

## Rebase 2019 Year in review

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

**Affiliation:** Genetic Information Research Institute, 20380 Town Center Lane Suite 240, Cupertino, CA 95014, USA

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

**Published:** 2020/02/14

### *Update of Rebase 2019*

In the year 2019, Rebase increased the number of entries by 3,103, including 2,745 sequences published on Rebase Reports. The total number of entries reached 52,503 at the end of the year 2019. These numbers include not only consensus and representative sequences of transposable elements (TEs), but also reference sequences for satellite repeats and microsatellites (547 entries), multicopy genes (165 entries), integrated viruses (179 entries), and uncharacterized repeats. The total number of TE sequences is 51,121.

Rebase transitioned to the subscription-based model on April 12, 2019. Since then, we have (1) increased the number of repeats published in each issue of Rebase Reports, and (2) covered more diverse organisms each month (Table 1). The choice of organisms is based on (1) the economic and scientific importance, (2) the quality of genome sequence, and (3) the phylogenetical distance from well-analyzed organisms.

At the end of the year 2019, Rebase includes >100 entries from >100 species or >10 entries from >300 species. The numbers of entries for three TE-rich species (Asian rice, yellow fever mosquito, and zebrafish) exceed 2,000 (Table 2). Table 3 shows the top 10 species whose entries increased in the year 2019, and Table 4 shows the updates of some model organisms. Even though the genomes of model organisms are very well analyzed, they still contain uncharacterized TE families to be discovered. The characteristics of several species we focused on in the year 2019 are described below.

### *Aedes aegypti* (yellow fever mosquito)

The yellow fever mosquito is the pest transmitting devastating disease yellow fever. The genome of this pest insect is very TE family-rich; especially LTR retrotransposons are very

diverse. Reported 1,999 LTR retrotransposons from *A. aegypti* are now classified into 725 *Gypsy*, 366 *Copia*, and 908 *BEL* LTR retrotransposons. Since LTR retrotransposons deposited in Repbase are split into 2 parts, LTR and internal portion, it means that there are ~1,000 LTR retrotransposons are or have been recently active in the genome.

#### *Cyprinus carpio* (common carp)

The common carp belongs to the Cypriniformes, which includes the model organism zebrafish *Danio rerio*, and goldfish *Carassius auratus*. Amur carp (*Cyprinus rubrofasciatus*), whose colored varieties are known as Koi, is closely related to the common carp, too. Almost all published repeat sequences from the carp in Repbase are DNA transposons. It is very likely that many TEs, including non-LTR retrotransposons and endogenous retroviruses, are still to be characterized.

#### *Saccharum* hybrid cultivar SP80-3280 (sugarcane)

Sugarcane is one of the most important energy source crops in the world. It fulfills 70% of world sugar needs, as well as a source of bioethanol. The published entries of Repbase from this cultivar are a subset of characterized repeat sequences, and all are DNA transposons. Repbase also includes 120 entries from another hybrid cultivar of sugarcane, R570, and they are all LTR retrotransposons. All sugarcane species interbreed, and most of common commercial cultivars are complex hybrids. Many more sequences from this cultivar are planned to be published in the coming years.

#### *Beta vulgaris* (beet)

*Beta vulgaris* includes varieties of vegetables called beet, sugar beet, and Swiss chard (leaf beet). Sugar beets accounted for 20% of the world's sugar production. Beet belongs to the family Amaranthaceae, which includes spinach, *Amaranthus*, and quinoa. Repbase does not contain sequences from these crop species other than beet at present. Many more repeat sequences from beet and quinoa are planned to be published in the coming years.

#### *Xenopus laevis* (African clawed frog)

*Xenopus laevis*, as well as its close relative, *Xenopus tropicalis*, are widely used in the studies of developmental biology. *X. laevis* is an allotetraploid, generated by a hybrid event of two closely related species. Repbase contains 287 and 534 repeat sequences from *X. laevis* and *X. tropicalis*, respectively.

### *Drosophila melanogaster* (fruit fly)

In the year 2019, 44 sequences from *Drosophila melanogaster* were added to Repbase, as well as 71 TE families from *Drosophila serrata*. In total, Repbase contains 2,424 repeat entries from the genus *Drosophila*. *Drosophila* species with >100 entries in Repbase are *D. melanogaster* (248 entries), *D. willistoni* (240), *D. ananassae* (240), *D. takahashii* (144), *D. yakuba* (136), *D. ficusphila* (129), *D. rhopaloa* (116), and *D. persimilis* (106).

