

is located, and the starting and ending position for the cluster. We provided each tool with the necessary files. For the WG tool, this is the TE locations found in the annotation file and total genome length. For the ROI tool, we provided the piRNA cluster locations, the TE annotation information and total length for the ROIs. For the MXROI tool, we used the same information that was used for the ROI tool with the exception of the information regarding the length of the ROI.

We ran the tools on the piRNA data and generated table files representing the densities for the TEs within the whole genome for both mouse and rat using the WG tool. Using the ROI tool we obtained the densities for the TEs within piRNA clusters. We also created a matrix representing the densities for the TEs for each family and sub-family of transposable elements within each distinct piRNA cluster using the MXROI tool. To verify the accuracy of our calculation we compared the densities for the TEs within the genome calculated using our tool to the densities calculated using Repeat Masker. Figures 2 and Figure 3 below show the different densities for TEs across different regions within the mouse and rat genome.

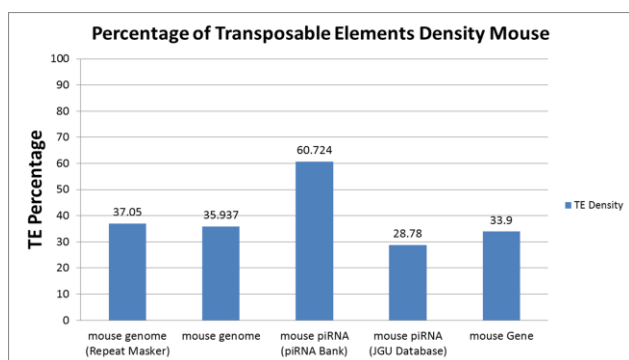


Figure 2: Different Densities of TE in Mouse

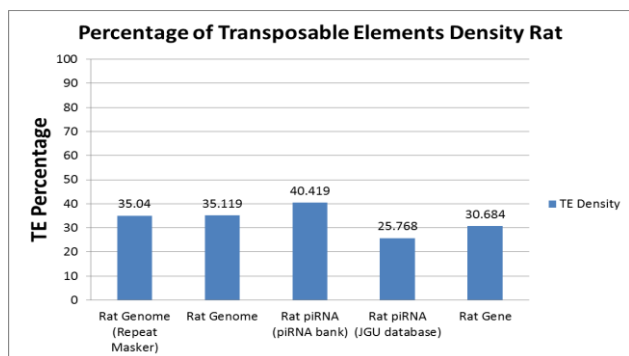


Figure 3: Different Densities of TE in Rat

5. Conclusions

RepCalc is a tool designed to be used for generating detailed quantitative information about the distribution of certain classes of TE families and their subfamilies at the genome level, within specific ROIs and across individual sub-regions, utilizing TE annotation information available throughout the various databases and annotation tools designed for this purpose. The tool is comprised of three tools, each one of them provides the user with a different level of quantitative detail regarding the TE densities across different segments of the genome.

To demonstrate how the tool works, we calculated the densities of TE families and sub-families within piRNA clusters and compared them to the TE densities within the genes and across the genomes of both mouse and rat utilizing two different piRNA databases. We found that the piRNA clusters that were extracted from piRNA bank exhibit a higher TE density than the JGU database for both the mouse and rat genes as well as across the whole genome.

Availability And Requirements

Project name: RepCalc
 Project home page: <https://github.com/eenblam/repcalc>
 Operating system: Windows, Linux, Mac OS X
 Programming languages: Python 3.4
 Other requirements: Python 3.4, Tkinter

Competing Interests

The authors declare that they have no competing interest.

Authors' Contributions

TA, AP and FH developed the concept. TA and BE designed the software. BE implemented the software. TA and BE tested and evaluated the software. TA, AP and BE wrote the manuscript. All the authors read and approved the final manuscript.

Acknowledgements

This work was supported by the National Science Foundation under award EPS-0903787. In addition we would like also to thank Dr. David Ray for his helpful suggestions throughout the work on this project.

References

- [1] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*, 409:860–921, 2001.

