SUPPLEMENTARY INFORMATION

The timing of pigmentation lightening in Europeans

Sandra Beleza, António Múrias dos Santos, Brian McEvoy, Isabel Alves, Cláudia Martinho, Emily Cameron, Mark D. Shriver, Esteban J. Parra, Jorge Rocha

CONTENTS

Supplementary Protocol: Tag-SNPs and microsatellite marker genotyping

rs642742A>G and rs2733831A>G were typed by PCR-Restriction Fragment Length Polymorphism (RFLP). The rs642752A>G polymorphism was amplified within a 231 bp fragment with primers 5'- CTGGTTCTTATCTGAAGGACT-3' (forward) and 5'-AAATCCATTCTGTGATAATAACATG-3' (reverse). The underlined base of the reverse primer is a mismatched base, intentionally inserted to create a *BseG*I restriction site. PCR reactions contained 0.5-5 ng of genomic DNA, 0.3 µM of each primer, 0.2 mM of each deoxynucleotide triphosphate (dNTP), 10 mM Tris–HCl (pH 8.8), 50 mM KCl, 0.08% Nonidet, 1.5 mM MgCl2 and 1 U Taq polymerase. Samples were denatured for 3 min at 95ºC, followed by 20 cycles of 94ºC for 30 s, 56ºC for 30 s, and 72ºC for 30 s, by 15 cycles of 94ºC for 30 s, 54ºC for 30 s, and 72ºC for 30 s, and finally by a 10 min-extension at 72ºC. Digestions (1 U/µl) were performed at 55ºC for 1 h and DNA fragments were visualized by silver staining after non-denaturing electrophoresis separation in 9% polyacrylamide gels. Digestion with *BseG*I generated 205- and 26 bp fragments in the presence of the rs642742*G allele.

The rs2733831A>G polymorphism was amplified within a 124 bp fragment with primers 5'-CTACCTGAATGCACTGATTTGAA-3' (forward) and 5'-GCATTAAGAAGACACAATCTGTAGAT-3' (reverse). The underlined base of the reverse primer is a mismatched base, intentionally inserted to create a *Bsp143I* restriction site. PCR reactions were identical to rs642742A>G, except for the thermocycling conditions. In this case, samples were denatured for 3 min at 95ºC, followed by 35 cycles of 94ºC for 30 s, 59ºC for 30 s, and 72ºC for 30 s, and finally by a 10 min-extension at 72ºC. Digestions (1 U/µl) were performed at 37ºC for 1 h and DNA fragments were visualized by silver staining after non-denaturing electrophoresis separation in 9% polyacrylamide gels. Digestion with *Bsp143I* generated 97- and 27 bp fragments in the presence of the rs2733831*G allele.

rs1426654G>A and rs16891982C>G were genotyped using allele-specific PCR with universal energy transfer-labelled primers as described in (Myakishev et al. 2001) by the company Prevention Genetics (Marshfield, Wisconsin, USA).

Microsatellites associated with *KITLG*, *TYRP1*, *SLC24A5* and *SLC45A2* were genotyped by PCR amplification of independent multiplex reactions. Multiplexes were put together based on non-overlapping marker size ranges and/or unique fluorescent dyes. All primer sequences were designed using PRIMER3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and are listed in Table SP1 (see below). *KITLG* and *TYRP1* multiplex-PCR amplifications were carried out using the QIAGEN Multiplex PCR kit (Qiagen) at 1x Qiagen multiplex PCR master mix and 0.5–5 ng of genomic DNA in a 5 µl final reaction volume. Final primer concentration in the reaction was 0.2 µM for all primers. Thermocycling conditions of both reactions included a pre-incubation for 15 min at 95°C, followed by 30 cycles of 94°C for 30 s, 60°C for 90 s, and 72ºC for 60 s, and a final incubation of 60°C for 30 min.

