in the population size between EFs and HGs is observable. The resolution at recent times was much improved when using two individuals per population (Fig. S2), in which case we found that all the HG populations and the population from Iran (WC1) exhibit a small effective population size of roughly N_e = 3,000 up to 35 kya. In contrast, we estimated a constant increase of population size after 35 kya in all western EF populations. This increase resulted in population sizes around N_e = 10,000 - 15,000 about 10 kya, with NW Anatolia showing the smallest population size, Northern Greece and Serbia showing similar sizes also around N_e = 10,000 and smaller than the Central Europe EF populations (Austria, Germany1, Germany2).

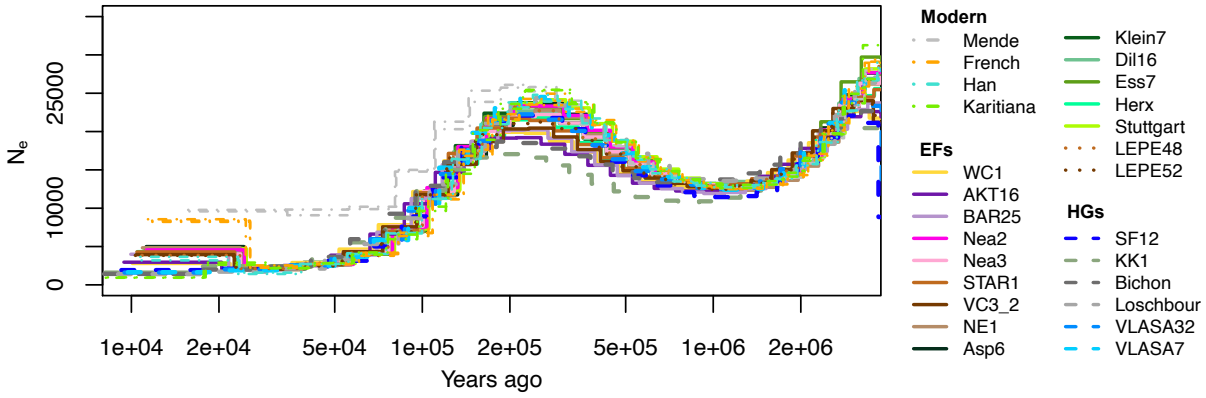

Figure M1_12 - *MSMC2* population size scaled using a mutation rate of 1.25×10^{-8} per generation per site and a generation time of 29 years. Past demography obtained for single individuals; n_{pop} = 30. The analysis suggests smaller population sizes in the most recent times for HGs compared to EFs.

Table M1_3 - Samples considered in the *MSMC2* analyses (n = 30).

Population	Period	Samples
Western Africa	Modern	Mende-1, Mende-2
Western Europe	Modern	French-1, French-2
Eastern Asia	Modern	Han-1, Han-2
Southern America	Modern	Karitiana-1, Karitiana-2
Iran	Neolithic	WC1
NW Anatolia	Neolithic	AKT16, Bar25
Northern Greece	Neolithic	Nea2, Nea3
Serbia1	Neolithic	STAR1, VC3-2
Hungary	Neolithic	NE1
Austria	Neolithic	Asp6, Klein7
Germany1	Neolithic	Dil16, Ess7
Germany2	Neolithic	Herx, Stuttgart
Serbia2	Transformational - Neolithic	LEPE48, LEPE52
Scandinavia HG	Mesolithic	SF12
Caucasus HG	Late Mesolithic	KK1
West 1 HG	Upper Palaeolithic - Mesolithic	Bichon, Loschbour
West 2 HG	Mesolithic	VLASA7, VLASA32

Divergence time between populations

The relative cross coalescent rate (CRR) was also calculated to estimate roughly the split times between populations (Fig. M1_13). As a first observation, we note that split time estimates involving populations with a single sample (Scandinavia HG, Caucasus HG, Iran and Hungary) tend to be much older. Indeed, CCR estimates from one-sample populations drop to zero more quickly than corresponding estimates from two-sample populations among Neolithic and Mesolithic populations. We therefore caution against interpreting the older split time estimates of Scandinavia HG from all other populations as strong evidence for the influence of Eastern HGs, which diverged from Western European HGs at an earlier period.

Focusing on two-sample EF populations, the estimated split times capture the old split time between HG and EF populations, and also show that all EF populations split at roughly the same time. However, no clear order of split times emerges among the EF populations, with large variations depending on the exact comparison. Among the most consistent results across several comparisons is a relatively old split of NW Anatolia, followed by populations from what is today Greece, the Balkans and finally Austria/Germany, in accordance with a stepwise migration between neighbouring regions. Serbia2 for instance, split first from NW Anatolia and Northern Greece (~24 kya), then from Serbia1 (~21 kya) and from the Austro-German Neolithic populations (~19-22 kya). But we note that several individual estimates are at odds

with this sequence of events: NW Anatolia, for instance, was inferred to have split from Austria and Serbia2 (~24 kya) earlier than from Northern Greece and Serbia (~22 kya). Similarly, Northern Greece split from NW Anatolia earlier than from Serbia and Germany1 (~21 kya). Our CCR results may thus not provide sufficient resolution at this scale. In addition, they are likely affected by recent admixture. The relatively old split time we estimated between NW Anatolia and the other Neolithic populations, along with the more recent split time we estimated between NW Anatolia and the two HG populations (West 1 and West 2 HG), might be well explained by recent HG admixture into NW Anatolia, in particular AKT16 (Fig. S1).

Interestingly, the split time between West 1 HG and West 2 HG is estimated to be the most recent (~15 kya) among all populations, suggesting they share more recent common ancestors. While this observation is in line with their clustering in our MDS analysis (Fig. M1_5 and Fig. 2A), it is somewhat younger than estimates obtained with *fastsimcoal2* (see below), and potentially a result of the very low estimates of their recent population sizes (and hence low intra-coalescent rates) and the uncertainty associated with cross-coalescent rates.

In line with the final demographic scenario inferred from *fastsimcoal2*, *MSMC2* infers split times of ~22 kya among EF populations and ~35 kya between EF and HG samples (Fig. M1_14). We note, however, that Bichon and Loschbour were estimated to have split from the Serbia HGs more recently (~15 kya) than what is inferred from the site frequency spectrum (23.3 kya). These estimates obtained from the analysis of whole genomes are older than what is inferred with *fastsimcoal2* on the neutrally evolving sites only, but are not incompatible given that they assume no gene flow or admixture between samples. Previous analyses have also shown that while *MSMC* and SFS-based inference programs were giving very congruent results on simulated data, they could lead to inconsistent results when applied to real genomes (Beichman *et al.*, 2017), implying that the two approaches were differentially affected by genomic factors not included in their assumptions.

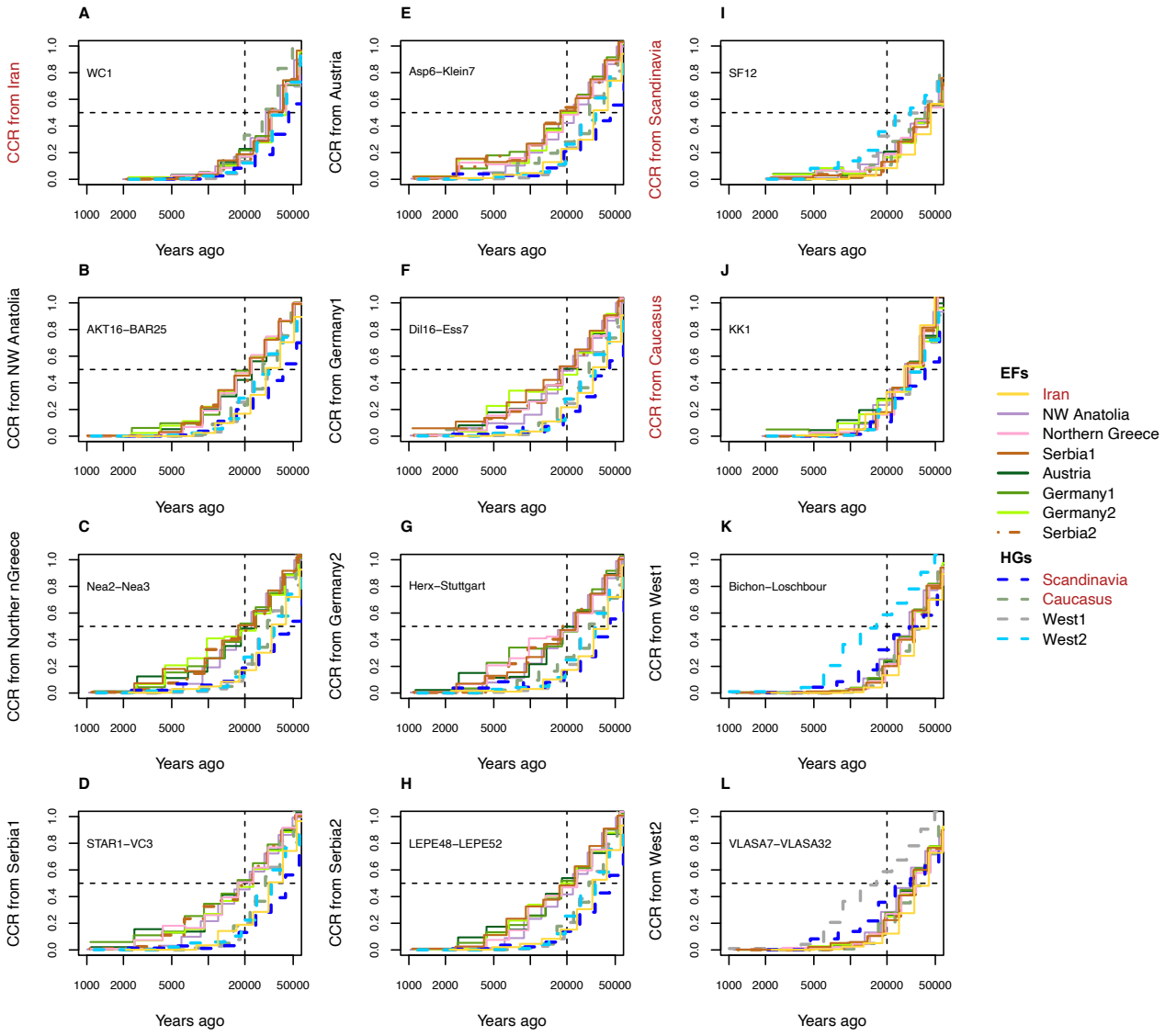

Figure M1_13 - Relative cross coalescence rate (CCR) estimated with *MSMC2* using pairs of individuals per population when available (WC1, SF12 and KK1 are the unique representative of one population each). We estimated the relative CCR for all possible pairwise group combinations. The results were scaled using a mutation rate of 1.25×10^{-8} per generation per site and a generation time of 29 years. Each subplot shows the pairwise relative CCR estimates of one population against all others; **first two columns** (A - G) EF populations from the Neolithic period and (H) transitional and Neolithic layers at Lepenski Vir in the Danube Gorges; **third column** (I - L) HG populations from the Upper Palaeo-Mesolithic period.

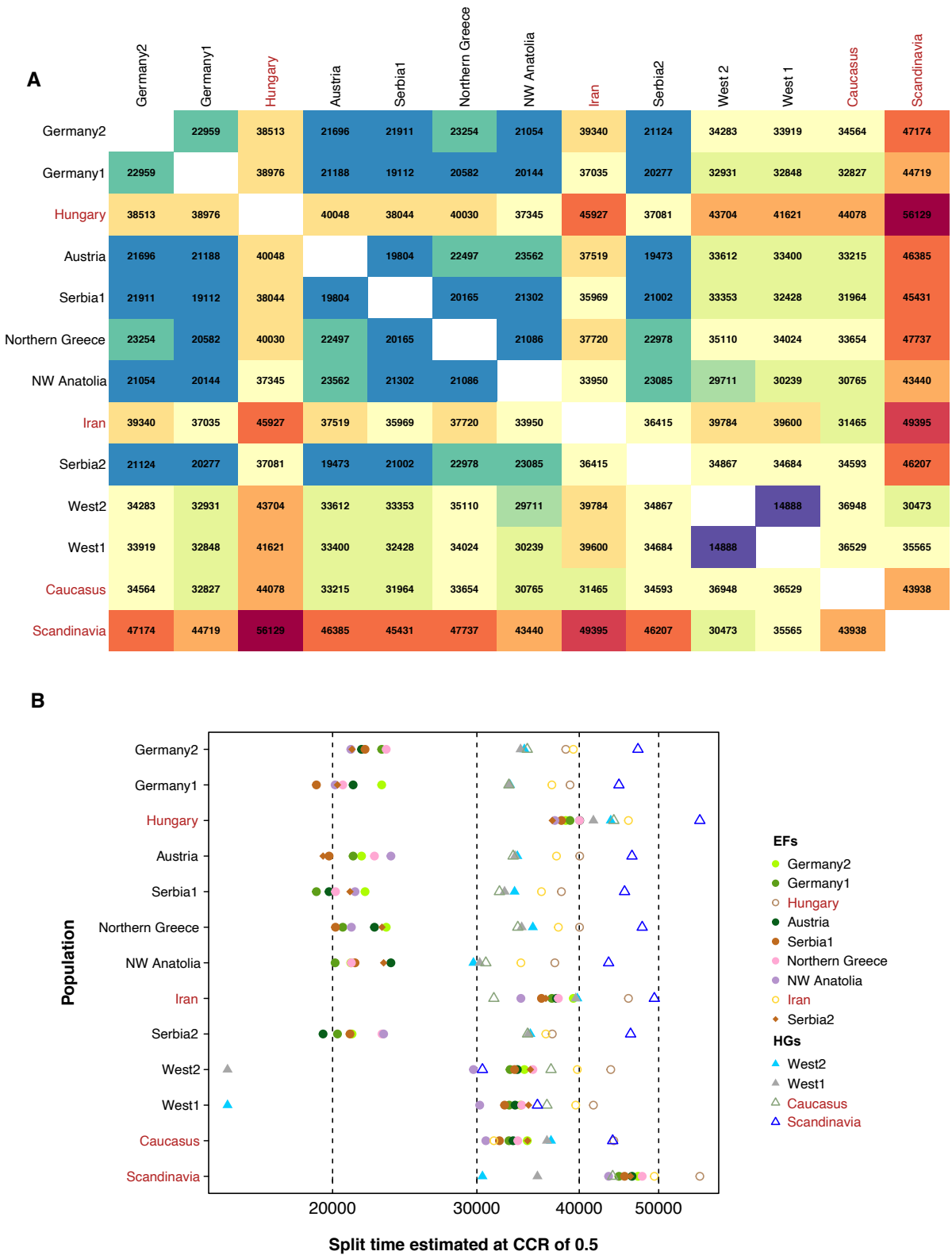

Figure M1_14 - Divergence time estimates obtained with *MSMC2* at the relative CCR of 0.5 for all pairwise comparisons.

(**A**) Square matrix showing the corresponding split time values between all pairs.

(**B**) Point split time estimates for all pairwise comparisons. Populations with only one individual (Hungarian and Iranian EFs, Caucasus and Scandinavian HGs) are marked in red and those comparisons are plotted using symbols with outlines only (not filled).

Bootstrap analysis for the N_e and CCR estimates

We show the bootstrap estimates together with the original estimates of the effective population size for all ancient populations (Fig. M1_15) as well as the bootstrap estimates and original values of the relative CCR between NW Anatolia and all other ancient populations (Fig. M1_16). We observed that there is little uncertainty around the estimates of the original values based on the bootstrap estimates, although the uncertainty increases for the most recent times (after 10 kya), for the cases where there are fewer haplotypes per run, or samples of lower depth (Fig. M1_16), as expected (Schiffels and Wang, 2020).

