



**HAL**  
open science

## A new tool to cross and analyze TE and gene annotations

Emmanuelle Lerat

► **To cite this version:**

Emmanuelle Lerat. A new tool to cross and analyze TE and gene annotations. 2021, pp.100010. 10.24072/pci.genomics.100010 . hal-03279813

**HAL Id: hal-03279813**

**<https://hal.science/hal-03279813v1>**

Submitted on 6 Jul 2021

**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.

## A new tool to cross and analyze TE and gene annotations

*Emmanuelle Lerat* based on reviews by 2 anonymous reviewers

A recommendation of:

A pipeline to detect the relationship between transposable elements and adjacent genes in host genomes

Caroline Meguerditchian, Ayse Ergun, Veronique Decroocq, Marie Lefebvre, Quynh-Trang Bui (2021), *bioRxiv*, 2021.02.25.432867, ver. 4 peer-reviewed and recommended by Peer Community In Genomics [10.1101/2021.02.25.432867](https://doi.org/10.1101/2021.02.25.432867)

Open Access

Submitted: 03 March 2021, Recommended: 01 July 2021

### Cite this recommendation as:

Emmanuelle Lerat (2021) A new tool to cross and analyze TE and gene annotations. *Peer Community in Genomics*, 100010. [10.24072/pci.genomics.100010](https://doi.org/10.24072/pci.genomics.100010)

Published: 6 July 2021

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

### Recommendation

Transposable elements (TEs) are important components of genomes. Indeed, they are now recognized as having a major role in gene and genome evolution (Biémont 2010). In particular, several examples have shown that the presence of TEs near genes may influence their functioning, either by recruiting particular epigenetic modifications (Guio et al. 2018) or by directly providing new regulatory sequences allowing new expression patterns (Chung et al. 2007; Sundaram et al. 2014). Therefore, the study of the interaction between TEs and their host genome requires tools to easily cross-annotate both types of entities. In particular, one needs to be able to identify all TEs located in the close vicinity of genes or inside them. Such task may not always be obvious for many biologists, as it requires informatics knowledge to develop their own script codes.

In their work, Meguerditchian et al. (2021) propose a command-line pipeline that takes as input the annotations of both genes and TEs for a given genome, then detects and reports the positional relationships between each TE insertion and their closest genes. The results are processed into an R script to provide tables displaying some statistics and graphs to visualize these relationships.

This tool has the potential to be very useful for performing preliminary analyses before studying the impact of TEs on gene functioning, especially for biologists. Indeed, it makes it possible to identify genes close to TE insertions. These identified genes could then be specifically considered in order to study in more detail the link between the presence of TEs and their functioning. For example, the identification of TEs close to genes may allow to determine their potential role on gene expression.

## References

- Biémont C (2010). A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics*, 186, 1085–1093. <https://doi.org/10.1534/genetics.110.124180>
- Chung H, Bogwitz MR, McCart C, Andrianopoulos A, ffrench-Constant RH, Batterham P, Daborn PJ (2007). Cis-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics*, 175, 1071–1077. <https://doi.org/10.1534/genetics.106.066597>
- Guio L, Vieira C, González J (2018). Stress affects the epigenetic marks added by natural transposable element insertions in *Drosophila melanogaster*. *Scientific Reports*, 8, 12197. <https://doi.org/10.1038/s41598-018-30491-w>
- Meguerditchian C, Ergun A, Decroocq V, Lefebvre M, Bui Q-T (2021). A pipeline to detect the relationship between transposable elements and adjacent genes in host genomes. *bioRxiv*, 2021.02.25.432867, ver. 4 peer-reviewed and recommended by Peer Community In Genomics. <https://doi.org/10.1101/2021.02.25.432867>
- Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research*, 24, 1963–1976. <https://doi.org/10.1101/gr.168872.113>

## Reviews

---

### ***Revision round #2***

2021-05-31

#### ***Author's Reply***

[Download author's reply \(PDF file\)](#)[Download tracked changes file](#)

#### ***Decision round #2***

I have received comments from one reviewer concerning your revised manuscript. Some points remain to be corrected, with which I totally agree; especially, the general comments of the reviewer need to be addressed.

