

Rebase 2022 Year in review

Authors: Kenji K. Kojima, Weidong Bao, Noriko F. Kojima, 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: 2023/02/14

Update of Rebase 2022

In the year 2022, Rebase increased the number of entries by 15,024, including 7,884 sequences published on Rebase Reports. The total number of entries reached 81,594 at the end of the year 2022. The majority (80,044: 98%) of these entries are transposable elements (TEs), either consensus or representative sequences. The remaining are satellite repeats and microsatellites (601 entries), multicopy genes (197 entries), integrated viruses (208 entries), and uncharacterized repeats.

Rebase transitioned to a subscription-based model on April 12, 2019. The last version of Rebase before the transition was Rebase Update 24.03 which included 50,356 entries. 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 the genome sequence, and (3) the phylogenetic distance from well-analyzed organisms. We have continually updated, corrected, and refined existing repeats that contain sequence flaws or are ambiguously classified. We updated over 500 such entries in the year 2022.

A selected portion of the mammalian repeat dataset generated in the Zoonomia project (Zoonomia Consortium 2020) was incorporated into Rebase in 2022. It is noteworthy that although they passed through our preliminary quality check, the imported dataset is not as of high quality as the regular entries in terms of sequence completeness and consensus reconstruction. There also may be duplicates. [These sequences will be subject to Rebase routine scrutiny and updating in the future.](#) These sequences can be recognized through their names; for example, *CapPil-1.131* is from *Capromys pilorides* (Desmarest's hutia). Sometimes,

the decimal point has been lost, such as *CavTsc-4472* (its original name is *CavTsc-4.472*, from *Cavia tschudii*) due to the conflict against the nomenclature system in Repbase.

The incorporation of Zoonomia dataset results in a significant increase in the number of mammalian repeat entries in Repbase. Repbase has been expanding to be a large dataset covering various eukaryotic repeats. The three major phylogenetic groups of Repbase entries are vertebrates, arthropods, and green plants (Figure 1). Vertebrate includes humans, lab animals, and almost all domestic or farmed animals for meat, labor/transportation, and pets (cow, sheep, goat, pig, camel, llama, horse, cat, dog, Guinea pig, mice, hamsters, rabbits, chicken, duck, goldfish, carp, salmon, tilapia, catfish, etc.). Major crops are classified into only a few orders in green plants: Poales (wheat, barley, rye, oat, maize, sorghum, millets, sugarcane, rice), Brassicales (cabbage/kale/broccoli, Chinese cabbage, radish), Fabales (pea, common bean, soybean, peanut, chickpea, alfalfa), Rosales (apple, pear, cherry, plum, strawberry), and Solanales (potato, tomato, sweet potato). Arthropod is the largest group of animals and counts for >80% of all living animal species. Although arthropod includes only a few domestic or farmed animals (such as honeybee, silkworm, shrimp), it includes a vast majority of pests for humans and domestic animals (mosquitoes, flies, flea, bugs, lice, ticks, etc.), and for crops (caterpillars, flies, sawflies, leaf beetles, aphids, whiteflies, locusts, etc.). Other noticeable groups are mollusks (including oyster, mussel, clam, and abalone), fungi (including yeasts, edible mushrooms, and plant pathogens such as *Puccinia* and *Blumeria*), and stramenopiles (including plant pathogens such as *Phytophthora*).

At the end of the year 2022, Repbase contains repeat sequences from over 1,600 species. Four TE-rich species (yellow fever mosquito, Asian rice, zebrafish, and maize) each have over 2,000 entries (Table 2). Sometimes, multiple genomes in the same genus, such as *Drosophila* and *Arabidopsis*, have been sequenced and analyzed. Such analysis helps characterize very low-copy TE families or single-copy TE families in model organisms and reveal the genome evolution over a longer time scale. Table 3 shows the total numbers of Repbase entries for each genus. In some cases, such as *Chondrus* and *Locusta*, only one species has been analyzed, while in other cases, such as *Drosophila* and *Anopheles*, more than 2 species have been examined (Tables 4 and 5). Table 6 shows the top 20 species whose entries increased in the year 2022. Even though the genomes of model organisms are very well analyzed, they still contain TE families to be discovered. The characteristics of several species we focused on in the year 2022 are described below.

