

FOREST TREE TRANSCRIPTOME DATABASE



FABI

Rivoningo Mageza
Supervisors: Nanette Christie & Prof Zander Myburg
Forest Molecular Genetics Programme, Department of Biochemistry, Genetics and Microbiology,
University of Pretoria

BACKGROUND

The Forest Molecular Genetics programme (FMG) studies economically and ecologically significant Forest trees, namely Eucalyptus, Pinus and Populus. It has collated large-scale transcriptome and will generate more RNA-seq and other omics data as it continues conduct molecular breeding and tree crop improvement research. This project adds an RNA-seq database, and an application that makes the hosted data accessible to researchers in the programme, to an existing FMG Information Resource.

OBJECTIVES

- Design and develop a database to host transcriptome data
- Design a user interface (UI) that is accessible to FMG researchers
- Develop an R shiny application to implement the UI
- Demonstrate the efficacy of the database and application

METHODOLOGY



1. Database development

- Data was processed with r (4.2.1) and python (3.10.4)
- Processed data was uploaded to Google Cloud Storage
- BigQuery tables were created in according to schema

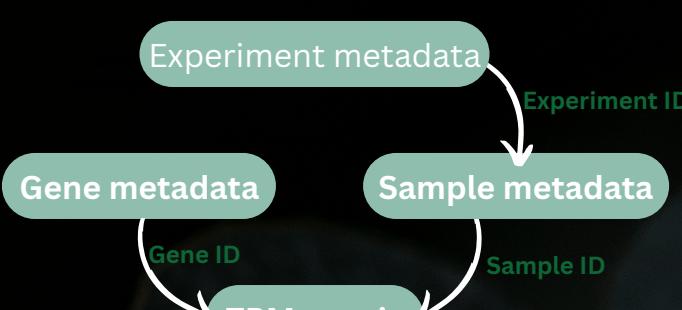


figure 1: Database schema



2. User Interface design

- 4 panels were designed for users to query the database and visualise the results



figure 2: Illustration of designed panels



3. User Interface development

- R Studio (version 2.2.1) was used to develop the user interface as an R Shiny



4. Case Study



RESULTS

1. Database implementation

- Data uploaded to Google Cloud Storage as .csv files
- Tables created in BigQuery database

Experiment table	Sample table	Gene table	TPM matrix
Exp_ID Castrato_a	RNA_ID M500	Gene_ID Eucgr.14	Gene_ID M501
Exp_name E. nitens	Exp_ID E1	Descrip Kinase	M501 o.571

figure 3: Illustration of created BigQuery tables



2. UI design and development

- Accessible UI
- Users can query the database using selectors
- List creation, list management and visualisation

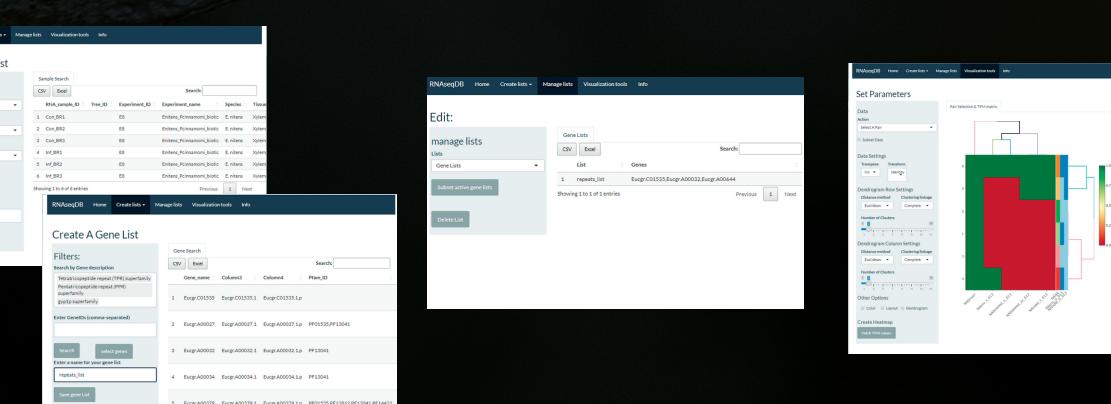


figure 4: R Shiny app tab panels



3. Case study

- Searched for samples in water stress experiments
- Generated a Heatmap to determine highly expressed genes in these samples
- Searched for samples in nitrate experiments
- Generated a heatmap to determine if the same genes are also highly expressed in the nitrate samples