Only for one suggestion, I would use « LTR retrotransposon » rather than « LTR transposon » contrary to what is suggested by the reviewer, although I totally agree with the fact that using only the word “LTR” is misleading.

Other minor points in the introduction:

- 1) you should add as a mechanism to modify expression of neighboring genes the fact that TE can bring regulatory sequences;
- 2) the sentence concerning the fact that the tool can be used with a “custom TE annotation” and a “de novo assembled genome” is not clear.

The latin name of the Apricot should be specified not only in the figure legends.

E. Lerat

Preprint DOI: [10.1101/2021.02.25.432867](https://doi.org/10.1101/2021.02.25.432867)

**Reviewed by anonymous reviewer, 2021-05-26 00:58**

[Download the review \(PDF file\)](#)

---

## **Revision round #1**

2021-04-05

### **Author's Reply**

[Download author's reply \(PDF file\)](#)[Download tracked changes file](#)

#### **Decision round #1**

I have received the comments of two reviewers for your manuscript. As you will see, they both consider your work interesting. However one reviewer points out that already some known tools exist that could perform similar analyses. I would thus recommend you to perform comparative analyses with other similar tools to evaluate the added value of your pipeline. Similarly, reviewer 2 points out the fact that you should make it clear that your pipeline may be used with other genomes.

Sincerely,

E. Lerat

Preprint DOI: [10.1101/2021.02.25.432867](https://doi.org/10.1101/2021.02.25.432867)

**Reviewed by anonymous reviewer, 2021-04-01 18:53**

[Download the review \(PDF file\)](#)

**Reviewed by anonymous reviewer, 2021-03-23 09:37**

Title : Pipeline to detect the relationship between transposable elements and adjacent genes in host genome

In this work, C. Meguerditchian and colleagues propose a pipeline to retrieve, from a list of transposable elements (TEs) coordinates and a list of gene annotations, a list of overlapping and closest upstream / downstream genes. It is true that this step is probably the first one to apply when searching for TEs with a potential impact on gene expression. However, different tools are already available to manage such analysis, and I would only cite the Bedtools suite (<https://bedtools.readthedocs.io/en/latest/>), with the tools "intersect" or "closest". In addition, and contrary to the postulate of the authors that "none existing tools can reveal the relationship between TEs and host coding sequences", several works attempted to address this question and even went further by taking into account expression data or functional data such as GO terms (for example, LIONS (Babaian et al. 2019), or GREAM (Chandrashekar et al. 2015)). Thus I do not think that this tool, in spite of the fact that it seems to function, constitutes neither a novelty nor a useful adding to the existing programs. I still list some propositions to improve the manuscript.

#### Major comment

- The authors retrieve upstream and downstream genes of TEs, but how do they deal with non-coding TEs, for instance MITEs ? Does it have a sense to distinguish downstream and upstream genes in such case?

#### Minor comments

- In the title, the kind of “relationship” between TEs and genes should be precised. Similarly, the authors should precise what they mean by “TEs associated genes” in the abstract, as well as in the sentence “We implemented a pipeline which is capable to reveal the relationship between TEs and adjacent gene distribution in the host genome”.
- Introduction: Please provide references for human and maize genome TE coverages.
- The sentence “Due to their role in transposition [...], TEs can regulate [...]” should be rephrased → “Due to their transposition...”
- “will help determine the important role of TEs” → “will help determine the role of TEs”
- How do the authors define “upstream” and “downstream” parts of TEs ?

#### Typos and grammar

- abstract: because of its transcriptional activity → because of their transcriptional activity
- Implementation:
  - in an downstream location → in a downstream location
  - this function returns gene with → this function returns genes with
  - this function searches for gene, which is... → this function searches for genes, which are...
  - what type of TE are present → what types of TEs are present
  - the number of TE → the number of TEs
- Conclusion:
  - running on two different TE annotation software → running on two different TE annotations
- Fig 3: specie → species
- Ref 1: M. Barbara → B. McClintock