Supplemental Data

Antiretroelement Activity of APOBEC3H

Was Lost Twice in Recent Human Evolution

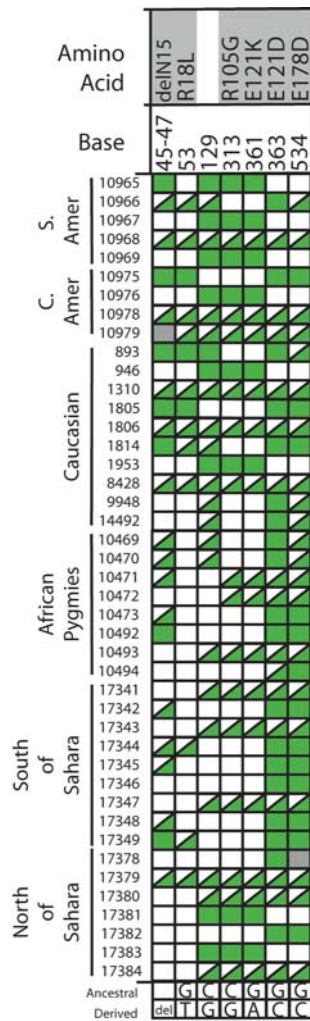Molly Oh Ainle, Julie A. Kerns, Melody M.H. Li, Harmit S. Malik, and Michael Emerman

Figure S1. Genotyping the APOBEC3H Coding Sequence in a Panel of Diverse Humans
The *APOBEC3H* coding sequence was sequenced from genomic DNA from 5 South American, 4 Central American, 10 Caucasian, 8 African Pygmy, 9 South of Sahara and 7 North of Sahara individuals. The seven polymorphisms typed in the panel are detailed at the top of the figure (the six non-synonymous polymorphisms are shown in gray – delN15, R18L, R105G, E121K, E121D and E178D). All polymorphisms are shown as changes from the ancestral sequence (ancestral and derived bases at each position are shown at bottom of figure) and do not necessarily represent minor alleles. Derived changes are shown in green and ancestral states are shown in white with heterozygotes represented by split squares. Missing data is shown in gray. Haplotypes that encode stable proteins contain G105 and do not have the deletion at position N15 (delN15).

```
CLUSTAL 2.0.5 multiple sequence alignment


hapIII              MALLTAETFRLQFN-KRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFI 59
hapIV               MALLTAETFRLQFN-KRLLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFI 59
hapII               MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFI 60
hapI                MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFI 60
human_ancestor      MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFI 60
chimpanzee          MALLTAETFRLQFNNRRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFI 60
hu_ch_ancestor      MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFI 60
gorilla             MALLTAETFRLQFNNKLRLRRPYYRRKTLLCYQLTPQNGSMPTRGYFKNKKKCHAEICFI 60
                    ************* :  ****** **:*********** ******:************


hapIII              NEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLRIFASRLYYHWCKPQQ 119
hapIV               NEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLRIFASRLYYHWCKPQQ 119
hapII               NEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLRIFASRLYYHWCKPQQ 120
hapI                NEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQ 120
human_ancestor      NEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLRIFASRLYYHWCKPQQ 120
chimpanzee          NEIKSMGLDETQCYQVTCYLTWSPCSSCAWKLVDFIQAHDHLNLRIFASRLYYHWCKPQQ 120
hu_ch_ancestor      NEIKSMGLDETQCYQVTCYLTWSPCSSCAWKLVDFIKAHDHLNLRIFASRLYYHWCKPQQ 120
gorilla             NEIKSMGLDETQCYQVTCYLTWSPCSSCAWKLVDFIKAHDHLNLRIFASRLYYHWCKRQQ 120
                    ******************************:*****:****** *********** **


hapIII              DGLRLLCGSQVPVEVMGFPEFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLDRI 179
hapIV               DGLRLLCGSQVPVEVMGFPEFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLDRI 179
hapII               DGLRLLCGSQVPVEVMGFPEFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLDRI 180
hapI                KGLRLLCGSQVPVEVMGFPEFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERI 180
human_ancestor      EGLRLLCGSQVPVEVMGFPEFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERI 180
chimpanzee          EGLRLLCGSQVPVEVMGLPEFNDCWENFVDHEKPLSFDPCKMLEELDKNSRAIKRRLERI 180
hu_ch_ancestor      EGLRLLCGSQVPVEVMGFPEFADCWENFVDHEKPLSFDPSKMLEELDKNSRAIKRRLERI 180
gorilla             EGLRLLCGSQVPVEVMGFPEFADCWENFVDHEKPLSFDPSKMLEELDKNSRAIKRRLERI 180
                    .***************:*** ***************:* *****************:**


hapIII              KQS--------------------------- 182
hapIV               KQS--------------------------- 182
hapII               KQS--------------------------- 183
hapI                KQS--------------------------- 183
human_ancestor      KQS--------------------------- 183
chimpanzee          KQS--------------------------- 183
hu_ch_ancestor      KQS--------------------------- 183
gorilla             KQSWSVDVLENDLRSLQLGPVSHSLSRSNSR 211
                    ***
```

