

Supplementary Table S3. Polymorphism and divergence in hominids at homologous positions in the 903 bp sequence including the 870 bp *NAT1* coding exon.

The screened sequence spans from 18'079'545 to 18'080'447 on chromosome 8 in the human reference sequence GRCh37/hg19. Positions upstream or downstream the *NAT1* coding exon are reported in italic font. Non-synonymous mutations are shown in bold type. The 21 substitutions corresponding to inter-species divergence (between two species at least) are highlighted by stars to the left part of the table. Polymorphisms shared between species or sub-species are highlighted in blue font. (N) indicates that undefined positions were also observed in some individuals. Variants also recorded in sequenced ancient genomes of hominins (Denisova, Neanderthal or 45'000 years old *Homo sapiens* from Ust'-Ishim) are boxed (see Supplementary Table S8).

	Position			Common chimpanzee ²						Gorilla ⁴		Orangutan ⁵	
	human reference sequence	human cds	SNP ID in Ensembl	Human ¹	Western	Nigeria-Cameroon	Eastern	Central	Bonobo ³	<i>Gorilla gorilla</i>	<i>Gorilla beringei</i>	<i>Pongo abelii</i>	<i>Pongo pygmaeus</i>
* 1	<i>18079548</i>	-9		<i>T</i>	<i>C</i>	<i>C</i>	<i>C</i>	<i>C</i>	<i>C</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>
2	18079577	21	rs4986992	T/G	T	T	T	T	T	T	T	T	T
* 3	18079595	39		G	A	A	A	A	A	G	G	G	G
4	18079631	75	rs201603160	T/C	T	T	T	T	T	T	T	T	T
5	18079632	76		G	G/A	G/A	G/A	G/A	G	G	G	G	G
* 6	18079643	87		A	A	A	A	A	A	A	A	G	G
7	18079653	97	rs56318881	C/T	C	C	C	C	C	C	C	C	C
* 8	18079664	108		C	C	C	C	C	C	C	C	A	A
9	18079688	132		T	T	T	T	T	T	T	T	C	C/T (N)
10	18079689	133	rs181298696	G/T	G	G	G	G	G	G	G	G/A	G
* 11	18079694	138		T	A	A	A	A	A	A	A	A	A
12	<i>18079703</i>	<i>147</i>		<i>C</i>	<i>C/T</i>	<i>C/T</i>	<i>C/T</i>	<i>C/T</i>	<i>C</i>	<i>C</i>	<i>C</i>	<i>C</i>	<i>C</i>
13	18079737	181		A	A	A	A	A	A	A	A/C(N)	A	A
14	18079746	190	rs56379106	C/T	C	C	C	C	C	C	C	C	C
15	18079757	201	rs200617057	G/C	G	G	G	G	G	G	G	G	G
16	18079775	219	rs765390690	T	T	T	T	T	T	T/C(N)	T	T	T
17	18079793	237	rs751824544	G/C	G	G	G	G	G	G	G	G	G
18	18079797	241		A	A	A	A	A	A	A	A	T/A	T
* 19	18079809	253		G	G	G	G	G	G	C	C	G	G
* 20	18079817	261	rs368205381	G	A	A	A	A	A	G	G	G	G

*	21	18079821	265		T	T	T	T	T	T	C	C	T	T
*	22	18079829	273		G	T	T	T	T	T	G	G	G	G
	23	18079859	303		C	C	C	C	C	C/T	C	C	C	C
	24	18079883	327		C	C	C	C	C	C	C	C	T/C	T
	25	18079892	336		G	G	G	G	G	G	G	G	G/C	G/C(N)
	26	18079897	341	rs145975713	T	T	T	T	T/C	T	T	T	T	T
	27	18079898	342	rs371752075	T/C	T	T	T	T	T	T	T	T	T
	28	18079901	345	rs768399724	T	T	T	T	T	T	T	T	C/T(N)	C
	29	18079902	346		G	G	G	G	G	G	G	G	G/A(N)	G
30 ⁶	18079906-18079907	350-351		rs72554606	G/C	G	G	G	G	G	G	G	G	G
	31	18079924	368	rs771202638	C	C	C	C	C	C	C	C	C	C
	32	18079925	369		T	T/C	T	T	T	T/C	T	T	T	T
	33	18079943	387		C	C	C	C	C	C	C/T(N)	C	C	C
*	34	18079946	390	rs376036465	G	G	G	G	G	G	G	G	A	A
	35	18079958	402	rs146727732	T/C	T	T	T	T	T	T	T	T	T
	36	18079976	420		G	G	G	G	G	G	G	G	G/C	G
*	37	18079994	438		G	G	G	G	G	G	T	T	G	G
*	38	18080000	444		T	T	T	T	T	T	T	T	C	C
	39	18080001	445	rs4987076	A/G	A	A	A	A	A	A/G(N)	A/G	A/G	A/G
	40	18080007	451	rs760970670	C	C	C	C	C	C	T	T	C/T	C
*	41	18080009	453		T	T	T	T	T	T	T	T	C	C
	42	18080014	458	rs374226986	C	C	C	C/T	C	C	C	C	C	C
	43	18080015	459	rs4986990	G/A	G	G	G	G	G	G	G	G/A	G
	44	18080024	468	rs141809132	T/C	T	T	T	T	T	T	T	T	T
45 ⁷	18080053-18080055	497-499		rs72554608/rs56328478/rs745892749	G/C	G	G	G	G	G	G	G	G	G
	46	18080074	518		A	A	A	A/C	A	A	A	A	A	A
*	47	18080085	529		C	A	A	A	A	A	C	C	A	A
	48	18080089	533		C	C	C	C	C	C	C	C	G/C	G
*	49	18080099	543		A	G	G	G	G	G	A	A	A	A
	50	18080107	551		G	G	G	G	G	G	G	G	G/A(N)	G

