HAL
open science

The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?

Peter Arensburger, Benoit Piegu, Yves Bigot

To cite this version:

Peter Arensburger, Benoit Piegu, Yves Bigot. The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 2016, 6 (6), pp.e1256852. 10.1080/2159256X.2016.1256852 . hal-02385838

HAL Id: hal-02385838

https://hal.science/hal-02385838v1

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Version postprint

The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?

Peter Arensburger[1], Benoît Piégu[2], and Yves Bigot[2]

[1] Biological Sciences Department, California State Polytechnic University, Pomona, CA 91768 - United States of America.

[2] Physiologie de la reproduction et des Comportements, UMR INRA-CNRS 7247, PRC, 37380 Nouzilly – France

Corresponding author address: Biological Sciences Department, California State Polytechnic University, Pomona, CA 91768 - United States of America. E-mail: arensburger@gmail.com

keywords: mobile genetic element, repeat, genome, taxonomy, ontology

Abbreviations:

ICE, integrative conjugative elements

GIRI, Genetic Information Institute

MGE, mobile genetic element

TE, transposable element

Abstract

Transposable element (TE) science has been significantly influenced by the pioneering ideas of David Finnegan near the end of the last century, as well as by the classification systems that were subsequently developed. Today, whole genome TE annotation is mostly done using tools that were developed to aid gene annotation rather than to specifically study TEs. We argue that further progress in the TE field is impeded both by current TE classification schemes and by a

Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

failure to recognize that TE biology is fundamentally different from that of multicellular organisms. Novel genome wide TE annotation methods are helping to redefine our understanding of TE sequence origins and evolution. We briefly discuss some of these new methods as well as ideas for possible alternative classification schemes. Our hope is to encourage the formation of a society to organize a larger debate on these questions and to promote the adoption of standards for annotation and an improved TE classification.

Version postprint

Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

Jean de la Fontaine, the most famous of the French poets of the seventeenth century, is well known for his fables involving animal protagonists that examine the organization of human society. Among these the "Wolf and Lamb"[1] and the "Lion and Rat"[2] describe two views of human society that may be summarized respectively as "might makes right" and "even the smallest can help the greater". In scientific circles sometimes it is the most powerful members of the community that dictate the dominant ideas in the field, "might makes right", which is an efficient system but often comes at the expense of collegiality. Alternatively, some scientific communities emphasize input from all members, along the lines of the second fable, but this comes at a cost of time to allow for contradictory debates. Communities in this second category often form societies where elected representatives organize the flow of information within the community and set procedures for making scientific decisions on the future of the field. We argue that the field of eukaryotic transposable elements (TEs) is currently organized around several concepts that are championed by a minority and accepted without argument by a majority. These concepts include what the nature of a TE is, how TEs are identified in assembled genomes, and how they are classified taxonomically. Recent publications (see discussions in[3,4]) reflect these views and it is time for the TE community to re-examine the bases upon which its science is organized. We call for the emergence of an international society for the biology of TEs to address these questions by including all the voices in the community, it is time for the "Lion and Rat".

A definition of TEs

Haren et al.[5] defined TEs as "discrete segments of DNA capable of moving from one locus to another in their host genome or between different genomes". Recently, we proposed that this definition needs to be broadened to "TEs are discrete segments of DNA capable of moving within a host genome from one chromosome or plasmid location to another, or between hosts by using parasitic vectors that they use for lateral transfers"[3]. An important aspect of this definition is that it includes mobile DNA sequences that are primarily maintained by vertical transmission as copies integrated into the chromosomes or plasmids of their hosts. Therefore, according to this definition, while viruses, phages, and integrative conjugative elements (ICE) have similar features to TEs, they are not considered to be TEs because they are able to move between hosts independently of transmission vectors. While we find this definition to be useful it is important

Version postprint
Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

to note that it is more of a practical definition that helps to deal with the diversity and complexity of TEs rather than a formal definition based only on scientific evidence. One of the reasons such a formal definition is difficult is that viruses, phages, ICEs (all three are grouped as MGEs by prokaryote researchers) and TEs, can recombine and exchange genetic material through lateral transfer both within closely related sequences but also between very divergent groups.