REFERENCES

- 1.BIGQUERY. (N.D.). GOOGLE. [HTTPS://CLOUD.GOOGLE.COM/BIGQUERY/DOCS](https://cloud.google.com/bigquery/docs)
- 2.CHANG, W., CHENG, J., ALLAIRE, J., SIEVERT, C., SCHLOERKE, B., XIE, Y., ALLEN, J., MCPHERSON, J., DIPERT, A., & BORGES, B. (2022). SHINY: WEB APPLICATION FRAMEWORK FOR R (R PACKAGE VERSION 1.7.3.9001) [R]. [HTTPS://SHINY.RSTUDIO.COM/](https://shiny.rstudio.com/)
- 3.CHRISTIE, N., MANNAPPERUMA, C., PLOYET, R., VAN DER MERWE, K., MÄHLER, N., DELHOMME, N., NAIDOO, S., MIZRACHI, E., STREET, N. R., & MYBURG, A. A. (2021). QTLXPLORER: AN ONLINE SYSTEMS GENETICS BROWSER IN THE EUCALEYPTUS GENOME INTEGRATIVE EXPLORER (EUCGENIE). *BMC BIOINFORMATICS*, 22(1), [HTTPS://DOI.ORG/10.1186/S12859-021-04514-9](https://doi.org/10.1186/s12859-021-04514-9)
- 4.CHURCH, P. C., & GOSCIŃSKI, A. M. (2014). A SURVEY OF CLOUD-BASED SERVICE COMPUTING SOLUTIONS FOR MAMMALIAN GENOMICS. *IEEE TRANSACTIONS ON SERVICES COMPUTING*, 7(4). [WORLDCAT.ORG. HTTPS://DOI.ORG/10.1109/TSC.2014.2353645](https://doi.org/10.1109/TSC.2014.2353645)
- 5.GOOGLE CLOUD COMPUTING SERVICES. (N.D.). GOOGLE. [HTTPS://CLOUD.GOOGLE.COM/DOCS](https://cloud.google.com/docs)
- 6.HEFER, C., MIZRACHI, E., JOUBERT, F., & MYBURG, A. (2011). THE EUCALEYPTUS GENOME INTEGRATIVE EXPLORER (EUCGENIE): A RESOURCE FOR EUCALEYPTUS GENOMICS AND TRANSCRIPTOMICS. *BMC PROCEEDINGS*, 5(7), 1. [WORLDCAT.ORG. HTTPS://DOI.ORG/10.1186/1753-6561-5-S7-049](https://doi.org/10.1186/1753-6561-5-S7-049)
- 7.MA, T., & ZHANG, AIDONG. (2017). OMICS INFORMATICS: FROM SCATTERED INDIVIDUAL SOFTWARE TOOLS TO INTEGRATED WORKFLOW MANAGEMENT SYSTEMS. *IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS*, 14(4). [WORLDCAT.ORG. HTTPS://DOI.ORG/10.1109/TB.2016.2535251](https://doi.org/10.1109/TB.2016.2535251)
- 8.SUNDELL, D., STREET, N. R., KUMAR, M., MELLEROWICZ, E. J., KUCUKOGLU, M., JOHNSSON, C., KUMAR, V., MANNAPPERUMA, C., DELHOMME, N., NILSSON, O., TUOMINEN, H., PESQUET, E., FISCHER, U., NIITTYLÄ, T., SUNDBERG, B., & HVIDSTEN, T. R. (2017). ASPWOOD: HIGH-SPATIAL-RESOLUTION TRANSCRIPTOME PROFILES REVEAL UNCHARACTERIZED MODULARITY OF WOOD FORMATION IN POPULUS TREMULA. *THE PLANT CELL*, 29(7), 1585–1604. [WORLDCAT.ORG. HTTPS://DOI.ORG/10.1105/TPC.17.00153](https://doi.org/10.1105/tpc.17.00153)