

Supplementary materials

Table S1. TEs inserted into simulated data. Transposon ids after Dfam. These TEs were inserted at random positions into a random sequence with 42% GC content.

TE_id	Count	Average identity	Standard deviation	Number of indels	Target site duplication	Length	Number of truncated elements	Number of nested element
AluY	730	87	12	16	y	311	67	11
BEL12-I_DR	550	89	12	10	y	7730	80	12
Copia1_DM	680	87	14	10	y	4530	65	10
DIRS-1_DR	710	85	15	12	n	5634	70	5
G5_DM-Jockey	590	91	10	15	y	4856	75	16
GYPSY8_DM	470	84	10	15	y	4955	74	12
hAT-9_DR	480	82	12	11	y	2386	65	13
Helitron-1N3_DR	550	83	10	11	n	1843	65	10
HERVL	590	88	14	14	y	6542	50	7
L1-4_DR	620	78	9	16	y	5548	72	13
Merlin1_HS	580	80	11	13	y	1175	60	18
Ngaro1_DR	560	86	12	11	n	6578	73	9
Penelope-1_DR	810	85	11	20	y	4446	60	11
piggyBac-N1_DR	610	80	10	19	y	1005	64	20
Polinton-1_DR	420	81	11	20	n	18485	70	10
RTE1	570	90	13	18	y	3291	71	11
SINE3-1	580	85	13	10	y	590	70	7
TC1-2_DM	590	83	14	12	y	1644	70	16
TRANSIB1	670	81	10	12	y	3014	68	9
Turmoil2	650	79	9	10	y	6999	69	15

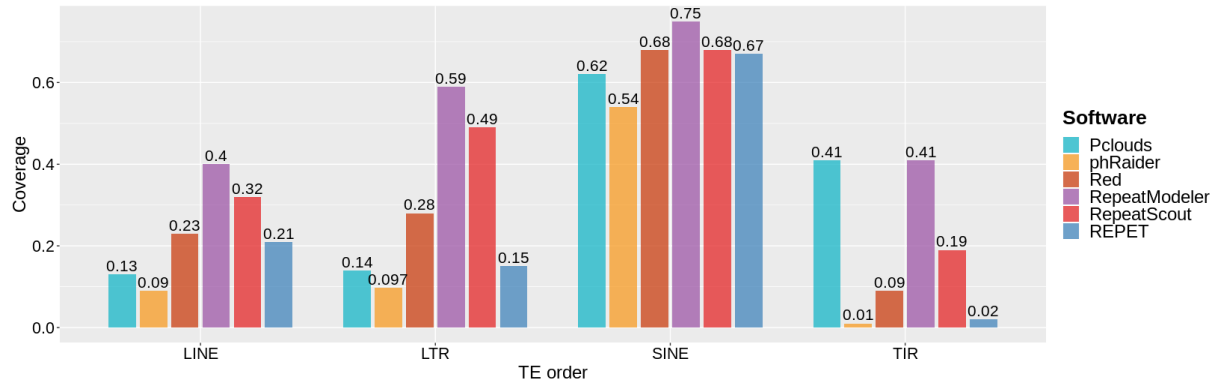
Table S2. Comparison of the consensus models obtained for each single TE inserted in the simulated sequence. With * are indicated models were the models include a duplication or an extension of the same TE and with # are indicated models that are the combination of two TEs which originated from nested TEs.

Original TE-families		Length of the models created		
TE	Length	RepeatScout	RepeatModeler	REPET
AluY	311	308	345 320 310	-
BEL12	7730	7724	7751 7730 6413	11148* 7750 2769
Copia	4530	8843*	4532 4514 4326 4076	4652 4314
DIRS	5654	5644 9385*	5672 5653 5610 5445 4633 4062	7134# 5401 1824 1727
G5	4856	4854	4851 4843 4836 3151	7502 * 5232 5197 4904 3967 2359
GYPSY	4955	4360 3920	4938 4012 3951 3373	5068 4682 2061 1445 1256
hAT	2386	2380	2380 2380	2426
Helitron	1843	1839 1472	1832 1823 1234	1907 1906 1744 871 633

Original TE-families		Length of the models created		
TE	Length	RepeatScout	RepeatModeler	REPET
HERVL	6542	12634*	6549 6523 6094	11495# 6553 3886 3169
L1	5548	4460 3379 2415	5496 5488 3386 2588 2478 1289	5784 5605 5443 2855 2142 1794 1639 1179 667
Merlin	1175	1168	1172 1167 744	1231 1216 1089 629
Ngaro	6578	6571 6020	6568 6566 6563 6156	7798# 6652 6379 2487 2445
Penelope	4446	4437 2276	4432 4407 4189 3910 3198 3908	5077 3251 812
piggyBac	1005	997	1002 990 840	1067 1055 966 641 477
Polinton	18485	14288 10127 4536 1513	17626 13482 11066 9356 8330 6655 372 169	16912 16015 15368 10418 6631 2793 1690 1216

Original TE-families		Length of the models created		
TE	Length	RepeatScout	RepeatModeler	REPET
RTE1	3291	3287	3286 3270 3212	3343 3200 3128 1222
SINE3	590	587	587 581 578	624 559
TC1	1644	1634	1636 1627 1472	1689 771
TRANSIB	3014	3002	2995 2546 2378	3094 2858 1166
Turmoil	6999	6676	7016 6985 6974 6970 4201	7113 6598 5956 5834# 1757 1381 1100

a)



b)

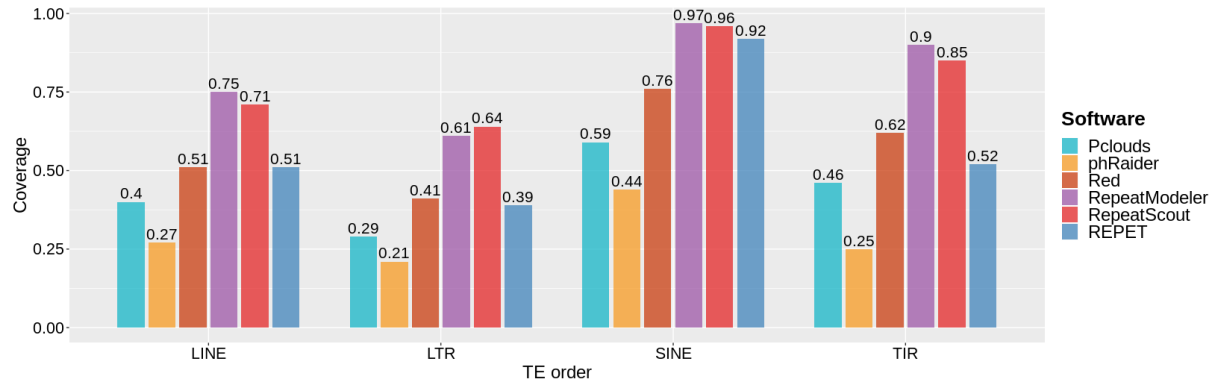


Figure S1. Coverage of each TE order in the human (a) and zebrafish (b) dataset.