

Year-end summary of Repbase 2017

Authors: Kenji K. Kojima, Weidong Bao, and Oleksiy Kohany

Affiliation: Genetic Information Research Institute, 465 Fairchild Drive, Suite 201, Mountain View, CA 94043, USA

Correspondence: Kenji K. Kojima, kojima@girinst.org

Published: 2018/01/29

Update of Repbase 2017

In the year 2017, *Repbase* increased the number of entries by 2,113 including 1,547 sequences published on *Repbase Reports*. The total number of entries reached 47,599 at the end of year 2017. They include not only consensus and representative sequences of transposable elements (TEs), but also reference sequences for multicopy genes and integrated viruses, and uncharacterized repeats. The total number of TEs sequences is 46,248.

Species expansion

Table 1 shows the top 10 species whose entries increased in the year 2017. We have evenly worked on vertebrates, invertebrates, and plants. Protists and fungi are underrepresented in the numbers of entries partly because of their smaller genome sizes. The large increase in viral entries is thanks to the submission by Zhang et al. They reported many TEs from large DNA viruses, especially from Pandraviruses and Bracoviruses.

The characteristics of several species we focused on last year are described below.

Chrysemys picta bellii (western painted turtle)

This is one of the first three turtle species sequenced. The other two are *Pelodiscus sinensis* (soft-shell turtle) and *Chelonia mydas* (green sea turtle). The genome evolution rates of turtles were revealed to be very low and thus, many ancient TE families would be present in the turtle genomes (Green et al. 2014). In fact, TEs from *C. picta bellii* enabled us to determine the classification of several unknown ancient TEs in human. Now the number of entries from *Chrysemys picta bellii* is 843, the 7th repeat

family-rich species in Repbase (Table 2).

Oryza sativa (Asian rice)

By the end of year 2017, Repbase had accumulated 1,980 repeat (1,935 TE) families from rice. Still we have published 358 repeat (354 TE) families from rice last year. The exhaustive characterization of TEs from this economically important crop & model plant organism definitely contributes to the human welfare.

Xenopus laevis (African clawed frog)

Xenopus laevis is a pseudotetraploid, and thus, it has around a double genome size of its sister species *Xenopus tropicalis* (western clawed frog). The genome sequence analysis of *X. laevis* reported that some TE families are distributed unevenly among chromosomes. It indicates that these TE families were mobilized before the hybridization of two parent species of *X. laevis* and after the split of these two species (Session et al. 2016).

Nicotiana sylvestris (wood tobacco)

Nicotiana sylvestris is known as a parental species of commercial tobacco *Nicotiana tabacum*, together with *Nicotiana tomentosiformis*. This is a traditional example of tetraploidy and is a good model of gene duplication and subsequent genome size reduction. TEs play a central role in genome size reduction as well as expansion, and so, the characterization of TEs in *N. sylvestris* enhances the genome evolution studies.

Parhyale hawaiensis (amphipod crustacean)

Recent phylogenomic researches indicate that insects have evolved from a lineage of crustaceans. Despite of their divergence and importance, crustacean genomes have not been well analyzed. This amphipod can be a good model organism for the evo-geno analysis.

TE classification

We have basically not updated our classification scheme since Bao et al. (2015). TEs are classified into three classes and further sub-classified into 61 superfamilies/clades (Table 3), or 74 divisions if the subgroups under Endogenous retrovirus, SINE, *Sola* and

Crypton are considered as so. Besides the TEs shown in Table 3, Repbase contains 260 unclassified TEs, which are non-autonomous and do not have signs for characterization, even whether DNA transposon or retrotransposon, and 157 “Integrated viruses”, which are mostly Caulimoviruses. It is noteworthy that the category “Integrated virus” in Repbase excludes endogenous retroviruses and “Polintoviruses”.

Large increase of entries in unclassified endogenous retroviruses and LTR retrotransposons are caused by the deposition of old LTR retrotransposons from the turtle and the frog. Large increase of entries in unclassified non-LTR retrotransposons is, on the other hand, caused by the inclusion of an unclassified non-LTR retrotransposon group *Odin* from *Oikopleura dioica*. The number of entries of *Ginger1* was almost doubled thanks to the characterization of 19 families from the ciliate *Oxytricha trifallax*.

References

Bao W, Kojima KK, Kohany O.

Repbase Update, a database of repetitive elements in eukaryotic genomes.

Mob DNA, 2015;6:11.

Table 1. Top 10 species whose entries increased in 2017.

Species	2017 Increase	Total
<i>Chrysemys picta bellii</i> (painted turtle)	472	843
<i>Oryza sativa</i> (Asian rice)	358	2333
<i>Xenopus laevis</i> (African clawed frog)	216	237
<i>Nicotiana sylvestris</i> (wood tobacco)	211	213
<i>Parhyale hawaiiensis</i> (white shrimp)	151	151
<i>Adineta vaga</i> and <i>Adineta sp.</i> (bdelloid rotifer)	118	124
<i>Oikopleura dioica</i> (tunicate)	97	184
<i>Gossypium raimondii</i> (cotton)	74	246
<i>Hydra vulgaris</i> (hydra)	55	745
<i>Brassica rapa</i> (Chinese cabbage)	48	381

Table 2. Top 20 species of the number of Rebase entries.