SLC24A5 and *SLC45A2* multiplex-PCR amplifications were performed in 96-well plates in 6 µl reactions containing ~45 ng DNA, 0.075 µM of all primers, 0.03 U Platinum Taq, 100 µM each dNTP, and 1.5 mM $MgCl_2$. Thermocycling conditions started with a 2-min denaturation at 95 ^{o}C, followed by 27 cycles of 95ºC for 40 s, 55ºC for 75 s, and 72ºC for 40 s, and by a final incubation of 6 min at 72^{o}C. These genotypings were carried out at the company Prevention genetics (Marshfield, Wisconsin).

Fragment analysis and weight determination were performed with automated high-throughput scanning fluorescence detectors that had the ability to detect and separate FAM, TET, NED, HEX, and ROX dyes.

Table SP1. Sequence of the primers constructed for multiplex amplification of the microsatellites associated with the pigmentation genes studied.

Gene	**Marker**	**Forward and Reverse Primer (5'-3')**
TYRP1	$(AT)_n1$	ACAGAAAAGCAAGCAAAGAAA
		GGTACGAAAGAAAATAAAACTATAGCA
	$(TC)_n$	AAAGCTGTCATTCTTACCTTGG
		GGGCATATTTGGTCATGTCA
	$(AT)_n2$	TGGCAATCTGCTTTTAGACA
		AGATCAGGGCTGCCTATCTTA
	$(AT)_n3$	CCTTGGTCCTCAGCTTTCA
		ATTAGCCAATGGGATGTCCA
SLC24A5	$(TG)_n$	CATTAAGGGCTGCTTGCATT
		TTACCCACAAGCCATGTGAA
	$(AT)_n$	TGGGAAAGGAAGACCTACCC

		CCCATTAATTCTTTGGGAGTTG
	$(AC)_n1$	AAAGGAGGGCTAAATGTCCAA
		TTCCACCTAGTGGGATGAGTG
	$(CA)_n$	AAAGTTTAAATGCCTTCCCAAA
		CAGCAACTACAATAGCATTCTTATCA
	$(AC)_n2$	GCCACATGCCCAGATAAAAT
		CACTGTCCCCAGTTTCCCTA
SLC45A2	$(TG)_n1$	GAGCCCCATAAATGACACCA
		CCCGATGTGCTAGCAATTTT
	$(TC)_n$	CGTAGCAATTATTGTCCCCATT
		TGCTGAGCTTCTGGAGCATA
	$(TG)_n2$	ATTTGTCTGGGGAAAGGAAA
		GGAAAGGTGACCCACAGCTA
	$(AC)_n$	AACCCCTTGTCTTGTAGAAGGTT
		AAGGAACCAATGGGGAAGAT
	$(TG)_n3$	GAGAAAGTCGTTCTGATTCTTTTC
		GGTGAAGATAGGTGCTCTAACAA
KITLG	$(AC)_n$	TGCCTCCTTCATTTATAGCTCT
		TGTGCGAATTTGTGTTCAG
	$(TG)_n$	CCTGAGTAGCTGGGACTACAG
		AGGGCTTGATTTCAGGAGTT
	$(GATA)_n$	GCTGGGGAACACTCATGTAT
		AGACTGGAATGTATACCGTTGG
	$(GT)_n$	TGCACCCTAATCGTCTACC
		CACCACATAATCTCCTGCAC

Supplementary text: Modelling haplotype variation

In the forward simulations, as the frequency of the derived SNP allele at each pigmentation gene rises, the ancestral microsatellite haplotype to which the derived SNP allele was originally associated is broken down by recombination and mutation at the linked microsatellite markers. During this process, it is also possible that the ancestral haplotype is regenerated by mutation or recombination. However, when the a relatively high number of marker microsatellite is used, as in the present work, the probability of regenerating the ancestral haplotype by mutation becomes increasingly small and may be neglected. When this simplification is used, each microsatellite locus may be defined by one of two states, corresponding to the presence or absence of the allele that became originally associated with the derived SNP variant.

