The Journal of Open Source Software

delfies: a Python package for the detection of DNA breakpoints with neo-telomere addition

Brice Letcher [1]

1 Laboratory of Biology and Modelling of the Cell, Ecole Normale Supérieure de Lyon, CNRS UMR 5239, Inserm U1293, University Claude Bernard Lyon 1, Lyon, France

DOI: 10.21105/joss.07385

Software

- Review
- Repository
- Archive

Editor: AHM Mahfuzur Rahman

Reviewers:

- @KatyBrown
- @natir

Submitted: 09 September 2024
Published: 12 January 2025

Summary

In multicellular organisms, all cells generally carry an identical genome, faithfully transmitted through cell divisions from the founding zygote. This is not the case in species that undergo Programmed DNA Elimination (PDE), the systematic destruction of portions of the genome in somatic cells during early development. In these species, somatic cells carry a reduced genome, while germline cells maintain an intact genome.

PDE was first documented in 1887 in a parasitic nematode (Boveri, 1887), and since then various forms of PDE have been found in a wide variety of organisms including birds, fish, insects, mammals, crustaceans, other nematodes and ciliates (Dedukh & Krasikova, 2021; Drotos et al., 2022). Some species eliminate entire chromosomes (birds, fish, insects, mammals) while others eliminate portions of chromosomes, with or without changes in chromosome number (copepod crustaceans, nematodes, ciliates).

In species that eliminate portions of chromosomes, two main types of elimination have been documented. The first is the elimination of small sequences (~100s of bp) called 'IESs', by a splicing process: a double-strand break is produced at each IES extremity, the IES is excised, and the two extremities are rejoined. This form has so far been documented in ciliates only. The second type is the elimination of large fragments of chromosomes (up to >1Mbp): a single double-strand break is produced, one side is eliminated, and telomeres on the retained side allow the new 'mini-chromosome' to be maintained in the soma. This form occurs in ciliates (Yu & Blackburn, 1991), nematodes (Gonzalez de la Rosa et al., 2020; Rey et al., 2023), and probably also in copepods (Beermann, 1977). While IES elimination in ciliates has been well-characterised genomically and functionally, chromosome fragmentation with neo-telomere addition has not.

Here, we present a tool called `delfies` to systematically detect sites of chromosome breakage and neo-telomere addition. `delfies` enables rapidly and comprehensively mapping the locations of elimination breakpoints in all species in which this form of DNA elimination occurs.

Statement of need

Several other tools for the detection of DNA elimination breakpoints have been developed and tested, all in the context of ciliates: `parTIES` (Denby Wilkes et al., 2015), `SIGAR` (Feng et al., 2020), `ADFinder` (Zheng et al., 2020) and `bleTIES` (Seah & Swart, 2021). Of these, `parTIES`, `ADFinder` and `SIGAR` allow the detection of IESs only, not sites of chromosome breakage with neo-telomere addition, and were primarily designed for short-read sequencing data. `bleTIES` was designed to detect and reconstruct IESs in the context of long-read sequencing data, and also includes a module for detecting chromosome breakage sites with telomere addition called MILTEL (Seah & Swart, 2021).

`delfies` was developed when studying PDE in nematodes, and presents several new features compared to MILTEL. Both tools output the locations of breakpoints in standard bioinformatics formats: MILTEL in a GFF3-formatted file, `delfies` in a BED-formatted file. While MILTEL expresses each putative breakpoint in isolation, `delfies` can merge multiple breakpoints occurring in close proximity in a user-configurable way. This allows for directly detecting more or less sharply-defined breakpoints, a feature that is known to vary in both ciliates and nematodes (Bétermier et al., 2023; Dockendorff et al., 2022; Estrem & Wang, 2023; Gonzalez de la Rosa et al., 2020). `delfies` also outputs the strand of breakpoints in the appropriate BED column, enabling subsequently classifying the genome into 'retained' and 'eliminated' compartments (details in the software repository).

