

Getting useful information from RNA-seq contaminants: A case of study in *Tetrapedia diversipes* transcriptome



Natalia **Araujo**, Alexandre **Zuntini** & Maria Cristina **Arias**
Universidade de São Paulo - Brazil

na.araujo@usp.br
@nat2bee

Introduction

The RNA-Seq is a second-generation sequencing approach widely used in the study of gene expression, especially for non-model species. Because this technique is not a species-specific approach and species are constantly interacting it is highly probably that one can find alien genes in a RNA-Seq transcriptome dataset. Herein we report the use of these “contaminant” transcripts from plants as a source of information about the floral resources used for nest provisioning by *Tetrapedia diversipes* (Fig. 1), an oil-collecting solitary bee native from the Neotropical region. This species presents two main reproductive generations during the year: one in the hot and wet season (generation one - G1) and the second during the cold and dry season (generation two - G2). The developmental cycle since egg till adult varies significantly between the two generations (Fig. 2) because the pre-pupal larvae from G2 enter in diapause.



Fig. 1 *Tetrapedia diversipes* by Cordeiro GD.

Material and Methods

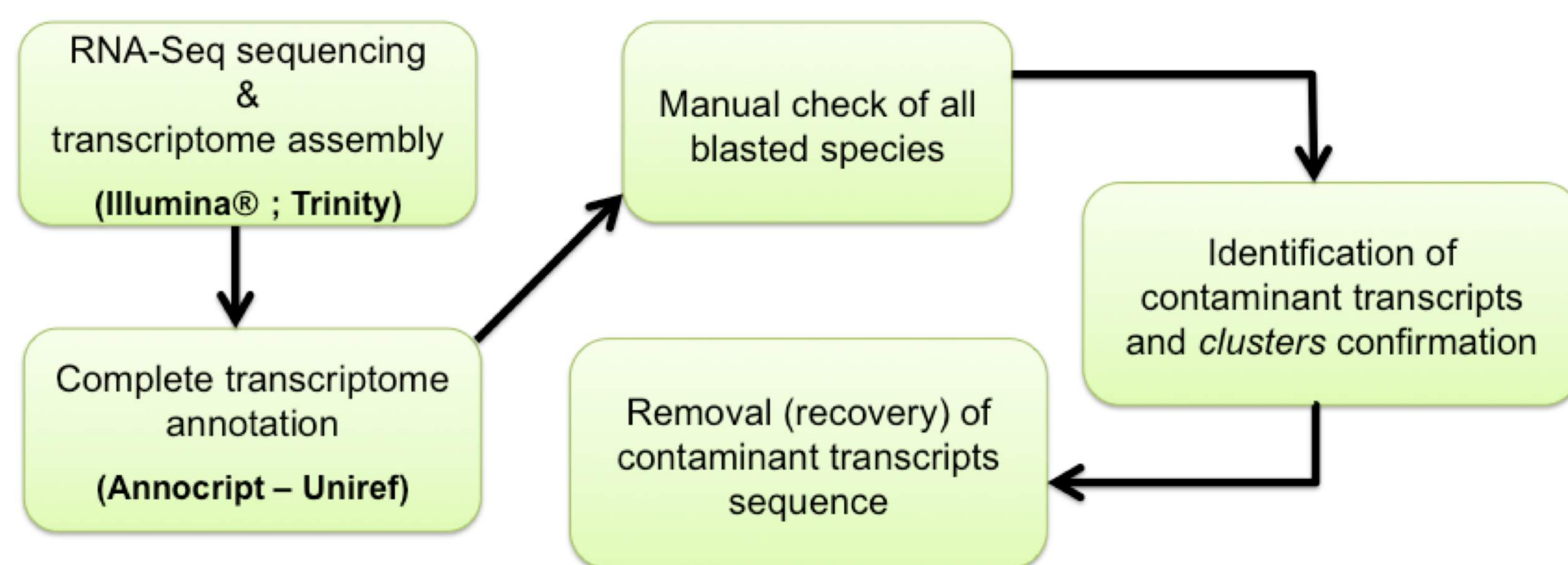


Fig. 3 Pipeline of analyses for “contaminants” identification. RNA from 9 bees from each reproductive generation were sequenced by RNA-Seq and assembled. Programs used in each step of the analyses are indicated in bold otherwise custom scripts in R or Python were used (available at <https://github.com/nat2bee>)