### *Mus musculus* (domestic mouse)

The domestic mouse *Mus musculus* is the widely used model mammalian species. Despite the fact that the draft mouse genome was published 18 years ago, there remain uncharacterized repeats in the mouse genome, and 35 new repeat sequences were added in the year 2019. Besides 348 entries just for *Mus musculus*, there are many entries of repeats distributed more widely in rodents, such as 49 for the genus *Mus*, 99 for Muridae, 44 for Rodentia, 229 for Eutheria, and 173 for Mammalia. In total, 1,230 repeat families deposited in Repbase might be found in the mouse genome, whereas some ancient families may have been eliminated from the genome completely.

### *Arabidopsis thaliana* (thale cress)

*Arabidopsis thaliana* is the most well-studied land plant species, and belongs to the family Brassicaceae, which includes several economically important crops, such as cabbage, Chinese cabbage, rapeseed, and radish. Repbase now contains 536 and 553 repeat entries from *A. thaliana* and its close relative *A. lyrata*, respectively.

### ***TE classification***

We have not updated our classification scheme since Bao et al. (2015). For more information and comparison with other classification schemes, please read our most recent review article (Kojima 2019). The number of Repbase entries in each category is shown in Table 5, as well as the increase in the number of entries in the last three years. It is noteworthy that our continuous efforts for classification result in the decrease in the number of unclassified LTR retrotransposons and DNA transposons. The increases in numbers of LTR retrotransposons are largely due to those from the yellow fever mosquito *Aedes aegypti*. The increases in numbers of ERVs are due to those from mammals and birds. Regarding non-LTR

retrotransposons, the increase of *Tad1* is caused by the presence of diverse families of *Tad1* from the plant parasite fungus *Erysiphe pisi*. Finally, the increases in DNA transposons are mostly in superfamilies abundant in plant genomes, such as *hAT*, *MuDR*, *Mariner*, *Harbinger*, and *Helitron*.

### **References**

Bao W, Kojima KK, Kohany O.

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

Mob DNA, 2015;6:11.

Kojima KK

Structural and sequence diversity of eukaryotic transposable elements

Genes and Genetic Systems, 2018 advanced online publication on Nov. 9

**Table 1. Organisms published in each issue of *Repbse Reports*.**

RR issue	Published entries	Vertebrate	Invertebrate	Plant	Fungi	Others
19(1)	156		<i>Aedes</i> (mosquito)	<i>Saccharum</i> (sugarcane)	<i>Erysiphe pisi</i>	
19(2)	166	<i>Pelodiscus</i> (turtle)	<i>Aedes</i> (mosquito)	<i>Brassica</i> (cabbage)		
19(3)	159		<i>Aedes</i> (mosquito)	<i>Arabidopsis lyrata</i>	<i>Erysiphe pisi</i>	
19(4)	159	<i>Microtus</i> (vole)	<i>Aedes</i> (mosquito)	<i>Musa</i> (banana)		
19(5)	225	<i>Mus musculus</i> (mouse)	<i>Acropora</i> (coral)	<i>Vaccinium</i> (cranberry)	<i>Chaetomium</i>	<i>Hyaloperonospora</i> (oomycete)
19(6)	217	<i>Cyprinus</i> (carp)	<i>Drosophila melanogaster</i> (fly)	<i>Beta</i> (beet)	<i>Diversispora</i> , <i>Rhizophagus</i>	
19(7)	202	<i>Orycteropus</i> (aardbark), <i>Eptesicus</i> (bat), <i>Dipodomys</i> (rat)	<i>Aplysia</i> (sea hare)	<i>Arabidopsis thaliana</i> , <i>Arabidopsis lyrata</i>	<i>Ajaromyces</i> , <i>Talaromyces</i>	<i>Albugo</i> (oomycete)
19(8)	276	<i>Xenopus laevis</i> (frog), <i>Uraeginthus</i> (Cordon-bleu)	<i>Aedes</i> (mosquito)	<i>Trifolium</i> (clover), <i>Saccharum</i> (sugarcane)	<i>Tricholoma</i>	
19(9)	248	<i>Parus</i> (tit)	<i>Tribolium</i> (beetle)	<i>Daucus</i> (carrot), <i>Solanum</i> (potato)		
19(10)	254	<i>Chinchilla</i> (chinchilla), <i>Trichechus</i> (manatee)	<i>Drosophila serrata</i> (fly), <i>Schmidtea</i> (planaria), <i>Litopenaeus</i> (shrimp)	<i>Dendrobium</i> (orchid), <i>Zea</i> (maize)	<i>Botryobasidium</i>	
19(11)	363	<i>Mustela</i> (ferret), <i>Oryzias</i> (medaka)	<i>Copidosoma</i> (wasp), <i>Litopenaeus</i> (shrimp)	<i>Olea</i> (olive), <i>Beta</i> (beet)	<i>Agaricus</i> (mushroom)	
19(12)	313	<i>Aquila</i> (eagle), <i>Anser</i> (goose), <i>Cyprinus</i> (carp)	<i>Litopenaeus</i> (shrimp), <i>Folsomia</i> (springtail)	<i>Nicotiana</i> (tobacco)	<i>Talaromyces</i>	