Uses and rationale for high quality TE annotations

Initially, interest in TE annotation was linked to gene annotation. When genes are annotated in a newly sequenced eukaryotic genome a common strategy is to mask repeated genomic regions in order to simplify the task of gene prediction. This masking is nearly always performed using the program RepeatMasker[6]. In addition to outputting a genome with masked repeats, RepeatMasker can also provide a list of annotated repeats and their location in the genome. Such lists have been used to study the composition and abundance of repeated sequences in many eukaryotic genomes (e.g. [7,8]). RepeatMasker works on the principle of library-based searching, matching sections of the newly sequenced genome to a preestablished library of known repeats (a.k.a. homology based searching). Typically, the libraries of repeats are the ones hosted at Repbase by the Genetic Information Institute (GIRI) (http://www.girinst.org/repbase/) a private, non profit institution supported in part by private funds, donations and US federal grants. These libraries are partially composed of submissions by outside academic groups. While academic use of these databases is free, GIRI reserves the right to charge a fee for the use of these databases by commercial entities.

Unfortunately, for those interested not in gene annotation but in studying the repeated portions of the genome, RepeatMasker results often lacks accuracy (i.e. not all repeats are annotated) and precision (i.e. repeats are taxonomically misidentified). These problematic annotations are typically the result of three issues: 1) poorly constructed and/or taxonomically mislabeled consensuses in Repbase, 2) absence from the library of consensuses matching a true repeat in the genome, and/or 3) genomic repeats that are too divergent from the library consensus to establish a match. The effect of these on the reliability of RepeatMasker and Repbase annotations varies by genome, but in a recent article we showed that in the model chicken (*Gallus gallus*) genome, RepeatMasker and Repbase annotated only about half of the existing repeats (11% vs.

Version postprint
Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

20% repeats) and dramatically underestimated the diversity of TEs[9]. Since then, our findings have been independently confirmed by researchers at the Roslin Institute[10].

The first two of the problematic issues for RepeatMasker are probably due, at least in part, to a lack of transparency at GIRI. The methodology by which many consensus sequences are generated and annotated in Repbase has never been published in detail, and this in turn leads to a lack of accountability to the community. The authors of RepeatMasker have attempted to address the third problematic issue (very divergent repeats) by introducing Dfam[11] which uses hidden Markov models and sequence alignments instead of simple consensus sequences. However, we find Dfam to be unsatisfying because: 1) it appears to use a flawed model for TE sequences causing it in some cases to detect fewer TEs in the human genome than RepeatMasker, 2) as of this writing it is available for only five species, 3) just as with RepeatMasker, its methodology for creating models is incompletely described in the literature.

An alternative way to generate a repeat library is to create one *de novo* based on the genome sequence assembly. Numerous programs have been developed for this and include programs that search for patterns of repetition (e.g. Tandem Repeat Finder[12]) and those that identify repeats from pairwise alignments of the genome to itself (e.g. RECON[13], dnaPipeTE[14]) or conserved k-mers (e.g. RepeatExplorer[15]). An attractive feature of these methods is that they have the potential to identify repeats even when these bear little or no sequence similarity to previously described repeats, or when repeats have diverged substantially from the consensus sequence. The disadvantages of *de novo* methods include: 1) a lack of studies testing their precision and accuracy, 2) no inherent way to place discovered repeats into taxonomic groups, and 3) many of these programs have historically required substantial computational resources. This last point has become much less of a problem in the last few years as the computation speed of personal computers has increased and computing clusters have become more accessible to researchers at public institutions (e.g. XSEDE allocations https://www.xsede.org). In order for these methods to become widely adopted in the community they need not only rapid calculation speed, but must both produce consistent and accurate results and be easy enough to be used by researchers without extensive computer expertise.

