LINK TO ORIGINAL ARTICLE
LINK TO AUTHOR'S REPLY

A universal classification of eukaryotic transposable elements implemented in Repbase

Vladimir V. Kapitonov and Jerzy Jurka

In their Perspective (A unified classification system for eukaryotic transposable elements. *Nature Rev. Genet.* **8**, 973–982 (2007))[1], Wicker *et al*. attempt to introduce the 'first universal classification scheme'. Here, we would like to point out a similar universal hierarchical classification system that was developed earlier by us and implemented in Repbase[2].

Repbase is a database of eukaryotic repetitive and transposable elements (TEs), developed since 1990. It was first published as a collection of the consensus sequences and sequence fragments of human TEs, as well as satellite DNA, that was available at the time[3]. Subsequently, repetitive and transposable sequences from other animal and plant species were added, and in the mid-1990s Repbase became available online for downloading and sequence analysis using the computer tools Censor server[4] and RepeatMasker. At the same time, a systematic classification of repetitive elements based on their origin from different classes of TEs was developed and implemented in parallel in Repbase and RepeatMasker. Since 2001, Repbase has been routinely used in conjunction with RepeatMasker to analyze and annotate entire genomes. The resulting information is available through major international genome browsers, including the University of California, Santa Cruz (UCSC) browser and the Ensembl browser.

The current version of Repbase, known as Repbase Update, contains >7,600 sequences of TEs and other repeats, including those that are reported in the literature and those that are only reported in Repbase. Since 2001, all new information on TEs compiled in Repbase is first published in an electronic journal, Repbase Reports[2,4].

In 2005, Repbase was converted to a relational database, which permitted us to implement our universal classification of TEs. According to this classification (FIG. 1), all eukaryotic TEs belong to two types (retrotransposons and DNA transposons) and are composed of five major classes: long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, cut-and-paste DNA transposons, rolling-circle DNA transposons (*Helitrons*) and self-synthesizing DNA transposons (*Polintons*). This classification is based on enzymology, structural similarities and sequence relationships[5–14]. Each class of TE is composed of a small number of superfamilies or clades[5,6,8–11,15] (see the 40 superfamilies in FIG. 1). Each superfamily consists of numerous families of TEs. Ancient families are represented in Repbase by consensus sequences approximating active TEs from which these families were derived (consensus sequences of any two families are less than 75% identical to each other).

For instance, the class of LTR retrotransposons is composed of the *Gypsy*, *Copia*, *BEL* and *DIRS* superfamilies, plus the *ERV1*, *ERV2* and *ERV3* superfamilies of endogenous retroviruses[6,13,15]. The class of non-LTR retrotransposons is composed of the *CR1*, *CRE*, *I*, *Jockey*, *L1*, *NeSL*, *Penelope*, *R2*, *R4*, *RandI*, *Rex1*, *RTE* and *Tx1* superfamilies (also known as clades)[8,15]. It also includes the *SINE1*, *SINE2*, and *SINE3* superfamilies of short interspersed nuclear elements (SINEs), which are viewed as non-autonomous non-LTR retrotransposons[7]. The class of cut-and-paste DNA transposons consists of 15 superfamilies, including those described only in Repbase (*Mirage*, *Rehavkus*, *Nobosib*, *Kolobok*, *ISL2EU* and *Chapaev*). Autonomous TEs from each of these superfamilies encode superfamily-specific transposases when transposases from different superfamilies are not similar to each other (that is, when the E-value in BLASTP or PSI-BLAST is greater than 0.01).

Based on a system that was established over a decade ago by Smit and ourselves[13,16], non-autonomous DNA transposons are routinely classified based on significant similarities of their terminal inverted repeats and target-site duplications to those in known autonomous DNA transposons. Analogously, structural and sequence similarities are used for the classification of non-autonomous LTR and non-LTR retrotransposons.

Although the Repbase interface does not directly display the hierarchical classification scheme, it reflects and corresponds to this scheme published in literature. According to the published information, eukaryotic DNA transposons identified so far in eukaryotes belong to three classes characterized by the so-called cut-and-paste, rolling-circle and self-synthesizing mechanisms of transposition, reflecting three different mechanisms of transposition[11,15].