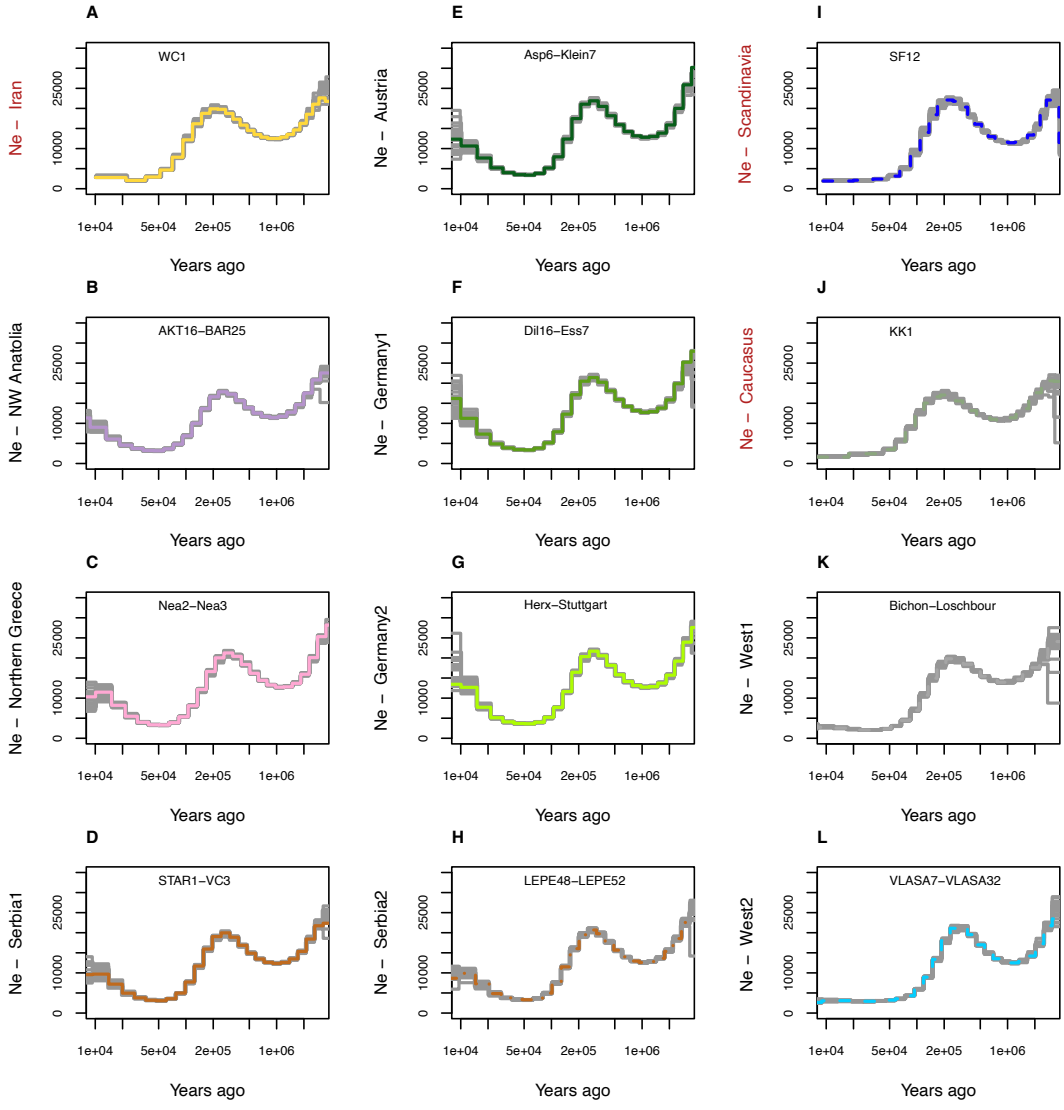

Figure M1_15 - Variation in estimates of past effective population sizes for all ancient populations obtained from 20 artificial datasets per population generated by bootstrapping 5Mb blocks.

First two columns: EF populations from the Neolithic period (A - G) and transitional and Neolithic layers at Lepenski Vir in the Danube Gorges (H).

Last column: HG populations from the Upper Palaeo-Mesolithic period (I - L). Grey lines represent the bootstrapping values and coloured lines their original value.

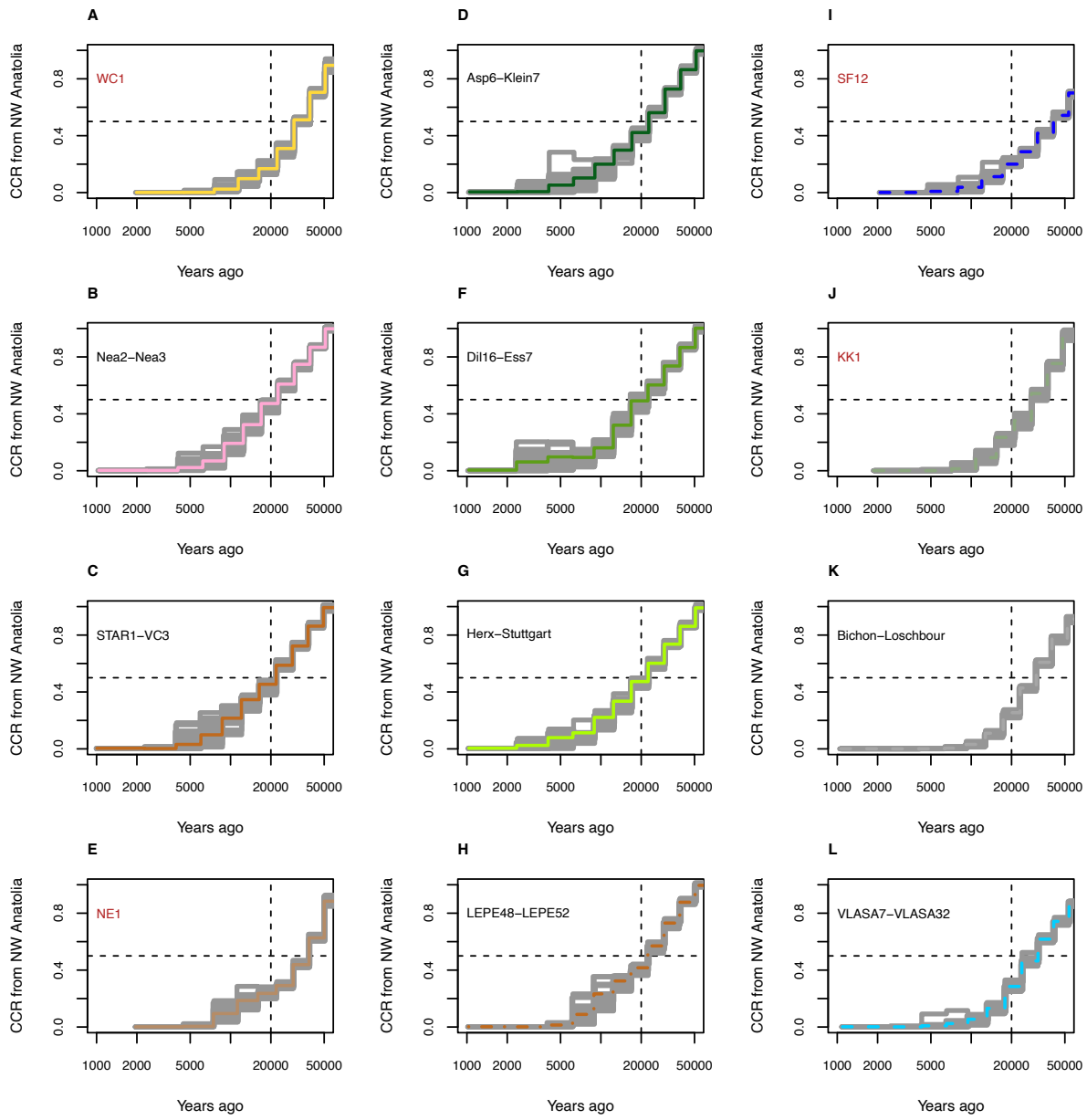

Figure M1_16 - Variation in estimates of coalescence rates (CCR) for all comparisons involving NW Anatolia and obtained from 20 artificial datasets per population generated by bootstrapping 5Mb blocks.

First two columns: EF populations from the Neolithic period (A - G) and transitional and Neolithic layers at Lepenski Vir in the Danube Gorges (H).

Last column: HG populations from the Upper Palaeo-Mesolithic period (I - L). Grey lines represent the bootstrapping values and coloured lines their original value.

Demogenomic inference with fastsimcoal2

Framework

Given the relatively high depth (>10X) of our newly sequenced individuals, it was possible to perform demogenomic inference (i.e. demographic modelling using genomic information and based on coalescent theory (Marchi *et al.*, 2021)) on ancient genomes in a way similar to what is done on modern individuals with *fastsimcoal2* (Excoffier *et al.*, 2021). By computing the site frequency spectrum (SFS) on unascertained polymorphic "neutral" sites (Table M1_4), we could evaluate the likelihood of various historical scenarios and infer parameter values under these models. We nevertheless had to make several adjustments to take into account the temporal and spatial heterogeneity of the collected samples. First, in order to perform population level analyses, we had to pool geographically and culturally similar individuals into the same population, even though they would have been living centuries apart (Table M1_4). Since this temporal heterogeneity can lead to a Wahlund effect (Wahlund, 1928) that translates into a faster rate of coalescence between the two homologous gene copies of an individual relative to two gene copies from different individuals, we introduced population inbreeding coefficients in our modelling as nuisance parameters to specifically account for this population subdivision effect (de Manuel *et al.*, 2016). Second, we introduced unsampled populations from which sampled populations could receive migrants, to reflect the fact that human populations are seldom isolated when not living on islands, which has a direct effect on the shape of the SFS within populations (Marchi and Excoffier, 2020) and could thus introduce biases in demographic inferences if not properly taken into account. These unsampled populations can also be considered as a set of populations (a metapopulation) surrounding sampled populations (i.e. other unsampled farmer populations still exchanging genes with sampled EF populations, or surrounding HG groups contributing to the EF gene pool by admixture) (Beerli, 2004; Excoffier, 2004; Slatkin, 2005). Note also that we have modelled gene flow between metapopulations and sampled populations as a single pulse for modelling convenience, even though gene flow might have been continuous over several generations.

Table M1_4 - Genetic properties and samples composition of the six different panels used for the demographic modelling. Populations that are archaeologically defined as HGs are highlighted in light grey; the other populations are of EFs belonging to Neolithic cultures. The age of the populations, given in generations BP with a generation time of 29 years (Fenner, 2005), were taken within the range of the ages of their constituent samples. The genomic sites were subsetting within the *Ancient* dataset as described in the STAR Methods.

			Panels					
			Core	***HG***	***Aegean***	***Aegean+AKT***	***CAnatolia***	***EurEF***
	Population (Genomes)	**Age**	*Stage I*	*Stage II*	*Stage II*	*Stage III*	*Stage III*	*Stage III*
Samples Composition	Iran (WC1)	317	**X**	**X**	**X**	**X**	**X**	
	C Anatolia (Bon002)	347					**X**	
	NW Anatolia (AKT16)	295				**X**		
	NW Anatolia (Bar25)	295			**X**	**X**		
	N Greece (Nea2; Nea3)	280	**X**		**X**	**X**	**X**	**X**
	Serbia (STAR1; VC3-2)	260						**X**
	Austria (Klein7; Asp6)	245						**X**
	Germany (Dil16; Ess7; Herx)	240						**X**
	Caucasus (KK1)	335	**X**	**X**	**X**	**X**		**X**
	West 2 (VLASA7; VLASA32)	295	**X**	**X**	**X**	**X**	**X**	**X**
	West 1 (Bichon)	472		**X**				
	West 1 (Loschbour)	278		**X**				
Genetic Counts[a]	T_X (Total number of sites passing filtering criteria[b])		262,581,434	167,849,230	243,505,860	200,141,597	35,751,107	175,791,292
	S_x (Polymorphic among T_x)		436,966	256,668	410,13	331,51	56,823	309,817
	$S_{neutral\_x}$ (WWSS S_x sites)		86,557	52,025	81,31	65,534	12,15	60,761
	M_x (Monomorphic among T_x)		262,144,468	167,592,562	243,095,730	199,810,087	35,694,284	175,481,475
	$M_{neutral\_x} = M_x/3$		87,381,489	55,864,187	81,031,910	66,603,362	11,898,095	58,493,825
	$T_{neutral\_x} = M_{neutral\_x}+S_{neutral\_x}$		87,468,046	55,916,212	81,113,220	66,668,896	11,910,245	58,554,586
	$\alpha_x = S_{neutral\_x}/S_x = r_x/3 =$		0.1981	0.2027	0.1982	0.1977	0.2138	0.1961
	$\mu_{neutral\_x} = r_x*\mu tot = r_x*1.25e\text{-}8$		7.43e-09	7.60e-09	7.43e-09	7.41e-09	8.02e-09	7.35e-09

[a] X indicates a given panel. [b] All sites found without missing data in all genomes of a given panel, that are non CpG sites nor in CpG islands, with similar reference allele in the chimpanzee and gorilla reference genomes and a recombination rate ≥ 1 cM/Mb

Demographic models and parameter estimation

Relationship between major groups

Tested models

In order to study the genetic relationships between the populations that occupied SW Asia and Southern Europe before the Neolithic period, we have investigated the fit of data for several models (Fig. M1_17).

The data consisted of the multidimensional unfolded SFS computed on the *Core* panel with samples representative of the four groups that are clearly distinguishable on the two first axes of the MDS analysis (Fig. 2A): the two genomes from the *West 2* population to represent the European HG cluster, the two genomes from the *Northern Greece* population for the European EFs cluster, the genomes of the *Iran* EF and of the *Caucasus* HG.

We assumed in our models (except F which is inspired from models found in the literature (Broushaki *et al.*, 2016; Jones *et al.*, 2015)) that the sampled populations belonged to larger pools of unsampled populations from which they diverged and could receive some migrants (we modelled such migrations as single pulses of gene flow occurring 10 generations before the populations were sampled). Here we model these sets of large and unsampled populations as metapopulations (called *Western, Central* and *Eastern*), sometimes described as ghost populations (Beerli, 2004; Slatkin, 2005) or continents (Excoffier, 2004; Excoffier *et al.*, 2013). We assumed that the *Eastern* and *Western* metapopulations diverged first, and then the *Central* metapopulation split from the *Eastern*. We explored alternative scenarios in which the *Central* metapopulation could receive some gene flow from the *Western* metapopulation, modelled as a single pulse after the divergence events.

Based on the MDS analyses (Fig. 2A) and geographic information, we considered the *West 2* population to be related to the *Western* metapopulation, the Neolithic *Iran* population closer to the *Eastern* metapopulation, and the EFs from *Northern Greece* related to the *Central* metapopulation. One class of models (A) assumed that the *Caucasus* Mesolithic population related to the KK1 genome was closer to the *Central* metapopulation (i.e. as a sister population of the *Northern Greece* Neolithic population), while in model B), the *Caucasus* population was closer to the *Eastern* metapopulation and to the *Iran* EFs as it is classically described (Lazaridis *et al.*, 2016).

Furthermore, while A) and B) models include bottlenecks occurring in the three

metapopulations one generation after the split events, we explored a model C) without any bottleneck as done in former *G-Phocs* analyses (Broushaki *et al.*, 2016; Jones *et al.*, 2015). Furthermore, we explored models in which the split time and the bottlenecks are not necessarily associated. In a climatic change-inspired model D), the *Eastern* and *Western* metapopulations were constrained to split before the Last Glacial Maximum (LGM, i.e. >26.5 kya) and to undergo a bottleneck 23 kya during the LGM. In the model E), it is the bottleneck occurring on the *Western* metapopulation branch that is let free to occur any time after the divergence between the *Western* and *Eastern* metapopulation.

Finally, we designed a relatively simple model F) without bottlenecks and only recent admixture between European EFs and HGs, inspired from (Broushaki *et al.*, 2016; Jones *et al.*, 2015).

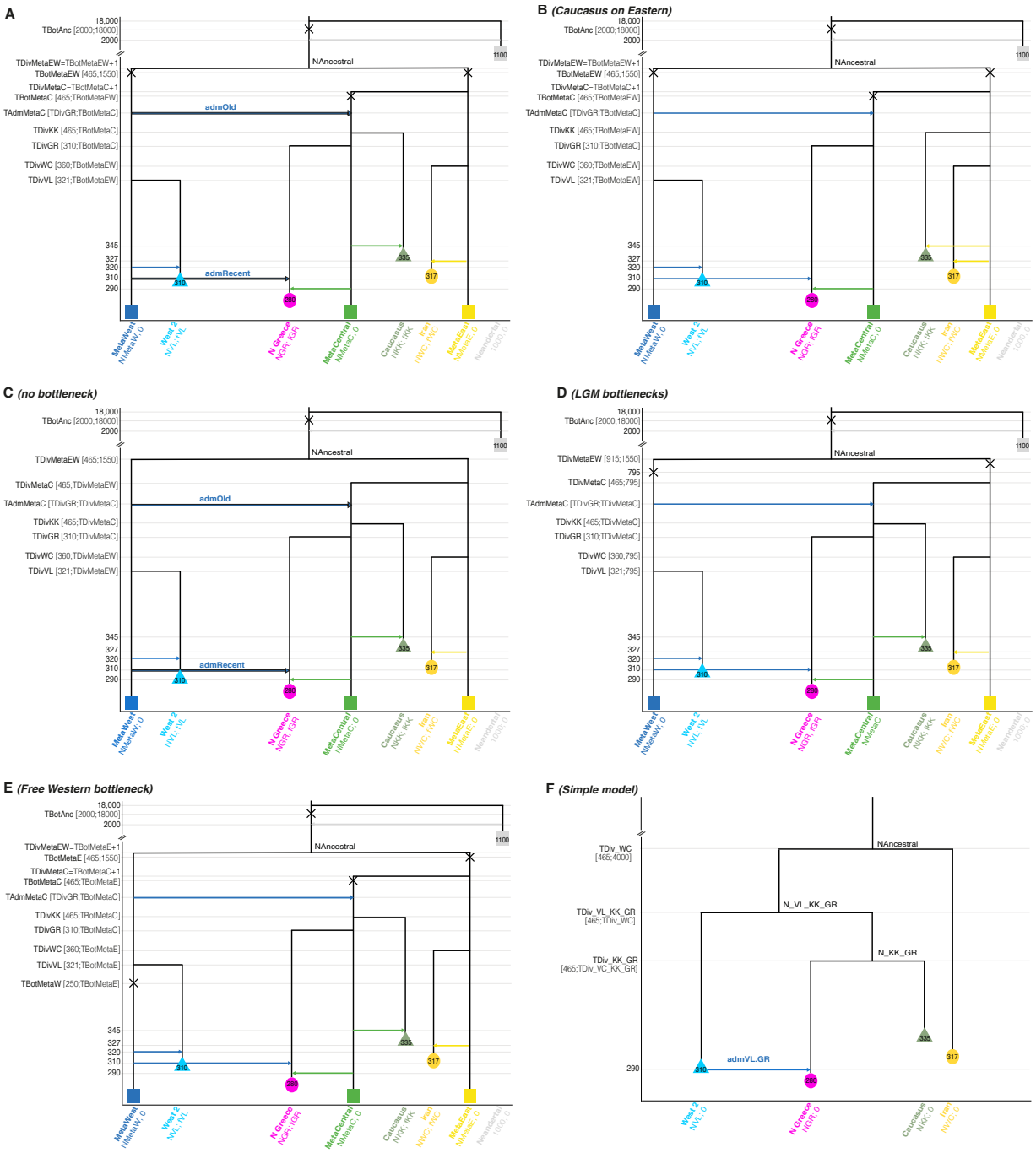

Figure M1_17 - Schematic description of models tested for the *Core* panel.
(**A**) ***Caucasus* HG population branching on the *Central* metapopulation.** The *Eastern* and *Western* metapopulations diverged TDivMetaEW generations ago; the *Central* metapopulation split TDivMetaC generation ago; the divergences are followed by bottlenecks (X) one generation later to model founder effects. The Mesolithic *West 2* population then diverged from the *Western* metapopulation TDivVL generations ago; the EF population from *Iran* diverged from the *Eastern* metapopulation TDivWC generations ago; the *Caucasus* HG and the *Northern Greece* EF populations diverged from the *Central* metapopulation respectively TDivKK and TDivGR generations ago. Each sampled population received some admixture from its source metapopulation 10 generations before its sampling time (indicated in the geometric shape representing the population). Three admixture scenarios were tested: i) no

admixture from the *Western* metapopulation toward the *Central* metapopulation and to the *Northern Greece* EF population, ii) an admixture "admOld" from the *Western* to the *Central* metapopulation TAdmMetaC generations ago, iii) both the "admOld" admixture between the *Western* and *Central* metapopulations and a more recent admixture from the *Western* metapopulation directly into the *Northern Greece* EF population 310 generations ago which corresponds to the onset of farming in this area (Weiberg *et al.*, 2019; Weninger *et al.*, 2014). All migration events are modelled as single pulses of gene flow.