Because it is usually desirable combine the results of multiple *de novo* repeat searching programs, several packages have been developed to do just that. Two of the more popular ones are RepeatModeler[16] and TEdenovo from the REPET package[17]. RepeatModeler, the older of the two, relies mostly on the output of the programs RECON[13] and RepeatScout[18] but has such poor

Version postprint
Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

performance in discovering repeats compared to TEdenovo that it should be considered of dubious value to modern research efforts[17]. This leaves TEdenovo as the new standard for *de novo* repeat discovery in assembled genomes.

An important assumption in the *de novo* repeat discovery methods we have discussed so far is the existence of a high quality genome assembly that contains the sequences of most repeated portions of the genome. In cases when such an assembly is not available it may be possible to discover some repeat sequences using the unassembled sequence data. Programs designed for this purpose include RepeatExplorer[15] and Red[12] among others. It is still early days in the development of these programs, but at the moment they appear to be limited to discovering only highly repeated sequences.

We are on the cusp of a sea change in the field of genome repeat annotation. As the importance of the non genic portion of the genome becomes better understood for its role in gene regulation and genome architecture in nuclei, the need for in-depth repeat annotation will become more important. As the computing power available to most researcher is reaching a stage where they are able to run *de novo* repeat discovery programs on their own, it will soon no longer be acceptable to only superficially annotate repeated genome sequences. The next issue for the scientific community will be to decide if, in parallel with these developments, gold standards are required for what constitutes a proper genome annotation. This will be of great importance not only for genome sciences, but for other disciplines that depend upon high quality genome annotations, including the medical field[20].

The description of a TE species is changing

The development of *de novo* TE annotation tools is changing the way TE species are described. Indeed, much of the literature and many of the databases (including Repbase) define a TE species by a single consensus sequence that attempts to represent an averaged sequence of multiple TE copies in the genome. Clearly, a single sequence cannot adequately fulfill this function which is why the creators of the RepeatExplorer and REPET programs have replaced it with a "TE model". In this representation the "TE model" is not only composed of a main consensus sequence (the most complete version of the TE) but also of all the consensuses detected as variants due to indels and/or highly divergent sequences. Furthermore, a TE that is found in several host species may be represented by a set of variants specific to each host.

Version postprint
Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

Rationale for revisiting the taxonomy of TEs

In the late 1980's David Finnegan pioneered TE systematics[21,22]. His conception was that TEs could, at their phylogenetic base, be classified into two classes based their presumed mechanism of transposition. His class I elements can transpose by reverse transcription of an RNA intermediate using a DNA-RNA-DNA mechanism, while the class II elements (a.k.a. DNA transposons) can transpose directly from DNA to DNA. Finnegan's basal dichotomy has been accepted by a large swath of the scientific community and has been the basis of two subsequent TE taxonomy updates ([23,24] see [3] for review of this issue). However, this is an issue where different TE communities have diverged. The eukaryotic transposon community adopted the Finnegan (and subsequent updates) model, while the prokaryotic community focused on a taxonomy based on transposition models. The prokaryotic view was outlined in a 2002 paper by Curcio and Derbyshire that described the diversity of the TE world based on the diversity of enzymatic machineries that trigger TE integration into their host DNA, including the nature of transposition intermediate[25]. This view more accurately captures many of the evolutionary adaptations that were acquired by TEs, making it a much better basis for a TE taxonomy that reflects evolutionary history.

Over the last few years we have critically analysed the bases of various TE taxonomy schemes, and have attempted to integrate the most recent discoveries (both among eukaryotes and prokaryotes) into our analysis. Based on out findings we concluded that 1) the basal dichotomy view of the TE world advocated by Finnegan is no longer justified by modern science, 2) the Curcio and Derbyshire view appears to us to fit the current science best, but requires careful attention to evolutionary convergences at the molecular level, and 3) the addition of several new TE classes and orders are necessary to keep up with recent discoveries in this field[3]. In Table 1 we outline our proposal for a new taxonomic scheme which has the advantage of allowing the integrating new classes, new orders and new-families, such as the recently described Casposons[26,27]. However, using the same standards as we have applied to previous taxonomic proposal we recognize that our scheme is deficient in several respects. The first is that our scheme has not (yet) been the subject of large-scale debate in the field, partially because the TE field lacks an organized social structure for such a debate. A second weakness is that we continue to use taxonomic divisions such as classes, orders and super-families. While practical for grouping purposes, such divisions are not flexible and make it difficult for novel taxonomic groups (i.e. class, family, and species) of TEs to emerge. Finally, our definition of what a TE is

Version postprint
Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

(see above) may be too strict since there is likely a continuum between certain viruses, phages, ICE and TEs.