Zea mays (maize)

Maize (corn) is a cereal grain first domesticated in southern Mexico about 9,000 years ago from its wild ancestor, called teosinte. Maize has become a primary food in many areas of the world, with the total production of maize surpassing that of wheat or rice. Besides the consumption as food for humans, maize is used for food for domesticated animals and is transformed to ethanol. Barbara McClintock used the knob markers of maize to validate the transposon theory. Now Repbase contains 2,037 entries from maize. Sugarcane, sorghum, and millets are closely related to maize, and together belong to the subfamily Panicoideae in the family Poales. Repbase contains 971 entries from sorghum (*Sorghum bicolor*), 360 from sugarcane (240 from *Saccharum* hybrid cultivar SP-80-3280 and 120 from *Saccharum* hybrid cultivar R570), 207 from proso millet (*Panicum miliaceum*), 165 from Japanese millet (*Echinochloa crus-galli*), and 137 from foxtail millet (*Setaria italica*).

Arabidopsis thaliana (thale cress)

The thale cress *Arabidopsis thaliana* is the best-studied land plant species. It belongs to the family Brassicaceae, which includes several economically important crops, such as cabbage, Chinese cabbage, rapeseed, and radish. In Repbase, there are 537 entries from radish (*Raphanus sativus*), 381 from Chinese cabbage (*Brassica rapa*), 132 from cabbage (*Brassica oleracea*), and 84 from *Schrenkiella parvula/Eutrema parvum*. Repbase now contains 1,136 and 1,567 repeat entries from *A. thaliana* and *A. lyrata*, respectively.

Medicago truncatula (barrel medic)

The barrel medic, *Medicago truncatula* is studied as a model organism for legume biology. Legume (Fabaceae) includes many important crops such as soybeans and peanuts. Most legumes have symbiotic nitrogen-fixing bacteria in structures called root nodules. Legumes can be consumed as food for humans (grain legumes), such as soybeans, peas and peanuts, as well as for animals (forage legumes), such as alfalfa (*Medicago sativa*) and clover. Now Repbase contains 1,001 entries from *M. truncatula*, 558 from *Glycine max* (soybean), 208 from *Arachis ipaensis* (a wild progenitor of peanuts), 119 from *Arachis hypogaea* (peanut), 83 from *Trifolium pratense* (red clover), 76 from *Cicer arietinum* (chickpea), 24 from *Senna tora* (sickle senna), 17 from *Lotus corniculatus* (common bird's-foot trefoil), 13 from *Pisum sativum* (pea), 9 from *Trifolium repens* (white clover), 8 from *Lotus japonicus* (miyakogusa), 4 from

Vicia faba (fava bean), 2 from *Phaseolus vulgaris* (common bean), and 1 from *Cajanus cajan* (pigeon pea).

Crassostrea gigas (Pacific oyster)

The Pacific oyster (*Crassostrea gigas*) is now the most widely farmed and the most commercially important oyster globally. It is native to Asia. It is also considered an invasive species where it has been introduced intentionally or accidentally. In the year 2022, the entries from *C. gigas* increased by 176, and the total number of entries is 1,646.

Corbicula fluminea (Asian clam)

The Asian clam *Corbicula fluminea* is a species of freshwater clam native to eastern Asia and it has successfully invaded throughout North America, South America, and Europe. In the year 2022, the entries from *C. fluminea* is 568. The class Bivalvia are now classified into 4 major subclasses (Heterodonta, Palaeoheterodonta, Protobranchia, and Pteriomorphia). *Corbicula* belongs to Heterodonta, and is distant from Pteriomorphia, which includes oysters and mussels.

Drosophila (fruit flies)

The genus *Drosophila* contains over 1,450 species, among which ~1,100 species belong to the subgenus *Drosophila*. *D. melanogaster* belongs to the subgenus *Sophophora*. In the year 2022, 947 TE families from the genus *Drosophila* were added to Repbase (Table 4). In total, Repbase contains 4,250 repeat entries from the genus *Drosophila*. Besides them, 92 TE families from *Scaptodrosophila lebanonensis* were added. *Scaptodrosophila* is a genus closely related to the genus *Drosophila*.

Anopheles (malaria mosquitoes)

Malaria is one of the most severe public health problems worldwide. Malaria is caused by parasitic protists of the genus *Plasmodium*. Malaria parasites are transmitted to humans by female mosquitoes of the genus *Anopheles*. In the year 2022, two mosquito genomes (*A. funestus* and *A. stephensi*) of the subgenus *Cellia* were analyzed and 251 repeats are added to Repbase (Table 5). *A. funestus* is the major vector of malaria in sub-Saharan Africa. *A. stephensi* is the primary vector of malaria in urban India. In total, 886 entries from the genus *Anopheles* are now available in Repbase.