Figure S2. Alignment of Gorilla, Chimpanzee and Predicted Human and Ancestral Human APOBEC3H Protein Sequences

Amino acid sequences of gorilla, chimpanzee, the PHASE-predicted human APOBEC3H haplotypes (hapI, hapII, hapIII and hapIV) as well as the parsimony-predicted human ancestor APOBEC3H (human ancestor) and human/chimpanzee APOBEC3H (hu_ch_ancestor) proteins were aligned using CLUSTAL_X (Thompson et al., 1997). Underlined in the human/chimpanzee ancestor sequence is the site (aa 160) at which parsimony could not yield a simple prediction of the ancestral state; therefore, three versions of the protein, 160C, 160Y and 160S (data shown is for 160S version), were constructed and found to encode proteins of similar stabilities and antiviral activities.

Table S1. Human APOBEC3H Haplotype Frequencies
Frequencies of the predicted *APOBEC3H* haplotypes as inferred using PHASE 2.1.1 software
(Stephens and Donnelly, 2003; Stephens et al., 2001). Haplotypes found at frequencies > 10% are
highlighted in bold.

index	haplotype	E (freq)	S.E.
1	1TGCGCC	0.00148	0.00388
2	1TGCGCG	0.001593	0.004926
3	1TGGAGC	0.000463	0.002274
4	1TGGAGG	0.002247	0.005927
5	**1TCCGCC**	**0.177715**	**0.01187**
6	1TCCGCG	0.000234	0.001634
7	1GGCGCC	0.000117	0.00116
8	1GGCGCG	0.001856	0.005638
9	1GGGAGG	0.02652	0.005603
10	**1GCCGCC**	**0.113544**	**0.009882**
11	1GCCGCG	0.00023	0.001624
12	0TGCGCC	0.00081	0.003389
13	0TGCGCG	0.01218	0.005034
14	0TGGAGG	0.00012	0.001183
15	0TCCGCC	0.00083	0.003008
16	0GGCGCC	0.010153	0.005361
17	0GGCGCG	0.053672	0.007507
18	0GGGAGC	0.011164	0.002282
19	**0GGGAGG**	**0.30762**	**0.008177**
20	**0GCCGCC**	**0.264882**	**0.014036**
21	0GCCGCG	0.000239	0.001654
22	0GCGAGG	0.0121	0.002301

Table S2. Human APOBEC3H Haplotype Counts
Numbers of each haplotype identified in the panel according to the "best reconstruction" by PHASE. The single haplotype in which both the delN15 and the R105G polymorphisms are found on the same chromosome is shown in bold (2/86 chromosomes).