During the last 4 years, thousands of families of transposable elements in genomes of several eukaryotic species have been identified, classified and named based on the classification scheme and nomenclature shown in FIG. 1, including those from protozoans (diatom *Thalassiosira pseudonana* and green alga *Chlamydomonas reinhardtii*)[17,18], fungi (*Aspergillus nidulans, Aspergillus oryzae* and *Aspergillus fumigatus*)[19], cnidarians (starlet sea anemone *Nematostella vectensis*)[20] and mammals (opossum *Monodelphis domestica*)[21].

In April 2006, the above classification scheme was presented by us during the first international conference and workshop named Genomic Impact of Eukaryotic Transposable Elements, which also included a session devoted to the unified classification and nomenclature of TEs. During this conference, which was attended by 150 scientists working in the field, an International Committee on the Classification of Transposable Elements was constituted.

Vladimir V. Kapitonov and Jerzy Jurka are at the Genetic Information Research Institute, 1925 Landings Drive, Mountain View, California 94043, USA.

e-mails: vladimir@girinst.org; jurka@girinst.org

1. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nature Rev. Genet.* **8**, 973–982 (2007).
2. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
3. Jurka, J., Walichiewicz, J. & Milosavljevic, A. Prototypic sequences for human repetitive DNA. *J. Mol. Evol.* **35**, 286–291 (1992).
4. Jurka, J. Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420 (2000).
5. Kapitonov, V. V. & Jurka, J. Rolling-circle transposons in eukaryotes. *Proc. Natl Acad. Sci. USA* **98**, 8714–8719 (2001).
6. Kapitonov, V. V. & Jurka, J. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl Acad. Sci. USA* **100**, 6569–6574 (2003).
7. Kapitonov, V. V. & Jurka, J. A novel class of SINE elements derived from 5S rRNA. *Mol. Biol. Evol.* **20**, 694–702 (2003).
8. Kapitonov, V. V. & Jurka, J. The esterase and PHD domains in CR1-like non-LTR retrotransposons. *Mol. Biol. Evol.* **20**, 38–46 (2003).
9. Kapitonov, V. V. & Jurka, J. *Harbinger* transposons and an ancient HARBI1 gene derived from a transposase. *DNA Cell Biol.* **23**, 311–324 (2004).
10. Kapitonov, V. V. & Jurka, J. RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. *PLoS Biol.* **3**, e181 (2005).
11. Kapitonov, V. V. & Jurka, J. Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl Acad. Sci. USA* **103**, 4540–4545 (2006).

12. Kapitonov, V. V. & Jurka, J. *Helitrons* on a roll: eukaryotic rolling-circle transposons. *Trends Genet.* **23**, 521–529 (2007).
13. Kapitonov, V. V., Pavlicek, A. & Jurka, J. in *Encyclopedia of Molecular Cell Biology and Molecular Medicine* (ed. Meyers, R. A.) 251–305 (Wiley-VCH, Weinheim, 2004).
14. Jurka, J. & Kapitonov, V. V. *PIFs* meet *Tourists* and *Harbingers*: a superfamily reunion. *Proc. Natl Acad. Sci. USA* **98**, 12315–12316 (2001).
15. Jurka, J., Kapitonov, V. V., Kohany, O. & Jurka, M. V. Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* **8**, 241–259 (2007).
16. Smit, A. F. & Riggs, A. D. *Tiggers* and other DNA transposon fossils in the human genome. *Proc. Natl Acad. Sci. USA* **93**, 1443–1448 (1996).
17. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
18. Merchant, S. S. *et al.* The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318**, 245–250 (2007).
19. Galagan, J. E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**, 1105–1115 (2005).
20. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
21. Mikkelsen, T. S. *et al.* Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).

Acknowledgements
This work was supported by the National Institutes of Health grant 5 P41 LM06252-09.

