

## Table S4. The 300 murine proviruses of groups G1-G3.

The most intact proviruses were selected from the *Mus musculus* (C57Black/6J) genome assembly mm8 portion of a RetroBank prototype by selecting for a RetroTector score of 2000 (essentially complete proviruses, with no or a few frameshifts or stops) and genus gammaretrovirus (estimated by RetroTector). 306 proviruses fulfilled these criteria. Of them, six were duplicates. RetroTector may produce duplicates when alternative proviral interpretations occur. The lowest scoring of each duplicate was removed, giving 300 proviruses. Three major branches were evident in all phylogenies. They were named G1, G2 and G3 (for gamma 1, gamma 2 and gamma 3). Some of their properties are listed below. Although some exogenous gammaretroviruses cluster with the groups (2-3 rat endogenous proviruses cluster with G2 and many ecotropic exogenous MLVs cluster with G3), only endogenous proviruses from assembly mm8 are shown below. The envelopes of a few G1 and G2 proviruses cluster differently from the rest of those proviruses, probably due to recombination. The subject cannot be treated extensively here.

The columns are:

1. Group: The allocation of groups G1-G3, based on *gag* and *pol* Clustal alignments and phylogenetic trees based on them.
2. Jern subgroups. These are from Jern et al, Plos Genetics 2007. We used BLAT at the UCSC site to search the most highly similar sequence in the mm8 assembly. Not all G3 proviruses could be given a Jern subgroup assignment, for unknown reasons. Several of the concatenated *gag/pro/pol/env* sequences were not sufficiently different from each other, and yielded ambiguous calls. This uncertainty is reflected in the column.
3. THIS, the G3 subgroup inferred from the clustering in *gag*, *pol* and *env* based trees in this paper.
4. Chromosome in mm8.
5. Chainstart, the start position of the provirus. The 5' end is always given as the start.
6. Chainend, the end position of the provirus. The 3' end is always given as the end.
7. Sense, which can be the primary (P) or secondary (S) strand. The primary strand is the one used for numbering the chromosome in question.
8. Score, the score calculated by RetroTector<sup>©</sup>. It is a rough representation of proviral completeness.
9. Genus, in this case always C (gammaretrovirus-like).
10. GagOrf, the sum of frameshifts and stops in the *gag* gene. -1 means that the *gag* gene could not be interpreted.
11. ProOrf, annotated in the same way as GagOrf.
12. PolOrf, annotated in the same way as GagOrf.
13. EnvOrf, annotated in the same way as GagOrf.
14. Estimated PBS. Upper case means "good fit to a PBS sequence", lower case means "moderate fit to a PBS sequence". All PBS entries were checked against the tRNAdb in Leipzig.
15. The PBS sequence.
16. LTRdivergence, the % divergence between the two LTRs, a rough measure of time since integration.







G3		MPMV	8	126052021	126060437	P	2473	C	0	0	2	12			
G3		MPMV	11	60402203	60394839	S	2265	C	1	0	2	-1	Q	tggagggtcccaccgagat	0
G3		MPMV	10	22424071	22432738	P	2929	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3	MPMV7	MPMV	5	43496527	43505213	P	2960	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3		MPMV	6	73223771	73213956	S	2547	C	0	0	2	-1			43,519
G3	MPMV6	MPMV	1	133470452	133479146	P	2862	C	1	0	0	0	Q	tggagggtcccaccgagat	0
G3		MPMV	5	110148316	110140851	S	2459	C	1	0	2	-1	Q	tggagggtcccaccgagat	0
G3	MPMV8	MPMV	11	102900091	102908778	P	2865	C	1	0	1	0	Q	tggagggtcccaccgagat	0
G3	MPMV4	MPMV	11	86698849	86707535	P	2944	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3	MPMV1	MPMV	7	64005512	64014552	P	2947	C	0	0	0	0	T	tggagggtcccaccgagat	0
G3	MPMV11	MPMV	12	70465730	70474228	P	3000	C	1	0	0	0			0
G3		MPMV	5	122453776	122460838	P	2230	C	2	0	3	-1	Q	tggagggtcccaccgagat	1,357
G3	MPMV5	MPMV	10	4628507	4620011	S	2728	C	1	0	3	2	Q	tggagggtcccaccgagat	0
G3	XMV18	XMV	19	60988354	60980003	S	2906	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3	XMV15	XMV	9	62237522	62245905	P	2929	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3	XMV12?	XMV	8	44818939	44810580	S	2936	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3	XMV19	XMV	14	53484913	53492360	P	2707	C	3	0	0	0	Q	tggagggtcccaccgagat	0
G3	XMV9	XMV	4	146062965	146071316	P	3063	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3	XMV13	XMV	13	68392368	68384024	S	2914	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3	XMV16	XMV	9	41663510	41656480	S	2282	C	10	0	0	0			0
G3	XMV8	XMV	4	145371516	145363002	S	2924	C	0	0	0	0	Q	tggagggtcccaccgagat	0
G3	XMV41	XMV	1	173317855	173326267	P	2985	C	0	0	2	0	P	tgggggtctgtccggat	0
G3	XMV17 ?	XMV	5	23212035	23212626	S	3147	C	0	0	0	0	P	tgggggtctgtccggat	0
G3	XMV42 43?	XMV	1	172778230	172786635	P	3099	C	0	0	0	0	P	tgggggtctgtccggat	0
G3	<eco>	EMV	8	126320498	126311126	S	2984	C	0	0	0	0	P	tgggggtctgtccggat	0