Litopenaeus vannamei (whiteleg shrimp) and *Penaeus monodon* (Asian tiger shrimp)

These two related shrimp species are farmed in a large scale; *L. vannamei* and *P. monodon* account for roughly 80% of all farmed shrimp. *L. vannamei* is native to the Pacific coast from Mexico to Peru, and is a major farmed shrimp in Latin America. *L. vannamei* was introduced into Asia too. *P. monodon* is native in Indian Ocean and the Pacific Ocean, and mainly farmed in Asia. Repbase now contains 438 entries from *L. vannamei* and 320 entries from *P. monodon*.

Carassius auratus (goldfish)

The goldfish (*Carassius auratus*) is one of the most popular aquarium fish and belongs to the family Cyprinidae of the order Cypriniformes. Goldfish have been selectively bred for color, size, body shape, and fin configuration for 1,000 years. Goldfish and zebrafish *Danio rerio* belong to the same order. Repbase currently contains 631 entries from goldfish, 110 from its relative, the common carp *Cyprinus carpio*, and 2,323 from zebrafish.

Puccinia (rust)

All species in the genus *Puccinia* are obligate plant pathogens and are known as rusts. The genus *Puccinia*, which contains ~4,000 species, is considered the most economically destructive genus of biotrophic fungi; members of this genus are serious pathogens on all major cereal crop species except rice. The entries from three *Puccinia* species (*P. hordei*, *P. triticina*, and *P. coronata*) were added to Repbase in the year 2022 (Table 7).

TE classification

The number of Repbase entries in each TE category is shown in Table 8, accompanied with the increase in the number of entries in the last three years. In the year 2022, we have no update in our classification scheme.

The large increases in *ERV1*, *ERV2*, *ERV3*, *L1*, *RTE*, *SINE1*, *SINE2*, *Mariner*, and *piggyBac* are mostly due to the incorporation of the Zoonomia dataset, as they are major components of mammalian mobilome. The increase in *Naiad/Chlamys* is due to the incorporation of the dataset reported in Craig et al. (2021). The increases in *Kolobok* and *Dada* are the output reported in Kojima and Bao (2022) and Kojima (2022).

References

Bao W, Kojima KK, Kohany O.

Repbase 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, 2020; 94(6):233-252.

Zoonomia Consortium.

A comparative genomics multitool for scientific discovery and conservation.
Nature. 2020 Nov;587(7833):240-245. doi: 10.1038/s41586-020-2876-6.

Craig RJ, Yushenova IA, Rodriguez F, Arkhipova IR.

An ancient clade of *Penelope*-like retroelements with permuted domains is present in the green lineage and protists, and dominates many invertebrate genomes.
Mol Biol Evol., 2021; 38(11):5005-5020.

Kojima KK

Diversity and evolution of DNA transposons targeting multicopy small RNA genes from actinopterygian fish.
Biology (Basel), 2022 Jan 20;11(2)

Kojima KK, Bao W

A unique eukaryotic lineage of composite-like DNA transposons encoding a DDD/E transposase and a His-Me finger homing endonuclease.
Mob DNA, 2022 Oct 22;13(1):24

Table 1. Organisms published in each issue of Rebase Reports.