`delfies` also explicitly models and outputs two types of breakpoints: in the first, the assembled genome is 'complete' and reads from the 'reduced' genome contain telomeres after the breakpoint. In the second, the assembled genome is 'reduced' and contains telomeres, and reads from the 'complete' genome contain unique non-telomeric sequence. These two types can be treated separately. For example, in the case of a reduced assembled genome, reads coming from the reduced genome can be specifically depleted at breakpoints and the new read-set used to assemble the complete genome (instructions provided in the software repository).

In addition to breakpoint locations, `delfies` extracts and outputs the sequences around the breakpoints in a Fasta-formatted file. This enables searching for motifs specifying breakpoints, e.g. using MEME (Bailey et al., 2015).

In practical terms, `delfies` has a highly configurable command-line interface, enabling specifying how much to filter read alignments, which regions of the genome to analyse and the types of breakpoints to look for. On a nematode genome of size 240Mbp sequenced at 85X average coverage with PacBio HiFi data, `delfies` finds all breakpoints in less than 2 minutes, using a single thread. For further speed, `delfies` also supports multi-threading.

`delfies` has already been used to successfully characterise the breakpoints, motifs, and retained/eliminated genomes of several nematode genera in the family *Rhabditidae*, supporting two upcoming publications (Letcher *et al.* and Stevens *et al.*, in preparation). For testing purposes, the author has prepared a subset of publicly-available data from the nematode *Oscheius onirici*, whose elimination breakpoint motif has been previously described (Estrem & Wang, 2023). The data are available on Zenodo (Letcher, 2024), and consist of a small genome region containing a single elimination breakpoint and alignment files for reads sequenced using three distinct technologies: Illumina NovaSeq, Oxford Nanopore Technologies PromethION and Pacific Biosciences Sequel II. The reads span a range of average lengths (151bp to 11.9kbp) and per-base qualities (Q11 to Q28). `delfies` recovered a single, identical breakpoint across all three datasets.

We anticipate this tool can be of broad use to researchers studying Programmed DNA Elimination, to characterise species known to eliminate but also to discover or screen for elimination in new species. This is especially relevant as new long-read and high-coverage sequencing data (of both germline and somatic cells) of eukaryotic species become increasingly available (Blaxter et al., 2022; Lewin et al., 2022). `delfies` may also be useful in other fields of research in which modified chromosomes with neo-telomeres are formed and maintained, such as cancer biology.

Acknowledgements

We acknowledge the many interactions with Lewis Stevens and Pablo Manuel Gonzalez de la Rosa at the Wellcome Sanger Institute and Marie Delattre at the Laboratory of Biology and Modelling of the Cell, which helped foster the development of `delfies`.

This work was supported by a grant from the Agence Nationale de la Rercherche: ANR-22-CE12-0027.

References

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research*, *43*(W1), W39–W49. https://doi.org/10.1093/nar/gkv416

Beermann, S. (1977). The diminution of heterochromatic chromosomal segments in cyclops (crustacea, copepoda). *Chromosoma*, *60*(4), 297–344. https://doi.org/10.1007/BF00292858

Bétermier, M., Klobutcher, L. A., & Orias, E. (2023). Programmed chromosome fragmentation in ciliated protozoa: Multiple means to chromosome ends. *Microbiology and Molecular Biology Reviews*, *87*(4). https://doi.org/10.1128/mmbr.00184-22

Blaxter, M., Mieszkowska, N., Di Palma, F., Holland, P., Durbin, R., Richards, T., Berriman, M., Kersey, P., Hollingsworth, P., Wilson, W., Twyford, A., Gaya, E., Lawniczak, M., Lewis, O., Broad, G., Howe, K., Hart, M., Flicek, P., & Barnes, I. (2022). Sequence locally, think globally: The darwin tree of life project. *Proceedings of the National Academy of Sciences*, *119*(4). https://doi.org/10.1073/pnas.2115642118

Boveri, T. (1887). Über differenzierung der zellkerne während der furchung des eies von ascaris megalocephala. *Anatomischer Anzeiger*, *2*, 688–693.

