Table 1
Geographic coordinates (decimal degrees) of *Papio* sampling sites and sample sizes.

No.	Taxon	Country	Site	Code	Sample size	Longitude	Latitude
1	Ph	Saudi Arabia	Abha	Abh	25	42.505228	18.216389
2	Ph	Saudi Arabia	Al Akhal	Akl	6	39.859444	23.315556
3	Ph	Saudi Arabia	Baha	Bah	15	41.466667	20.016667
4	Ph	Saudi Arabia	Dhilafa Escp.	Dhi	4	42.466667	17.933333
5	Ph	Saudi Arabia	Taif	Tif	15	40.415833	21.270278
6	Ph	Yemen	Bura'a Forest A	BuH	4	43.416667	14.866667
7	Ph	Yemen	Bura'a Forest B	BuL	5	43.866944	14.867222
8	Ph	Yemen	Jebel Iraf	Ira	1	44.250000	13.116667
9	Ph	Yemen	Jebel Raymah	Ray	1	43.433333	14.666667
10	Ph	Yemen	Jebel Sabir	Sab	1	44.200000	13.583333
11	Ph	Eritrea	Mt. Abagamsei	Aba	14	39.018620	15.349100
12	Ph	Eritrea	Abdur	Abd	11	39.845850	15.128570
13	Ph	Eritrea	Afabet	Afb	3	38.749583	16.120166
14	Ph	Eritrea	Barka Bridge	Bbr	7	38.020380	15.555120
15	Ph	Eritrea	R. Baeat	Bea	2	38.094270	15.671570
16	Ph	Eritrea	Dada (Bolo)	Dad	13	42.508889	13.129630
17	Ph	Eritrea	Debresina	Deb	3	38.825930	15.705350
18	Ph	Eritrea	Dogali	Dog	6	39.284730	15.579080
19	Ph	Eritrea	Durfo	Dur	7	38.964580	15.373700
20	Ph	Eritrea	Filfil Bridge	Fil	6	38.944450	15.614420
21	Ph	Eritrea	Furrus	Fur	9	38.971150	15.011480
22	Ph	Eritrea	Geleb	Gel	7	38.824070	15.821430
23	Ph	Eritrea	Halhal	Hal	7	38.314330	15.941370
24	Ph	Eritrea	Af Himbol	Him	9	37.397100	15.945050
25	Ph	Eritrea	Kubkub	Kub	11	38.632170	16.344820
26	Ph	Eritrea	Mensura	Men	5	38.351230	15.445980
27	Ph	Eritrea	Molki	Mol	7	38.221700	14.909080
28	PX	Eritrea	R. Shackat	Sha	4	37.499350	14.983100
29	Pa	Eritrea	R. Griset	Gri	8	36.760180	14.883220
30	Pa	Eritrea	R. Hadejemi	Had	6	36.907100	14.358270
31	Pa	Eritrea	Haykota	Hay	17	37.066000	15.156950
32	Pa	Eritrea	Tesseney	Tes	9	36.701420	15.145100
33	Ph	Ethiopia	Awash Station	ASt	5	40.177750	8.992683
34	Ph	Ethiopia	Gerba Luku	Ger	10	41.534000	9.587400
35	Ph	Ethiopia	Mieso	Mie	7	40.764083	9.203533
36	PX	Ethiopia	Awash Falls	AFa	5	40.019167	8.842683
37	PX	Ethiopia	Wolenkiti	Wol	5	39.487883	8.694583
38	Pa	Ethiopia	Adami Tulu	Ada	4	38.714933	7.825583
39	Pa	Ethiopia	Alambada	Ala	3	38.747683	7.504633
40	Pa	Ethiopia	Managasha 1	Mng	1	38.583333	9.083333
41	Pa	Ethiopia	Managasha 2	Man	6	38.571250	8.968383
42	Pa	Ethiopia	Wendo Genet	Wen	1	38.649650	7.071267
43	Pc	Somalia	Webi Shebelli	Web	1	45.433333	2.420833

Ph = *Papio hamadryas*; Pa = *P. anubis*; PX = phenotypic hybrids between *P. hamadryas* and *P. anubis*; Pc = *P. cynocephalus*. Longitude and latitude in decimal degrees.

To investigate whether the Arabian baboon population expanded after the colonization event, we calculated mismatch distributions for both Arabian clades in Arlequin 3.5.1.3 (Excoffier and Lischer, 2010) with 1000 bootstraps. We tested both the model for demographic expansion and the model for spatial expansion. We then calculated the time since expansion with $\tau = 2\mu t$ (μ: mutation rate, t: number of generations since expansion). Here we applied a generation time of 12 years (Rogers and Kidd, 1996) and the specific mutation rate of primate HVRI of 15–20% per million years (Jensen-Seaman and Kidd, 2001).