RR issue	Published entries	Vertebrates	Arthropods	Other invertebrates	Plants	Fungi	Other eukaryotes
22(1)	647	Carassius (goldfish), Coilia (anchovy) , Oreochromis (tilapia)	Drosophila (fruit fly), Trialeurodes (whitefly)	Cyclina (clam)	Corbicula (clam), Patiria (starfish)	Ascochyta , Cordyceps , Grosmannia , Pichia , Phizopus	
22(2)	643	Clupea (herring) , Xenopus (frog)	Anopheles (mosquito), Monomorium (ant) , Trichoplusia (moth)	Oscheius (nematode)	Medicago (barrel medic)	Aquanectria , Articulospora , Aspergillus , Monilinia	
22(3)	657	Clupea (herring), Xenopus (frog)	Anopheles (mosquito), Lymantria (moth)	Crassostrea (oyster)	Zea (maize), Raphanus (radish)	Kazachstania , Lactarius , Trichoderma , Xylaria	
22(4)	692	Carassius (goldfish), Sebastes (rockfish)	Danaus (monarch), Drosophila (fruit fly), Sitophilus (weevil), Penaeus (shrimp)	Bursaphelenchus (nematode) , Helobdella (leech)	Arabidopsis (thale cress), Cannabis (hemp), Cucumis (cucumber)	Mrakia , Paxillus , Penicillium , Phyllosticta	
22(5)	719	Cottoperca (blenny), Amblyraja (skate)	Tinea , Blastobasis , Zygaena , Pieris , Vanessa , Hypolimnas , Endotricha , Habrosyne , Hyles , Manduca, Deilephila , Biston , Hylaea , Spodoptera, Helicoverpa, Lymantria (all moths and butterflies), Drosophila (fruit fly), Trichogramma (wasp), Empoasca (leafhopper) , Penaeus (shrimp)	Owenia (annelid)	Oryza (rice), Brassica (cabbage), Ilex (Yerba mate)	Pyricularia, Tolypocladium , Torulaspora	Trichomonas
22(6)	572	Sciurus (squirrel), Doichotis (cavy), Chinchilla (chinchilla), Dasyprocta (agouti), Thalassophryne (teleost fish), Amblyraja (skate)	Drosophila (fruit fly), Habrosyne (moth), Tenthredo (sawfly)	Lineus (bootlace worm) , Acanthopleura (mollusk)	Saccharum (sugarcane), Raphanus (radish), Apostasia , Brachypodium, Brassica (cabbage), Cannabis (hemp), Cucumis (melon), Dendrobium (orchid) , Humulus (hop) , Linum (flax), Persea (avocado) , Prunus (apricot), Ricinus, Trifolium	Aspergillus, Metschnikowia , Pisolithus , Psilocybe , Reticulascus , Trichophyton , Verticillium	

					(clover), <i>Vigna</i> (bean), <i>Vaccinium</i> (cranberry)		
22(7)	715	<i>Gasterosteus</i> (stickleback), <i>Archocentrus</i> (cichlid)	<i>Drosophila</i> (fruit fly), <i>Trialeurodes</i> (whitefly)	<i>Corbicula</i> (clam), <i>Metaphire</i> (earthworm), <i>Hirudo</i> (leech)	<i>Nicotiana</i> (tobacco), <i>Vigna</i> (cowpea), <i>Gerbera</i> (daisy)	<i>Puccinia</i> (rust)	<i>Trichomonas</i>
22(8)	634	<i>Archocentrus</i> (cichlid)	<i>Drosophila</i> (fruit fly), <i>Sitophilus</i> (weevil), <i>Litopenaeus</i> (shrimp)	<i>Marthasterias</i> (starfish), <i>Crassostrea</i> (oyster)	<i>Zea</i> (maize)	<i>Acidomyces</i> , <i>Meliniomyces</i> , <i>Bsidiobolus</i> , <i>Fistulina</i> , <i>Curvularia</i>	
22(9)	654	<i>Leptobrachium</i> (toad)	<i>Anopheles</i> (mosquito), <i>Penaeus</i> (shrimp)	<i>Corbicula</i> (clam), <i>Fasciola</i> (flatworm)	<i>Gossypium</i> (cotton), <i>Lactuca</i> (lettuce), <i>Zea</i> (maize)	<i>Puccinia</i> (rust)	
22(10)	672	<i>Leptobrachium</i> (toad), <i>Pachypanchax</i> (panchax), <i>Scleropages</i> (arowana)	<i>Drosophila</i> (fruit fly), <i>Ceratitis</i> (fruit fly), <i>Bemisia</i> (whitefly)	<i>Corbicula</i> (clam), <i>Lytechinus</i> (sea urchin)	<i>Zea</i> (maize)	<i>Piromyces</i>	
22(11)	704	<i>Carassius</i> (goldfish), <i>Amphiprion</i> (clownfish)	<i>Scaptodrosophila</i> (fruit fly), <i>Sogaella</i> (planthopper)	<i>Corbicula</i> (clam), <i>Patiria</i> (starfish)	<i>Trifolium</i> (clover), <i>Dioscorea</i> (yam), <i>Capsicum</i> (pepper)	<i>Alternaria</i> , <i>Amylostereum</i> , <i>Aspergillus</i> , <i>Choironomyces</i> , <i>Colletotrichum</i> , <i>Fusarium</i> , <i>Sclerotinia</i>	
22(12)	575	<i>Ilyophis</i> (eel)	<i>Drosophila</i> (fruit fly), <i>Sacculina</i> (barnacle)	<i>Corbicula</i> (clam), <i>Fasciola</i> (trematode), <i>Clonorchis</i> (fluke)		<i>Puccinia</i> (rust)	

Species whose repeats are published first in Repbase Reports are colored in red.

