Transcriptomic Analysis of *Protea* Hybrids Using HybridExpress

Pavan Mudrageda, Dr. Nora Mitchell | Department of Biology, University of Wisconsin-Eau Claire

pavan.mudrageda9@gmail.com | https://noramitchell.weebly.com/ | www.linkedin.com/in/pavan-mudrageda-799803274

University of Wisconsin Eau Claire

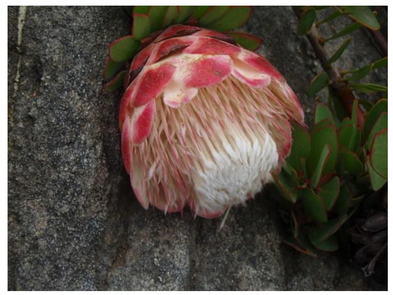
U.S. National Science Foundation

Introduction

Hybridization in plants is known to produce novel characteristics unseen in any of its progenitor species, with broad implications for evolutionary biology, habitat restoration, and crop development. This study aims to explore the genetic basis of hybridization by comparing transcriptomic data of two hybrid strains between *Protea punctata* and *P. venusta* grown from seeds collected from the Cape Floristic Region of South Africa. The genus *Protea* is a well-established model in evolutionary research, making it an ideal subject for this work. RNA-seq data from one *P. venusta*, two *P. punctata*, and three putative hybrid individuals were processed using the Trinity assembly pipeline and analyzed with the HybridExpress R package to characterize gene expression patterns associated with hybridization.

P. puncata

Observation ©Sam McCarren, some rights reserved (CC BY)
https://www.inaturalist.org/observations/151022800



P. venusta

Observation ©Di Turner, no rights reserved (CC0)
https://www.inaturalist.org/observations/10787440

Materials and Methods

Seeds of wild *Protea* plants were collected in South Africa and grown as part of Carlson et al. (2011). Hybrids were designated a *P. punctata* hybrid or *P. venusta* hybrid based on closest morphological resemblance to parent. The sample label prefix corresponds to the species as shown in Figure 1.

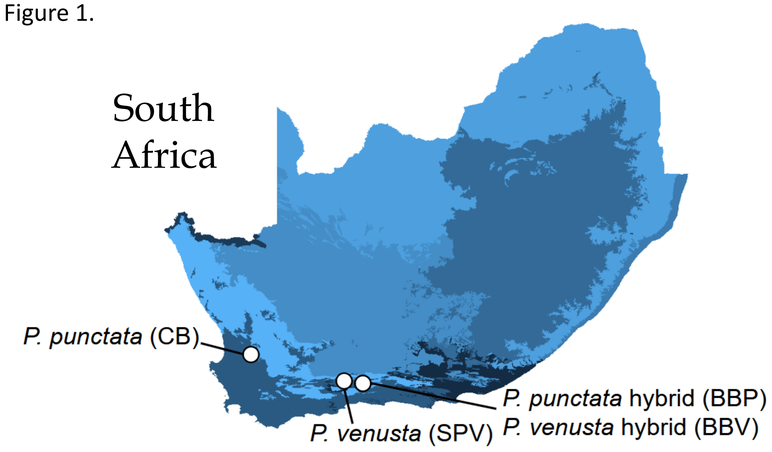


Figure 1: Locations where sample seeds were collected.

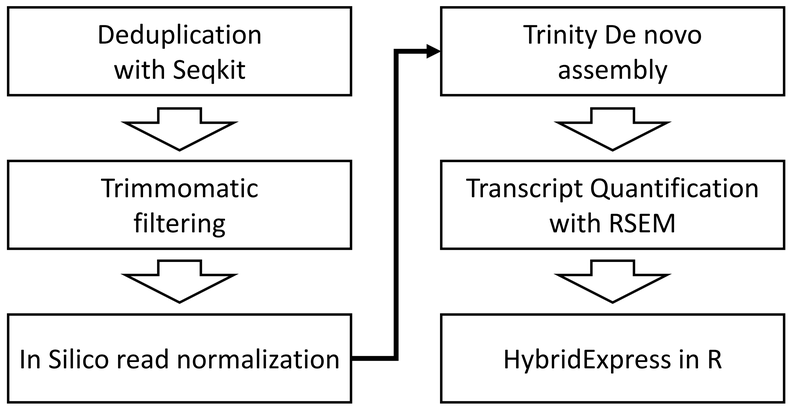


Figure 2: RNA-seq data processing workflow

RNA was extracted from leaf samples using the Qiagen RNeasy mini kit and prepared for mRNA-Sequencing using the Illumina TruSeq Stranded mRNA Sample Preparation kit. Sample libraries were prepared using version 2 sequencing chemistry for NextSeq 500 sequencing. The RNA-seq data was processed as shown in Figure 2. FastQC was used for quality control assessment. The raw data was deduplicated to allow for transcript quantification using RSEM. Trimmomatic filtering, In Silico read normalization, and RSEM were run with Trinity using the default inputs.

