

## Year-end summary of Repbase 2016

**Authors:** Kenji K. Kojima, Weidong Bao, Oleksiy Kohany, Jolanta Walichiewicz, and Elzbieta Jurka

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

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

**Published:** 2017/01/31

### *Update of Repbase 2016*

In the year 2016, *Repbase Update* increased the number of entries by 2716 including 2351 sequences published on *Repbase Reports*. The total number of entries reached 45483 at the end of year 2016. They include repeat sequences published elsewhere, representative sequences characterized by automated pipelines, reference sequences for multicopy genes, and uncharacterized repeats. The total number of transposable elements (TEs) sequences is 44207.

### *Species expansion*

Table 1 shows the top 10 species whose entries increased in the year 2016. Our extensive effort has characterized almost 2000 TE families from the domestic rice *Oryza sativa*. Around a half of repeat entries in 2016 were from rice. Now, rice is the third repeat family-rich species in Repbase (Table 2). The number of zebrafish *Danio rerio* repeat entries is the highest due to the addition of 375 sequences in 2016. Almost all repeat sequences from *Helobdella robusta*, *Setaria italica*, *Callorhinchus milii*, *Condylura cristata*, *Cicer arietinum*, *Mesobuthus martensii*, and *Mnemiopsis leidyi* in Repbase were reported in 2016.

### *Current 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. *Sola* (Bao et al. 2009) and *Crypton* (Kojima and Jurka 2011) has been sub-classified extensively and is reflected on the first update of 2017.

*SINEU*, which was proposed by Kojima (2015), had included TEs only from crocodylians, but SINEs with U2 snRNA-derived sequences were also found from *C. milii* (*SINEU-1\_CM* and *SINEU-1B\_CM*) in 2016 (Bao 2016). Besides the TEs shown in Table 3, Repbase contains 264 unclassified TEs, which are non-autonomous and do not have signs for characterization, even whether DNA transposon or retrotransposon, and 90 “Integrated viruses”, which are mostly Caulimoviruses. It is noteworthy that the category “Integrated virus” in Repbase excludes endogenous retroviruses and “Polintoviruses”. Polintovirus was proposed based on the finding of capsid proteins encoded by *Polintons* (Krupovic et al. 2014; Krupovic and Koonin 2015). Outside of TEs, we replaced the “Pseudogene” class and by “Multicopy gene” to indicate more clearly that these sequences are representatives of multicopy genes, which are not always non-functional.

Since many sequences published in 2016 were TEs from rice, families abundant in rice, such as *MuDR*, *Helitron*, *Gypsy*, *Harbinger*, *hAT*, *Copia* and *L1* increased by more than 100 in 2016. The increase of *R2* is due to our article regarding the evolution of *R2* (Kojima et al. 2016). The increase in *L2*, *Daphne* and *CR1* is mainly due to the TEs from *H. robusta*.

## **References**

Bao W, Kojima KK, Kohany O.

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

Mob DNA, 2015;6:11.

Kojima KK.

A new class of SINEs with snRNA gene-derived heads.

Genome Biology and Evolution, 2015;7:1702-1712.

Bao W.

Non-LTR retrotransposons from the elephant shark genome.

Repbase Reports, 2016; 16: 2313-2346

Bao W, Jurka MG, Kapitonov VV, Jurka J

New superfamilies of eukaryotic DNA transposons and their internal divisions.

Mol Biol Evol, 2009 May;26(5):983-93

Kojima KK, Jurka J

Crypton transposons: identification of new diverse families and ancient domestication events.

Mob DNA, 2011 Oct 19;2(1):12

Krupovic M, Bamford DH, Koonin EV.

Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses.

Biol Direct. 2014; 9:6.

Krupovic M, Koonin EV.

Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution.

Nat Rev Microbiol. 2015; 13:105-15.

Kojima KK, Seto Y, Fujiwara H.

The wide distribution and change of target specificity of R2 non-LTR retrotransposons in animals.

PLoS ONE., 2016 Sep 23;11(9): e0163496.

**Table 1. Top 10 species whose entries increased in 2016.**

<b>Species</b>	<b>2016 Increase</b>	<b>Total</b>
<i>Oryza sativa</i> (Asian rice)	1446	1980
<i>Danio rerio</i> (zebrafish)	375	2313
<i>Helobdella robusta</i> (Californian leech)	156	189
<i>Setaria italica</i> (foxtail millet)	134	137
<i>Callorhynchus milii</i> (elephant shark)	128	140
<i>Condylura cristata</i> (star-nosed mole)	73	73
<i>Cicer arietinum</i> (chickpea)	67	76
<i>Mesobuthus martensii</i> (Chinese armor-tail scorpion)	56	56
<i>Mnemiopsis leidyi</i> (sea walnut)	30	30
<i>Anopheles gambiae</i> (malaria mosquito)	19	86