Table 2. Top 25 species based on the number of Rebase entries.

Species	Total	2022 Increase
<i>Aedes aegypti</i> (yellow fever mosquito)	3648	0
<i>Oryza sativa</i> (Asian rice)	3248	172
<i>Danio rerio</i> (zebrafish)	2323	0
<i>Zea mays</i> (maize)	2037	558
<i>Crassostrea gigas</i> (Pacific oyster)	1646	176
<i>Arabidopsis lyrata</i> (lyrate rockcress)	1567	0
<i>Chondrus crispus</i> (Irish moss)	1183	0
<i>Arabidopsis thaliana</i> (thale cress)	1136	180
<i>Locusta migratoria</i> (migratory locust)	1127	0
<i>Medicago truncatula</i> (barrel medic)	1001	129
<i>Sorghum bicolor</i> (sorghum)	971	0
<i>Chrysemys picta bellii</i> (painted turtle)	843	0
<i>Nematostella vectensis</i> (starlet sea anemone)	757	0
<i>Hydra vulgaris</i> (hydra)	745	0
<i>Drosophila takahashii</i> (fruit fly)	736	162
<i>Xenopus tropicalis</i> (western clawed frog)	697	69
<i>Culex quinquefasciatus</i> (southern house mosquito)	660	0
<i>Lepeophtheirus salmonis</i> (salmon louse)	654	0
<i>Drosophila willistoni</i> (fruit fly)	636	395
<i>Carassius auratus</i> (goldfish)	631	325
<i>Homo sapiens</i> (human)	584	0
<i>Corbicula fluminea</i> (Asian clam)	568	568
<i>Glycine max</i> (soybean)	558	0
<i>Raphanus sativus</i> (radish)	537	21
<i>Salmo salar</i> (Atlantic salmon)	535	0

Table 3. Top 20 genera based on the number of Rebase entries.

Genus (Species with >10 entries)	Total	2022 Increase
<i>Drosophila</i> ¹	4250	946
<i>Aedes</i> (<i>A. aegypti</i>)	3649	0
<i>Oryza</i> (<i>O. sativa</i>)	3257	172
<i>Arabidopsis</i> (<i>A. thaliana</i> , <i>A. lyrata</i>)	2703	180
<i>Danio</i> (<i>D. rerio</i>)	2337	0
<i>Zea</i> (<i>Z. mays</i>)	2041	558
<i>Crassostrea</i> (<i>C. gigas</i> , <i>C. virginica</i>)	1832	190
<i>Puccinia</i> (<i>P. striiformis</i> , <i>P. graminis</i> , <i>P. hordei</i> , <i>P. triticina</i> , <i>P. coronata</i> , <i>P. horiana</i>)	1296	412
<i>Chondrus</i> (<i>C. crispus</i>)	1183	0
<i>Locusta</i> (<i>L. migratoria</i>)	1127	0
<i>Xenopus</i> (<i>X. tropicalis</i> , <i>X. laevis</i>)	1033	113
<i>Medicago</i> (<i>M. truncatula</i>)	1007	129
<i>Sorghum</i> (<i>S. bicolor</i>)	971	0
<i>Anopheles</i> ²	886	251
<i>Chrysemys</i> (<i>C. picta</i>)	843	0
<i>Phytophthora</i> (<i>P. infestans</i> , <i>P. sojae</i> , <i>P. ramorum</i> , <i>P.</i> <i>parasitica</i>)	786	3
<i>Nematostella</i> (<i>N. vectensis</i>)	758	0
<i>Hydra</i> (<i>H. vulgaris</i>)	745	0
<i>Culex</i> (<i>C. quinquefasciatus</i>)	669	0
<i>Lepeophtheirus</i> (<i>L. salmonis</i>)	654	0

1 Species are shown in Table 4.

2 Species are shown in Table 5.

Table 4. Repbase entries from the genus *Drosophila*.