For example, suppose there are four haplotypes consisting of one SNP with one derived (D) and one ancestral allele (A), and five microsatellite loci, each with several tetranucleotide repeat alleles of different size:

H1: 100-200-300-D-400-500
H2: 100-204-300-D-400-500
H3: 100-200-312-A-416-508
H4: 116-212-320-A-424-528

with H1 representing the haplotype where the derived SNP allele originally arose. While a 204→200 one-step mutation in the second (from the left) microsatellite locus of haplotype H2 may regenerate the ancestral haplotype, we note that the waiting time for this mutation is substantially larger than the time necessary for mutation to hit other microsatellites at the H2 haplotype, or for H2 to be modified by recombination. In this setting, the allelic states at the microsatellite loci may be re-coded as:

H1: 1-1-1-D-1-1
H2: 1-0-1-D-1-1

H3: 1-1-0-A-0-0
H4: 0-0-0-A-0-0

where 1 represents the allele that became originally associated with the derived SNP variant at each microsatellite locus, and 0 pools all microsatellite alleles that were not originally present in the ancestral haplotype. Using this approximation (not taking into account the mutational regeneration of the ancestral alleles 0→1), the relevant mutation events are 1→0. Mutations involving non-ancestral-associated microsatellite alleles (0→0) do not change the allelic state and do not need to be taken into account to assess the proportion of derived lineages that preserved the ancestral haplotype (P_{anc}). This simplification has the advantage of considerably reducing computation time and avoiding assumptions about unknown details of the microsatellite mutation model.

Note that regeneration through recombination keeps being taken into account. For example, recombination between the third and the second microsatellite loci involving haplotypes H2 and H3 will regenerate haplotype H1

H2: 1-0-1-D-1-1
H3: 1-1-0-A-0-0

Recombinant 1: 1-0-0-A-0-0
Recombinant 2: 1-1-1-D-1-1 (=H1)

Table S1. Features of tag-SNPs and microsatellite markers used to study haplotype variation at the four pigmentation loci.

Gene/Marker name	Chrom[a]	Physical position (Mb)[b]	Repeat Structure in UCSC	nº of alleles[c]	Mutation rate	Recombination rate[d]
KITLG						
$(AC)_n$	12	86,912,262	$(AC)_3$ag$(AC)_8$atacatgc$(AC)_6$	3	0.000003	0.00149
$(TG)_n$	12	87,264,517	$(TG)_{20}$	11	0.00041	0.00083
$(GATA)_n$	12	87,529,775	$(GATA)_{10}$	6	0.00009	0.00070
$(GT)_n$	12	87,793,170	$(GT)_{24}$	13	0.00015	0.000009
rs642742	12	87,823,877	-	2	0.0	-
TYRP1						
$(AT)_n1$	9	12,510,605	$(AT)_{18}$	19	0.00176	0.00006
$(TC)_n$	9	12,629,663	$(TC)_{14}$	15	0.00021	0.00022
$(AT)_n2$	9	12,673,207	$(AT)_{24}$	23	0.00154	0.00063
rs2733831	9	12,693,484	-	2	0.0	0.00038
$(AT)_n3$	9	12,746,781	$(AT)_{27}$	28	0.00015	-
SLC24A5						
$(TG)_n$	15	46,035,800	$(TG)_{24}$	12	0.00039	0.000006
$(AT)_n$	15	46,197,288	$(AT)_{12}$	10	0.00016	0.00205
rs1426654	15	46,213,786	-	2	0.0	0.00001
$(AC)_n1$	15	46,239,976	$(AC)_{16}$	11	0.00011	0.00003
$(CA)_n$	15	46,258,895	$(CA)_6$ta$(CA)_6$tacacaa$(CA)_6$	4	0.00002	0.00122
$(AC)_n2$	15	46,315,582	$(AC)_{21}$	12	0.00035	-
SLC45A2						
$(TG)_n1$	5	33,796,742	$(TG)_{21}$	13	0.00030	0.00139
$(TC)_n$	5	33,871,770	$(TC)_{23}(TG)_8$	25	0.00095	0.00068
$(TG)_n2$	5	33,977,160	$(TG)_{10}(CG)_3$	3	0.00003	0.00053
rs16891982	5	33,987,450	-	2	0.0	0.00045
$(AC)_n$	5	34,003,646	$(AC)_{23}$	14	0.00032	0.00045
$(TG)_n3$	5	34,004,928	$(TC)_{10}(TG)_{16}$	19	0.00139	-