(**B**) ***Caucasus* HG population branching on the *Eastern* metapopulation.** Same as (A) with both "admOld" and "admRecent" migrations authorised but the *Caucasus* HG population branched off directly from the *Eastern* metapopulation.

(**C**) **No bottlenecks.** Same as the (A) but without bottlenecks in the metapopulations. We tested having both "admOld" and "admRecent" migrations, only "admRecent" migration or no migration at all.

(**D**) **EW bottlenecks during the LGM.** Same as (A) with both "admOld" and "admRecent" migrations allowed, but the divergence between the *Eastern* and *Western* metapopulations had to occur before the LGM and these two metapopulations undergo a bottleneck during the LGM 23 kya (= 795 generations).

(**E**) **Free *Western* bottleneck.** Same as (A) with both "admOld" and "admRecent" migrations allowed, but the bottleneck occurring in the *Western* metapopulation branch is now decorrelated of the divergence between the *Eastern* and *Western* metapopulations.

(F) Simple model without metapopulations, bottlenecks, F_{IS}, ancestral nor archaic admixture. Only population size changes are allowed after each divergence event, as well as a single admixture event occurring 290 generations ago from *West 2* population toward *Northern Greece* EFs. The topology is also fixed: the more recent divergence is between *Northern Greece* EFs and *Caucasus* HGs; their ancestors split up with *West 2* ancestors; the oldest divergence is with the *Iran* EF ancestors.

On each figure, we show the parameters to estimate with their search ranges within brackets. Further information about the parameters can be found in Table S4 and the properties of the sampled populations in Table M1_4.

Best model parameters estimate and SFS fit

Within the tested models for the *Core* panel, we find that the model of class A) with both an old admixture from the *Western* metapopulation towards the *Central* metapopulation and a recent admixture towards the *Northern Greece* EF population is best supported (*A_admOld&Recent* in Fig. M1_18A, Table S4), with excellent fits between the observed and estimated SFS (Fig. M1_18C).

Effective population sizes. In this model, the effective population sizes of the *West 2* and *Caucasus* HG populations are in the range of the EF population sizes. However, there is a significant difference in the bottlenecks modelled as lasting a single generation occurring on the metapopulations' branches: the bottlenecks on the *Eastern* and *Central* metapopulation branches are relatively weak (~50 haploid individuals) while the ancestral bottleneck and the bottleneck occurring on the *Western* metapopulation branch are stronger (~6 haploid individuals [4-10]).

Divergence times between metapopulations. For the best model *A_admOld&Recent*, we find a deep divergence between the *Eastern* and *Western* metapopulations (~25.6 kya [17.3-31.3]) and a more recent divergence between the *Central* and *Eastern* metapopulations (~15.8 kya [14.3-25.6 kya]). However, these divergences are much younger than the previously inferred times for the divergence between Iranian and European EFs (46-77 kya) (Broushaki *et al.*, 2016) or that between European EFs and Western European HGs (46 kya) (Jones *et al.*, 2015). Both these estimates were obtained under simple models without metapopulations, bottlenecks nor ancestral admixture. We managed to obtain older divergence times (~39 kya) between the ancestors of *Western* and *Eastern* metapopulations with the models *C_admOld&Recent* and *C_admRecent* than for model *A_admOld&Recent*, which does not include any bottleneck in the metapopulations. But their fit is much worse as for *A_admOld&Recent* (Fig. M1_18A, Table S4). Interestingly, the model *C_noadm* which is without admixture gives recent divergence times (~16.4 kya), with a fit much worse than the original model *A_admOld&Recent* or the models *C* with admixture described above. Thus, it appears that both admixture and bottlenecks are necessary to have a correct fit to the data and that by modelling bottlenecks we indeed obtain younger divergence times than those published earlier. Note that with our model *F* that is inspired from those of (Broushaki *et al.*, 2016; Jones *et al.*, 2015) we obtained recent divergence time too (~14.2 kya) despite the absence of bottlenecks but the fit is bad and results are therefore difficult to interpret.

Bottlenecks on metapopulation branches. In the best model *A_admOld&Recent*, we find that the divergence between the *Eastern* and *Western* metapopulations occurred during the Last Glacial Maximum (LGM). By construction, the bottlenecks have to occur in this model one generation after the split, i.e. also during the LGM. In the model *D*, we constraint the bottlenecks on the *Eastern* and *Western* metapopulation branches to occur during the LGM and their split to be before the LGM. As expected, this model leads to an older divergence time between these metapopulations (32.7 kya) than for the best model *A_admOld&Recent*. However, model *D* being the second best supported model (Fig. M1_18A, Table S4), it strengthens our hypothesis of bottlenecks during the LGM. To further test this result, we decoupled the strong bottleneck observed on the *Western* metapopulation branch from its divergence from the *Eastern* metapopulation in model *E*. In that model, we find a slightly older metapopulation divergence (27 kya) and a more recent bottleneck (23 kya) than in *A_admOld&Recent* but during the LGM. Furthermore, this LGM bottleneck is found to be strong, as observed in the best model (~13 haploid individuals in model *E*, 8 in model *D*).

Divergence from the source metapopulations. In the best model *A_admOld&Recent*, the divergence of the sampled populations from their source metapopulation are found to be relatively old but with quite broad confidence intervals: *West 2* HG (16.0 kya [9.7-24.5]); *Iran* EF (13.6 kya [11.0-24.6]); *Northern Greece* EF (13.5 kya [9.6-21.7]). The *Caucasus* HG population is found to have diverged from the *Central* metapopulation about 14.2 kya [13.7-19.0]. Indeed, the scenario where Caucasus HGs are connected to the *Central* metapopulation (model *A*) is better supported than model *B* where they are directly connected to the *Eastern* metapopulation (Fig. M1_18A, Table S4). This split from the *Central* metapopulation happens just after the ancestral population received some admixture from the *Western* metapopulation (14% [8-26]). Therefore, both the *Northern Greece* EF and the *Caucasus* HG populations descend from the ancestral admixture between the *Western* and *Central* metapopulations. Indeed, models *A_admOld&Recent* and *A_admOld* in which the *Caucasus* HG population receives some *Western* metapopulation admixture are more likely than the models without migration *A_noadm* or *B* (Fig. M1_18A, Table S4). In addition to this ancestral admixture, the *Northern Greece* EF population seems to have received an extra admixture from the *Central* metapopulation (about 15% [11-24]). Interestingly, our best model requires a significant amount of gene flow (5-11%) from metapopulations to sampled populations, showing that human populations are seldom isolated (Marchi and Excoffier, 2020), also validating the need to model surrounding unsampled populations.

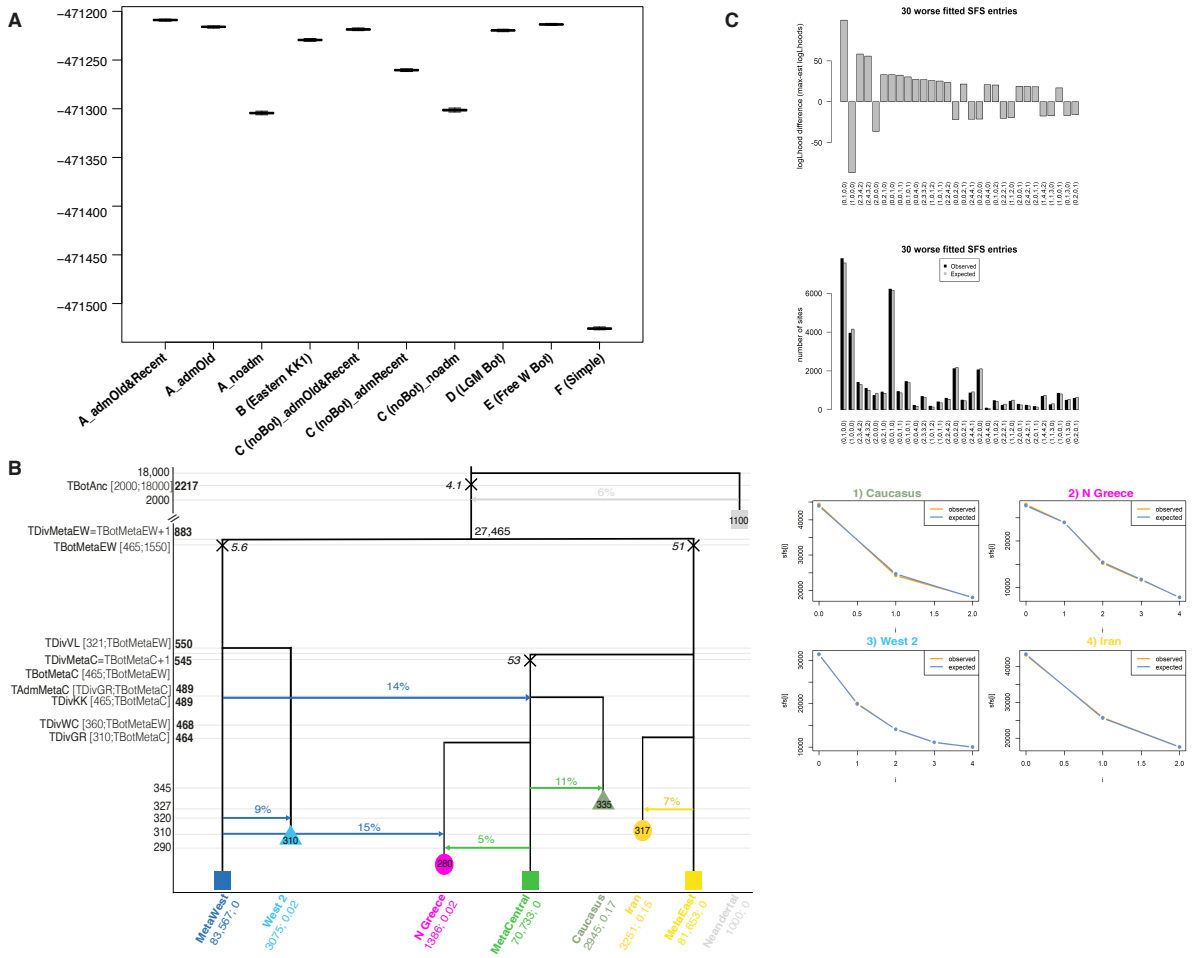

Figure M1_18 - Exploration of models best explaining the observed multi-dimensional *Core* SFS.

(**A**) **Distribution of the log10-likelihoods** obtained by simulation from the maximum likelihood point estimates for the tested models described in Fig. M1_17. Higher likelihoods (top of the plot) indicate better fit.

(**B**) **Schematic representation of the ML demographic model *A_admOld&Recent.*** Values of various parameters are reported with their 95% CI estimated from 100 parametric bootstraps. Close to bottleneck symbols (X), we report the size of the bottleneck, conveniently modelled as a single generation bottleneck (*instbot fastsimcoal2* option).

(**C**) **Fit between the observed and estimated SFS under the ML model** (multidimensional SFS on the upper panel; marginal SFS for each population on the bottom).

Neolithic Aegean structure

Tested models

In a second step, we extended the best model inferred from the *Core* panel by including an individual from *NW Anatolia* (Bar25, *Aegean* panel) in order to estimate the divergence times between the sampled Aegean populations, i.e. from *Northern Greece* and *NW Anatolia*. We tested two classes of models: A) models in which a population ancestral to Aegean EFs would have diverged from the *Central* metapopulation and then split into *NW Anatolia* and *Northern Greece* populations; B) a model in which *NW Anatolia* and *Northern Greece* populations directly and separately diverged from the *Central* metapopulation, without the existence of an ancestral shared genetic structure (Fig. M1_19). Furthermore, in class A), we authorised different admixture events from the *Western* metapopulation: one occurring directly in the population ancestral to EFs one generation after its founding and/or separate admixture events into the *Northern Greece* and *NW Anatolia* populations 310 generations ago (i.e. approximately when Neolithic culture was introduced in these areas and, therefore, when the Neolithic populations were founded (Weninger *et al.*, 2014)).

The divergence times, population sizes, inbreeding coefficients and admixture rates related to the Aegean populations were left to be freely estimated while the other parameters were inherited from the best model inferred for the *Core* panel.

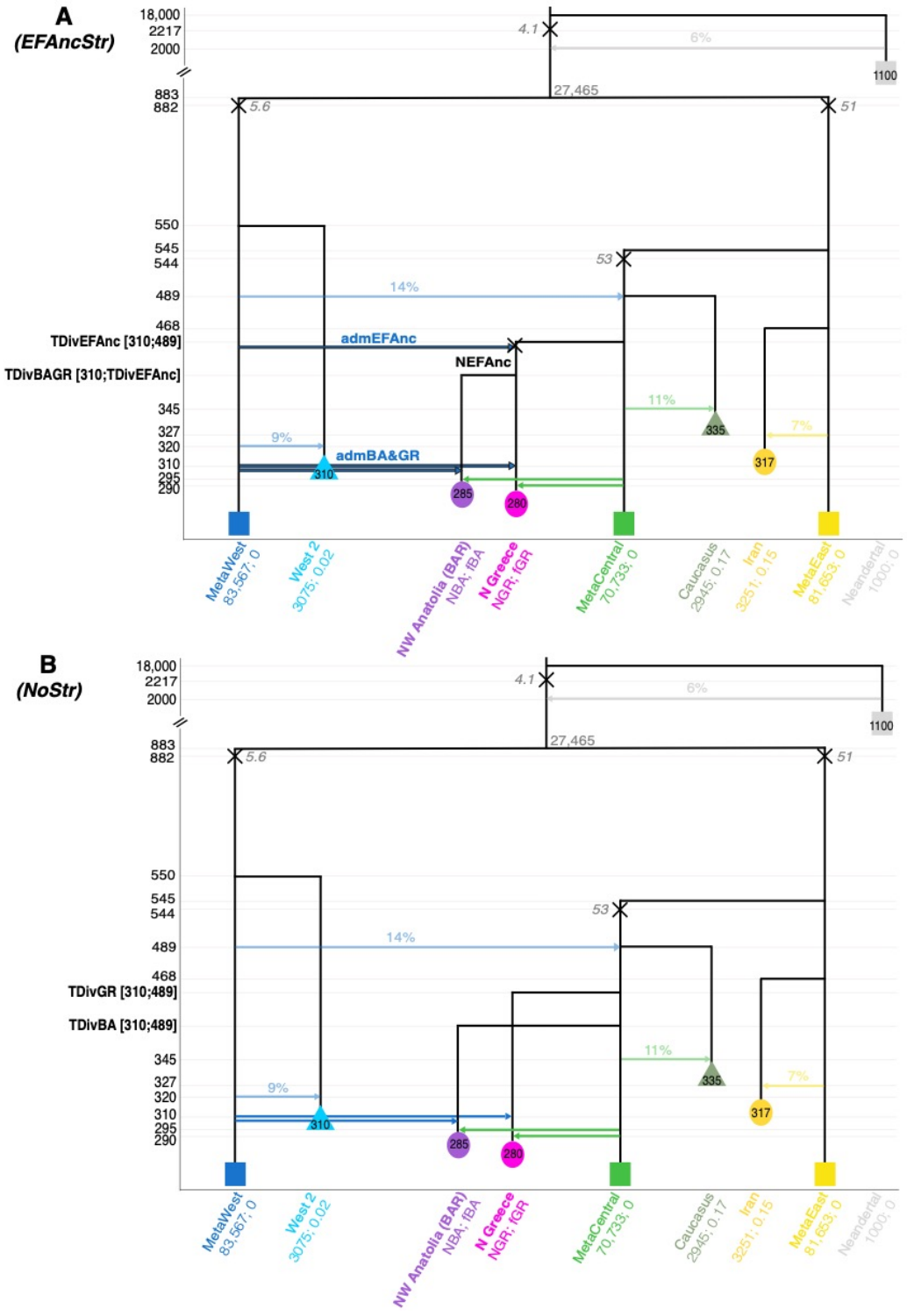

Figure M1_19 - Schematic description of models tested for the *Aegean* panel.
(A) An ancestral EF structure. Under this model, the population ancestral to Aegean EFs diverged TDivEFAnc generations ago from the *Central* metapopulation followed by an instantaneous bottleneck one generation after; then, the ancestral population split into *Northern Greece* and *NW Anatolia* populations TDivBAGR generations ago. We tested four scenarios: i) allowing a single pulse of gene flow coming from the *Western* metapopulation to the ancestral EF branch one generation after its founding (*admEFAnc*), ii) with independent admixture events into the *Northern Greece* and *NW Anatolia* populations from the *Western*

metapopulation 310 generations ago (*admBA&GR*), iii) allowing both *admEFAnc* and *admBA&GR*, iv) without any admixture.

