

Pipeline to detect the relationship between transposable elements and host's coding genome

Caroline Meguerditchian, Ayse Ergun, Veronique Decroocq, Marie Lefebvre, Quynh-Trang Bui

[10.1101/2021.02.25.432867](#)

Submitted by Marie Lefebvre 2021-03-03 15:08

Abstract

Understanding the relationship between transposable elements (TEs) and their associated genes in the host genome is a key point to explore their potential role in genome evolution. Transposable elements can regulate and affect gene expression not only because of their mobility within the genome but also because of its transcriptional activity. Gene expression can be suppressed, decreased or increased and cellular signalling pathways can be activated through the act of the nearby TE expression itself or subsequent TE replication intermediates. We implemented a pipeline, which is capable to reveal the relationship between TEs and adjacent gene distribution in the host genome. Our tool is freely available here:

https://github.com/marieBvr/TEsgenesrelationship_pipeline

Keywords: Transposable element, Gene, Genome, Bioinformatics, Pipeline

Round #2

Author's Reply:

by Emmanuelle Lerat, 2021-05-31 08:41

Manuscript: [10.1101/2021.02.25.432867](#)

Minor revision needed

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 adressed.

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;

The information was added in the Introduction.

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 sentence was corrected.

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

The species name was added in the text.

E. Lerat

Reviews

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

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

Review of manuscript entitled

" Pipeline to detect the positional and directional relationship between transposable elements and adjacent genes in host genome "
round #2

General comments:

The new manuscript is entitled "Pipeline to detect the positional and directional relationship between transposable elements and adjacent genes in host genome" but I don't see anything in the manuscript that concerns the orientation of the genes. It seems to me that only the orientation of the TEs is taken into account. Even in the new figure 1, this is not clear because the orientation is not shown, neither for TEs nor for genes. This point needs to be clarified.

The title was corrected as :

"Pipeline to detect the positional relationship between transposable elements and adjacent genes in host genome"

Regarding the same figure 1, I don't understand how genes can be inside transposable elements (gene 7 / TE9) and why many TEs are shown as partially overlapping genes (TEs 1, 3, 4, 5, 6, 9, 10, 11, and 13). TEs may insert into a gene, generally within introns or eventually 3'-UTRs, or they may insert up- or downstream of a gene. In the first case, the TE should be entirely comprised within the gene, in the second case the TE will be outside of the gene but never partially overlapping with a gene.

We reported all physical possible scenarios that we could imagine in the Figure 1 to be sure our pipeline can generate correctly all complicated cases occur in the genome and it does not mean all these scenarios will happen in the genome.

Errors that have not been corrected: **Done**

what type of TE are present → what types of TEs are present
replace " distance between LTRs and genes as shown on Figure 5, ..." with distance between LTRs and genes as shown in Figure 5,

Other errors to correct: **Done**

An error that I didn't see during the first round of review:

Transposable elements (TEs) were first discovered in maize by Barbara McClintock in 1948 [1]. (not in 1944)

citation errors, correct is:

PIRATE: ref. 10

REPET: ref. 11 and 12

Reference 14 is cited before reference 13.

sequences can be applied for any species → sequences that can be applied for any species

Legends to Figures 4 and 5:

"Number of TEs" has been replaced with "Number of sense LTRs (strand+) and antisense LTRs (strand-)" in the new manuscript.

Do the authors really mean LTRs, i.e. Long Terminal Repeats that are only part of a LTR

transposon, or do they mean "LTR transposons" ? I already pointed to this confusion in

the 1st round of review. If the authors mean the TEs, "LTRs" should be replaced either

with "TEs" or with "LTR transposons" in the legends of figures 4 and 5.

Round #1

Author's Reply:

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

[Download tracked changes file](#)

Decision

by *Emmanuelle Lerat*, 2021-04-05 12:28

Manuscript: [10.1101/2021.02.25.432867](#)

Revision needed

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

Reviews

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