In the course of discussion of a recent international meeting on these issues[28] a proposal was put forth that a system based on ontology terms could be developed to create a dynamic taxonomic system. TEs could be clustered based on their shared ontological terms. This would allow both characterization of novel sequences and discovery of new TE taxonomic groups. Such a proposal, while new to this field, has previously been used. Since the early 2000's Ariane Toussaint's team have used such a scheme to create ACLAME and PhiGO[29–31], tools to annotate and classify prokaryotic phages, ICEs and TEs.

Concluding remarks

In 2015 we called for the creation of a new organization called the International Committee for the Taxonomy of Transposable Element (ICTTE) that would gather TE researchers (both on the prokaryotic and eukaryotic sides) as well as virologists to take charge of the issues surrounding the taxonomy of TEs. However, a first step to the emergence of such a committee may be an international scientific society that gathers several communities focused on various aspects of TE biology including their role in nuclear organization, regulatory networks, as well as TE taxonomy. The organization of such a society would certainly be challenging but to quote the La Fontaine fable the "Cat and Fox": "The dispute is of great help. Without it, we would always be sleeping!"[32].

Acknowledgments

This work was funded by the California State Polytechnic University, C.N.R.S., the I.N.R.A., the STUDIUM the Groupement de Recherche CNRS 2157, and the Ministère de l'Education Nationale, de la Recherche et de la Technologie.

Version postprint

Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

References

1. La Fontaine, J. 'The Wolf and Lamb' Book 1, Fable 10.
2. La Fontaine, J. 'The Lion and Rat' Book 2, Fable 11.
3. Piégu, B., Bire, S., Arensburger, P. & Bigot, Y. A survey of transposable element classification systems - A call for a fundamental update to meet the challenge of their diversity and complexity. *Mol. Phylogenet. Evol.* **86,** 90–109 (2015).
4. Hoen, D. R. *et al.* A call for benchmarking transposable element annotation methods. *Mob. DNA* **6,** (2015).
5. Haren, L., Ton-Hoang, B. & Chandler, M. Integrating DNA: Transposases and Retroviral Integrases. *Annu. Rev. Microbiol.* **53,** 245–281 (1999).
6. Smit, A F A, Hubley, R. & Green, P. RepeatMasker Open-3.0. 1996-2010. (1996).
7. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
8. Holt, R. A. *et al.* The genome sequence of the malaria mosquito Anopheles gambiae. *Science* **298,** 129 (2002).
9. Guizard, S., Piégu, B., Arensburger, P., Guillou, F. & Bigot, Y. Deep landscape update of dispersed and tandem repeats in the genome model of the red jungle fowl, Gallus gallus, using a series of de novo investigating tools. *BMC Genomics* **17,** 659 (2016).
10. Mason, A. S., Fulton, J. E., Hocking, P. M. & Burt, D. W. A new look at the LTR retrotransposon content of the chicken genome. *BMC Genomics* **17,** 688 (2016).
11. Hubley, R. *et al.* The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44,** D81–D89 (2016).
12. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–580 (1999).
13. Bao, Z. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. *Genome Res.* **12,** 1269–1276 (2002).
14. Goubert, C. *et al.* De Novo Assembly and Annotation of the Asian Tiger Mosquito (Aedes albopictus) Repeatome with dnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (Aedes aegypti). *Genome Biol. Evol.* **7,** 1192–1205 (2015).
15. Novak, P., Neumann, P., Pech, J., Steinhaisl, J. & Macas, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29,** 792–793 (2013).