(B) No EF ancestral structure, i.e. independent split events within the Aegean region. Under this model, *Northern Greece* and *NW Anatolia* populations independently split from the *Central* metapopulation TDivGR and TDivBA generations ago, and received independent pulses of gene flow from the *Western* metapopulation 310 generations ago. As is the case for the *Northern Greece* population, the *NW Anatolia* population received some gene flow from the *Central* metapopulation 10 generations before sampling.

These models are derived from the best model inferred for the *Core* panel: parameter values estimated in the previous scenario were fixed in this analysis and shown in light fonts while newly estimated parameters are shown in bold fonts with search ranges shown within brackets. Further information about the parameters can be found in Table S4 and the properties of the sampled populations are listed in Table M1_4.

Best model parameters estimate and SFS fit

We find that all models of class *A* very significantly outperformed model *B*, implying that the EFs from the Aegean region diverged from the same ancestral population (Fig. M1_20A, Table S4). This ancestral population received some gene flow from the *Western* metapopulation, as shown by the better fit of models *A_admEFAnc+BA&GR* and *A_admEFAnc* as compared to *A_admBA&GR*, and even more importantly than *A_noadm*.

Under the model *A_admEFAnc+BA&GR*, which is the most complex and best supported scenario, the Barcın individual Bar25 from *NW Anatolia* seems to be drawn from a smaller population (2Ne = 1491 [358-5737]) than the *Northern Greece* population (2Ne = 4437 [1720-7689]), from which it would have diverged recently (320 generations ago, corresponding to ~9.3 kya [313-414]) (Fig. M1_20B, Table S4). The population ancestral to EFs is found to have been of small size (2Ne = 1242 [143-4300]), after a mild bottleneck (during one generation, 2Ne = 250 [9.5-333]) and to have diverged from the *Central* metapopulation 126 generations before (i.e. 12.3 kya [9.4-13.9]). Interestingly, this modelling allows us to refine the admixture from the *Western* metapopulation received by the *Northern Greece* population found for the *Core* panel: it confirms that about 15% [6-17] of the gene pool of all EFs were received from the *Western* metapopulation (modelled here as a single pulse occurring when the EF ancestral population split from the *Central* metapopulation). There were also additional pulses of gene flow in the *NW Anatolia* population (12% [6-16]) and in the *Northern Greeks* (3% [1-11]) that might have contributed to their differentiation into distinct populations. Note that this approach of extending a former model with some fixed and newly estimated parameters leads to a very good fit between the observed and simulated SFS (Fig. M1_20C).

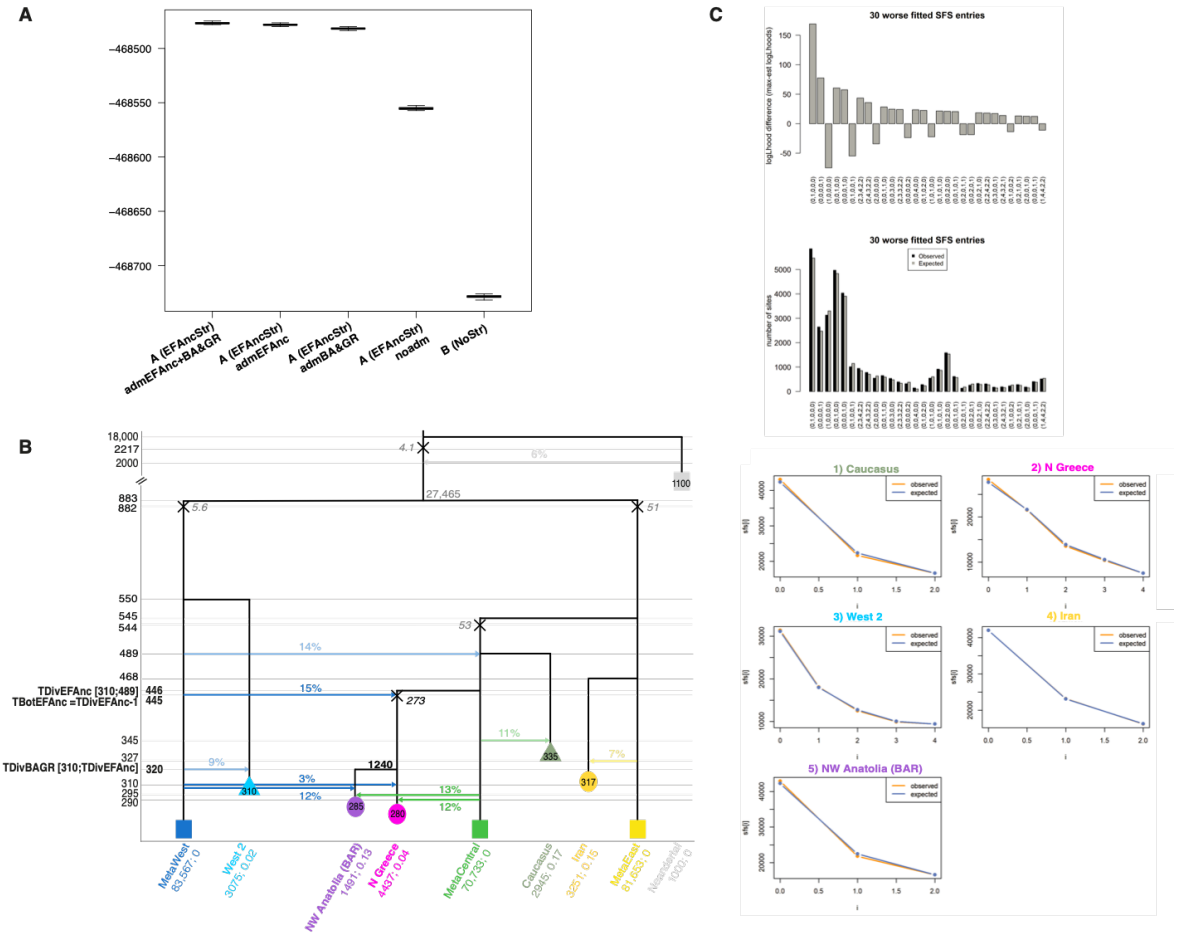

Figure M1_20 - Exploration of models best explaining the observed multi-dimensional *Aegean* SFS.

(**A**) **Distribution of the log10-likelihoods** obtained by simulation from the maximum likelihood point estimates for the four tested models described in Fig. M1_19. Higher likelihoods (top of the plot) reflect better fit.

(**B**) **Schematic representation of the ML demographic model *A_admEFAnc+BA&GR*.** Values of various parameters are reported with their 95% CI estimated from 100 parametric bootstrap estimations. Close to bottleneck symbols (X), we report the size of the bottleneck, conveniently modelled as a single generation bottleneck (*instbot fastsimcoal2* option).

(**C**) **Fit between the observed and estimated SFS under the ML model** (multidimensional SFS on the upper panel; marginal SFS for each population on the bottom).

NW Anatolia genetic structure

Tested models

In a third step, we used the *Aegean* best model parameters (Fig. M1_20B) as a backbone to examine more complex models involving an additional sample (AKT16) from the site of Aktopraklık, which in addition to being one of the oldest Neolithic sites in NW Anatolia, belongs to the local 'Fikirtepe culture', thought to have been influenced by both Mesolithic and Neolithic traditions (Özdoğan, 2011). In order to investigate the possibly complex settlement history of this site, we tested four models allowing AKT16 to branch off from the *NW Anatolia (BAR)* population, from the *Northern Greece* population, from the ancestral EF population or from the *Central* metapopulation (Fig. M1_21).

Best model parameters estimate and SFS fit

The model best explaining the *Aegean+AKT* panel (Fig. M1_22, Table S4) is when the Aktopraklık individual is drawn from a small population (2Ne = 687 [145-4933]) emerging recently as a sister population from the *Northern Greece* population ~9.1 kya [9.1-9.2], 6 generations after the split between the *Barcın* and *Northern Greece* population, suggesting the existence of an Aegean cluster to which Aktopraklık, Barcın and Northern Greece populations belong as they have diverged from each other at about the same time. Despite this recent split within the Aegeans, the Aktopraklık-related population has 17% [11-18] of its genome drawn from the *Western* metapopulation, and another 13% [5-17] from surrounding EF populations (modelled by the *Central* metapopulation), i.e. a proportion similar to that received by the two other Aegean populations which might suggest a single ancestral pulse of gene flow before their split. The quite important *Western* metapopulation contribution to the Aktopraklık genome is in line with the admixture analysis (Fig. 2B) and *f*-statistics (Fig. S1). Note that we observed a good fit between the observed and simulated SFS using the backbone approach (Fig. M1_22C).

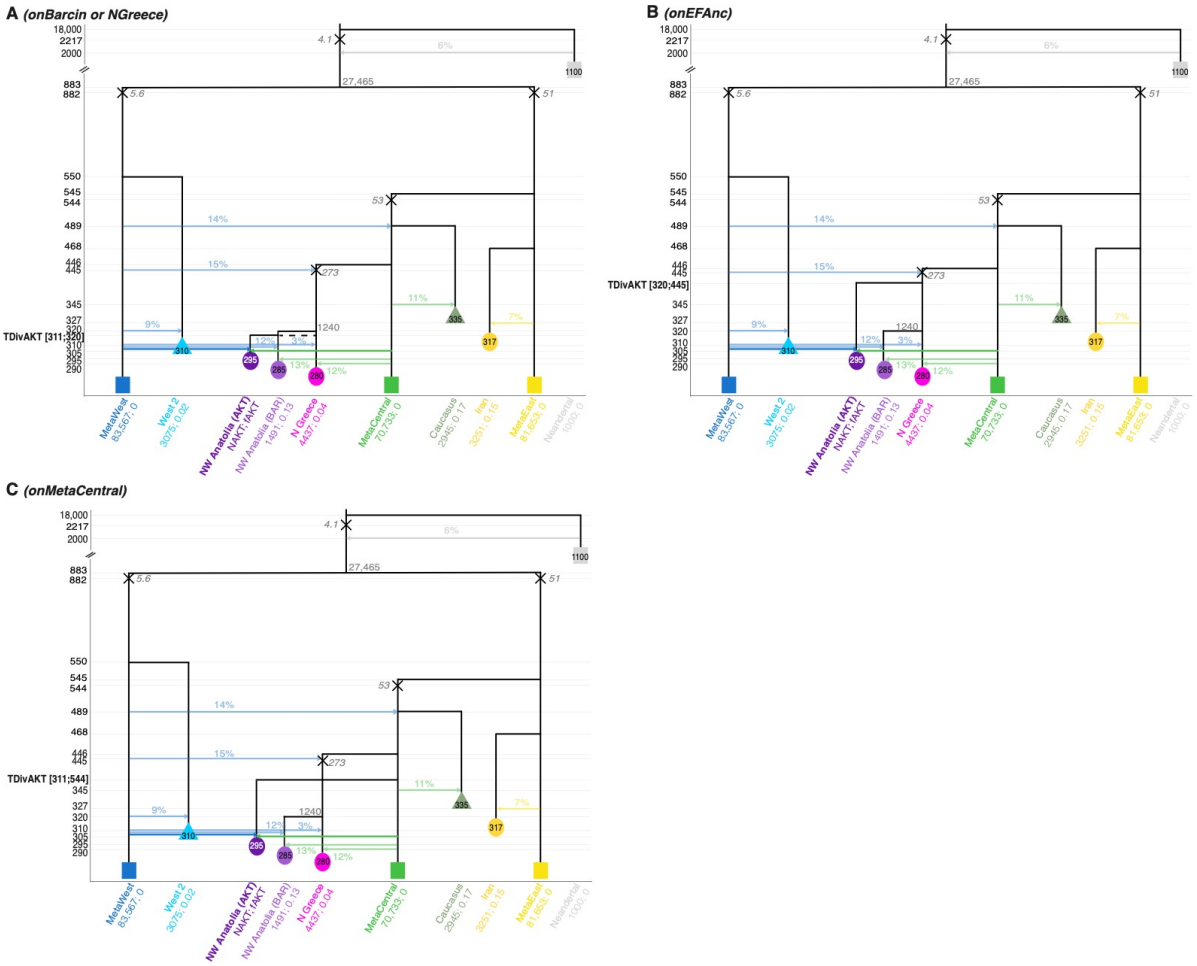

Figure M1_21 - Schematic description of models tested for the *Aegean+AKT* panel.
(**A**) **As a sister population of *Barcın* or of the *Northern Greece* populations.** Under this model, the AKT-related population has diverged TDivAKT generations ago from the population from *NW Anatolia* or *Northern Greece*, necessarily after their split 320 generations ago.
(**B**) **From the Ancestors of EFs.** Under this model, the AKT-related population has diverged TDivAKT generations ago from the population ancestral to EFs, i.e. before the split of the *Northwest Anatolia* and *Northern Greece* EFs 320 generations ago but after the divergence of the branch from the *Central* metapopulation 445 generations ago.
(**C**) **From the *Central* metapopulation.** Under this model, the AKT-related population has diverged TDivAKT generations ago from the *Central* metapopulation as did the ancestral EF branch. The AKT-related population received some gene flow from the *Western* metapopulation 310 generations ago and from the *Central* metapopulation 10 generations before sampling.
These models are derived from the best model inferred for the *Aegean* panel: parameter values estimated in the previous scenario were fixed in this analysis and shown in light fonts while newly estimated parameters are shown in bold fonts with search ranges shown between brackets. Further information about the parameters can be found in Table S4 and the properties of the sampled populations are listed in Table M1_4.

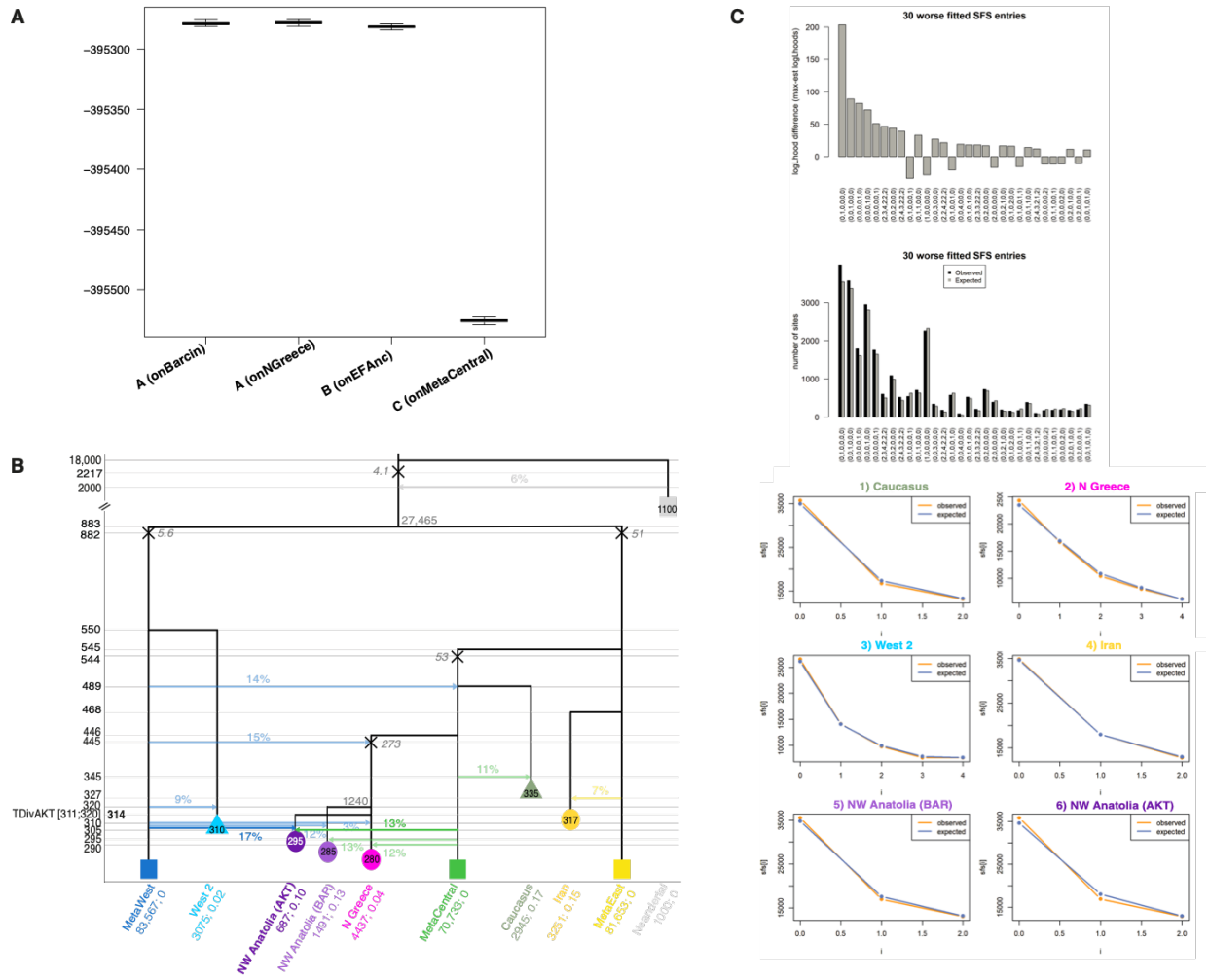

Figure M1_22 - Exploration of models best explaining the observed multi-dimensional *Aegean+AKT* SFS.