Subgenus - Group (subgroup)	Species	Abbr.	Repbase entries	2022 Increase	
<i>Sophophora</i> - <i>melanogaster</i> (<i>melanogaster</i> subgroup)	<i>D. melanogaster</i>	DM	296	0	
	<i>D. simulans</i>	DSim	43	0	
	<i>D. sechellia</i>	DSe	33	0	
	<i>D. yakuba</i>	DY	139	0	
	<i>D. erecta</i>	DEre	18	0	
	(<i>suzukii</i> subgroup)	<i>D. suzukii</i>	DSuz	207	207
		<i>D. biarmipes</i>	DBi	80	0
	(<i>ananassae</i> subgroup)	<i>D. ananassae</i>	DAn	240	0
		<i>D. bipectinata</i>	DBp	158	0
	(<i>elegans</i> subgroup)	<i>D. elegans</i>	DEl	314	0
	(<i>takahashii</i> subgroup)	<i>D. takahashii</i>	DTa	736	162
	(<i>rhopaloa</i> subgroup)	<i>D. rhopaloa</i>	DRh	116	0
	(<i>montium</i> subgroup)	<i>D. kikkawai</i>	DKi	97	0
		<i>D. serrata</i>	DSer	71	0
<i>Sophophora</i> - <i>ficuspbila</i>	<i>D. ficuspbila</i>	DF	129	0	
<i>Sophophora</i> - <i>eugracilis</i>	<i>D. eugracilis</i>	DEu	200	122	
<i>Sophophora</i> - <i>obscura</i>	<i>D. azteca</i>	DAzt	102	0	
	<i>D. pseudoobscura</i>	DPse	80	0	
	<i>D. persimilis</i>	DPer	106	0	
<i>Sophophora</i> - <i>willistoni</i>	<i>D. willistoni</i>	DWil	636	395	
<i>Siphlodora</i> - <i>repleta</i>	<i>D. mojagensis</i>	DMoj	154	0	
	<i>D. hydei</i>	DHyd	39	0	
<i>Siphlodora</i> - <i>virilis</i>	<i>D. virilis</i>	DVi	148	60	
<i>Idiomyia</i>	<i>D. grimshawi</i>	DGri	41	0	

¹ Only species with >10 entries are shown.

² Subgroups are shown only in the *melanogaster* group.

Table 5. Repbase entries from the genus *Anopheles*.

Species	Abbr.	Repbase entries	2022 Increase
<i>A. gambiae</i> str. PEST	AG	379	0
<i>A. funestus</i>	AnFu	340	200
<i>A. merus</i>	AnMe	68	0
<i>A. stephensi</i>	AnSt	95	51

¹ Only species with >10 entries are shown.

Table 6. Top 20 species whose entries increased in 2022.

Species	2022 Increase	Total
<i>Corbicula fluminea</i> (Asian clam)	568	568
<i>Zea mays</i> (maize)	558	2037
<i>Drosophila willistoni</i> (fruit fly)	395	636
<i>Carassius auratus</i> (goldfish)	325	631
<i>Penaeus monodon</i> (Asian tiger shrimp)	320	320
<i>Archocentrus centrarchus</i> (flier cichlid)	251	251
<i>Trialeurodes vaporariorum</i> (greenhouse whitefly)	227	227
<i>Drosophila suzukii</i> (fruit fly)	207	207
<i>Anopheles funestus</i> (mosquito)	200	340
<i>Clupea harengus</i> (Atlantic herring)	198	198
<i>Amblyraja radiata</i> (thorny skate)	181	467
<i>Arabidopsis thaliana</i> (thale cress)	180	1136
<i>Crassostrea gigas</i> (Pacific oyster)	176	1646
<i>Oryza sativa</i> (Asian rice)	172	3248
<i>Puccinia triticina</i> (wheat leaf rust)	167	288
<i>Drosophila takahashii</i> (fruit fly)	162	736
<i>Sacculina carcini</i> (crab hacker barnacle)	161	161
<i>Leptobrachium leishanense</i> (Leishan spiny toad)	156	218
<i>Owenia fusiformis</i>	152	152
<i>Monomorium pharaonis</i> (pharaoh ant)	136	136

Table 7. Rebase entries from the genus *Puccinia*.

