Saharan African ancestry in Ashkenazi Jews, with our best estimate of mixture date. Here, we analyze genome-wide single nucleotide polymorphism (SNP) genotype data from 35 West Eurasian groups to obtain a high resolution assessment of the impact of African admixture in terms of its geographic distribution, in quantitative terms to modern populations, and its date. We show that almost all Southern Europeans have inherited 1-3% African ancestry with a tight estimate of the mixture date of 54 generations, consistent with Ashkenazi gene flow and migrations at the end of the Roman Empire. The Middle Eastern groups are estimated to harbor 4-15% African ancestry with an average date of 54 generations, consistent with the cold political, economic cultural links between the Middle East and Egypt in the late middle ages. A novel finding is the detection of ~5% sub-Saharan African ancestry in Ashkenazi Jews, with our best estimate of mixture date being 76 generations. The dates in Ashkenazi Jews are consistent with genetic input from populations with African ancestry during classical or biblical times.

Datasets
We used individuals of West Eurasian ancestry from several sources:

-Population Reference Sample (HapMap)4: n=1115; 11; Illumina 1 M; Pop: CEU
-GenomeHap (CEPH-HGDP): n=940; 51; Illumina 60K; Pop: CEU
-HapMap Phase III: n=1115; 11; Illumina 1 M; Pop: CEU
-IntHap (IHD): n=932; 01; Illumina 200K
-Genome HapMap4

The combined dataset overlapping markers from all four studies consisted of 78 985 SNPs. SNP positions were mapped to NCB1 Build 35 and the fine scale genetic map of Myer et al., 2005 was used.

Analysis
Evidence of sub-Saharan African ancestry in West Eurasians
To study the signal of African gene flow in West Eurasian populations, we computed principal components (PCs) using ancestral sub-Saharan Africans (YRI) and ancestral West Eurasians (CEU) only, and plotted the mean values of the samples from each other populations on the first PC, a predictor called "PCA projection".

ROLLOFF: Method to estimate a date of mixture
To develop a novel software called ROLLOFF that uses summary statistics of admixture linkage disequilibrium (LD) to estimate dates of mixture. ROLLOFF computes a (signed) statistic for LD between a pair of markers in the genome and a group of genes containing sub-Saharan ancestry. It is designed to calculate the LD between all pairs of markers in the genome, weighted allele frequency differences, in the supposed admixing populations at each of the two markers. It uses this signal of an approximately exponential decay of admixture LD to estimate the date of mixture. Simulations have shown that ROLLOFF gives unbiased estimates for the date of mixture even in cases of old mixture and is robust to substantial frequencies differences, in the superimposing admixing populations, as well as fine scale errors in genetic map.

ROLLOFF Simulation Results
We simulated 100 populations of 2000 individuals each of 2000 and 1000 generations, with admixture proportions of 10% and 20%. We performed ROLLOFF analysis and used the admixture proportion of 10% and 20% to estimate the date of mixture. The estimated date of mixture was between 0 and 1000 generations.

References