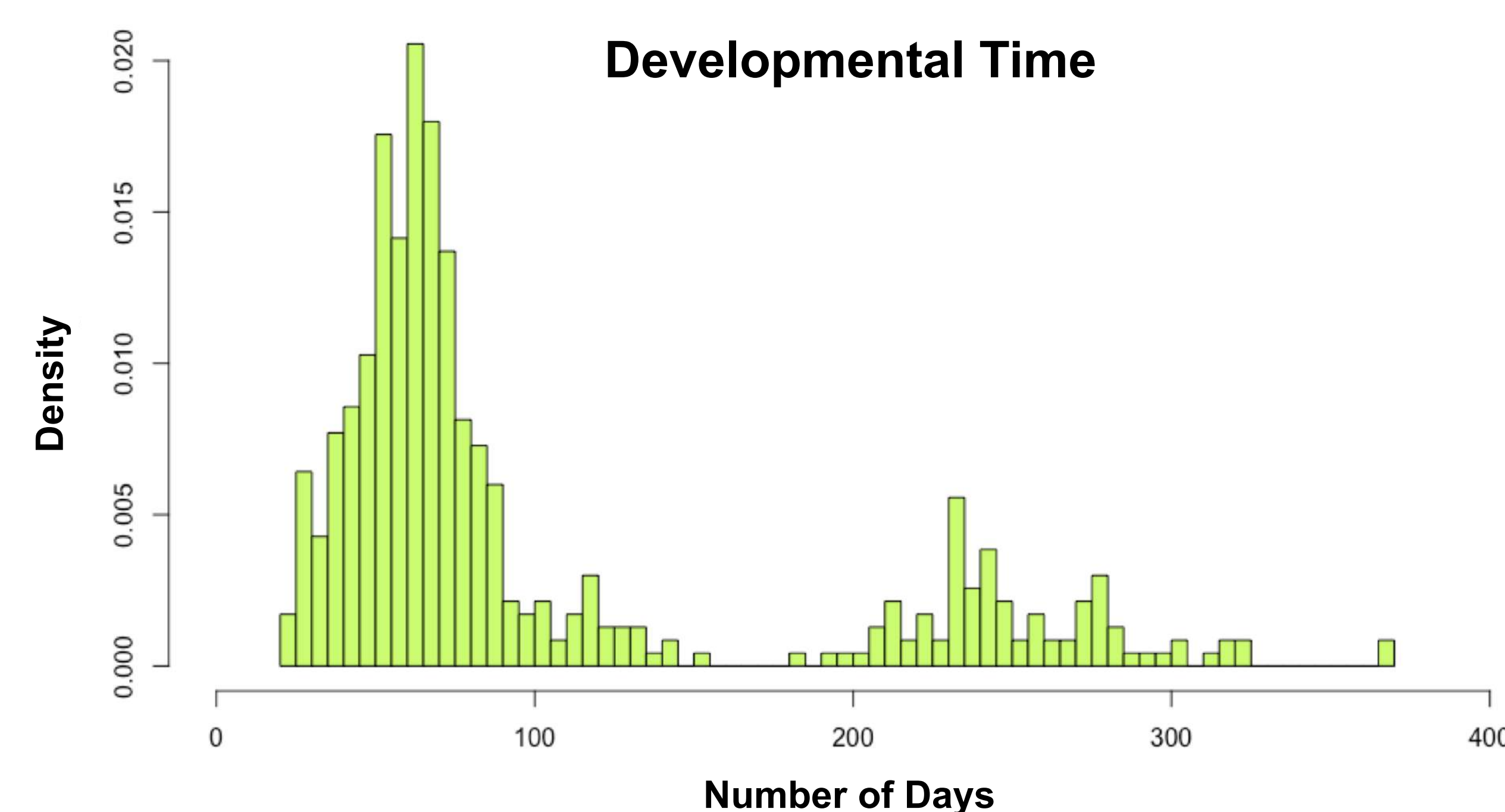


Fig. 2 *T. diversipes* developmental time in days. Data from: Cordeiro GD. (2009) Abelhas solitárias nidificantes em ninhos-armadilha em quatro áreas de Mata Atlântica do Estado de São Paulo. Dissertation. USP - Brazil

Results and Conclusions

In total 857 and 538 plant contaminants transcripts were identified in foundresses from G1 and G2, respectively. Most of the transcripts in G1 were from Euphorbiaceae (41.2%) and Amaranthaceae (31.5%) while in G2 65.7% were from Euphorbiaceae (Fig. 4).