(**A**) **Distribution of the log10-likelihoods** obtained by simulation from the maximum likelihood point estimates for the four tested models. Higher likelihoods (top of the plot) reflect better simulations.

(**B**) **Schematic representation of the ML demographic model *A (onGR)*.** Values of various parameters are reported with their 95% CI estimated from 100 parametric bootstrap estimations. Close to bottleneck symbols (X), we report the size of the bottleneck, conveniently modelled as a single generation bottleneck (*instbot fastsimcoal2* option).

(**C**) **Fit between the observed and estimated SFS under the ML model** (multidimensional SFS on the upper panel; marginal SFS for each population on the bottom).

Relationship between Aegean and Central Anatolian early farmers

Tested models

The significant Mesolithic European-like contribution (that we model as the *Western* metapopulation) found in the ancestors of all European and *NW Anatolia* EFs as well as *Caucasus* HGs raises the question of the geographic location of this HG input, and whether it can also be observed in Central Anatolia. We examined these questions by building on the best scenario found for the *Aegean* panel and including the Bon002 (Boncuklu) early Neolithic individual from the Konya Plain in Central Anatolia. In contrast to the *Aegean+AKT* panel, the lower number of genomic sites available in the Bon002 genome (due to its lower coverage) led us to reduce the dimensions of the SFS and only keep the *West 2* HGs, *Northern Greece* and *Iran* EFs in this panel (Table M1_4). Thus, we were only able to test two models: A) in which the *C Anatolia* population branches from the EF ancestral population, or B) from the *Central* metapopulation (Fig. M1_23).

Best model parameters estimate and SFS fit

Our analysis reveals that the Boncuklu individual has diverged from the EF ancestral population some 362 [361-378] generations ago (~10.5 kya) during the cold Younger Dryas period, and after the admixture with the *Western* metapopulation shared by all the EFs (Fig. M1_24, Table S4). Therefore, it appears that Bon002 from *C Anatolia* shares an ancestry with the *NW Anatolian* and European EFs, and that the *Iran* EFs seems to be the only sampled Neolithic population not to have received any European-like HG admixture. An additional *Western* metapopulation genetic input, forced to be received 10 generations before sampling, contributed to 10% [3-15] of the Boncuklu's genome. Furthermore, this population seems to be more isolated than other Aegean Neolithic populations since only 2% [1-11] of the Boncuklu gene pool comes from the *Central* metapopulation. However, the results obtained for this dataset should be considered with caution due to the small number of polymorphic sites available for these analyses.

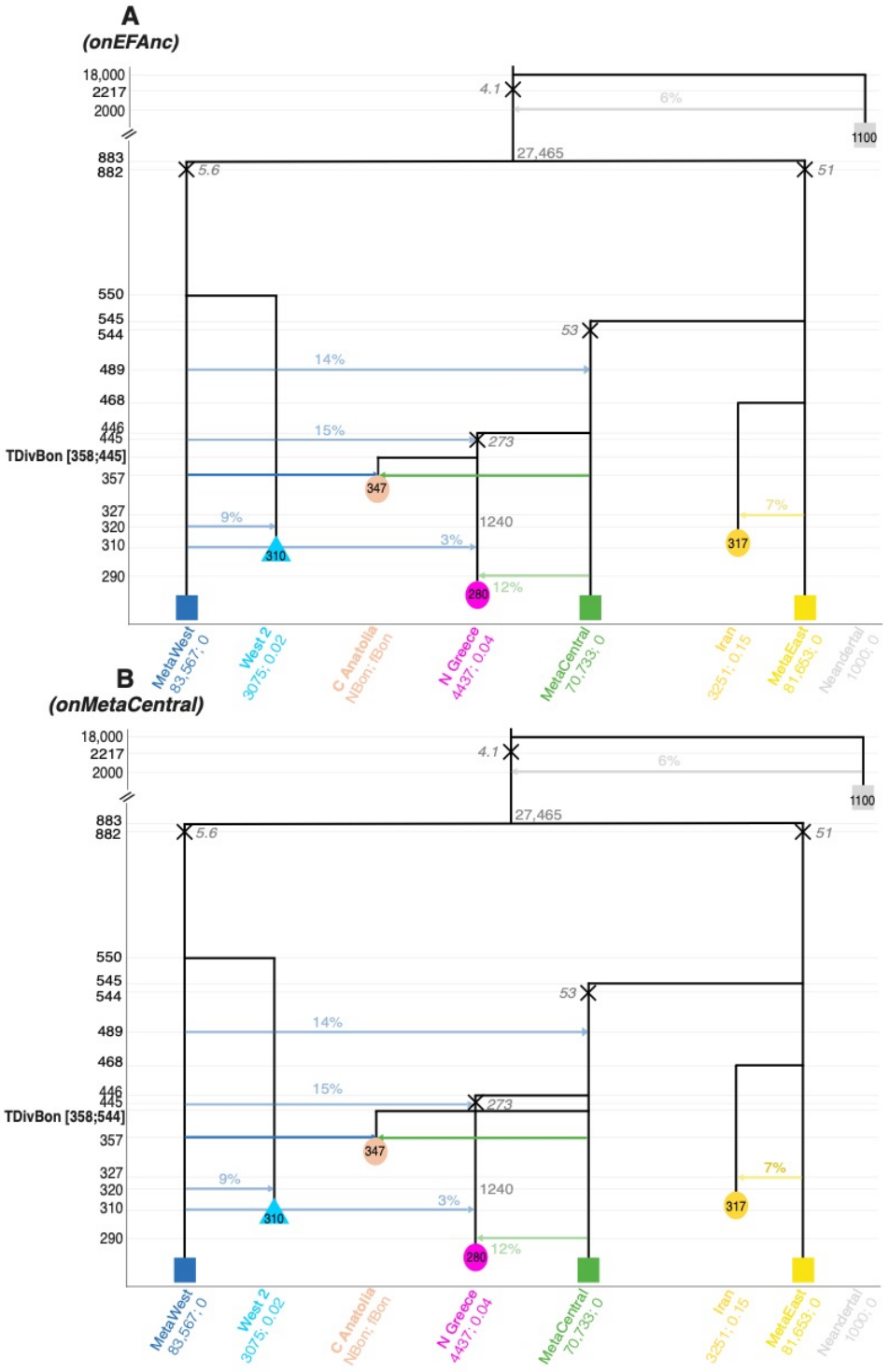

Figure M1_23 - Schematic description of models tested for the *CAnatolia* panel.

(**A**) **From the EF Ancestors.** Under this model, the *C Anatolia* population has diverged TDivBon generations ago from the ancestral EF population, i.e. before the split of the *NW Anatolia* and *Northern Greece* populations 320 generations ago, but after the divergence of the *Central* metapopulation 445 generations ago.

(**B**) **From the *Central* metapopulation.** Under this model, the *C Anatolia* population has diverged TDivBon generations ago from the *Central* metapopulation. The *C Anatolia* population received gene flow from the *Western* and *Central* metapopulations 10 generations

before sampling. These models are derived from the best model inferred for the *Aegean* panel: parameter values estimated in the previous scenario were fixed and shown in light fonts, while newly estimated parameters are shown in bold fonts with search ranges shown between brackets. Further information about the parameters can be found in Table S4 and the properties of the sampled populations are listed in Table M1_4.

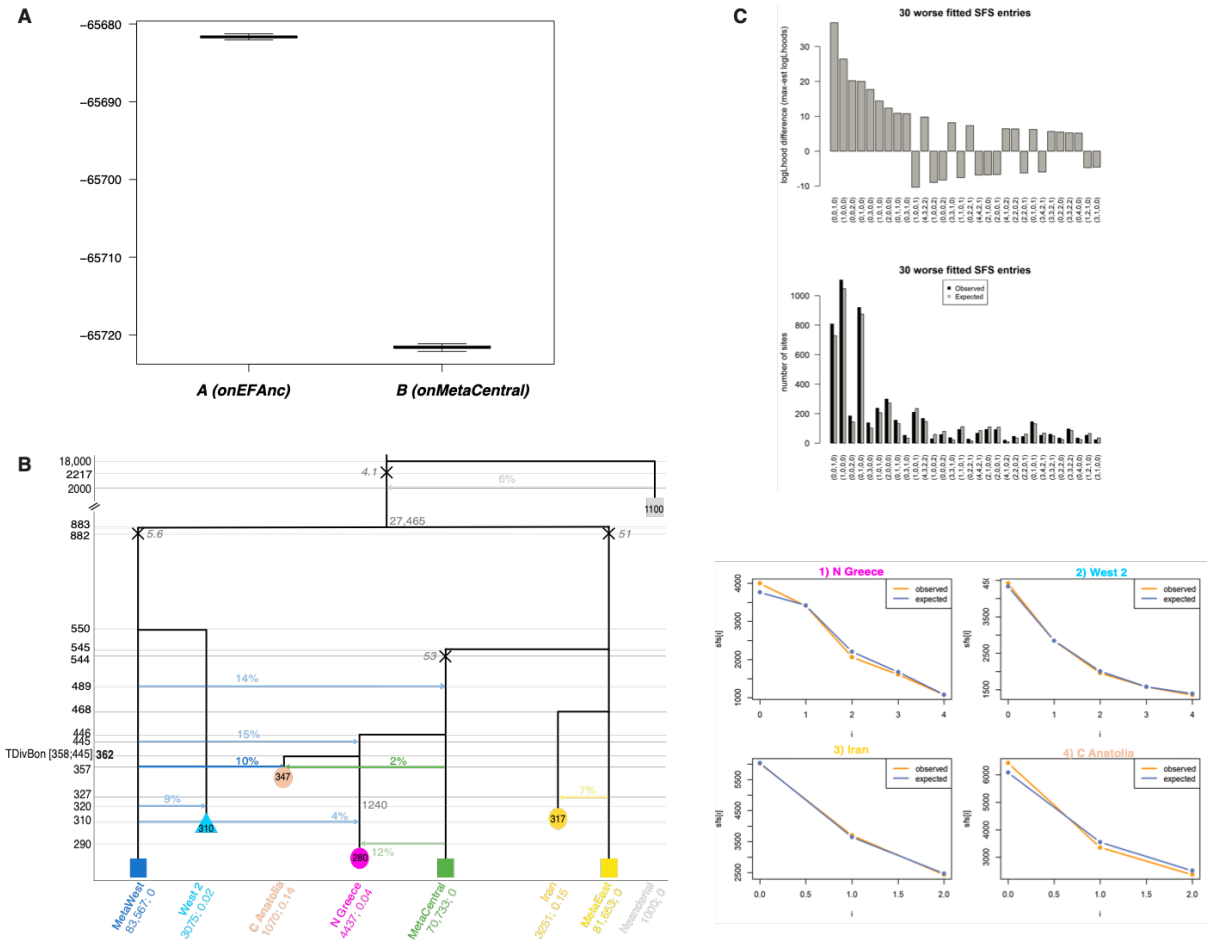

Figure M1_24 - Exploration of models best explaining the observed multi-dimensional *CAnatolia* SFS.

(**A**) **Distribution of the log10-likelihoods** obtained by simulation from the maximum likelihood point estimates for the two tested models described in Fig. M1_23. Higher likelihoods (top of the plot) reflect better fit.

(**B**) **Schematic representation of the ML demographic model *A (onEFAnc)*.** Values of various parameters are reported with their 95% CI estimated from 100 parametric bootstrap estimations. Close to bottleneck symbols (X), we report the size of the bottleneck, conveniently modelled as a single generation bottleneck (*instbot fastsimcoal2* option).

(**C**) **Fit between the observed and estimated SFS under the ML model** (multidimensional SFS on the upper panel; marginal SFS for each population on the bottom).

Spread of the Neolithic along the Danube

Tested models

In order to study the spread of Neolithic EF populations from the Aegean basin to Central Europe along the Danube corridor, we considered samples from *Serbia*, *Austria*, and *Germany* in addition to those from *Northern Greece, Caucasus* and *West 2* HGs (Table M1_4). In this analysis, we did not consider the Hungarian individual (NE1) who is a geographic outlier nor the two individuals from Lepenski Vir, even though they are located in the Danube corridor, as individuals from this site show recent admixture with Iron Gates HGs and display subsistence and burial practices more akin to the European Mesolithic (de Becdelièvre *et al.*, 2020).

We built two classes of models (Fig. M1_25): A) here we assume that the spread of EF populations consisted in a series of stepping-stone (*SS*) migration events from *Northern Greece* to *Serbia*, then from *Serbia* to *Austria*, and finally from *Austria* to *Germany*; B) here we explored a partial *SS* model, where *Austria* could be directly settled from *Northern Greece*, thus bypassing (*BP*) *Serbia*. For both types of models, we allowed EF populations to receive some input from local Mesolithic populations, modelled by admixture from the *Western* metapopulation occurring at the founding of the EF population. We re-estimated here the parameters for the population from *Northern Greece* too as this population is the source of the Danubian subsequent populations.

Best model parameters estimate and SFS fit

We find that *stepping-stone (SS)* models are better supported than their corresponding *bypassing (BP)* model (Fig. M1_26A, Table S4), suggesting some stepwise expansion of the EFs in Europe along the Danube between ~8 and 7.5 kya. Interestingly, admixture from the *Western* metapopulation always improves the fit, suggesting that the EF populations incorporated a few HG individuals (2-7%) at all stages of the dispersal along the Danubian corridor (Fig. M1_26B, Table S4). Finally, the EF populations were well connected to other farming communities, as modelled by single pulses of gene flow coming from the *Central* metapopulation with intensity between 4 and 8%. Interestingly, this continuous gene flow from less-mixed neighbouring farmer populations could counterbalance the rate of *Western* metapopulation admixture entering the farmer gene pool during the expansion. This complex pattern of gene flow might explain the apparent lack of genetic structure among EFs observed in the MDS plot (Fig. 2A) and the absence of increasing *Western* ancestry along the Danubian corridor in the admixture analysis (Fig. 2B).

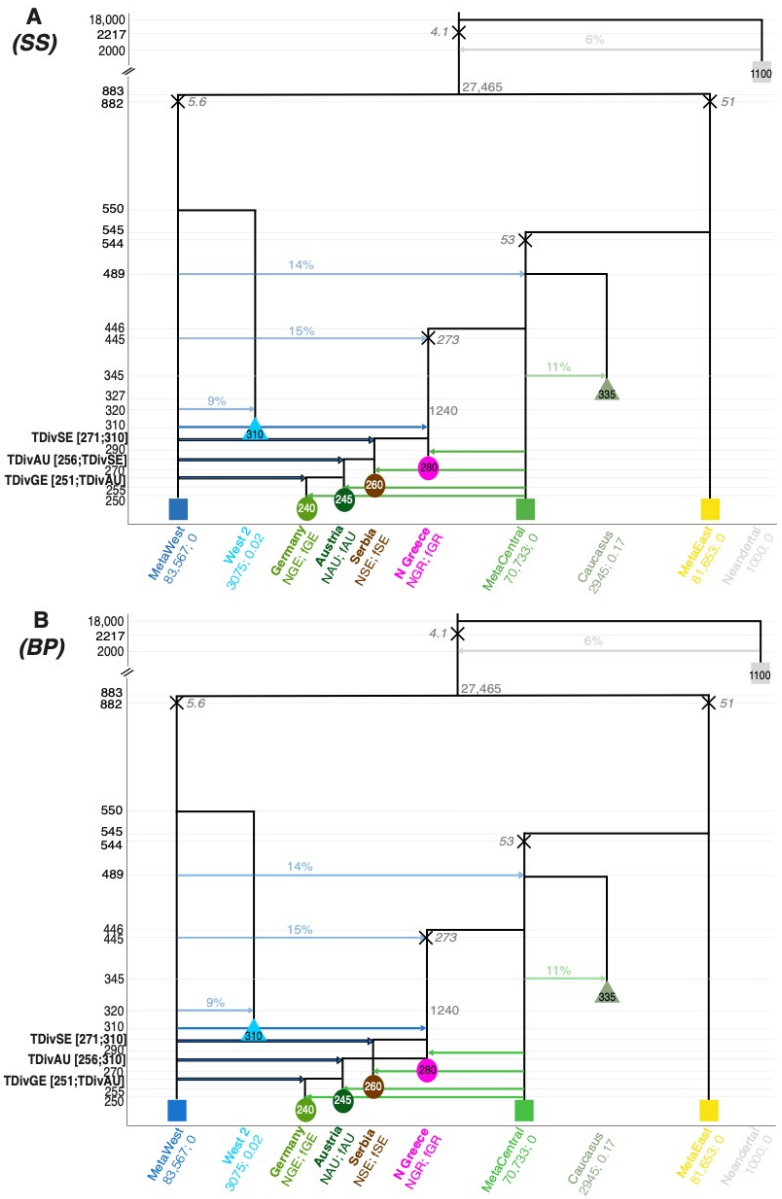

Figure M1_25 - Schematic description of models tested for the *EurEF* panel.

