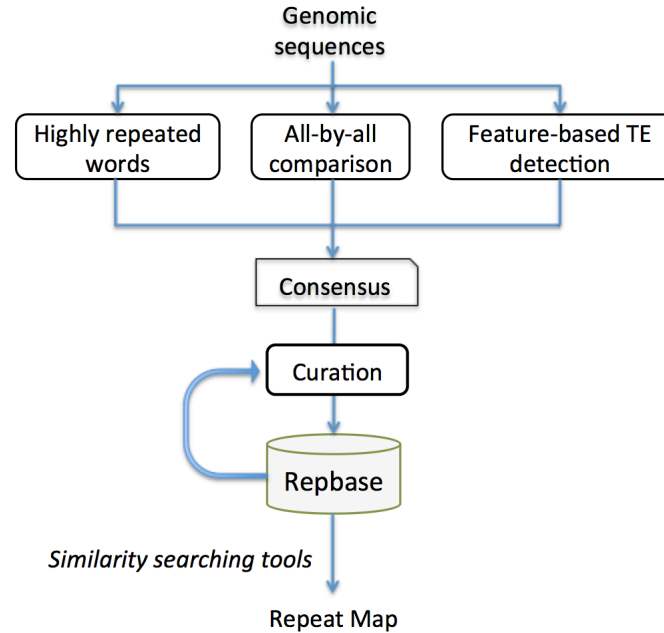## Broad species coverage

Repbase covers over a hundred species, including animals, plants and fungi. The top 40 species by the number of repeats.

No. of Repeats	Species
2141	*Aedes aegypti*
1904	*Danio rerio*
1181	*Chondrus crispus*
1128	*Locusta migratoria*
971	*Sorghum bicolor*
827	*Zea mays*
791	*Crassostrea gigas*
751	*Nematostella vectensis*
689	*Hydra vulgaris*
660	*Culex quinquefasciatus*
654	*Lepeophtheirus salmonis*
630	*Oryza sativa*
583	*Homo sapiens*
557	*Glycine max*
534	*Salmo salar*
534	*Xenopus (Silurana) tropicalis*
527	*Malus domestica*
525	*Arabidopsis thaliana*
508	*Alligator mississippiensis*
500	*Melampsora larici-populina*
461	*Phytophthora infestans*
442	*Blumeria graminis*
434	*Monodelphis domestica*
394	*Crocodylus porosus*
388	*Amborella trichopoda*
383	*Solanum tuberosum*
375	*Latimeria chalumnae*
373	*Anolis carolinensis*
368	*Chrysemys picta bellii*
355	*Anopheles gambiae*
346	*Vitis vinifera*
343	*Acyrthosiphon pisum*
333	*Brassica rapa*
331	*Populus trichocarpa*
322	*Morus notabilis*
313	*Mus musculus*
306	*Esox lucius*
295	*Rhodnius prolixus*
290	*Gavialis gangeticus*
283	*Strongylocentrotus purpuratus*

How we add new repeats to Repbase

Genomic sequences → Highly repeated words / All-by-all comparison / Feature-based TE detection → Consensus → Curation → Repbase

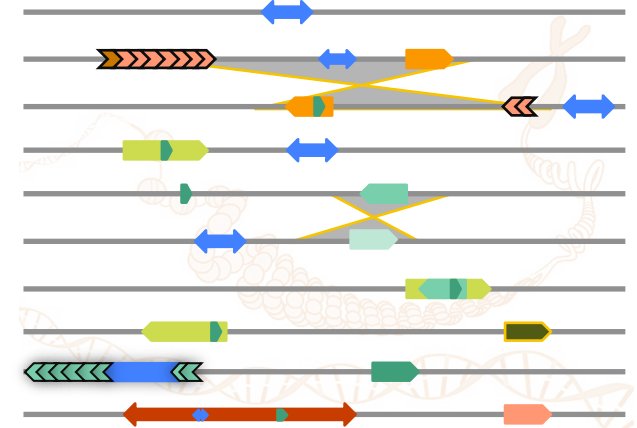*Similarity searching tools*

Repeat Map

BMC Bioinformatics, 2006;7:474; Mob DNA, 2015;6:11

giri Genetic Information Research Institute

The **Genetic Information Research Institute (GIRI)**, founded in 1994 by Dr. Jerzy Jurka, is a private, nonprofit institution devoted to the study and characterization of repetitive or mobile DNA elements in eukaryotic species. The Institute maintains and develops **Repbase Update (RU)**, a database of representative repetitive sequences from eukaryotic species. RU is being used worldwide as the reference standard for annotating the presence of repetitive DNA in genomic data. GIRI publishes **Repbase Reports**, a monthly e-journal that reports newly identified repetitive DNA elements, which are subsequently stored and shared in RU. GIRI also provides computational tools (**Censor** and **RepbaseSubmitter)** and online services for screening and submission of the repeats.

Inquiries contact: e-mail giri@girinst.org; tel: 650-961-4480

Working with a **repeat-rich** genome?

Repbase has what you need!

giri http://www.girinst.org/

Why Repbase?