Species whose repeats are published first in *Repbse Reports* are colored in red.

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

<b>Species</b>	<b>Total</b>	<b>2019 Increase</b>
<i>Oryza sativa</i> (Asian rice)	3012	1
<i>Aedes aegypti</i> (yellow fever mosquito)	2767	451
<i>Danio rerio</i> (zebrafish)	2323	1
<i>Chondrus crispus</i> (Irish moss)	1183	2
<i>Locusta migratoria</i> (migratory locust)	1127	0
<i>Sorghum bicolor</i> (sorghum)	971	0
<i>Zea mays</i> (maize)	888	64
<i>Chrysemys picta bellii</i> (painted turtle)	843	0
<i>Crassostrea gigas</i> (Pacific oyster)	797	0
<i>Nematostella vectensis</i> (starlet sea anemone)	757	0
<i>Hydra vulgaris</i> (hydra)	745	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)	558	0
<i>Arabidopsis lyrata</i> (lyrate rockcress)	553	100
<i>Arabidopsis thaliana</i> (thale cress)	536	9
<i>Salmo salar</i> (Atlantic salmon)	535	0
<i>Xenopus tropicalis</i> (western clawed frog)	534	0
<i>Malus domestica</i> (apple)	526	0
<i>Alligator mississippiensis</i> (American alligator)	508	0

**Table 3. Top 10 species whose entries increased in 2019.**

<b>Species</b>	<b>2019 Increase</b>	<b>Total</b>
<i>Aedes aegypti</i> (yellow fever mosquito)	451	2767
<i>Litopenaeus vannamei</i> (whiteleg shrimp)	202	317
<i>Vaccinium macrocarpon</i> (American cranberry)	146	187
<i>Folsomia candida</i> (springtail)	114	114
<i>Cyprinus carpio</i> (common carp)	105	106
<i>Arabidopsis lyrata</i> (lyrate rockcress)	100	553
<i>Uraeginthus cyanocephalus</i> (Blue-capped cordon-bleu)	98	98
<i>Saccharum</i> hybrid cultivar SP80-3280 (sugarcane)	87	142
<i>Copidosoma floridanum</i> (parasitic wasp)	87	87
<i>Beta vulgaris</i> (sugar beet)	77	82

**Table 4. Update of model organisms in 2019.**

<b>Species</b>	<b>2019 Increase</b>	<b>Total</b>
<i>Zea mays</i> (maize)	64	888
<i>Xenopus laevis</i> (African clawed frog)	50	287
<i>Drosophila melanogaster</i> (fruit fly)	44	248
<i>Mus musculus</i> (house mouse)	35	348
<i>Oryzias latipes</i> (medaka)	28	93
<i>Tribolium castaneum</i> (red flour beetle)	11	88
<i>Arabidopsis thaliana</i> (thale cress)	9	536



**Table 5. TE classification and the numbers of entries in Repbase.**

<i>A. LTR retrotransposon</i>				
<b>Superfamily</b>	<b>Total</b>	<b>2019 increase</b>	<b>2018 increase</b>	<b>2017 increase</b>
<i>Gypsy</i>	11617	432	590	286
<i>Copia</i>	7388	314	380	170
<i>BEL</i>	2183	311	17	2
<i>DIRS</i>	459	41	0	38
Endogenous Retrovirus (ERV)				
<i>ERV1</i>	2147	160	20	17
<i>ERV2</i>	1461	182	13	3
<i>ERV3</i>	799	129	13	0
<i>ERV4</i>	193	6	0	0
<i>Lentivirus</i>	4	0	0	0
Unclassified ERV	328	3	1	16
Unclassified LTR retrotransposon	684	-37	2	132
<b>Total</b>	<b>27263</b>	<b>1541</b>	<b>1036</b>	<b>664</b>