Species	Abbr.	Rebase entries	2022 Increase
<i>P. striiformis</i> (stripe rust, or yellow rust)	PSt	289	0
<i>P. triticina</i> (wheat leaf rust, or brown rust)	PTrit	287	167
<i>P. hordei</i> (barley brown rust)	PHord	286	129
<i>P. graminis</i> (stem rust, or black rust)	PG/PGra	233	0
<i>P. coronata</i> (barley crown rust)	PCor	175	116
<i>P. horiana</i> (chrysanthemum white rust)	PHor	25	0

¹ Only species with >10 entries are shown.

Table 8. TE classification and the numbers of entries in Repbase.

A. LTR retrotransposon

Superfamily	Total	2022 increase	2021 increase	2020 increase
<i>Gypsy</i>	22109	4508	4139	1843
<i>Copia</i>	9469	605	908	569
<i>BEL</i>	4022	974	432	432
<i>DIRS</i>	936	104	187	186
Endogenous Retrovirus (ERV)				
<i>ERV1</i>	3885	1578	45	115
<i>ERV2</i>	2440	877	0	102
<i>ERV3</i>	1192	375	0	18
<i>ERV4</i>	202	9	0	0
<i>Lentivirus</i>	4	0	0	0
<i>Lokiretrovirus</i>	55	4	-	-
<i>Spumaretrovirus</i>	7	1	-	-
Unclassified ERV	322	4	-10	0
Unclassified LTR retrotransposon	774	38	9	44
Total	45417	9077	5767	3310

B. Penelope-like retrotransposon

Superfamily	Total	2022 increase	2021 increase	2020 increase
<i>Penelope/Poseidon</i>	341	52	-	-
<i>Neptune</i>	225	5	-	-
<i>Nematis</i>	5	0	-	-
<i>Athena</i>	42	0	-	-
<i>Coprina</i>	15	3	-	-
<i>Hydra</i>	28	6	-	-
<i>Naiad/Chlamys</i>	119	110	-	-
Unclassified <i>Penelope</i>	22	5	-	-
Total	797	181	77	11

C. Non-LTR retrotransposon

Superfamily (clade)	Total	2022 increase	2021 increase	2020 increase
SINE				
<i>SINE1/7SL</i>	192	92	0	1
<i>SINE2/tRNA</i>	1788	977	133	60
<i>SINE3/5S</i>	32	6	-6	0
<i>SINEU/snRNA</i>	17	0	0	0
Unclassified SINE	75	2	-39	1
<i>CRE</i>	55	2	5	0
<i>R4</i>	56	3	0	1
<i>Hero</i>	29	1	0	4
<i>NeSL</i>	119	0	0	0
<i>R2</i>	181	21	1	0
<i>RandI/Dualen</i>	13	0	0	0
<i>Proto1</i>	10	0	4	0
<i>L1</i>	3401	1237	195	184
<i>Tx1</i>	371	66	1	21
<i>RTETP</i>	1	0	0	0
<i>Proto2</i>	53	1	0	5
<i>RTEX</i>	294	141	5	4
<i>RTE</i>	675	108	14	12
<i>Outcast</i>	38	7	8	0
<i>Ingi</i>	34	8	4	0
<i>Vingi</i>	153	0	0	10
<i>I</i>	236	28	4	1
<i>Nimb</i>	126	9	0	2
<i>Tad1</i>	519	32	177	19
<i>Loa</i>	90	2	1	6
<i>R1</i>	294	15	13	9
<i>Jockey</i>	343	44	19	18
<i>Rex1</i>	138	5	7	26
<i>CR1</i>	976	59	44	17
<i>Kiri</i>	108	3	13	1
<i>L2</i>	311	12	2	8
<i>L2A</i>	7	2	0	0
<i>L2B</i>	29	0	0	0
<i>Crack</i>	148	5	1	2
<i>Daphne</i>	245	6	1	2
<i>Ambal</i>	8		0	0
Unclassified non-LTR retrotransposon	225	23	23	5
Total	11390	2917	630	419