	Haplotype	Count
1	1TCCGCC	16
2	**1GGGAGG**	**2**
3	1GCCGCC	10
4	0TGCGCG	1
5	0GGCGCC	1
6	0GGCGCG	5
7	0GGGAGC	1
8	0GGGAGG	27
9	0GCCGCC	22
10	0GCGAGG	1

Table S3. Typing the APOBEC3B Deletion

Genotyping of the APOBEC3B deletion in the African American, European American and Han Chinese populations. Presence of deletion is denoted by 1; absence of deletion is denoted by 0.

sample ID (NA-----)	population	Genotype
17102	African American	0 0
17103	African American	1 0
17104	African American	0 0
17105	African American	0 0
17106	African American	0 0
17107	African American	0 0
17108	African American	0 0
17109	African American	0 0
17110	African American	0 0
17111	African American	0 0
17113	African American	0 0
17116	African American	0 0
17125	African American	0 0
17133	African American	0 0
17135	African American	0 0
17136	African American	0 0
17137	African American	0 0
17138	African American	0 0
17139	African American	1 0
17140	African American	0 0
6990	European (EC)	0 0
7019	European (EC)	0 0
7348	European (EC)	0 0
7349	European (EC)	0 0
10830	European (EC)	1 0
10831	European (EC)	0 0
10842	European (EC)	0 0
10843	European (EC)	0 0
10844	European (EC)	0 0
10845	European (EC)	0 0
10848	European (EC)	0 0
10850	European (EC)	0 0
10851	European (EC)	0 0
10852	European (EC)	0 0
10853	European (EC)	0 0
10854	European (EC)	0 0
10857	European (EC)	0 0
10858	European (EC)	1 0
10861	European (EC)	0 0
12547	European (EC)	0 0
12548	European (EC)	0 0
12560	European (EC)	0 0
17201	European (EC)	0 0
17733	Han Chinese (HL)	0 0

17734	Han Chinese (HL)	1 1
17735	Han Chinese (HL)	0 0
17736	Han Chinese (HL)	0 0
17737	Han Chinese (HL)	1 0
17739	Han Chinese (HL)	1 0
17740	Han Chinese (HL)	0 0
17741	Han Chinese (HL)	1 0
17742	Han Chinese (HL)	1 1
17743	Han Chinese (HL)	1 0
17744	Han Chinese (HL)	1 0
17745	Han Chinese (HL)	1 0
17746	Han Chinese (HL)	1 1
17747	Han Chinese (HL)	0 0
17749	Han Chinese (HL)	1 0
17752	Han Chinese (HL)	1 1
17753	Han Chinese (HL)	1 0
17754	Han Chinese (HL)	1 1
17755	Han Chinese (HL)	1 0
17756	Han Chinese (HL)	0 0
17757	Han Chinese (HL)	1 0
17759	Han Chinese (HL)	1 0
17761	Han Chinese (HL)	0 0

Table S4. APOBEC3H Primers

er e	uence (5' to 3')
)12	AACTGACATGCCCCAGGG
)15	ACACGATGGCTCTGTTAACAGCC
36	GCTCTGTTAACAGCCGAAAC
43	GGTACCGGCTCTGTTAACAGCCGAAACATTCC
73	ACACGATGGCTCTCCTAACAGCC
)8	GGGTGCGTGAGTGCAGGG
I0	TTGAAAAACTACCTATTGGGTGC
85	AM]GTTAACAGCCGAAACATTCCGCTTAC
86	ACCTTGTTTTCAAAGTAGCCTCT
For	ACACGATGGCTCTGTTAACAGCC
Rev	GCACCCGCTTCCTGCC
3_For	GGGTGCGTGAGTGCAGGG
3_Rev	TTGAAAAACTACCTATTGGGTGC
For	ACCCCAATAGGTAGTTTTTCAACCC
Rev	AACTGACATGCCCCAGGG
76	TCTGTTGCACAGAAACACGATGG
71	TTCCCGAAGTAGTGACTGAGC
72	GCCACGCACTAGAAAGTTCAC
73	CGGGTGGTGTCAGATCTTG
74	TCTGACACCACCCGGGAG
75	ACAGTGCCTCACCTTTATCC

Supplemental References

Stephens, M., and Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet *73*, 1162-1169.

Stephens, M., Smith, N. J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. Am J Hum Genet *68*, 978-989.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res *25*, 4876-4882.