These results indicate that contaminant transcripts might be used not only for the study of bee/ plant relation but also for other interaction analyses such as parasitism and symbiotic studies.

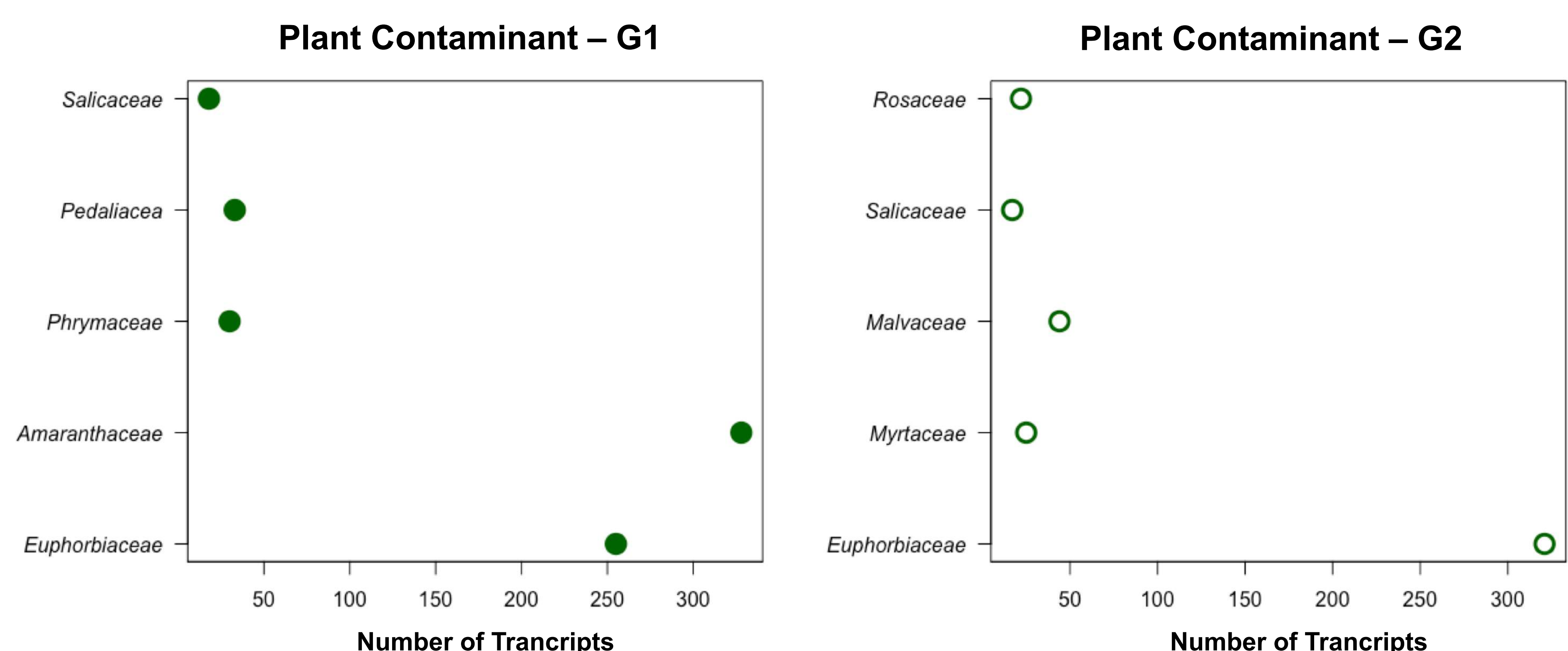


Fig. 4: The most common plant families detected among *T. diversipes* contaminant transcripts from reproductive generations 1 and 2 and the number of transcripts reported for each of them.

Acknowledgments:



LABORATÓRIO DE ABELHAS DA UNIVERSIDADE DE SÃO PAULO



Do you want to know more?