**B. Non-LTR retrotransposon**

<b>Superfamily (clade)</b>	<b>Total</b>	<b>2019 increase</b>	<b>2018 increase</b>	<b>2017 increase</b>
SINE				
<i>SINE1/7SL</i>	95	4	0	0
<i>SINE2/tRNA</i>	544	74	5	20
<i>SINE3/5S</i>	31	1	1	0
<i>SINEU/snRNA</i>	17	0	0	0
Unclassified SINE	111	0	-1	0
<i>CRE</i>	43	7	0	0
<i>R4</i>	46	6	0	1
<i>Hero</i>	23	1	0	0
<i>NeSL</i>	118	0	12	9
<i>R2</i>	159	1	0	0
<i>RandI/Dualen</i>	13	0	0	0
<i>Proto1</i>	6	0	0	0
<i>L1</i>	1731	56	41	42
<i>Tx1</i>	278	5	5	1
<i>RTETP</i>	1	0	0	0
<i>Proto2</i>	47	0	0	2
<i>RTEX</i>	140	4	2	3
<i>RTE</i>	527	14	40	12
<i>Outcast</i>	23	0	0	0
<i>Ingi</i>	17	5	0	3
<i>Vingi</i>	143	0	2	8
<i>I</i>	203	0	8	1
<i>Nimb</i>	112	3	4	3
<i>Tad1</i>	183	108	42	1
<i>Loa</i>	76	5	2	0
<i>R1</i>	247	10	10	0
<i>Jockey</i>	256	5	13	1
<i>Rex1</i>	95	5	0	0
<i>CR1</i>	851	5	48	32
<i>Kiri</i>	91	0	0	0
<i>L2</i>	286	3	1	9
<i>L2A</i>	5	0	0	0
<i>L2B</i>	29	0	2	0
<i>Crack</i>	140	0	0	0
<i>Daphne</i>	230	6	3	3
<i>Ambal</i>	8	0	0	0
<i>Penelope</i>	494	27	17	108
Unclassified non-LTR retrotransposon	175	0	-4	57
<b>Total</b>	<b>7954</b>	<b>360</b>	<b>253</b>	<b>316</b>

**C. DNA transposon**

<b>Superfamily</b>	<b>Total</b>	<b>2019 increase</b>	<b>2018 increase</b>	<b>2017 increase</b>
<i>EnSpm/CACTA</i>	757	10	32	29
<i>Transib</i>	135	13	-1	0
<i>hAT</i>	3640	529	156	280
<i>MuDR</i>	1548	166	37	284
<i>P</i>	199	9	1	0
<i>Kolobok</i>	328	34	8	35
<i>Dada</i>	36	0	0	0
<i>Mariner/Tc1</i>	2849	229	81	91
<i>Zator</i>	61	7	0	1
<i>piggyBac</i>	400	23	0	6
<i>Merlin</i>	89	14	0	1
<i>Harbinger</i>	1280	115	68	103
<i>ISL2EU</i>	113	3	22	0
<i>Ginger1</i>	39	0	0	19
<i>Ginger2/TDD</i>	21	12	1	0
<i>IS3EU</i>	35	8	4	0
<i>Sola</i>				
<i>Sola1</i>	102	2	0	0
<i>Sola2</i>	92	0	2	0
<i>Sola3</i>	29	1	0	0
Unclassified <i>Sola</i>	0	-1	0	0
<i>Academ</i>	90	0	0	1
<i>Novosib</i>	9	0	0	0
<i>Zisupton</i>	22	3	1	0
<i>Helitron</i>	1021	41	25	55
<i>Polinton</i>	132	13	11	4
<i>Crypton</i>				
<i>CryptonF</i>	23	0	0	0
<i>CryptonA</i>	18	1	0	0
<i>CryptonI</i>	10	1	0	0
<i>CryptonS</i>	59	0	0	0
<i>CryptonV</i>	50	3	4	0
Unclassified <i>Crypton</i>	87	0	7	0
Unclassified DNA transposon	2362	-53	58	135
<b>Total</b>	<b>15651</b>	<b>1183</b>	<b>508</b>	<b>1044</b>