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

<b>Species</b>	<b>Total</b>	<b>2016 Increase</b>
<i>Danio rerio</i> (zebrafish)	2313	375
<i>Aedes aegypti</i> (yellow fever mosquito)	2141	0
<i>Oryza sativa</i> (Asian rice)	1980	1446
<i>Chondrus crispus</i> (Irish moss)	1181	0
<i>Locusta migratoria</i> (migratory locust)	1128	0
<i>Sorghum bicolor</i> (sorghum)	971	0
<i>Zea mays</i> (maize)	827	0
<i>Crassostrea gigas</i> (Pacific oyster)	792	1
<i>Nematostella vectensis</i> (starlet sea anemone)	751	0
<i>Hydra vulgaris</i> (hydra)	690	0
<i>Culex quinquefasciatus</i> (southern house mosquito)	660	0
<i>Lepeophtheirus salmonis</i> (salmon louse)	654	0
<i>Homo sapiens</i> (human)	583	0
<i>Glycine max</i> (soybean)	557	0
<i>Salmo salar</i> (Atlantic salmon)	535	1
<i>Xenopus (Silurana) tropicalis</i> (western clawed frog)	534	0
<i>Arabidopsis thaliana</i> (thale cress)	525	0
<i>Malus domestica</i> (apple)	524	0
<i>Alligator mississippiensis</i> (American alligator)	508	0
<i>Melampsora larici-populina 98AG31</i> (poplar leaf rust fungus)	500	0

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

*A. LTR retrotransposon*

<b>Superfamily</b>	<b>Total</b>	<b>2016 increase</b>
<i>Gypsy</i>	10312	304
<i>Copia</i>	6523	239
<i>BEL</i>	1853	26
<i>DIRS</i>	380	4
Endogenous Retrovirus (ERV)		
<i>ERV1</i>	1950	71
<i>ERV2</i>	1263	35
<i>ERV3</i>	657	8
<i>ERV4</i>	187	0
<i>Lentivirus</i>	4	0
Unclassified endogenous retrovirus	309	0
Unclassified LTR retrotransposon	586	54
<b>Total</b>	<b>24024</b>	<b>741</b>

*B. Non-LTR retrotransposon*

<b>Superfamily (clade)</b>	<b>Total</b>	<b>2016 increase</b>
SINE		
<i>SINE1/7SL</i>	95	0
<i>SINE2/tRNA</i>	517	45
<i>SINE3/5S</i>	30	2
<i>SINEU/snRNA</i>	17	2
Unclassified SINE	112	3
<i>CRE</i>	43	7
<i>R4</i>	45	1
<i>Hero</i>	23	1
<i>NeSL</i>	97	0
<i>R2</i>	159	81
<i>Randi/Dualen</i>	13	5
<i>Proto1</i>	6	0
<i>L1</i>	1646	109
<i>Tx1</i>	272	9
<i>RTETP</i>	1	0
<i>Proto2</i>	44	1
<i>RTEX</i>	134	2
<i>RTE</i>	473	34
<i>Outcast</i>	23	0
<i>Ingi</i>	14	0
<i>Vingi</i>	133	0
<i>I</i>	194	0
<i>Nimb</i>	105	2
<i>Tad1</i>	140	0
<i>Loa</i>	74	5
<i>R1</i>	237	0
<i>Jockey</i>	242	0
<i>Rex1</i>	95	8
<i>CR1</i>	767	39
<i>Kiri</i>	91	0
<i>L2</i>	276	56
<i>L2A</i>	5	0
<i>L2B</i>	27	1
<i>Crack</i>	140	3
<i>Daphne</i>	224	43
<i>Ambal</i>	8	0
<i>Penelope</i>	366	18
Unclassified non-LTR retrotransposon	128	13
<b>Total</b>	<b>7016</b>	<b>490</b>

*C. DNA transposon*

<b>Superfamily</b>	<b>Total</b>	<b>2016 increase</b>
<i>EnSpm/CACTA</i>	681	90
<i>Transib</i>	123	0
<i>hAT</i>	2664	260
<i>MuDR</i>	1049	362
<i>P</i>	189	2
<i>Kolobok</i>	251	9
<i>Dada</i>	36	0
<i>Mariner/Tc1</i>	2442	79
<i>Zator</i>	53	0
<i>piggyBac</i>	371	18
<i>Merlin</i>	73	0
<i>Harbinger</i>	989	137
<i>ISL2EU</i>	88	0
<i>Ginger1</i>	20	0
<i>Ginger2/TDD</i>	20	1
<i>IS3EU</i>	23	1
<i>Sola</i>		
<i>Sola1</i>	100	0
<i>Sola2</i>	90	1
<i>Sola3</i>	28	0
Unclassified <i>Sola</i>	1	0
<i>Academ</i>	89	1
<i>Novosib</i>	9	0
<i>Zisupton</i>	18	2
<i>Helitron</i>	900	288
<i>Polinton</i>	103	3
<i>Crypton</i>		
<i>CryptonF</i>	23	0
<i>CryptonA</i>	17	0
<i>CryptonI</i>	9	0
<i>CryptonS</i>	59	0
<i>CryptonV</i>	46	2
Unclassified <i>Crypton</i>	80	0
Unclassified DNA transposon	2259	167
<b>Total</b>	<b>12903</b>	<b>1423</b>