[a]Chrom=Chromosome

[b] Data from UCSC genome web browser (http://genome.ucsc.edu).

[c] Total number of alleles observed in African, Asian and European populations.

[d] Recombination rate between adjacent markers, measured to the right-side of the first marker (see Figure 1).

Table S2. Sweep age and selection coefficients assuming that the European population sizes are recovering exponentially after a bottleneck

SNP, allele and demographic model	Dominant Model		Additive Model	
	s [b]	Age (years) [c]	s [b]	Age (years) [c]
rs1426654, *SLC24A5**A				
N_e=1,000; γ=0.0020 [a]	0.160	12,432	0.048	8,736
rs16891982, *SLC45A2**G				
N_e =1,000; γ=0.0020 [a]	0.049	15,036	0.031	14,560
rs2733831, *TYRP1**G				
N_e =1,000; γ=0.0020 [a]	0.015	15,652	0.021	14,028
rs642742, *KITLG**G				
N_e =1,000; γ=0.0020 [a]	0.011	27,160	0.011	30,884

[a] Growth (recover) phase in Europe after bottleneck. N_e = effective population size at the end of the bottleneck; γ = population growth rate (see Material and Methods).
[b] Selection coefficient.
[c] Assuming a generation interval of 28 years (Fenner 2005).

Table S3. Age and selection coefficient estimates for other *KITLG* candidate SNPs.

Tag SNP	Allele	Dominant model		Additive model	
		s[a]	**Age (years)**[b]	***s***[a]	**Age (years)**[b]
rs10732643	*KITLG* *G-Eur[c]	0.01	42,308	0.01	39,480
rs10732643	*KITLG* *G-EAsi[d]	0.009	29,456	0.01	34,720
rs1881227	*KITLG* *T-Eur[c]	0.015	42,251	0.01	39,424
rs1881227	*KITLG* *T-EAsi[d]	0.009	51,156	0.01	36,680
rs3782181	*KITLG* *A-Eur[c]	0.010	28,224	0.01	37,856
rs3782181	*KITLG* *A-EAsi[d]	0.009	23,324	0.01	33,124

[a] Selection coefficient (see Material and Methods).
[b] Assuming a generation interval of 28 years (Fenner 2005).
[c] *KITLG* *G lineages sampled in the European population.
[d] *KITLG* *G lineages sampled in the East Asian population.

Supplementary Figure Legends

Figure S1. Results from accuracy tests showing the fraction of known values (y axis) falling in different HDR intervals (x axis).

Figure S2. Microsatellite allele frequency distributions in the pooled sample of European, East Asian and African populations within ancestral (black bars) and derived (yellow bars) lineages defined by tag SNPs at four pigmentation genes. V=variance in repeat number; H=heterozigosity; n= number of chromosomes.

Figure S3. Median-joining network showing the relationships between SNP/microsatellite-haplotypes associated with the SNP rs642742 at the *KITLG* locus. Ancestral and derived lineages at rs642742 are separated by a dashed line. The network contains only the haplotypes with frequency ≥ 2.