Results and Discussion

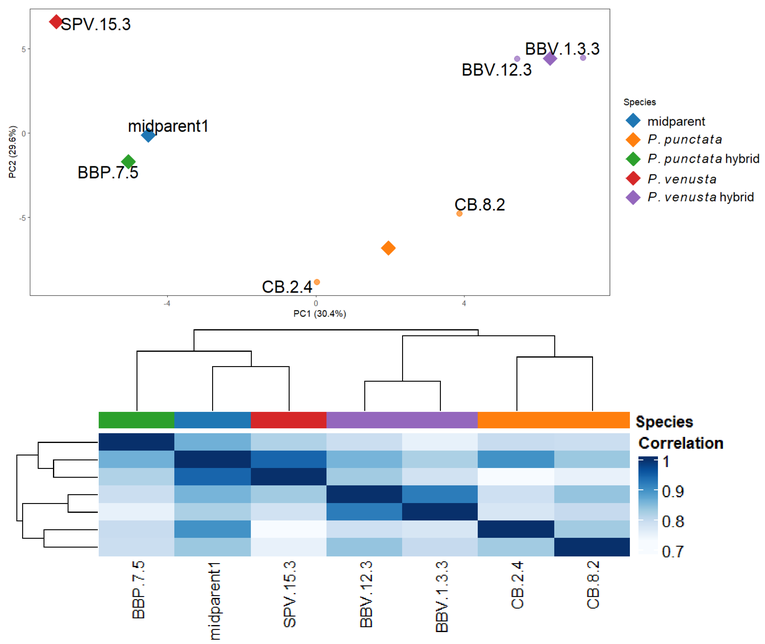


Figure 3: PCA plot (top) and heatmap (bottom) of gene expression data. Heatmap shows pairwise sample correlations with hierarchical clustering.

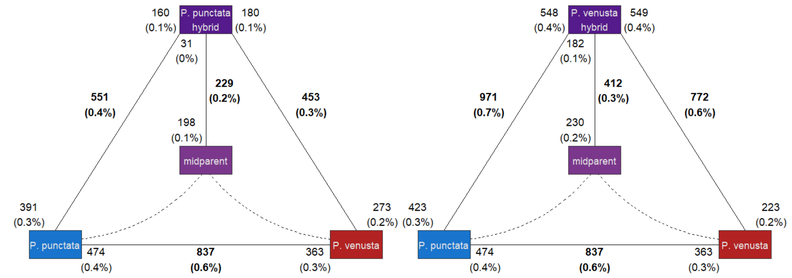


Figure 4: Expression triangle plot showing number of upregulated genes next to species name and total number of Differentially Expressed Genes (DEGs) between species.

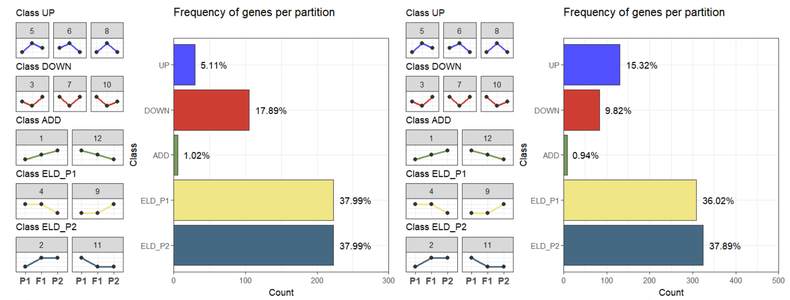


Figure 5: Expression partitions plot of *P. puncata* hybrid (left) and *P. venusta* hybrid (right)**.** UP – transgressive upregulation, DOWN – transgressive downregulation, ADD – additivity, ELD_P1 – expression level dominance (ELD) towards *P. punctata*, ELD_P2 – ELD towards *P. venusta*.

- The *P. punctata* hybrid is significantly closer to the parent species than the *P. venusta* hybrids in terms of expression (Figures 3 & 4).
- The *P. venusta* hybrids show notably higher transgressive upregulation and lower transgressive downregulation (Figure 5).
- Both hybrids show very low additivity and even amounts of ELD towards both parents, with the *P. venusta* hybrid displaying slightly higher ELD towards *P. venusta* that *P. punctata* (Figure 5).
- One possible future direction is to perform Gene Ontology analysis to determine the differences in biological processes.

References

- Mitchell, N., and K.E. Holsinger. 2018. Cryptic natural hybridization between two species of *Protea*. *South African Journal of Botany*. 118:306-314. http://doi.org/10.1016/j.sajb.2017.12.002
- Carlson JE, Holsinger KE, Prunier R. Plant responses to climate in the Cape Floristic Region of South Africa: evidence for adaptive differentiation in the Proteaceae. Evolution. 2011 Jan;65(1):108-24. doi: 10.1111/j.1558-5646.2010.01131.x. Epub 2010 Oct 7. PMID: 20840595.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011 May 15;29(7):644-52. doi: 10.1038/nbt.1883. PMID: 21572440; PMCID: PMC3571712.
- Almeida-Silva F, Prost-Boxoen L, Van de Peer Y (2025). *HybridExpress: Comparative analysis of RNA-seq data for hybrids and their progenitors*. doi:10.18129/B9.bioc.HybridExpress, R package version 1.4.0, https://bioconductor.org/packages/HybridExpress.

Acknowledgements

This research was funded by the National Science Foundation (NSF) Research Experience for Undergraduates (REU) grant OAC- 2447779. The computational resources of this study were provided by the Blugold Center for High-Performance Computing under NSF grant CNS-1920220. Special thanks to Dr. Nora Mitchell for her mentorship.

We thank the Office of Research and Sponsored Programs for supporting this research, and Learning & Technology Services for printing this poster.