Version postprint

Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

16. Smit, A F A, H., R. RepeatModeler Open-1.0. (2008).

17. Flutre, T., Duprat, E., Feuillet, C. & Quesneville, H. Considering Transposable Element Diversification in De Novo Annotation Approaches. *PLoS ONE* **6,** e16526 (2011).

18. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21,** i351–i358 (2005).

19. Girgis, H. Z. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics* **16,** (2015).

20. Ardeljan, D. *et al.* Meeting Report: The Role of the Mobilome in Cancer. *Cancer Res.* **76,** 4316–4319 (2016).

21. Finnegan, D. J. Transposable elements. *Curr. Opin. Genet. Dev.* **2,** 861–867 (1992).

22. Finnegan, D. J. Eukaryotic transposable elements and genome evolution. *Trends Genet. TIG* **5,** 103–107 (1989).

23. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8,** 973–982 (2007).

24. Kapitonov, V. V. & Jurka, J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **9,** 411–412 (2008).

25. Curcio, M. J. & Derbyshire, K. M. The outs and ins of transposition: from mu to kangaroo. *Nat. Rev. Mol. Cell Biol.* **4,** 865–877 (2003).

26. Krupovic, M., Shmakov, S., Makarova, K. S., Forterre, P. & Koonin, E. V. Recent Mobility of Casposons, Self-Synthesizing Transposons at the Origin of the CRISPR-Cas Immunity. *Genome Biol. Evol.* **8,** 375–386 (2016).

27. Hickman, A. B. & Dyda, F. The casposon-encoded Cas1 protein from *Aciduliprofundum boonei* is a DNA integrase that generates target site duplications. *Nucleic Acids Res.* **43,** 10576–10587 (2015).

28. Analysis and Annotation of DNA Repeats and Dark Matter in Eukaryotic Genomes. in (2015).

29. Leplae, R. ACLAME: A CLAssification of Mobile genetic Elements. *Nucleic Acids Res.* **32,** 45D–49 (2004).

30. Toussaint, A., Lima-Mendez, G. & Leplae, R. PhiGO, a phage ontology associated with the ACLAME database. *Res. Microbiol.* **158,** 567–571 (2007).

31. Leplae, R., Lima-Mendez, G. & Toussaint, A. ACLAME: A CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res.* **38,** D57–D61 (2010).

32. La Fontaine, J. 'The Cat and Fox' Book 9, Fable 14.

Version postprint

Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

33. Huang, S. *et al.* Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* **166,** 102–114 (2016).

Version postprint

Accepted Manuscript

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

Table 1. TE classification proposed in[3]
TE Classes with some members having a DNA transposon phenotype°

Class	Order	Superfamilies
Nuclease/Recombinase	*Transposition mechanism*	*Phylogenetic relationships between Nuclease/Recombinase*
DDE-transposons	*DDE transposons with no DNA-transposition intermediate*	*Mu*
	(Copy-in)	*Tn3*
	DDE/D transposons with a linear dsDNA transposition intermediate	*IS1, IS3, IS4, IS701, ISH3, IS1634, IS1182, IS6, IS21, IS30, IS66, IS110, IS630, IS982, IS1380, ISAs1, ISL3*
	(Cut-out/Paste in)	
		IS630/Tc1/mariner (ITm)/Zator
		IS1595-Merlin,

Version postprint

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI: 10.1080/2159256X.2016.1256852

		IS5/PIF/Harbinger,
		IS256/MuDR/Mutator/Rehavkus
		IS1380/PiggyBac,
		Academ, CACTA/Mirage/Chapaev (CMC), Dada, Hobo/Ac/Tam (hAT), Kolobok, P(?),Sola, Transib/ProtoRag[33]
	DDE/D transposons with a linear dsDNA transposition intermediate and using a heteromeric transposase (Cut-out/Paste in)	*Tn7*