- The largest collection of eukaryotic transposons and repetitive sequences
- Includes over 40,000 sequences (mostly family consensus) and continues to grow
- Covers over a hundred model organisms and species of interest including animals, plants and fungi
- Subject to extensive manual curation and ongoing updates with the aim of recovering all repeat families in each species
- Sequences are systematically classified according to the nature of the repeats
- Essential for well-known tools such as CENSOR, RepeatMasker, REPET, Dfam
- Recognized and used as a standard around the world

How is Repbase used to produce a repeat-masked genome?

Tools such as CENSOR and RepeatMasker use a library of repeats drawn from **Repbase** to recognize and mask repeat sequences.

As Repbase is enhanced with more repeat sequences the ability to accurately recognize, identify and mask repeats improves.

giri

Search

| Home | About GIRI | Conference | Repbase | Repeat Masking | Downloads | Research |

Submit sequence to CENSOR

CENSOR is a software tool which screens query sequences against a reference collection of repeats and "censors" (masks) symbols, as well as generating a report CENSOR as a tool in your published re

Kohany O, Gentles AJ, Hankus I Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*, 2006 Oct 25;7:474

Sequence source: All
Force translated search: ☐
Search for identity: ☐
Report simple repeats: ☐
Mask pseudogenes: ☐

Enter query file name:
(Up to 2MB; IG-Stanford, FASTA, GENBANK, EMBL fc
Choose File   No file
Submit File

CENSOR repeat-masking tool

Current release of Repbase Update

Here you will find the **Repbase Update** collection. You can either download the entire collection:

- EMBL format (78.89 MB) 10-28-2015: RepBase20.10.embl.tar.gz
- FASTA format (37.68 MB) 10-28-2015: RepBase20.10.fasta.tar.gz
- Repbase-derived RepeatMasker libraries: repeatmaskerlibraries-20150807.tar.gz (59.98 MB)
- REPET edition: RepBase20.05_REPET.embl.tar.gz (42.84 MB)

Frequent releases

SVG viewer is required to view graphical representation of the map as Scalable Vector Graphics (SVG plot).

con (SVG Plot; Alignments; Masked)

Repeat map & masked sequence

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
con	43	204	DNA-2-32_DR	6612	6761	DNA	c	0.6928	1.7273	287
con	253	310	EnSpm-11_STu	2675	2729	DNA/EnSpm/CACTA	c	0.8364	3.0000	247
con	411	520	Transib-6_DRh	1688	1777	DNA/Transib	c	0.7732	1.1667	233
con	1062	1173	EnSpm-20_HMa	3010	3122	DNA/EnSpm/CACTA	d	0.7345	1.9167	206
con	1202	1302	L1-7_Cis	4910	5000	NonLTR/L1	d	0.7959	2.5000	247
con	1534	1624	CR1-4_HM	4254	4343	NonLTR/CR1	d	0.8022	2.1667	347
con	1977	2175	Gypsy-57_GR-LTR	1414	1623	LTR/Gypsy	c	0.7487	2.0000	236
con	2462	2621	MuDR-13_GM	6495	6659	DNA/MuDR	d	0.7562	2.8000	239
con	2813	2940	Polinton-1N1_DR	499	628	DNA/Polinton	c	0.7559	2.5556	226
con	2942	2981	Zisupton-5_DR	6878	6919	DNA/Zisupton	c	0.8780	99.0000	200
con	3413	3460	DNA3-4_CGi	5622	5667	DNA	a	0.8298	2.0000	226

>con
CCCTTACGGAAAATTCCCATGAAAAAACCCATAAAAAAAACCXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCATAG
AAAAACCATGAGTTTTCCATGGAAAACCCATGAGTTTTCCAT
XXGTTTTTCATGGAAAAGTGTTCCAAATACCATGGATATTCA
TGGAATTTCCATAGGTATTAGGAACACCTTTCXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXACACTAATATTA

Repeat composition

Repeat landscape

Classification of transposable elements

Group	Class/Superfamily
DNA transposon	Academ, Crypton (CryptonA, CryptonF, CryptonI, CryptonS, CryptonV), Dada, EnSpm/CACTA, Ginger1, Ginger2, Harbinger, hAT, Helitron, IS3EU, ISL2EU, Kolobok, Mariner/Tc1, Merlin, MuDR, Novosib, P, piggyBac, Polinton, Sola (Sola1, Sola2, Sola3), Transib, Zator, Zisupton
LTR retrotransposon	BEL, Copia, DIRS, Gypsy, Endogenous retrovirus (ERV1, ERV2, ERV3, ERV4, Lentivirus)
Non-LTR retrotransposon	Ambal, CR1, CRE, Crack, Daphne, Hero, I, Ingi, Jockey, Kiri, L1, L2, L2A, L2B, Loa, NeSL, Nimb, Outcast, Penelope, Proto1, Proto2, R1, R2, R4, Randl/Dualen, Rex1, RTE, RTETP, RTEX, Tad1, Tx1, Vingi, SINE (SINE1/7SL, SINE2/tRNA, SINE3/5S, SINE4, SINEU)