(**A**) **Pure stepping-stone (SS) model.** The *Serbian* EF population derived from *Northern Greece* TDivSE generations ago, the *Austrian* from *Serbia* TDivAU generations ago and the *German* population from *Austria* TDivGE generations ago.

(**B**) **By-passing (BP) model.** Here the *Serbian* and *Austrian* EF populations are both settled from *Northern Greece* TDivSE and TDivAU generations ago, respectively, then the *German* population derived from *Austria* TDivGE generations ago. In the models with admixture, the EF populations can receive some gene flow from the *Western* metapopulation at the time of their founding (TDivSE, AU or GE). They receive in all models some gene flow from the *Central* metapopulation 10 generations before sampling.

These models are derived from the best model inferred for the *Aegean* panel: parameter values estimated in the previous scenario were fixed and shown in light fonts, while newly estimated parameters are shown in bold fonts with search ranges shown within brackets. Further

information about the parameters can be found in Table S4 and the properties of the sampled populations are listed in Table M1_4.

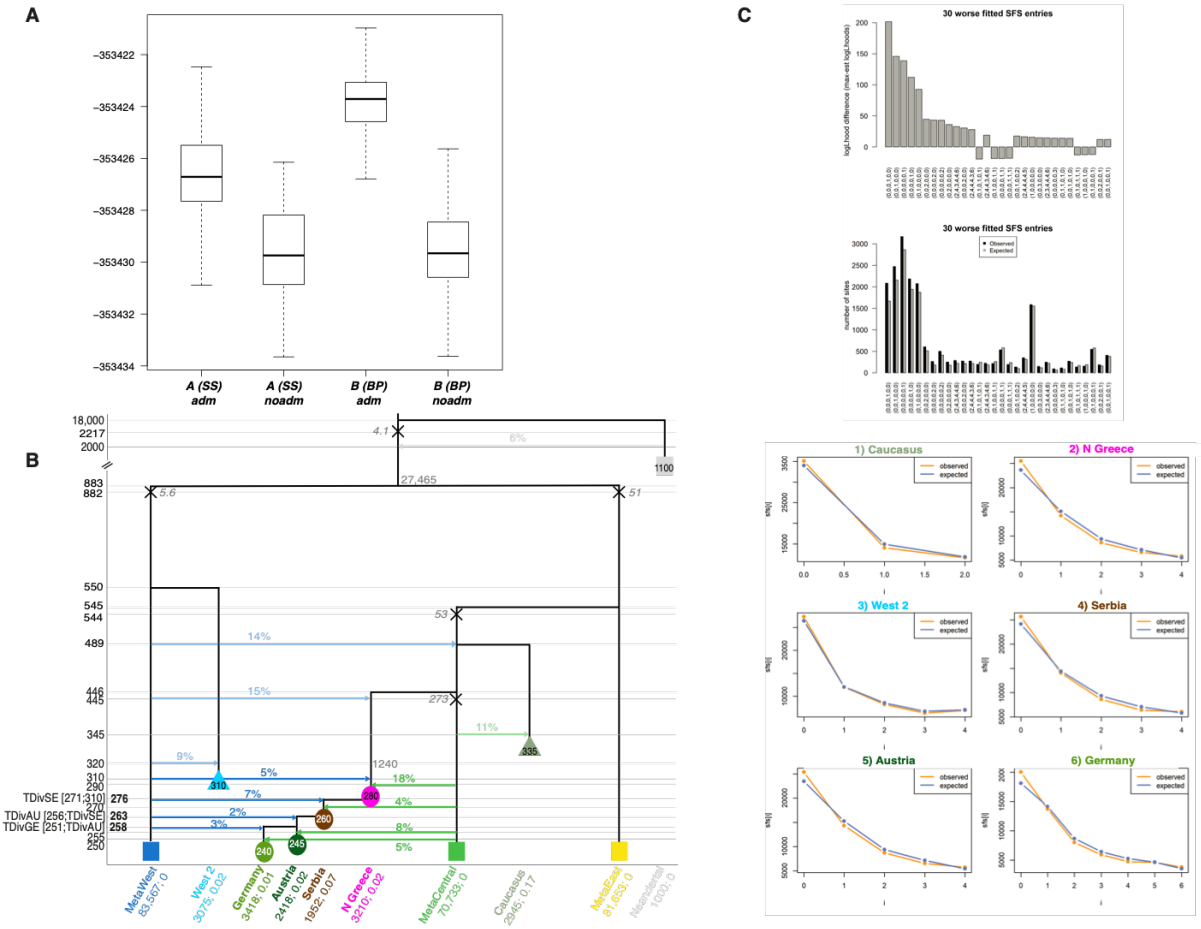

Figure M1_26 - Exploration of models best explaining the observed multi-dimensional *EurEF* SFS.

(**A**) **Distribution of the log10-likelihoods** obtained by simulation from the maximum likelihood point estimates for the four tested models described in Fig. M1_25. Higher likelihoods (top of the plot) reflect better simulations.

(**B**) **Schematic representation of the ML demographic model *SS_adm*.** Values of various parameters are reported with their 95% CI estimated from 100 parametric bootstrap estimations. Close to bottleneck symbols (X), we report the size of the bottleneck, conveniently modelled as a single generation bottleneck (*instbot fastsimcoal2* option).

(**C**) **Fit between the observed and estimated SFS under the ML model** (total on the upper panel; marginal for each population on the bottom).

Relationship between European HG populations

Tested models

In order to investigate the relationship between European HG populations, we included in the *HG* panel two previously published good quality genomes: the Upper Palaeolithic individual from Bichon and the Mesolithic individual from Loschbour (as *West 1* population), the two Mesolithic genomes from Vlasac (as *West 2*) as well as the genomes from the Mesolithic *Caucasus* HG and from *Iran* EF. We considered two models: a shared ancestry for Lochbour and Bichon (Fig. M1_27A) with a Western European ancestral population that splitted from the *Western* metapopulation before diverging into the Bichon and Loschbour-related populations, or two independent splits from the *Western* metapopulation for the populations ancestral to Bichon and Loschbour respectively (Fig. M1_27B). In both models, we estimated the demographic parameters related to the European populations (*West 1* and *2*) and used for the other parameters the values obtained for the best *Core* run.

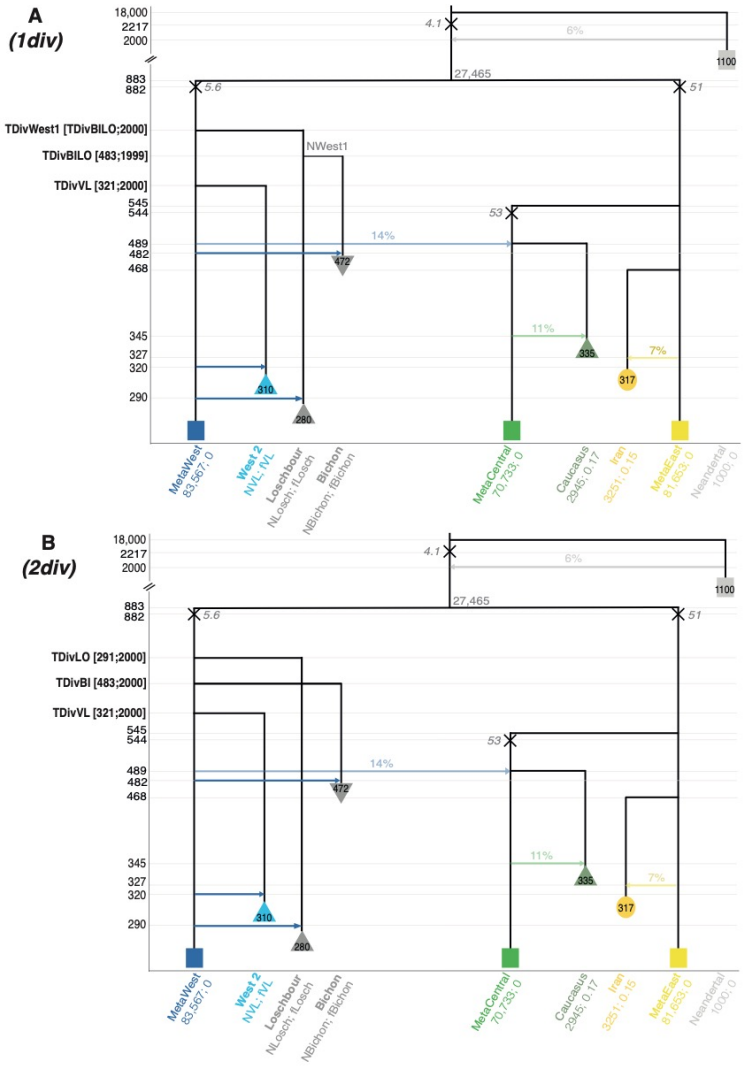

Figure M1_27 - Schematic description of models tested for the *HG* panel.

(**A**) **An ancestral *West 1* structure, i.e. a shared divergence event for the populations of *West 1* group.** Under this model, the *West 1* ancestral population has diverged TDivWest1 generations ago from the *Western* metapopulation with a resize event; then, the ancestral population split into populations related to Loschbour and Bichon respectively, TDivBILO generations ago; then both sampled populations received some gene flow from the *Western* metapopulation 10 generations before their sampling time.
(**B**) **Two independent divergence events for the populations of *West 1* group.** Under this model, populations related to Bichon and Loschbour independently split from the *Western* metapopulation TDivBI and TDivLO generations ago respectively, before they received independent pulses of gene flow from the *Western* metapopulation ten generations before sampling time.
These models are derived from the best model inferred for *Core*: parameter values estimated in the previous scenario were fixed and shown in light fonts, while newly estimated parameters are shown in bold fonts with search ranges shown within brackets. Note that all parameters of the *West 2* HG population (i.e. NVL, fVL, TDivVL, admMetaW.VL) are re-estimated here. Further information about the parameters can be found in Table S4 and the properties of the sampled populations are listed in Table M1_4.

Best model parameters estimate and SFS fit

We find that the model where Bichon and Loschbour have a common ancestry is best supported (Fig. M1_28A, Table S4). In this scenario, the ancestral population of these two individuals would have diverged from the *Western* metapopulation during the LGM, 787 [579-851] generations ago (~22.8 kya [16.7-24.7]), 100 generations after its divergence from the *Eastern* metapopulation. Note that our modelling did not constrain them to have diverged from the *Western* branch as they could have diverged before the LGM from the ancestors of the two metapopulations. It implies that during the LGM, the HG population split into (at least) three components (Western European i.e. *West 1*, Central European i.e. *West 2* and Near-Eastern/Caucasus HGs). Interestingly, the divergence of *West 1* and *2* HGs from the *Western* metapopulation has occurred after the very severe bottleneck linked to the divergence of the metapopulations (about two diploid individuals for one generation or 25 individuals for 10 generations). This bottleneck seems to have affected all European HGs and it explains their reduced diversity as compared to most EFs, e.g. HGs clustering all together on MDS plots, low heterozygosity, numerous short and sometimes long ROHs (Fig. 2). Then, the Loschbour and Bichon populations seem to have split during the LGM some 748 [501-748] generations ago (~21.7 kya [14.5-21.7]) and do not seem to have remained isolated after the LGM since they received about 8-14% [2-13] of admixture from the *Western* metapopulation. Contrastingly, the *West 2* HG population split more recently (~14.7 kya [11.3-22.9]) from the *Western*

metapopulation, suggesting some closer affinities with the local HG populations that admixed with the EFs of our study.

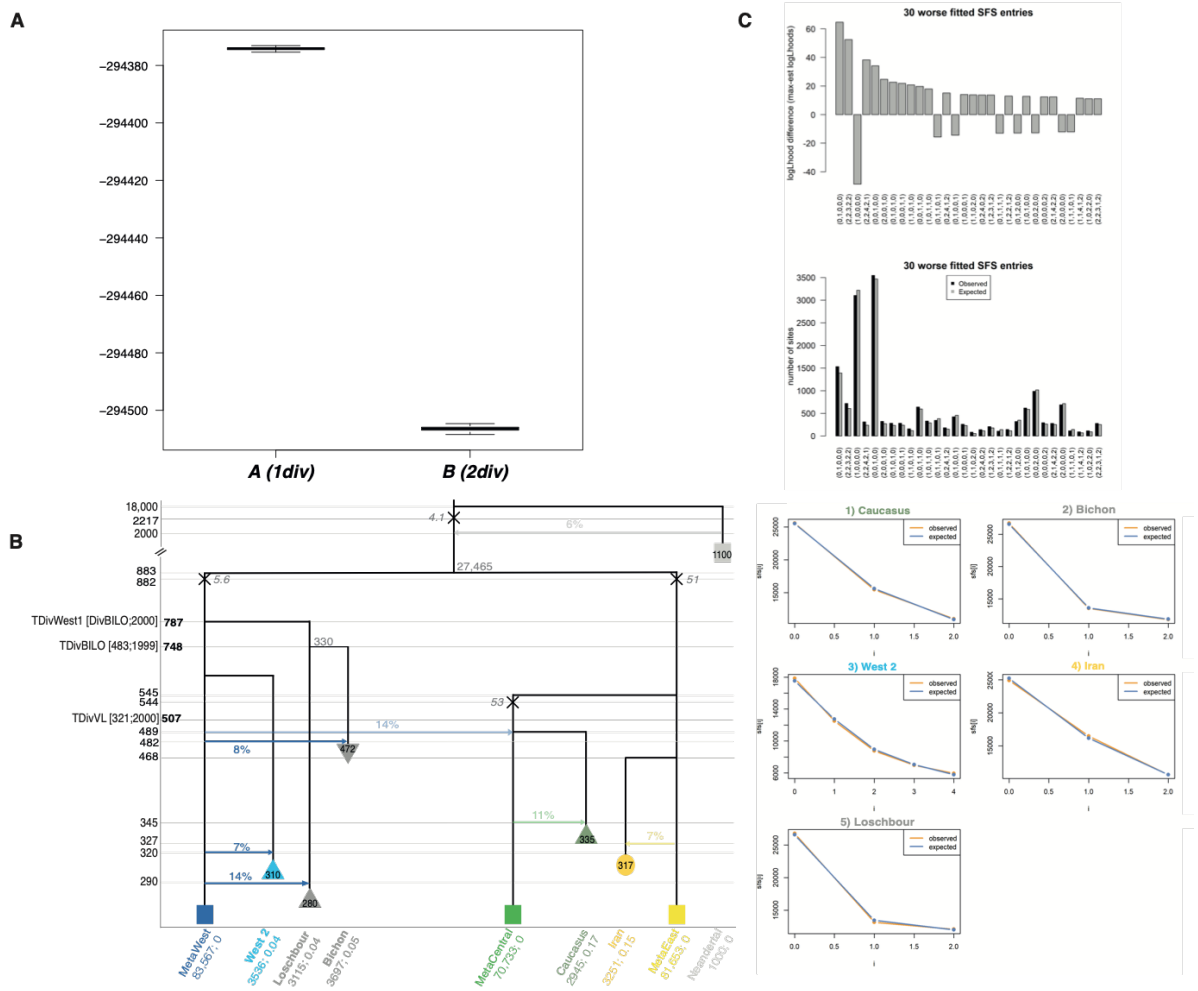

Figure M1_28 - Exploration of models best explaining the observed multi-dimensional *HG* SFS.

(**A**) **Distribution of the log10-likelihoods** obtained by simulation from the maximum likelihood point estimates for the two tested models described in Fig. M1_27. Higher likelihoods (top of the plot) reflect better simulations.

(**B**) **Schematic representation of the ML demographic model *A (1div)*.** Values of various parameters are reported with their 95% CI estimated from 100 parametric bootstrap estimations. Close to bottleneck symbols (X), we report the size of the bottleneck, conveniently modelled as a single generation bottleneck (*instbot fastsimcoal2* option).

(**C**) **Fit between the observed and estimated SFS under the ML model** (multidimensional SFS on the upper panel; marginal SFS for each population on the bottom).