FURTHER INFORMATION
Ensembl browser: http://www.ensembl.org
First international conference and workshop, Genomic Impact of Eukaryotic Transposable Elements: http://www.girinst.org/conference/Asilomar-2006/schedule.html
International Committee on the Classification of Transposable Elements: http://girinst.org/conference/ICCTE.html
Repbase: http://girinst.org/repbase
RepeatMasker: http://www.repeatmasker.org
University of California, Santa Cruz (UCSC) genome browser: http://genome.ucsc.edu

UNIVERSAL CLASSIFICATION SCHEME OF TEs

Type 1: DNA transposons

Superfamily	TSDs bp
Chapaev	4
En/Spm (*CACTA*)	3
hAT	5,6,8
Harbinger (*Pif*)	3
ISL2EU (*IS4EU*)	2
Kolobok	4
Mariner	2
Merlin	8,9
Mirage	2
MuDR (*MULE*)	9,10
Novosib	8
P	7,8
PiggyBac	4
Rehavkus	9
Transib	5
Helitron	–
Polinton (*Maverick*)	6

Type 2: retrotransposons

Non-LTR retrotransposons

Superfamily (Clade)	TSDs bp
CRE	22–50
NeSL	7–22
R4	~13
R2	0–30
L1	~15
RTE	0–100
Jockey	~10
CR1	0
Rex1	0
I	10–15
RandI (*Dualen*)	~10
Tx1	~15
SINE1	*
SINE2	*
SINE3	*
Penelope	0

LTR retrotransposons

Superfamily (Clade)	TSDs bp
Copia	5,6
Gypsy	3,5
BEL	4,5
ERV1	4
ERV2	6
ERV3	5
DIRS	–

UNIVERSAL NOMENCLATURE OF TEs

Name structure: **Prefix-Infix1{-Infix2}_Suffix**
Prefix — unique superfamily name (based on the universal classification scheme),
Infix1 — family identifier,
Infix2 — structural identifier (for example, LTR and internal portion of a retrovirus),
Suffix — species identifier (2–4 letters)

Examples of universal nomenclature for completely classified TEs:
Mariner-4_NV — family number 3 of autonomous *Mariner* DNA transposons in *Nematostella vectensis*;
Mariner-4N1_NV — family 1 of non-autonomous *Mariners* in *N. vectensis* (*Mariner-4_NV* is its closest autonomous counterpart);
Harbinger-N5_NV — family 5 of non-autonomous *Harbinger* DNA transposons in *N. vectensis* (its autonomous counterpart is uknown);
Harbinger-N5B_NV — subfamily A of *Harbinger-N5_NV*;
Gypsy-1-I_TP — internal portion of *Gypsy-1_TP*, which belongs to the family 1 of *Gypsy* LTR retrotransposons from *Aspergillus fumigatus*;
Gypsy-1-LTR_TP — LTR of *Gypsy-1_TP,* which belongs to the family 1 of *Gypsy* LTR retrotransposons from *A. fumigatus;*
RTE-1_TP — family 1 of *RTE* non-LTR retrotransposons in *Thalassiosira pseudonana;*
Helitron-1N1_DVir — family 1 of non-autonomous *Helitrons* in *Drosophila virilis* (*Helitron-1_DVir* is its autonomous counterpart);
Polinton-3_TC — family 3 of *Polinton* DNA transposons in *Tribolium costaneum.*

Examples of universal nomenclature for partially classified TEs:
DNA-TA-7_BF — family 7 of unclassified DNA transposons in *Branchiostoma floridae* that are characterized by the TA TSDs;
DNA-3-4_BF — family 4 of unclassified DNA transposons in *B. floridae* that are characterized by 3-bp TSDs;

Examples of universal nomenclature for unclassified TEs:
TE-TA-7_BF — family 7 of unclassified transposable elements in *B. floridae* that are characterized by the TA TSDs;
TE-3-4_BF — family 4 of unclassified transposable elements in *B. floridae* that are characterized by 3-bp TSDs.

Figure 1 | **The universal classification and nomenclature of eukaryotic transposable elements.** Different classes of transposable elements (TEs) are differently coloured. *Penelope* and *DIRS* can be viewed as two additional classes of retrotransposons. An asterisk indicates that the lengths of target-site duplications (TSDs) by short interspersed nuclear elements (SINEs) depend on non-LTR retrotransposons being involved in their transpositions. Alternative names for the superfamilies are shown in parentheses. LTR, long terminal repeat; TA, TpA dinucleotide.