Version postprint

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

	DDE transposons with a circular dsDNA transposition intermediate	*IS3*
	(Copy-out /Paste-in)	
	LTR retrotransposons	*Copia*
	(Copy-out/Paste-in)	*Gypsy*
		BEL
		ERV1
		ERV2
		ERV3
Y1-transposons	*Y1 transposons with a circular dsDNA transposition*	*IS200/IS605*
	(Cut-out/Paste in)	*Tn916*
		CTnDOT
		Crypton
	Y1 retrotransposons with a circular dsDNA transposition	*DIRS*

Version postprint

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

	(Copy-out/Paste-in)	*Ngaro*
		VIPER
Y2-transposons	*Y2 transposons with a circular ssDNA transposition*	*IS91*
	(Copy-in or -out/ Copy-in)	*Helitrons*
S-transposons	*S transposons with a circular dsDNA transposition*	*IS607*
	(Cut-out/paste-in)	*Tn5397*
Casposons[26,27]	*Casposase with a DNA intermediate in a configuration that remains to define*	*Casposons*
	(Copy-in or -out/ Copy-in or paste-in)	
TEs pending classification	*?*	*ISAs1*
	?	*Fanzor*

Version postprint

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI: 10.1080/2159256X.2016.1256852

	Polintons/Mavericks	*Mavirus (?)*
	DDE integrase	*Polintons/Mavericks*
	(Copy-in or -out/ Copy-in)	*Tlr1*
	Transposase putatively related to integrases of LTR retrotransposons	*Ginger1*
		Ginger2
	DDE-transposons with a DDE-transposase having another origin	*P(?)*
	Zisupton (Unknown transposition depending on a "Zisuptase")	*Zisupton*

TE Classes for TEs with a non-LTR retrotransposon phenotype °

Class *non-LTR retrotransposons*	**Order** *Endonuclease (En)*	**Superfamilies** *Phylogenetic relationships between endonuclease, then RT*
Retroposons	*LINEs*	*LINEs with an AP EN* *LINEs with a PD-(D/E)XK EN*

Version postprint

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852

		*LINEs with both AP and PD-(D/E)XK EN**
	Penelope-like elements (PLE)	*Athena, no GIY-YIG domain* *Coprina, no GIY-YIG domain* *Neptune, GIY-YIG domain* *Penelope, GIY-YIG domain*
	Group II introns	*Group II introns*
		Mobile lariat introns
		Introners-like elements

TE Classes for TEs with an SSE phenotype °

Class *Machinery for excision of host genes*	**Order** *Transposition mechanism*	**Superfamilies** *Phylogenetic relationships between HEN, and-or site into host genes*
Intein	*LAGLIDADG inteins* (HEN dependent HR)	*Host genes in which each intein specifically inserted could be used, as proposed in InBase (http://tools.neb.com/inbase/)*
	HNH inteins (HEN dependent HR)	
Group I intron (G1i)	*LAGLIDADG G1i* (HEN dependent HR)	*Host sites in which each group I intron specifically inserted could also be used*
	HNH G1i (HEN dependent HR)	
	His-Cys G1i (HEN dependent HR)	
	GIY-YIG G1i (HEN dependent HR)	
	PD-(D/E)XK G1i (HEN dependent HR)	
	Vsr G1i (?) (HEN dependent HR)	

TE Classes for rare prokaryotic TEs with a retroposon phenotype °

Class *RT features*	**Order** *Transposition mechanism*	**Superfamilies** *Phylogenetic relationships between RT*
Retron/msRNA	Retron/msRNA (retrotransposition)	*msRNA*

° With the exception of the Intein and Group I intron Classes, names of superfamilies found in prokaryotes are typed in black, those in eukaryotes being in blue. Both colors are used for mixed superfamilies. The criteria used are indicated in italics just below the levels of Class, Order and Superfamilies. Bibliographic references of updated points with respect to the previous version[3] are indicated.

Version postprint

Comment citer ce document :
Arensburger, P., Piégu, B., Bigot, Y. (2016). The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine?. Mobile Genetic Elements, 6 (6), 1-17. DOI : 10.1080/2159256X.2016.1256852