Species	Total	2017 Increase
<i>Oryza sativa</i> (Asian rice))	2333	358
<i>Danio rerio</i> (zebrafish)	2313	0
<i>Aedes aegypti</i> (yellow fever mosquito)	2141	0
<i>Chondrus crispus</i> (Irish moss)	1181	0
<i>Locusta migratoria</i> (migratory locust)	1128	0
<i>Sorghum bicolor</i> (sorghum)	971	0
<i>Chrysemys picta bellii</i> (painted turtle)	843	472
<i>Zea mays</i> (maize)	827	0
<i>Crassostrea gigas</i> (Pasific oyster)	798	6
<i>Nematostella vectensis</i> (starlet sea anemone)	757	6
<i>Hydra vulgaris</i> (hydra)	745	55
<i>Culex quinquefasciatus</i> (southern house mosquito)	660	0
<i>Lepeophtheirus salmonis</i> (salmon louse)	654	0
<i>Homo sapiens</i> (human)	581	0
<i>Glycine max</i> (soybean)	560	3
<i>Xenopus tropicalis</i> (western clawed frog)	535	1
<i>Salmo salar</i> (Atlantic salmon)	535	0
<i>Malus domestica</i> (apple)	526	2
<i>Arabidopsis thaliana</i> (thale cress)	525	0
<i>Alligator mississippiensis</i> (American alligator)	508	0

Table 3. TE classification and the numbers of entries in Repbase.

A. LTR retrotransposon

Superfamily	Total	2017 increase	2016 increase
<i>Gypsy</i>	10595	286	304
<i>Copia</i>	6694	170	239
<i>BEL</i>	1855	2	26
<i>DIRS</i>	418	38	4
Endogenous Retrovirus (ERV)			
<i>ERV1</i>	1967	17	71
<i>ERV2</i>	1266	3	35
<i>ERV3</i>	657	0	8
<i>ERV4</i>	187	0	0
<i>Lentivirus</i>	4	0	0
Unclassified ERV	325	16	0
Unclassified LTR retrotransposon	719	132	54
Total	24687	664	741

B. Non-LTR retrotransposon

Superfamily (clade)	Total	2017 increase	2016 increase
SINE			
<i>SINE1/7SL</i>	95	0	0
<i>SINE2/tRNA</i>	539	20	45
<i>SINE3/5S</i>	30	0	2
<i>SINEU/snRNA</i>	17	0	2
Unclassified SINE	112	0	3
<i>CRE</i>	43	0	7
<i>R4</i>	46	1	1
<i>Hero</i>	23	0	1
<i>NeSL</i>	106	9	0
<i>R2</i>	159	0	81
<i>RandI/Dualen</i>	13	0	5
<i>Proto1</i>	6	0	0
<i>L1</i>	1690	42	109
<i>Tx1</i>	273	1	9
<i>RTETP</i>	1	0	0
<i>Proto2</i>	47	2	1
<i>RTEX</i>	138	3	2
<i>RTE</i>	487	12	34
<i>Outcast</i>	23	0	0
<i>Ingi</i>	17	3	0
<i>Vingi</i>	141	8	0
<i>I</i>	195	1	0
<i>Nimb</i>	108	3	2
<i>Tad1</i>	141	1	0
<i>Loa</i>	74	0	5
<i>R1</i>	237	0	0
<i>Jockey</i>	243	1	0
<i>Rex1</i>	95	0	8
<i>CR1</i>	803	32	39
<i>Kiri</i>	91	0	0
<i>L2</i>	285	9	56
<i>L2A</i>	5	0	0
<i>L2B</i>	27	0	1
<i>Crack</i>	140	0	3
<i>Daphne</i>	227	3	43
<i>Ambal</i>	8	0	0
<i>Penelope</i>	477	108	18
Unclassified non-LTR retrotransposon	179	57	13
Total	7341	316	490

C. DNA transposon

Superfamily	Total	2017 increase	2016 increase
<i>EnSpm/CACTA</i>	715	29	90
<i>Transib</i>	123	0	0
<i>hAT</i>	2955	280	260
<i>MuDR</i>	1345	284	362
<i>P</i>	189	0	2
<i>Kolobok</i>	286	35	9
<i>Dada</i>	36	0	0
<i>Mariner/Tc1</i>	2539	91	79
<i>Zator</i>	54	1	0
<i>piggyBac</i>	377	6	18
<i>Merlin</i>	75	1	0
<i>Harbinger</i>	1097	103	137
<i>ISL2EU</i>	88	0	0
<i>Ginger1</i>	39	19	0
<i>Ginger2/TDD</i>	20	0	1
<i>IS3EU</i>	23	0	1
<i>Sola</i>			
<i>Sola1</i>	100	0	0
<i>Sola2</i>	90	0	1
<i>Sola3</i>	28	0	0
Unclassified <i>Sola</i>	1	0	0
<i>Academ</i>	90	1	1
<i>Novosib</i>	9	0	0
<i>Zisupton</i>	18	0	2
<i>Helitron</i>	955	55	288
<i>Polinton</i>	108	4	3
<i>Crypton</i>			
<i>CryptonF</i>	23	0	0
<i>CryptonA</i>	17	0	0
<i>CryptonI</i>	9	0	0
<i>CryptonS</i>	59	0	0
<i>CryptonV</i>	46	0	2
Unclassified <i>Crypton</i>	80	0	0
Unclassified DNA transposon	2357	135	167
Total	13960	1044	1423