Figure S1

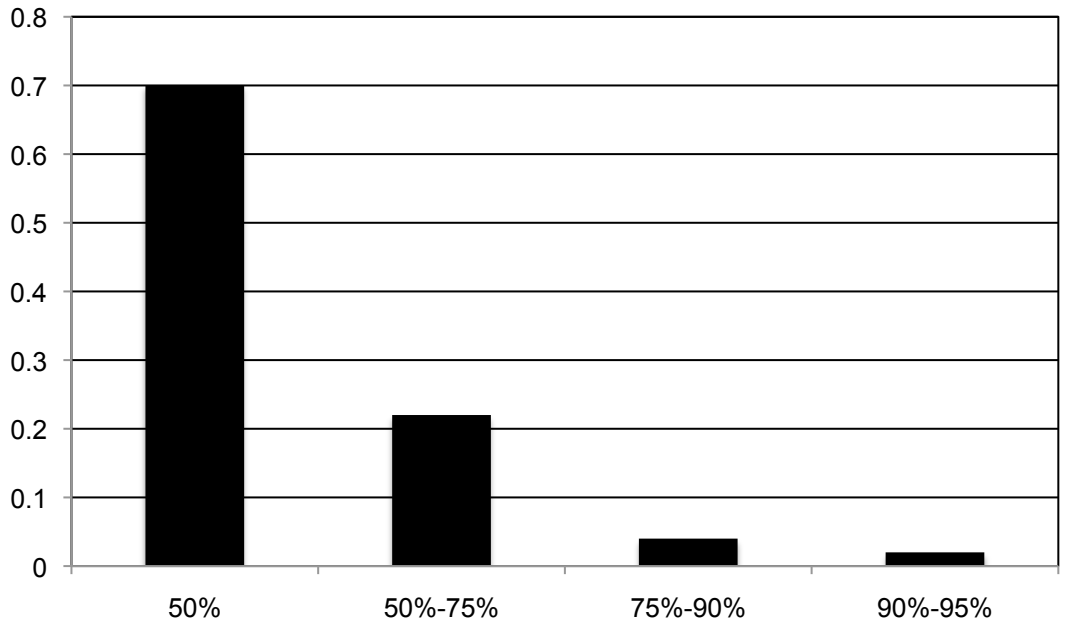


Figure S2

