perSVade: personalized Structural Variation detection in your species of interest

Miquel Àngel Schikora-Tamarit[#1], Toni Gabaldón[#]

[#]*Comparative genomics lab, Life Sciences Department, BSC*

[1]miquel.schikora@bsc.es

Keywords: **genomics, structural variation, variant calling**

EXTENDED ABSTRACT

Background

Structural variants (SV) such as translocations, inversions, deletions, and other genomic rearrangements can contribute significantly to genetic and phenotypic variability across many species. The role of SV has been traditionally overlooked due to the technical limitations of SV detection and interpretation from short-read sequencing datasets. Most available algorithms yield low recall when tested on humans, but few studies have investigated the performance in non-human genomes [1]. Similarly, there are no specific indications about what parameters should be used for SV calling for most species. It is unclear whether the accuracy of each algorithm and running parameters validated on model species work equivalently for other species.

Results

In order to fill this gap we have developed perSVade (personalized Structural Variation Detection), a pipeline that identifies and annotates SVs in a way that is optimized for any input sample. Starting from a set of paired-end whole-genome sequencing reads, perSVade uses simulations on the reference genome to choose the best SV calling parameters. The output includes the optimally-called SVs, a report of the accuracy and a friendly graphical interface that shows the SVs on a genome browser. In addition, perSVade allows the calling small variants and copy-number variation. In summary, this pipeline is useful to identify several types of genomic variation in short reads using a single bash command.

We validated that perSVade increases the SV calling accuracy on both simulated and real variants for five diverse eukaryotic organisms. Importantly, we find that there is no universal set of "optimal" parameters, which makes our method essential to yield accurate variant calls.

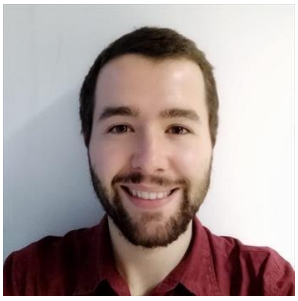
Conclusions

We consider that this tool will help to understand how SVs generate phenotypes across non-human organisms.

References

[1] D. L. Cameron, L. Di Stefano, and A. T. Papenfuss, "Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software.," *Nature communications*, vol. 10, no. 1, p. 3240, Jul. 2019.

Author biography



Miquel Àngel Schikora-Tamarit was born in Lleida in 1995. In 2017 he obtains a Bachelor Degree in Human Biology by the Universitat Pompeu Fabra (UPF) of Barcelona. He works on several projects focused on understanding single-cell behavior under the supervision of Dr. Lucas Carey in the Department of Experimental and Health Sciences of the UPF between 2014 and 2018. His research, involving experimental molecular biology techniques and computational analysis, yields four first-author publications in the journals Integrative Biology, Transcription, Genome Research and Cell Reports. He pursues a Master Degree in Bioinformatics at the UPF between 2017 and 2019. He develops the thesis project in the lab of Dr. Toni Gabaldón of the Center for Genomic Regulation (CRG) understanding the evolution of Complex I from fungal genomes. He is currently conducting his PhD thesis in Biomedicine at the University of Barcelona (UB), working on the evolution of drug resistance in fungal pathogens under the supervision of Dr. Toni Gabaldón at the Barcelona Supercomputing Center.