Final Model

We assembled as a single *fastsimcoal2* input file (.par, available in our GitHub repository: https://github.com/CMPG/originsEarlyFarmers) a complex historical scenario including the maximum-likelihood parameter values of the best models for the 6 panels, corresponding to what is described in Fig. 3. This model includes the 12 sampled populations (Table M1_4), the three metapopulations and a Neandertal population. The parameters of the final model were estimated in the different scenarios (in particular, *West 2*'s parameters have been taken from the *HG* best scenario and the *Northern Greeks*' from the *Aegean* best scenario, details given in an excel data in the GitHub repository). For later use for *f*-statistics below, we added an African population for which we chose a haploid population size of 20,000, a null F_{IS}, and a split time from the European/SW Asian group some 97.1 kya (i.e. 3350 generations ago). Eventually, we sampled 11 ancestral populations at key moments of the demography we reconstructed: on the ancestral branch before the split between *Western* and *Eastern* metapopulations 25.6 kya (i.e. 883 generations ago), on the *Western* metapopulation branch just before the admixture occurring 14.2 kya (i.e. 490 generations ago), on the *Central* metapopulation branch just before and after this admixture (i.e. 490 and 489 generations ago); on the *Eastern* metapopulation branch at the time of split of the *C Anatolia* population; on the ancestral branch just after its founding and its admixture with the *Western* metapopulation (445 generations ago = 12.9 kya), then every 25 generations until the split into Aegean distinct populations (related to Northern Greece, Aktopraklık and Barcın; 320 generations ago = 9.3 kya).

Validating the proposed demographic model

Validation of parameter estimates from parametric bootstraps

In addition to the clear good fit found between the expected and observed SFS for the best model of each panel, the parametric bootstrap approach used to compute parameters confidence intervals actually brings a different validation of our estimation procedure. The fact that we can infer confidence intervals around our estimated parameters shows that we can correctly recover the simulated parameters with our approach. Indeed, for every panel, all but four point estimates (except admMetaW.LO in *HG*; admMetaC.BA in *Aegean*; NAU, f_SE in *EurEF)* were found to lie within the 95% confidence intervals, showing that if the true model was the one we estimated, we should be able to recover it with our maximum-likelihood estimation procedure.

Predictive Simulations

We validated the scenario from Fig. 3 using several complementary approaches, as outlined below. First, we tested if we were able to reproduce the actual genetic diversity observed within ancient samples under the scenario shown in Fig. 3. We simulated some genetic data as 10 million independent segments of 100 bp for 10 different replicates (we obtained between 1,829,466 and 1,834,676 polymorphic sites among the replicates) and a mutation rate of $7.5e^{-9}$ (chosen within the range of rates of the 6 panels).

We then computed the nucleotide divergence π_{XY} between all diploid samples. We found a significant correlation between simulated and observed distances computed over the neutral portion of the genome (Mantel test: r = 0.8, p-value = 0.001, with similar values for the 10 replicates). We then performed an MDS analysis of the π_{XY} genetic distances computed among the European and SW Asian simulated samples (Fig. 4A). We see that the simulated MDS is very similar to the observed one (Fig. 2A), with a cluster of European HGs, a cluster of western EFs and the group of the Iranian and Caucasus samples, showing that our model can reproduce very well the observed genetic structure of the ancient samples. By simulating data for ancestral populations, we gain some interesting insights on the admixture process between the *Central* and the *Western* metapopulations and the effect of genetic drift in the history of the first farmers of Europe and SW Asia (Fig. 4A).

Admixture analyses performed on simulated data are also really able to replicate observed patterns (compare Fig. S3A and B). It is possible to analyse simulated data for ancestral

populations too (Fig. 4B), an approach that offers the opportunity to better understand the biological processes causing the observed pattern. Indeed, we can see the increasing proportion of the blue ancestry from *CentralMeta (14.2 kya, pre-admixture)* to *CentralMeta (post-admixture)*, i.e. just before and one generation after its admixture with the *Western* metapopulation; this signal of admixture is also visible in the EF ancestral population (*12.9 kya*) that is modelled as receiving a second gene flow in our model. Furthermore, we can see that the green ancestry found in the EFs seems to appear progressively through time with its proportion increasing in the EF ancestral populations sampled every 25 generations. Note that the EF ancestors are changing only because of drift with a small population size estimated for this population (Ne = 620, 95% CI 72-2150). Thus, drift makes the EF genomes look like they were from a completely independent gene pool and erases the signal of admixture from the *Western* metapopulation, explaining why the hybrid nature of EFs had been previously unnoticed.

f-statistics

Confirming population assignment

As shown in Fig. S1, only very few comparisons violated our assignment. First, AKT16 from the NW Anatolia region showed some Western European HG ancestry (represented by the *West 1* individuals, Loschbour and Bichon) not present in Bar25. This led us to model these two *NW Anatolia* individuals as independent populations in the *fastsimcoal2* demogenomic analysis. Second, more Central European HG ancestry (represented by the *West 2* genomes from Vlasac, VLASA7 and VLASA32) was evidenced for Asp6 than Klein7 from Austria. However, variation in HG ancestry is expected in early Neolithic populations due to the ongoing process of admixture. Since these samples did not show variation in their affinities to other Neolithic samples, modelling them as a single population seems justified (Table M1_4).

Global model-fit using admixture graphs

To validate the model inferred by *fastsimcoal2,* we aimed at comparing *f*-statistics predicted under this model against those calculated from the data. Since the full graph could not be fit (see STAR Methods), we created a simplified admixture graph matching the model shown in Fig. 3 in structure, but containing fewer admixture edges. For this graph, no major differences between the *f*-statistics predicted under the model and those calculated from the simulated data were detected. For the best among 100 runs on this simplified graph (with the smallest final

score), only one *f*-statistics showed a barely significant Z-score: *f4(NGreece, Germany; Iran, Caucasus)*, predicted 0.000265, observed 0.001644, Z-score 3.131. But we note that the admixture graph analysis did not result in accurate estimates of all admixture proportions, as those were inferred rather variably among the 20 best runs (Fig. M1_29).

We next tested for discrepancies between the *f*-statistics predicted under this simplified graph and those calculated on the 12 populations used in model fitting with *fastsimcoal2* (Table M1_4). For this purpose, we refitted the admixture graph 100 times with *qpGraph.* We first conducted this analysis on the set of neutral SNPs used to infer the SFS for model fitting (Fig. M1_30). On this data, no *f*-statistics was violated in the best graph, indicating that this graph is compatible with all *f*-statistics calculated on the real data. Since this graph was constructed to match the model inferred by *fastsimcoal2,* the *fastsimcoal2* model therefore is also fully compatible with these *f*-statistics.

To study the impact of using ascertained markers, we next repeated this analysis on the *1240k* dataset (Fig. M1_31). While the graph still generally holds for this set of markers, the best run resulted in twelve *f*-statistics that were significant, albeit mostly barely so:

1. *f4(Serbia, NGreece; Germany, NWA_AKT); pred. 0.002634, obs. -0.002462 Z-score -3.279*
2. *f4(NGreece, Germany; NGreece, Germany); pred. 0.260349, obs. 0.253955 Z-score -4.045*
3. *f4(NGreece, CAnatolia; Germany, NWA_AKT); pred. 0.002508, obs. 0.009600 Z-score 3.568*
4. *f4(NGreece, NWA_BAR; Germany, West 2); pred. 0.000562, obs. 0.006496 Z-score 3.425*
5. *f4(NGreece, NWA_AKT; Germany, CAnatolia); pred. 0.003642, obs. 0.010112 Z-score 3.307*
6. *f4(NGreece, West 2; Germany, NWA_BAR); pred. -0.005740, obs. 0.000405 Z-score 3.522*
7. *f4(NGreece, West 2; Germany, NWA_AKT); pred. 0.004104, obs. 0.010444 Z-score 3.749*
8. *f4(NGreece, Loschb; Germany, NWA_AKT); pred. 0.003970, obs. 0.010613 Z-score 3.139*
9. *f4(NGreece, Caucasus; Germany, NWA_AKT); pred. 0.001200, obs. 0.007007 Z-score 3.004*
10. *f4(Austria, CAntatolia; NWA_AKT, Bichon); pred. -0.001021, obs. -0.008986, Z-score -3.237*
11. *f4(Austria, NWA_BAR; West 2, Bichon); pred. 0.000162, obs. -0.006377, Z-score -3.641*
12. *f4(Germany, NWA_BAR; West 2, Bichon); pred. 0.000230, obs. -0.006031 Z-score -3.796.*

This indicates that the marker choice is important and suggests that ascertained SNPs may hint at affinities not found when using neutral markers. However, we note that the graph used is likely too complex for the information contained in *f*-statistics and thus many branches and particularly admixture proportions were not reliably fitted with both sets of markers (large 90% quantiles across the 20 best runs, Fig. M1_30, Fig. M1_31). An important exception was the admixture event resulting in the source population of western EFs, which was inferred to have received 25% and 29% from the *Western* metapopulation (*WMeta*) on the *Neutral* and *1240k* set of markers, respectively, in agreement with a substantial contribution as inferred with *fastsimcoal2* (8-26% 95% CI, Fig. M1_20).

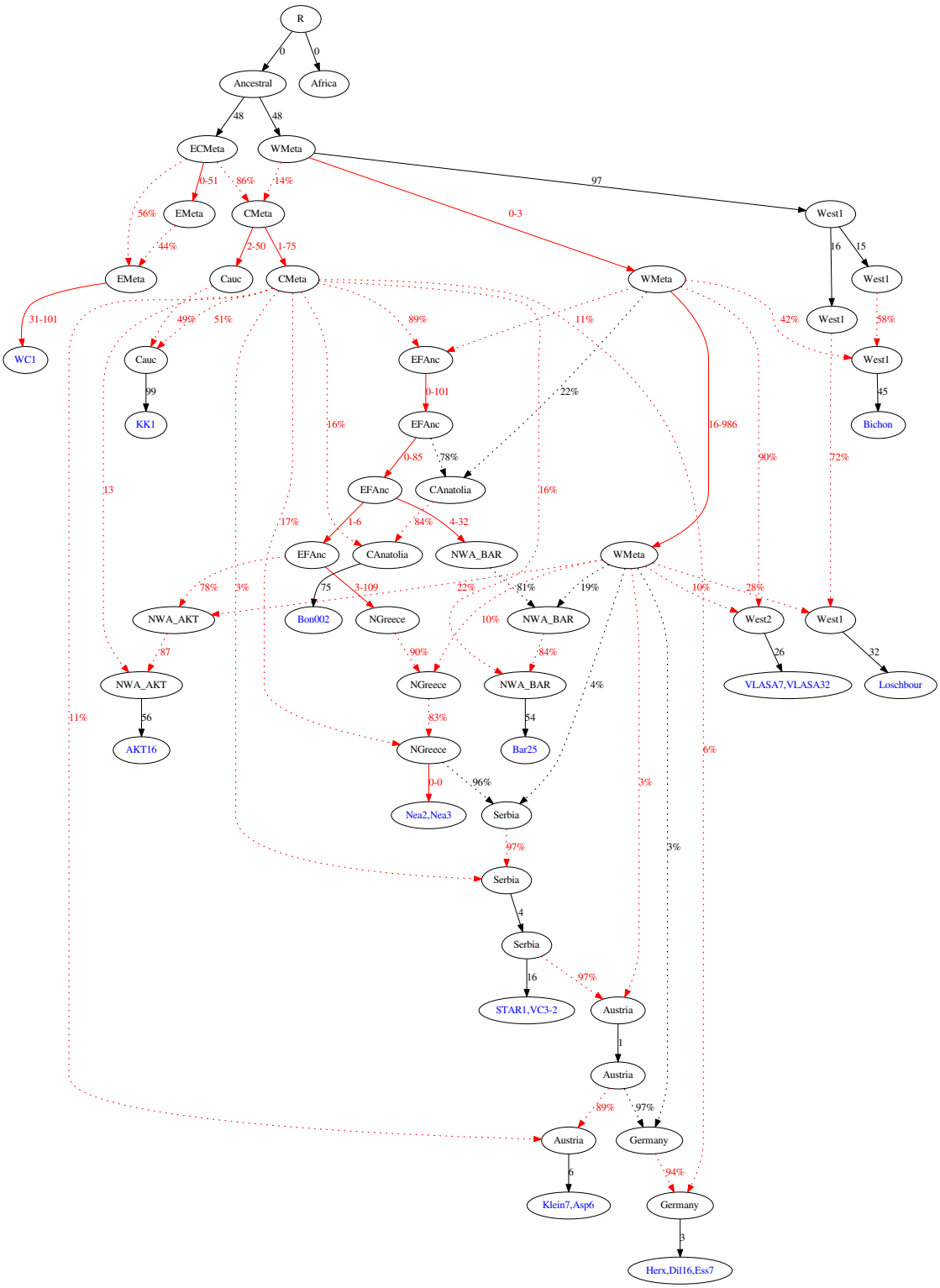

Figure M1_29 - Admixture graph fitted to data simulated under the *fastsimcoal2* model. Shown are the median branch lengths and mixture proportions among the best 20 graphs (lowest final score). Branches and mixture proportions are shown in red if the 90% quantile among the best 20 graphs was larger than given by half and twice the median (factor two) of the branch length or odds ratios, respectively. For outlier branches, the full range of estimates among the best 20 graphs are shown.

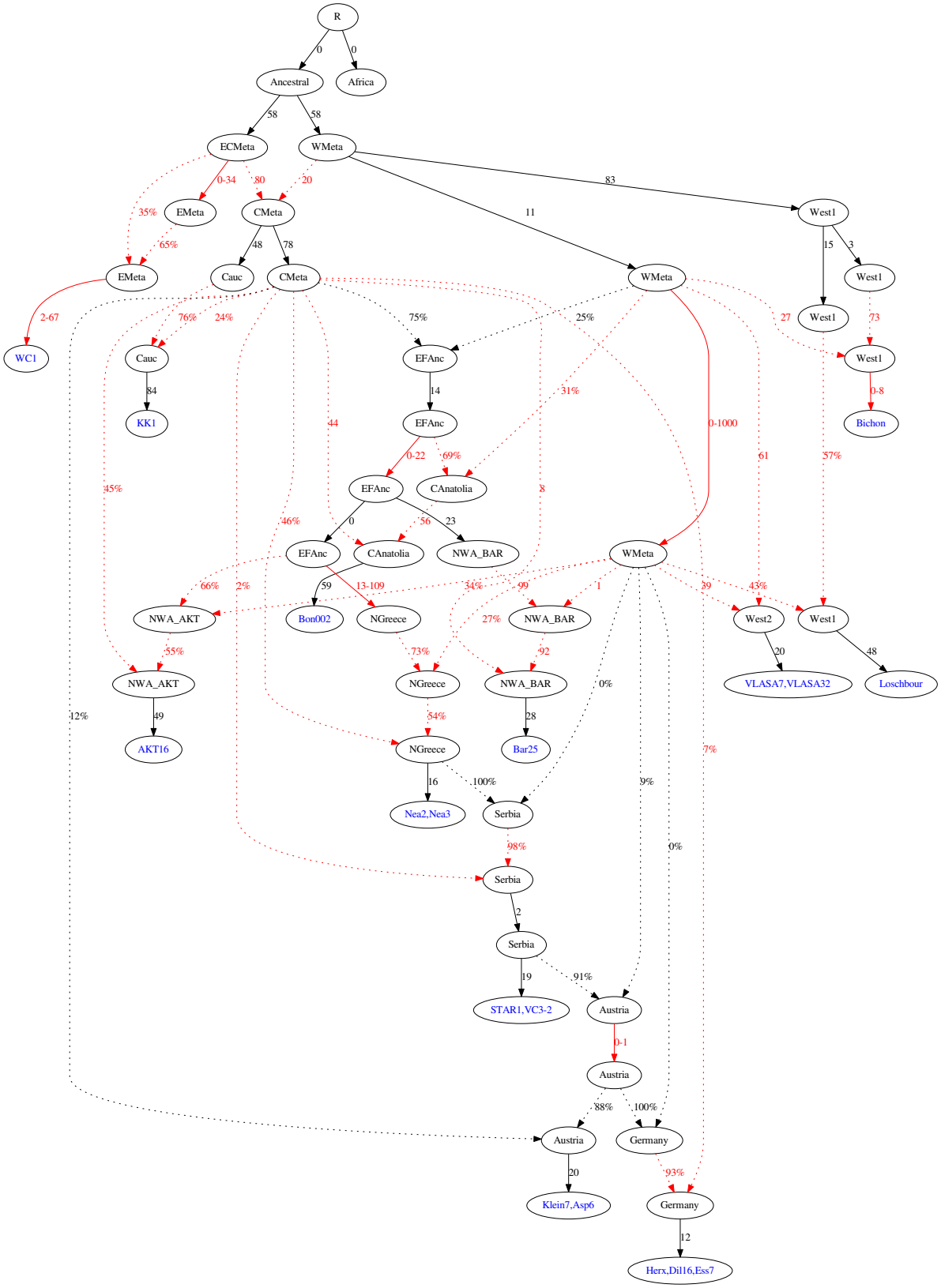

Figure M1_30 - Admixture graph fitted to the *EurEF* panel. Colour and labels as in Fig. M1_29.