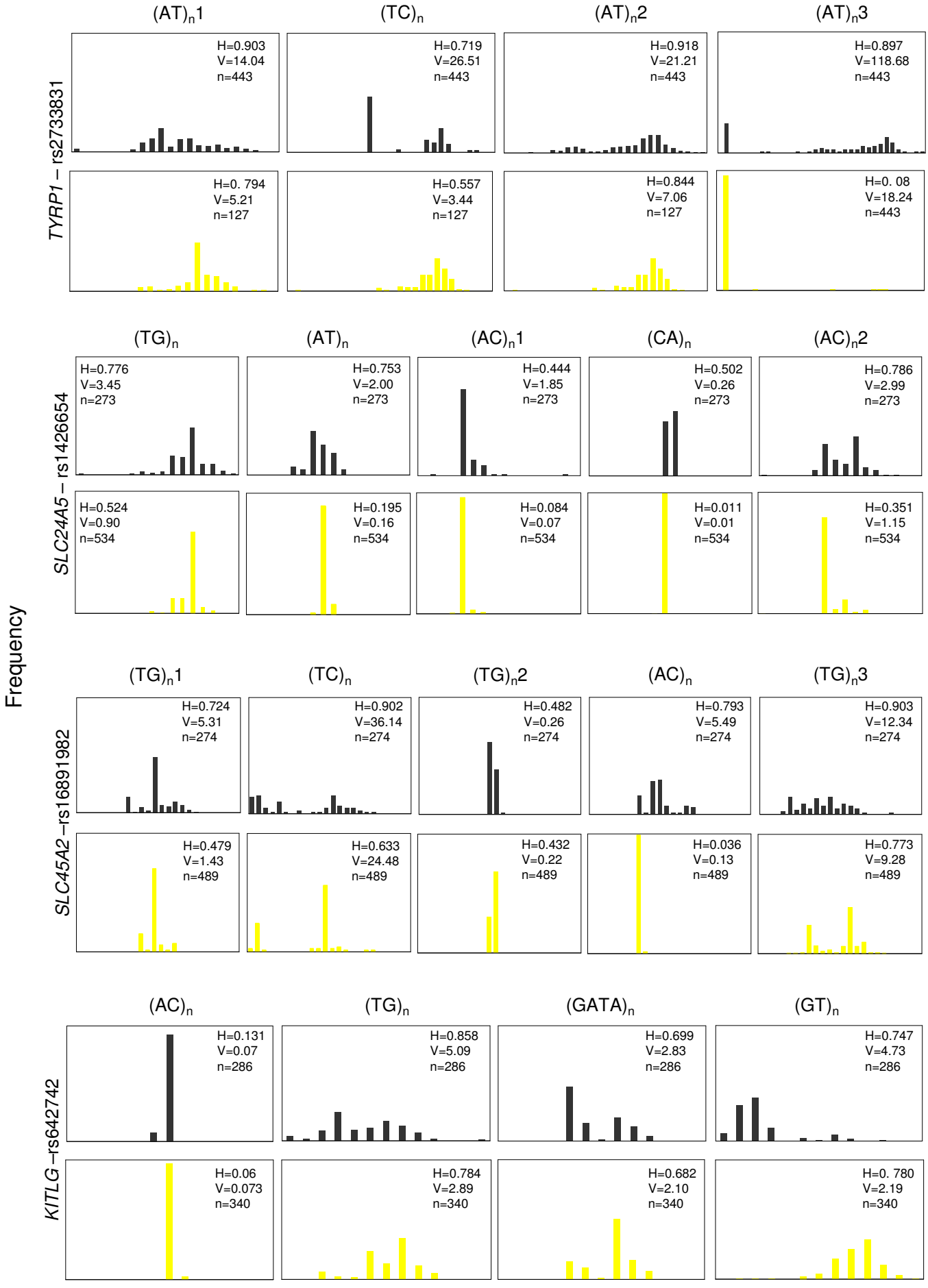
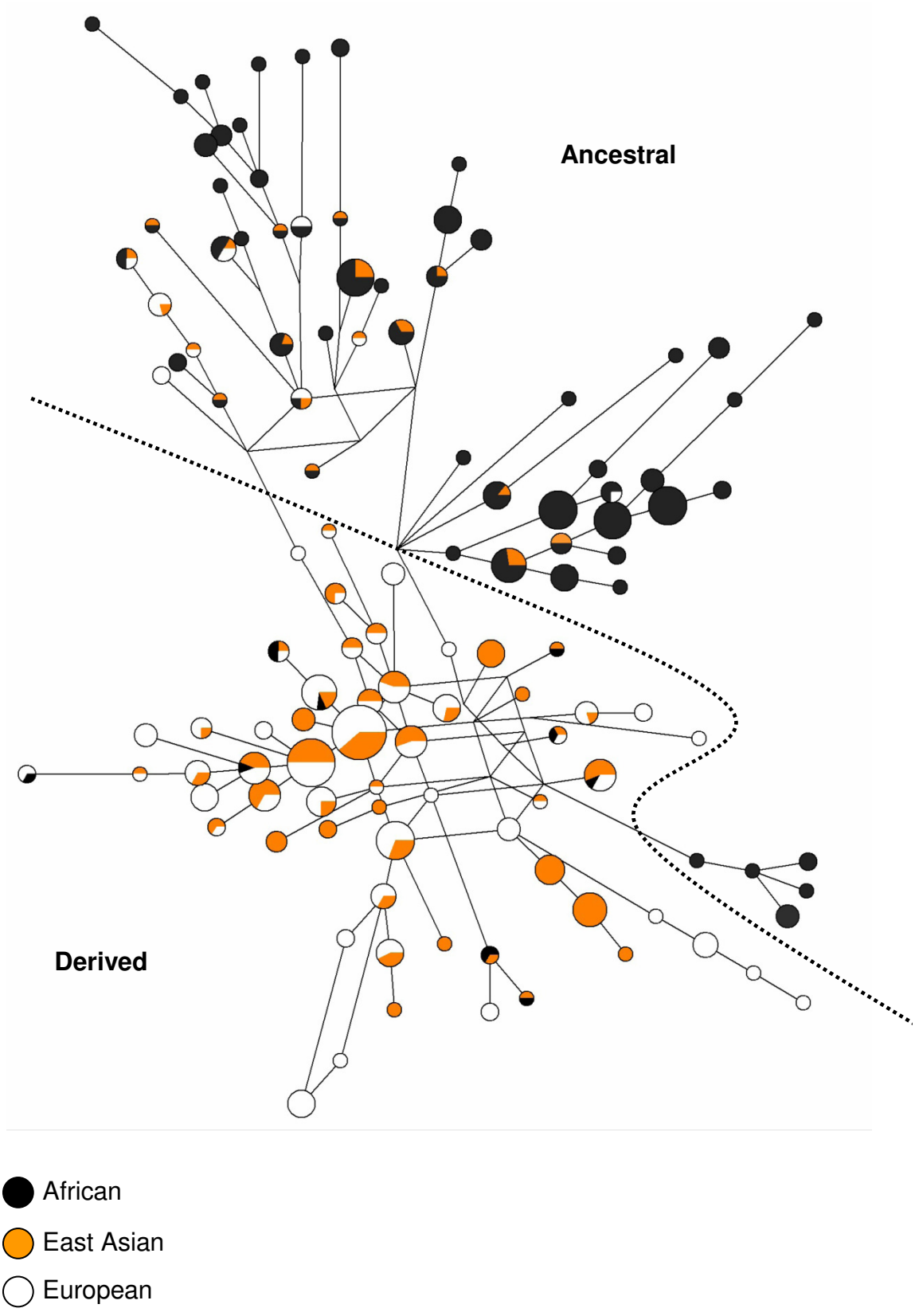