- [2] Pray L. Transposons: The jumping genes. *Nat. Educ*,1:204, 2008.
- [3] Fedoroff N V. Transposable Elements, Epigenetics , and Genome Evolution. *Science*, 338:758–767, 2012.
- [4] Malone CD, Hannon GJ. Molecular evolution of piRNA and transposon control pathways in *Drosophila*. Cold Spring Harb Symp *Quant Biol*, 74:225–234, 2009.
- [5] Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*, 8:272–285, 2007.
- [6] McClintock B. The Origin and Behavior of Mutable Loci in Maize. *Proc. Natl. Acad. Sci. U. S. A.*, 36:344–355, 1950.
- [7] Bergman CM, Quesneville H. Discovering and detecting transposable elements in genome sequences. *Brief. Bioinform.*,8:382 – 392, 2007.
- [8] Lerat E. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity*,104:520 – 533. 2010.
- [9] Smit, AFA, Hubley, R & Green P. RepeatMasker. <http://www.repeatmasker.org/>, March 2017.
- [10] Jurka J, Klonowski P, Dagman V, Pelton P. CENSOR--a program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem*, 20:119 –121.1996.
- [11] Volfovsky N, Haas BJ, Salzberg SL. A clustering method for repeat analysis in DNA sequences. *Genome Biol*, 2(8):1-11, 2001.
- [12] Achaz G, Boyer F, Rocha EPC, Viari A, Coissac E. Repseek, a tool to retrieve approximate repeats from large DNA sequences. *Bioinformatics*, 23:119–121, 2007.
- [13] Estill JC, Bennetzen JL. The DAWGPAWS pipeline for the annotation of genes and transposable elements in plant genomes. *Plant Methods*,5(1):1-11, 2009.
- [14] Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics*, 21:i351–i358, 2005.
- [15] Jurka J, Kapitonov V V., Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res*,110:462–467 ,2005.
- [16] Carène Rizzon, Gabriel Marais, Manolo Gouy, Christian Biémont. Recombination Rate and the Distribution of Transposable Elements in the *Drosophila melanogaster* Genome. *Genome Res*, 12: 400-407, 2002.
- [17] Cridland JM, Macdonald SJ, Long AD, Thornton KR. Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Mol Biol Evol*, 30(10):2311-2327, 2013.
- [18] Bartolomé C, Maside X, Charlesworth B. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol*. 19(6):926-937, 2002.
- [19] Jae-Sung Rhee, Beom-Soon Choi, Jaebum Kim, Bo-Mi Kim, Young-Mi Lee, Il-Chan Kim, Akira Kanamori, Ik-Young Choi, Manfred Scharl & Jae-Seong Lee. Diversity, distribution, and significance of transposable elements in the genome of the only selfing hermaphroditic vertebrate *Kryptolebias marmoratus*. *Scientific Reports*. 7, Article number:40121, 2017.
- [20] Leonardo Galindo González and Michael K Deyholos. Identification, characterization and distribution of transposable elements in the flax (*Linum usitatissimum* L.) genome. *BMC Genomics*. 13:644, 2012.
- [21] Xulio Maside, Carolina Bartolome, Stavroula Assimakopoulos, Brian Charlesworth. Rates of movement and distribution of transposable elements in *Drosophila melanogaster*: in situ hybridization vs Southern blotting data. *Genetics Research*. 78(2):121-136, 2001.
- [22] 24.1. Tkinter — Python interface to Tcl/Tk — Python v2.7.8 documentation. <https://docs.python.org/2/library/tkinter.html>, August 2014.
- [23] Lakshmi SS, Agrawal S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res*, 36:D173–D177, 2008.
- [24] Rosenkranz D, Zischler H. proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*,13:1-10, 2012.
- [25] Laurent Duret, Gabriel Marais , Christian Biémont. Transposons but Not Retrotransposons Are Located Preferentially in Regions of High Recombination Rate in *Caenorhabditis elegans*. *GENETICS*, 156(4): 1661-1669, 2000.
- [26] Cas Simons, Michael Pheasant, Igor V. Makunin, John S. Mattick. Transposon-free regions in mammalian genomes. *Genome Res*, 16(2): 164–172, 2006.