D. DNA transposon

Superfamily	Total	2022 increase	2021 increase	2020 increase
<i>EnSpm/CACTA</i>	1145	37	195	156
<i>Transib</i>	225	43	39	8
<i>hAT</i>	5275	557	643	432
<i>MuDR</i>	2113	68	313	182
<i>P</i>	273	27	32	15
<i>Kolobok</i>	802	411	40	6
<i>Dada</i>	171	120	0	15
<i>Mariner/Tc1</i>	4024	865	136	171
<i>Zator</i>	94	5	6	22
<i>piggyBac</i>	549	118	10	21
<i>Merlin</i>	113	20	2	1
<i>Harbinger</i>	1932	146	221	279
<i>ISL2EU</i>	192	28	22	29
<i>Ginger1</i>	20	-19	0	0
<i>Ginger2/TDD</i>	64	20	5	6
<i>IS3EU</i>	45	0	0	10
<i>Sola</i>				
<i>Sola1</i>	139	30	2	5
<i>Sola2</i>	124	21	5	6
<i>Sola3</i>	44	0	2	13
<i>Academ</i>	387	24	34	239
<i>Novosib</i>	9	0	0	0
<i>Zisupton</i>	43	1	1	19
<i>Helitron</i>	1690	250	218	201
<i>Polinton</i>	195	41	12	10
<i>Crypton</i>				
<i>CryptonF</i>	24	0	0	1
<i>CryptonA</i>	18	0	0	0
<i>CryptonI</i>	10	0	0	0
<i>CryptonS</i>	59	0	0	0
<i>CryptonV</i>	53	0	0	0
Unclassified <i>Crypton</i>	89	0	0	2
Unclassified DNA transposon	2333	22	35	-53
Total	22503	2816	1973	1796

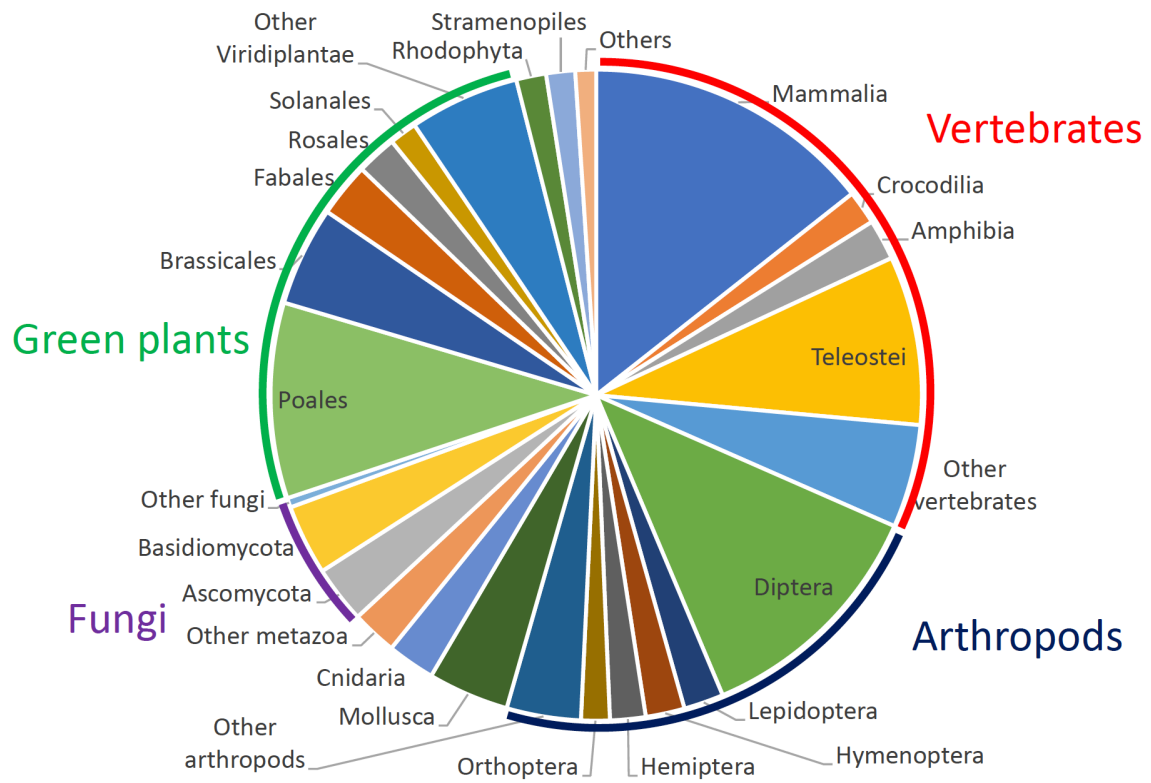


Figure 1. Proportions of Rebase entries in phylogenetic lineages. Lineages with over 1000 entries are shown.