To estimate divergence times between clades, we concatenated the Brown Region, cyt *b*, and HVRI sequences ($n = 70$), and applied a Bayesian Markov Chain Monte Carlo method, which employs a relaxed molecular clock approach (Drummond et al., 2006) as implemented in BEAST 1.6.1 (Drummond and Rambaut, 2007). The three loci were partitioned, each with its optimal nucleotide substitution model (Brown Region: TrN + G; cyt *b*: HKY + G; HVRI: HKY + I + G) as chosen with the Bayesian information criterion (BIC) in jModeltest 0.1.1 (Posada, 2008). We assumed a relaxed uncorrelated lognormal model of lineage variation and a Birth-Death Process prior for branching rates. As a calibration point, we applied the fossil-based split of *Theropithecus* and *Papio* 5.0 ± 1.0 mya (Jablonski et al., 2008; Frost et al., 2014). Four replicates were run for 25 million generations with tree and parameter sampling occurring every 100 generations. The adequacy of a 10% burn-in and convergence of all parameters was assessed by visual inspection of the trace of the parameters across generations using TRACER 1.5 (Rambaut et al., 2003). The sampling distributions were combined (25% burn-in) using LogCombiner 1.6.1 (Rambaut and Drummond, 2002a). A consensus chronogram with node height distribution was generated and visualized with TreeAnnotator 1.6.1 (Rambaut and Drummond, 2002b) and FigTree 1.3.1 (Rambaut, 2006).

Results

The 294 baboon samples comprised 109 HVRI haplotypes. The subset of 73 samples for which we analysed the Brown Region, cyt *b*, and HVRI, comprised 52 haplotypes.

Haplotype network

The HVRI haplotype network reveals three major clades (Fig. 3). Clade X is strictly African and consists of Eritrean and a few Ethiopian hamadryas baboons, and phenotypical *P. hamadryas* × *P. anubis* hybrids from Ethiopia. Clade Y is more complex, encompassing Eritrean hamadryas and olive baboons, Eritrean hybrids, and Arabian hamadryas baboons. Clade Z comprises Ethiopian, Eritrean, and Arabian hamadryas baboons. Two Arabian clades are identifiable. Clade Arab_Y comprises four haplotypes and clusters closely with Eritrean baboons. Clade Arab_Z consists mainly of haplotypes found in Arabia but also some haplotypes found in Eritrea from sampling locations closest to the Bab-el-Mandab Strait (Dad) and one haplotype from Gerba Luku, Ethiopia (0317PHGer). Clade Arab_Z clusters more closely with Ethiopian baboons.

Population genetics of Arabian baboons

Whereas the three northern Arabia sampling locations (Akla, Taif, and Baha) harbour only haplotypes of Clade Arab_Z, both Clades Arab_Z and Arab_Y are represented in all other locations in Arabia (Fig. 4). One haplotype (H1) of Clade Arab_Z is found in every sampling location in Arabia.

Haplotype diversity and nucleotide diversity are both significantly higher in the African than in the Arabian hamadryas baboon populations ($n_{Africa} = 149$, $n_{Arabia} = 77$, $Hd_{Africa} \pm SD = 0.983 \pm 0.003$, $Hd_{Arabia} \pm SD = 0.871 \pm 0.026$, $p < 0.001$; $\pi_{Africa} \pm SD = 0.04251 \pm 0.00088$, $\pi_{Arabia} \pm SD = 0.01920 \pm 0.00243$, $p < 0.001$). Haplotype diversity and nucleotide diversity are both significantly higher ($p < 0.001$) in Clade Arab_Z ($n = 61$) than in Clade Arab_Y ($n = 16$): ($Hd_Z \pm SD = 0.825 \pm 0.040$, $Hd_Y \pm SD = 0.533 \pm 0.142$) and $\pi_Z \pm SD = 0.00431 \pm S0.00046$, $\pi_Y \pm SD = 0.00218 \pm 0.00076$.

When genetic diversity for Arabian hamadryas baboons is depicted from south to north, a decrease is observed in nucleotide diversity but not in haplotype diversity (Fig. 5). Both Arabian clades probably underwent a population expansion, as neither the demographic nor the spatial expansion model is rejected at $\alpha = 5\%$ (Table 2). The expansion of Clade Arab_Z occurred twice as early as the expansion of Clade Arab_Y, as indicated by a τ value, which is twice as high (Table 2).

Phylogenetic tree and divergence time estimates

Similar to the network, the phylogenetic tree reconstruction, based on concatenated Brown + cyt *b* + HVRI sequences, reveals the three distinct Clades X, Y, and Z, all of which include African