

Development of a Snakemake “framework/templates” to study the whole-genome transcriptional profiling from a blood-flesh trait (*bf*) and non blood-flesh trait (*non-bf*) cultivars in *Prunus persica*

Laure Heurtevin¹, Thierry Pascal¹, Bénédicte Quilot-Turion¹,
Marie Laure Martin Magniette² and Jacques Lagnel¹

¹ INRAE, GAFL, 84140, Avignon, France

² INRAE, IPS2, 91190 Gif-sur-Yvette, France

Corresponding Author: laure.heurtevin@inrae.fr, jacques.lagnel@inrae.fr

Little is known about the mechanisms controlling anthocyanin biosynthesis in flesh of fruit. Here, we explored the genetic pathways related to the elaboration of the blood-flesh trait in peach (*Prunus persica*). For this purpose, we employed a RNAseq approach to collect genome wide expression data and identify differentially expressed genes.

A comparative RNAseq study was carried out on flesh from a blood-flesh cultivar and a non blood-flesh cultivar at 4 fruit development stages, from 60 days after blooming up to fruit maturity.

40 libraries including biological replicates were sequenced by Illumina platform (Get-PlaGe) which generated 145Gbp (2.5Greads PE 150pb reads).

The RNAseq pipeline was developed using Snakemake [1] and Singularity in order to ensure reproducibility and flexibility in the analysis, traceability of the samples, pipeline ease of use as well as facilitate the portability and the scalability to large data sets.

Here, we propose a Snakemake “framework” based on a set of interoperable Snakemake rules as well as a set of templates (config, Slurm and samples sheet).

This Snakemake framework/templates and Singularity recipes/images will be available on a public forge based on GitLab source code management software. In addition the GitLab-CI/CD (Continuous Integration/Delivery) was used in order to build and push singularity containers to a registry.

Statistical analysis was performed by DiCoExpress (R pipeline developed by the team of ML Martin-Magniette) embedded in a image singularity.

We found 4379 up- and 5027 down- regulated genes in peach flesh between blood-flesh and non blood-flesh cultivars in this development kinetics. We observed a cluster of genes linked to the photosystems I and II overexpressed at times 2 and 3 in the blood-flesh cultivar. This cluster follows the pattern of expression of a candidate gene and of genes involved in the pathway of anthocyanin biosynthesis along the stage of fruit development.

Finally, the use of the Snakemake “framework/template” proposed here will facilitate NGS bio-analyses by non developers in bioinformatics.

Acknowledgements

We are grateful to the Genotoul bioinformatics platform Toulouse for providing help and computing and storage resources.

References

- 1 Köster J and Rahmann S. Snakemake – a scalable bioinformatics workflow engine. *Bioinformatics*, 28:2520-2522, 2012.