51	18080115	559	rs5030839	C/T	C	C	C	C	C	C	C	C	C
52	18080116	560	rs4986782	G/A	G	G	G	G	G	G	G	G	G
53	18080128	572	rs141552883	C/T	C	C	C	C	C	C	C	C	C
* 54	18080139	583		A	C	C	C	C	C	G	G	G	G
55	18080153	597		T	T/G	T	T	T	T	T	T	T	T
56	18080169	613	rs72554609	A/G	A	A	A	A	A	A	A	A	A
57	18080196	640	rs4986783	G/T	G	G	G	G	G	G	G	G	G
58	18080219	663		A/G	A	A	A	A	A	A	A	A	A
58	18080247	691	rs201916736	G/T	G	G	G	G	G	G	G	G	G
* 60	18080277	721		C	C	C	C	C	C	C	C	T	T
61	18080293	737	rs201129487	A/G	A	A	A	A	A	A	A	A	A
62	18080308	752	rs56172717	A/T	A	A	A	A	A	A	A	A	A
63	18080312	756	rs112846752	A/T	A	A	A	A	A	A	A	G	G
64	18080316	760		G	G	G/C	G	G	G	G	G	G	G
* 65	18080321	765		C	T	T	T	T	T	C	C	C	C
* 66	18080332	776		G	G	G	G	G	G	G	G	A	A
67	18080333	777	rs4986991	T/C	T	T	T	T	T	T	T	T	T
68	18080336	780		G	G	G	G	G	G	G	G	G/T	G
69	18080337	781	rs72554610	G/A	G	G	G	G	G	G	G	G	G
70	18080343	787	rs72554611	A/G	A	A	A	A	A	A (N)	A (N)	A	A
71	18080345	789		A	A/G	A	A/G	A/G	A	A	A	A	A
72	18080363	807	rs201670530	T/C	T	T	T	T	T	T	T	T	T
* 73	18080375	819	rs746370481	T	A	A	A	A	A	T	T	T	T
* 74	18080382	826		C	G	G	G	G	G	G	G	G	G
75	18080432	876		T	T	T	T	T	T	T	T	C/T	C
76	18080440	884	rs55793712	A/G	A	A	A	A	A	A	A	A	A

¹ Human polymorphisms are those recorded by the consensus gene nomenclature of human *NAT* alleles (<http://nat.mbg.duth.gr/>), complemented with haplotype data from 1000Genomes Phase 1 (The Genomes Project Consortium 2012), (Patin et al. 2006a) and (Mortensen et al. 2011). Human positions not recorded as polymorphic but associated with a SNP identifier are reported with a highest population MAF < 0.01 in Ensembl (Yates et al. 2016) (http://www.ensembl.org/Homo_sapiens/Info/Index).

² Western, Niger-Cameroon, Eastern and Central chimpanzees are *Pan troglodytes verus*, *P. t. ellioti*, *P. t. schweinfurthii* and *P. t. troglodytes*, respectively. Polymorphism recording is based on the individuals of the present study (Supplementary Figure S1) and the chimpanzees of Prado-Martinez et al. (2013) cross-checked with the *P. t. verus* assembly reference sequence (panTro4, February 2011).

³ Based on the individual of this study (Bonobo), the bonobos of Prado-Martinez et al. (2013) and the *Pan paniscus* draft assembly reference sequence (panPan1, May 2012).

⁴ Based on the individuals of this study, the gorillas of Prado-Martinez et al. (2013) and the *Gorilla gorilla gorilla* draft assembly reference sequence (gorGor4, December 2014).

⁵ Based on the individuals of this study, the orangutans of Prado-Martinez et al. (2013) and the *Pongo pygmaeus abelii* draft assembly reference sequence (ponAbe2, July 2007).

⁶ Polymorphism (rs72554606) reported by the consensus gene nomenclature of human *NAT* alleles as occurring on haplotype *NAT1*5*, on one of these two positions: 18'079'906 or 18'079'907.

⁷ Polymorphism (rs72554608) reported in the consensus gene nomenclature of human *NAT* alleles as occurring also on haplotype *NAT1*5*, on one of these three positions: 18'080'053, 18'080'054, or 18'080'055. Two additional SNP identifiers are reported in Ensembl (rs56328478 and rs745892749).

References

- Mortensen HM, Froment A, Lema G, Bodo JM, Ibrahim M, Nyambo TB, Omar SA, and Tishkoff SA. 2011. Characterization of genetic variation and natural selection at the arylamine N-acetyltransferase genes in global human populations. *Pharmacogenomics* 12(11):1545-1558.
- Patin E, Barreiro LB, Sabeti PC, Austerlitz F, Luca F, Sajantila A, Behar DM, Semino O, Sakuntabhai A, Guiso N et al. . 2006. Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. *American journal of human genetics* 78(3):423-436.
- Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G et al. . 2013. Great ape genetic diversity and population history. *Nature* 499(7459):471-475.
- The Genomes Project C. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56.
- Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L et al. . 2016. Ensembl 2016. *Nucleic Acids Res* 44(D1):D710-D716.