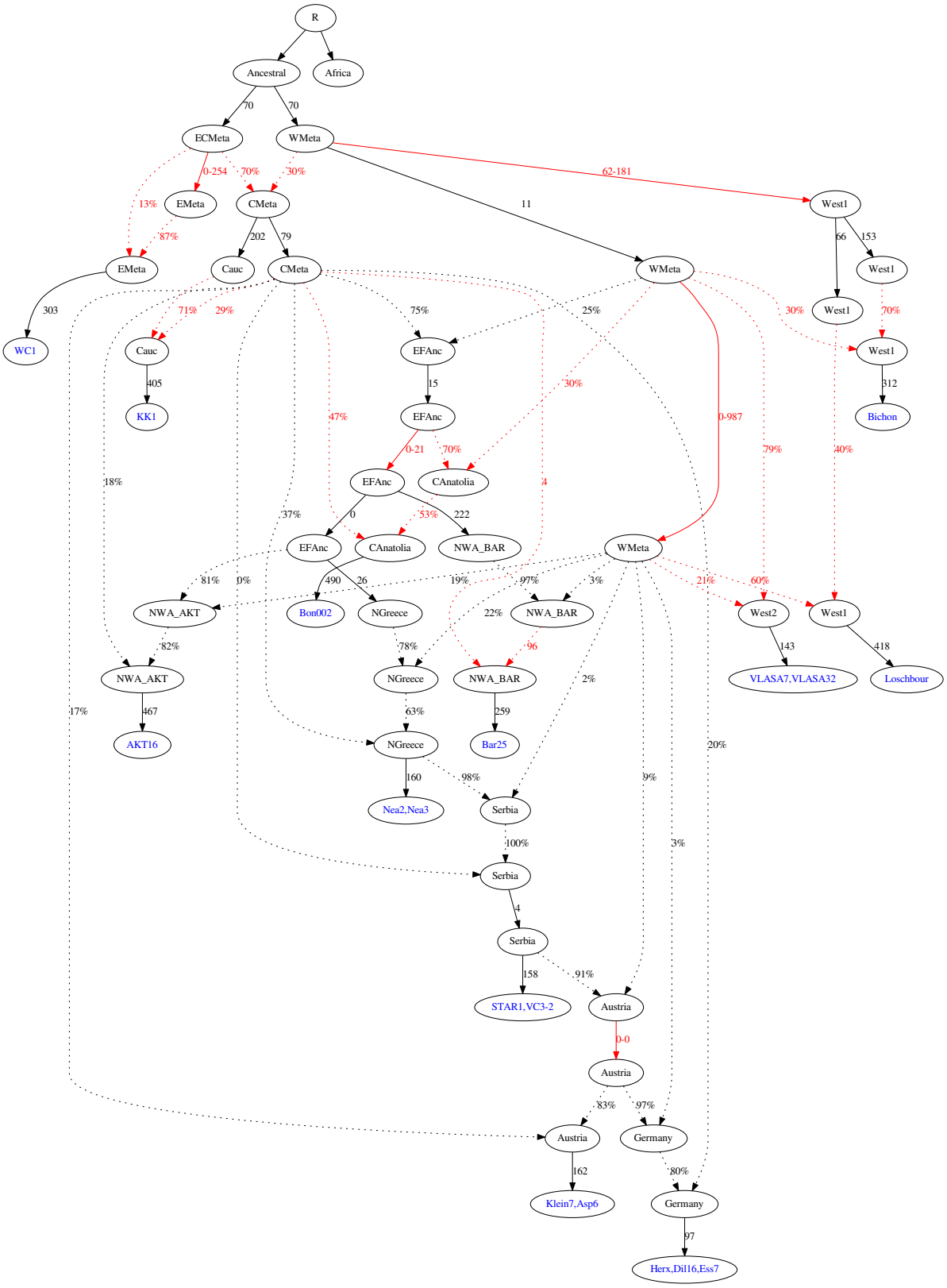

Figure M1_31 - Admixture graph fitted to the *1240k* dataset. Colour and labels as in Fig. M1_29.

Early separation of Central Anatolia from Northern Greece

The *fastsimcoal2* analysis revealed an early split between the population from the Konya Plain in *Central Anatolia* and the population ancestral to the Aegean EFs (Fig. M1_24). In the same analysis, admixture from the *Central* metapopulation was estimated to be substantially lower (2%) into *Central Anatolia* than into *Northern Greece* (9%). In contrast, the admixture from the *Western* metapopulation into *Central Anatolia* was estimated to be substantially larger (8%) than in *Northern Greece* (3%). Here we used two sets of *f*-statistics to support these findings.

Excess of Western ancestry in Central Anatolia

To confirm the excess of European "Western" HG ancestry in the Konya Plain, we used *f*-statistics of the form *D(NGreece/NWAnatolia, CAnatolia; HG_West, Outgroup)*. Here, *HG_West* is represented by the *West 2* individuals (*VLASA7* and *VLASA32*) from this study and the *West 1* individuals (*Bichon* and *Loschbour)*; *Central Anatolia* is represented by 1) the four *Anatolia_Boncuklu* samples available in *1240k* (Kılınç *et al.*, 2016) consisting of the *Bon002* sample used in the *fastsimcoal2* analysis and three additional low-depth samples, and 2) the additional sample *Pınarbaşı*, an Epipalaeolithic HG from the Konya Plain dating to 15,592-15,023 cal BP (Feldman *et al.*, 2019). As shown in Fig. S4, we observed a significant excess of shared drift between *HG_West* and *CAnatolia* compared to *NGreece/NWAnatolia*. This signal was particularly strong for *Pınarbaşı* (all comparisons involving > 800,000 SNPs, Table S5). For *NWAnatolia_Boncuklu*, the signal was only present when compared to the late Neolithic Greek samples (*Pal7* and *Klei10*).

A notable exception from the general pattern is *AKT16* that shows excess shared drift with *HG_West* compared to *CAnatolia* (Fig. S4), in line with elevated *Western* metapopulation admixture estimated by *fastsimcoal2* (Fig. 3) and also evidenced by the admixture analysis (Fig. 2B).

We conducted similar comparisons using *Neolithic Iranian* and *HG_Caucasus* samples from *1240k* of the form *D(NGreece/NWAnatolia, CAnatolia; Neolithic Iranian/HG_Caucasus, Outgroup).* Here, *Neolithic Iranian* are represented by *WC1*, *Iran_TepeAbdulHosein_N* and *Iran_GanjDareh_N* (Gallego-Llorente *et al.*, 2016) and *HG_Caucasus* is represented by *Georgia_Kotias* and *Georgia_Satsurblia* (Jones *et al.*, 2015). None of these comparisons revealed a significant difference between the samples from *NGreece/NWAnatolia* and those from *CAnatolia* in their level of shared drift with neither *Neolithic Iranian* nor *HG_Caucasus* (Table S5), in line with the *fastsimcoal2* model that places these populations as sister groups.

Scarcity of Eastern metapopulation ancestry in Central Anatolia

To confirm the scarcity of *Eastern* metapopulation ancestry in the Konya Plain in Central Anatolia, we used individuals from Israel from the Chalcolithic period (*Israel_C*), the PPNB period (*Israel_PPNB*) and the Natufian culture (*Israel_Natufian;* all Israel individuals were from (Lazaridis *et al.*, 2016)) as *Eastern* proxies and we performed *f*-statistics of the form *D(NGreece/NWAnatolia, CAnatolia; Israel, Outgroup)*. Most of the comparisons involving *NWAnatolia_Boncuklu* showed a strongly significant excess of shared drift between Israel and *NGreece/NWAnatolia* (Fig. S4), in line with the scarcity of *Eastern* metapopulation admixture in *Central Anatolia* inferred by *fastsimcoal2*. Interestingly, however, this signal was not replicated when representing *Central Anatolia* by *Pınarbaşı* (Fig. S4): in that case, none of the comparisons was significant. Similarly, no significant comparisons were found when using either *Neolithic Iranian* or *HG_Caucasus* instead of *Israel* (Table S5).

Structure among HG individuals

The model inferred by *fastsimcoal2* (Fig. 3) predicts a deep split between *Loschbour, Bichon* and the *Western* metapopulation. To validate these findings, we investigated the relationships among European HG samples using *f*-statistics. First, we calculated *f*-statistics of the form *D(Loschbour, Bichon; West 2, Outgroup)*. In line with the deep split inferred by *fastsimcoal2*, none of these comparisons was significant.

We next investigated if the European HG ancestry in *NW Anatolia* individuals can be attributed more to *West 2* than to *Bichon*, as predicted by the *fastsimcoal2* model. In line with these predictions, *f*-statistics of the form *D(Bichon, West 2; AKT16/Bar25, Outgroup)* resulted in significant, negative values (Z-score from -3.218 to -4.826).

Interestingly, the comparisons of the form *D(Loschbour, West 2; AKT16/Bar25, Outgroup)* were not significant and additional comparisons of the form *D(Loschbour, Bichon; AKT16/Bar25, Outgroup*) resulted in significant positive values. We interpret this as a result of sample age (Bichon is much older than Loschbour) and the admixture from *West 2* that was consequently estimated at a much older time for Bichon than Loschbour, allowing for shared drift in the *West 2* component in both Loschbour and AKT16/Bar25.

Noteworthy are further the many significant and positive *f*-statistics of the form *D(West 1, West 2; HG_Adriatic, Outgroup)*, where *HG_Adriatic* is represented by Italian and Croatian Mesolithic and Epipalaeolithic HGs (see Table S5). This may point to an enhanced level of drift in the Mesolithic samples from Vlasac (representing *West 2* here) and their relative uniqueness when compared to other HG populations in the region.

Differentiation within Greece and Central Anatolia

Peloponnese

Based on differences between Neolithic samples from the Peloponnese to those from other regions, it was suggested that the Peloponnese might constitute a wave of Neolithization other than the Balkan route (Mathieson *et al.*, 2018). Specifically, it was shown that Neolithic samples from the Peloponnese are shifted towards Caucasus HG-like ancestry and away from WHG-like ancestry (represented by *West 1* in this study, i.e. Bichon and Loschbour), potentially due to contacts with Central Anatolian populations like Kumtepe or Tepecik-Çiftlik that show similar patterns. We benefited from the additional samples generated in this study to shed additional light on the relationship between Neolithic samples from the Peloponnese and Northern Greece. We confirm that *Greece_Peloponnese* (four individuals from Diros Cave and one from Franchthi Cave labelled as *Greece_Peloponnese_N*, all Middle or Late Neolithic) are indeed more Caucasus HG-like (represented here by *Georgia_Kotias* and *Georgia_Satsurblia*) than *Rev5* from Northern Greece. However, this pattern is stronger when Caucasus HG is substituted with *Iran_GanjDareh* individuals (interestingly not WC1): *D(NGreece, Greece_Peloponnese; Iran_GanjDareh, Mbuti)* is significant and positive for all Greek Neolithic samples (Table S5). However, this signal is not obviously connected to Kumtepe or Tepecik-Çiftlik, as *Greece_Peloponnese* samples also show more influence from the Levant (with *Israel* samples as proxies) than *NWAnatolia*, *Greece* (Late Neolithic Greece) and *Central European* samples (Klein7), see Fig. S4. In addition, the same analyses conducted with the Peloponnese early Neolithic sample I5427 (labelled here *Diros_EN*) showed that a higher amount of Levantine ancestry was already present in the early Neolithic, at least compared to the Barcın samples (Bar8 and Bar25). The sample *Diros_EN* also shows more shared drift with *Iran_GanjDareh* than *Nea2* and more influence from *Pınarbaşı* than the later Peloponnese group (*Greece_Peloponnese_N*).

Northern Greece and NW Anatolia

In general, samples from *Northern Greece* and *NWAnatolia* show a relatively high level of heterogeneity in shared drift with samples from other populations (Fig. S4, Fig. M1_32). Particularly interesting differences include the excess shared drift with Levantine samples (with Israel samples as proxies) found for the Nea Nikomedeia sample Nea3 (the older) but not Nea2 (the younger) when comparing these to other samples from the region (Fig. S4). Rev5, which is of similar age as Nea3, also shows a bit more of the Levantine ancestry, suggesting a decrease

in that ancestry over time. While there is no significant f-statistics when directly comparing Rev5 to Nea2 or Nea3, it does seem that all comparisons with the 4 *HG_West* genomes in our dataset (VLASA7, VLASA32, Loschbour, Bichon) show a trend for an excess shared drift with HG samples for Rev5 and Nea3 (Table S5).

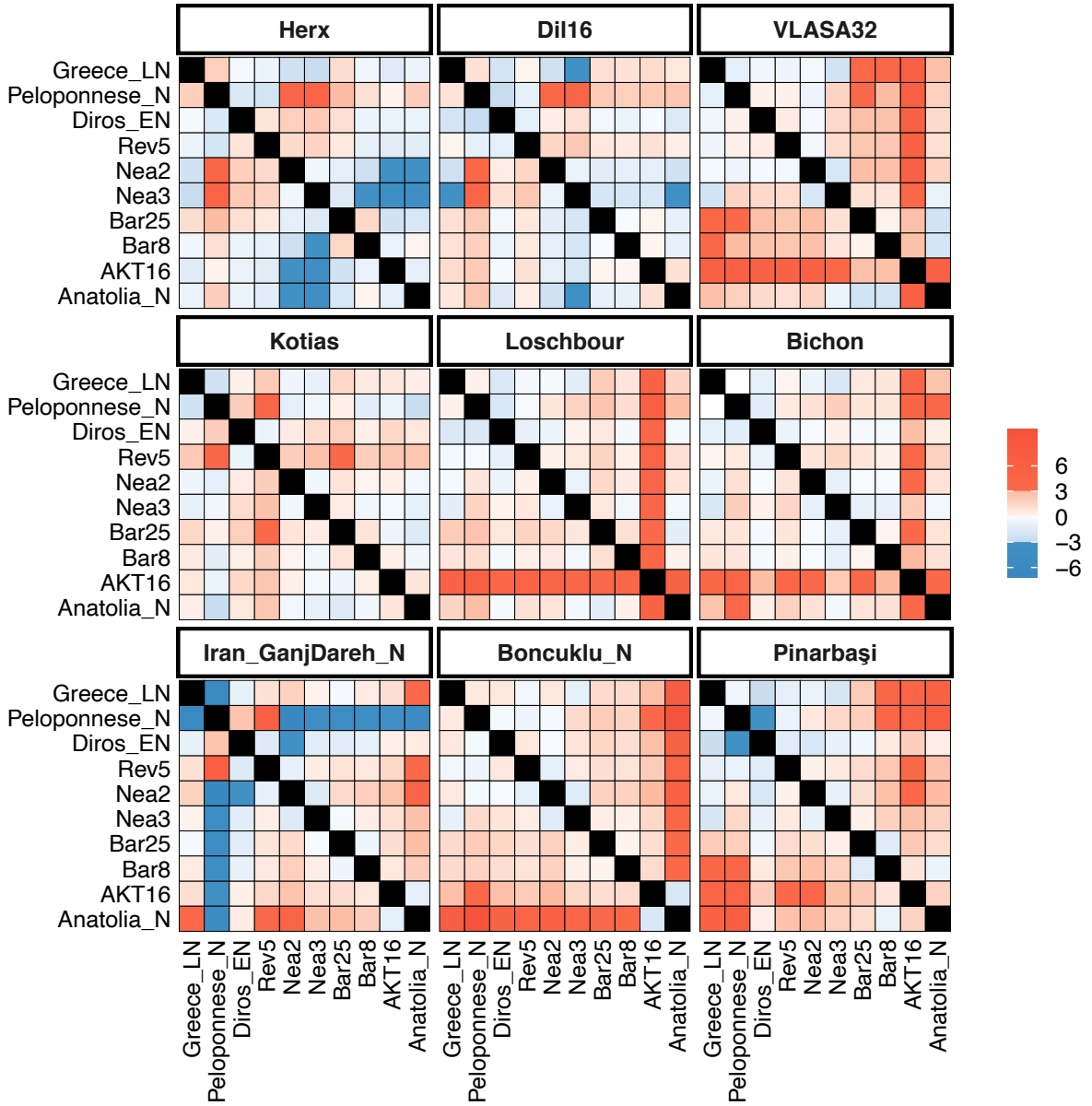

Figure M1_32 - Heatmap of Z-scores of *f*-statistics of the form *D(Greece/NWAnatolia1; Greece/NWAnatolia2; Test; Outgroup)* with Mbuti as an outgroup and *Test* noted as title of each graph. Red shades indicate positive Z-scores, blue shades indicate negative Z-scores. Non-significant Z-scores (below +-3) are shown as shades, while significant Z-scores are indicated with strong colours.

Central Anatolia

We investigated the relationships inside Anatolia by comparing our samples from *Northern Greece* and *NWAnatolia* to Central Anatolian genomes from *Pınarbaşı* (Epipalaeolithic), *Boncuklu* (Neolithic), *Kumtepe* (Chalcolithic) and *Tepecik-Çiftlik* (Neolithic). A first result was that *Tepecik-Çiftlik* is cladding outside the branch represented by *Boncuklu* and *Pınarbaşı* as *D(Boncuklu, Pınarbaşı, Tepecik-Çiftlik, Outgroup)* is not significant (Z-score 0.76), while the two alternative *f*-statistics with either *Boncuklu* or *Pınarbaşı* as a sister group were highly significant (Z-score < -6.29, Table S5). *Tepecik-Çiftlik* also seems to clad outside a branch leading to *Boncuklu*, *Pınarbaşı* and all Aegean EFs as all *f*-statistics of the form *D(Aegean, Boncuklu / Pınarbaşı, Tepecik-Çiftlik, Outgroup)* were not significant (Z-scores < |2.051|), but 8 of 24 *f*-statistics of the form *D(Aegean, Tepecik-Çiftlik, Boncuklu / Pınarbaşı, Outgroup)* and *D(Boncuklu / Pınarbaşı, Tepecik-Çiftlik, Aegean, Outgroup)* were significant, in particular where *Aegean* is represented by the *NWAnatolian* samples and an early Neolithic sample from Peloponnese (*AKT16, Bar8, Anatolia_N* and *Diros_EN*, Table S5).