

Supplementary Table S16. Predictions of the effect of substitutions between inferred hominid ancestral *NAT1* and *NAT2* coding sequences and *Pan* basal and human reference haplotypes according to PolyPhen, SIFT and PANTHER cSNP Scoring.

A description of the scores and predictions returned by each of the three online software tools is provided in Supplementary File S1. Inferred ancestral haplotypes are those reconstructed for the theoretical hominid common ancestor at the center of the *NAT1* and *NAT2* networks (Supplementary Figures S2 and S3 for, respectively, *NAT1* and *NAT2*). For *NAT1*, two ancestral haplotypes were considered: ancestral_1 bearing C at position 583, and ancestral_2, bearing A at position 583.

Haplotypes	cDNA	protein	PolyPhen		SIFT		PANTHER cSNP Scoring	
			Score ¹	Prediction ²	Score ³	Prediction ⁴	PSEP ⁵	Prediction ⁶
<i>NAT1</i>								
<i>NAT1</i> ancestral_1 to <i>Pan troglodytes</i>								
	C529A	H177N	0 (1.00-0.00)	B	1.00 (3.07, 81)	T	Not scored (invalid substitution) ⁷	
<i>NAT1</i> ancestral_1 to <i>NAT1</i> ancestral_2								
	C583A	Q195K	0.001 (0.99-0.15)	B	0.06 (3.07, 81)	T	30	B
<i>NAT1</i> ancestral_2 to <i>NAT1</i> ancestral_1								
	A583C	K195Q	0 (1.00-0.00)	B	0.17 (3.07, 81)	T	Not scored (invalid substitution) ⁷	
<i>NAT1</i> ancestral_2 to <i>Homo sapiens</i>								
	A138T	E46D	0.177 (0.92-0.87)	B	0.01 (3.07, 81)	A	908	PRD
	G826C	E276Q	0.046 (0.94-0.83)	B	0.24 (3.07, 78)	T	908	PRD
<i>NAT2</i>								
<i>NAT2</i> ancestral to <i>Pan troglodytes</i>								
	T293C	V98A	0 (1.00-0.00)	B	0.91 (3.07, 50)	T	6	B
	T664C	F222L	0.009 (0.96-0.77)	B	0.46 (3.07, 51)	T	176	B
	C345A	D115E	0.001 (0.99-0.15)	B	0.2 (3.07, 50)	T	6	B
	G443C	C148S	0 (1.00-0.00)	B	0.34 (3.07, 51)	T	6	B
	A595G	I199V	0.019 (0.95-0.80)	B	0.45 (3.07, 51)	T	6	B
<i>NAT2</i> ancestral to <i>Homo sapiens</i>								
	C451T	R151C	0.980 (0.75-0.96)	PRD	0 (3.07, 51)	A	Not scored (invalid substitution) ⁷	
	C511A	P171T	0.188 (0.92- 0.87)	B	0.27 (3.07, 51)	T	Not scored (invalid substitution) ⁷	
	C573A	F191L	0 (1.00-0.00)	B	0.47 (3.07, 51)	T	Not scored (invalid substitution) ⁷	
	G834T	K278N	1.00 (0.00-1.00)	PRD	0 (3.08, 49)	A	Not scored (invalid substitution) ⁷	

¹ PolyPhen score : probability that a substitution is damaging; sensibility and specificity in brackets.

² PolyPhen prediction : “benign” (B), “possibly damaging” (POD), “probably damaging” (PRD).

³ SIFT score : probability that a substitution is tolerated; median sequence information and number of sequences used for the prediction in brackets.

⁴ SIFT prediction : T: “tolerated” (T), A: “affect protein function” (A).

⁵ PANTHER cSNP Scoring PSEP (position-specific evolutionary preservation) : length of time (in millions of years) of preservation of a position.

⁶ PANTHER cSNP Scoring prediction : "probably damaging" (PRD), "possibly damaging" (POD), "probably benign" (B).

⁷ PANTHER cSNP Scoring returns the message “Not scored (invalid substitution)” when the input reference amino acid (e.g. H in H177N) does not match the amino acid at that site in the best-match sequence in the PANTHER library, and is likely to be a partial mismatch between the user-input sequence and the best match in PANTHER.