Dedukh, D., & Krasikova, A. (2021). Delete and survive: Strategies of programmed genetic material elimination in eukaryotes. *Biological Reviews*, *97*(1), 195–216. https://doi.org/10.1111/brv.12796

Denby Wilkes, C., Arnaiz, O., & Sperling, L. (2015). ParTIES: a toolbox for Paramecium interspersed DNA elimination studies. *Bioinformatics*, *32*(4), 599–601. https://doi.org/10.1093/bioinformatics/btv691

Dockendorff, T. C., Estrem, B., Reed, J., Simmons, J. R., Zadegan, S. B., Zagoskin, M. V., Terta, V., Villalobos, E., Seaberry, E. M., & Wang, J. (2022). The nematode oscheius tipulae as a genetic model for programmed DNA elimination. *Current Biology*, *32*(23), 5083–5098.e6. https://doi.org/10.1016/j.cub.2022.10.043

Drotos, K. H. I., Zagoskin, M. V., Kess, T., Gregory, T. R., & Wyngaard, G. A. (2022). Throwing away DNA: Programmed downsizing in somatic nuclei. *Trends in Genetics*, *38*(5), 483–500. https://doi.org/10.1016/j.tig.2022.02.003

Estrem, B., & Wang, J. (2023). Programmed DNA elimination in the parasitic nematode ascaris. *PLOS Pathogens*, *19*(2), e1011087. https://doi.org/10.1371/journal.ppat.1011087

Feng, Y., Beh, L. Y., Chang, W.-J., & Landweber, L. F. (2020). SIGAR: Inferring Features of Genome Architecture and DNA Rearrangements by Split-Read Mapping. *Genome Biology and Evolution*, *12*(10), 1711–1718. https://doi.org/10.1093/gbe/evaa147

Gonzalez de la Rosa, P. M., Thomson, M., Trivedi, U., Tracey, A., Tandonnet, S., & Blaxter, M. (2020). A telomere-to-telomere assembly of oscheius tipulae and the evolution of rhabditid nematode chromosomes. *G3 Genes|Genomes|Genetics*, *11*(1). https://doi.org/10.1093/g3journal/jkaa020

Letcher, B. (2024). *Dataset supporting the tool 'delfies: A python package for the detection of DNA breakpoints with neo-telomere addition'*. Zenodo. https://doi.org/10.5281/zenodo.14101797

Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M., Barker, K. B., Baumgartner, B., Belov, K., Bertorelle, G., Blaxter, M. L., Cai, J., Caperello, N. D., Carlson, K., Castilla-Rubio, J. C., Chaw, S.-M., Chen, L., Childers, A. K., Coddington, J. A., … Zhang, G. (2022). The earth BioGenome project 2020: Starting the clock. *Proceedings of the National Academy of Sciences*, *119*(4). https://doi.org/10.1073/pnas.2115635118

Rey, C., Launay, C., Wenger, E., & Delattre, M. (2023). Programmed DNA elimination in mesorhabditis nematodes. *Current Biology*, *33*(17), 3711–3721.e5. https://doi.org/10.1016/j.cub.2023.07.058

Seah, B. K. B., & Swart, E. C. (2021). BleTIES: annotation of natural genome editing in ciliates using long read sequencing. *Bioinformatics*, *37*(21), 3929–3931. https://doi.org/10.1093/bioinformatics/btab613

Yu, G.-L., & Blackburn, E. H. (1991). Developmentally programmed healing of chromosomes by telomerase in tetrahymena. *Cell*, *67*(4), 823–832. https://doi.org/10.1016/0092-8674(91)90077-c

Zheng, W., Chen, J., Doak, T. G., Song, W., & Yan, Y. (2020). ADFinder: accurate detection of programmed DNA elimination using NGS high-throughput sequencing data. *Bioinformatics*, *36*(12), 3632–3636. https://doi.org/10.1093/bioinformatics/btaa226