(AT)n1
(TC)n
(AT)n2
(AT)n3
TYRP1 – rs2733831
H=0.903 V=14.04 n=443
H=0.719 V=26.51 n=443
H=0.918 V=21.21 n=443
H=0.897 V=118.68 n=443
H=0. 794 V=5.21 n=127
H=0.557 V=3.44 n=127
H=0.844 V=7.06 n=127
H=0. 08 V=18.24 n=443
(TG)n
(AT)n
(AC)n1
(CA)n
(AC)n2
SLC24A5 – rs1426654
H=0.776 V=3.45 n=273
H=0.753 V=2.00 n=273
H=0.444 V=1.85 n=273
H=0.502 V=0.26 n=273
H=0.786 V=2.99 n=273
H=0.524 V=0.90 n=534
H=0.195 V=0.16 n=534
H=0.084 V=0.07 n=534
H=0.011 V=0.01 n=534
H=0.351 V=1.15 n=534
(TG)n1
(TC)n
(TG)n2
(AC)n
(TG)n3
SLC45A2 –rs16891982
H=0.724 V=5.31 n=274
H=0.902 V=36.14 n=274
H=0.482 V=0.26 n=274
H=0.793 V=5.49 n=274
H=0.903 V=12.34 n=274
H=0.479 V=1.43 n=489
H=0.633 V=24.48 n=489
H=0.432 V=0.22 n=489
H=0.036 V=0.13 n=489
H=0.773 V=9.28 n=489
(AC)n
(TG)n
(GATA)n
(GT)n
KITLG –rs642742
H=0.131 V=0.07 n=286
H=0.858 V=5.09 n=286
H=0.699 V=2.83 n=286
H=0.747 V=4.73 n=286
H=0.06 V=0.073 n=340
H=0.784 V=2.89 n=340
H=0.682 V=2.10 n=340
H=0. 780 V=2.19 n=340
Frequency
Alleles

Figure S3



References

Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am J Phys Anthropol 128:415-423.

Myakishev MV, Khripin Y, Hu S, Hamer DH. 2001. High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. Genome Res 11:163-169.