The Genetics of Chocolate: Transcriptome Atlas of *T. cacao*

Evelyn Kulesza

The Pennsylvania State University

USA

Outline

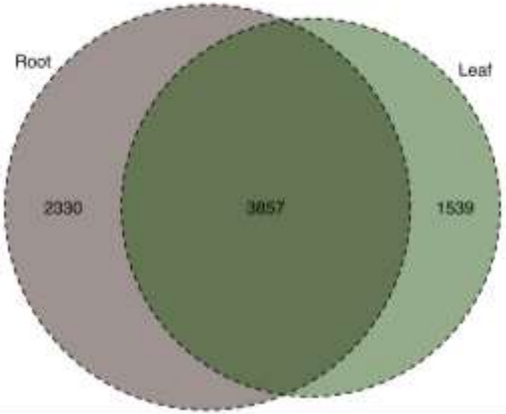
- What is a Gene Atlas?
- Why Use a Gene Atlas?
- Project Workflow
- Atlas Validation Case Studies
- Resources and Conclusions
- Acknowledgements

What is a Gene Atlas?

- The cacao genome has ~28,000 genes
- Each tissue and cell type expresses a different subset of genes (gene expression profile)

What are the differences in gene expression?

Root



Leaf

Genes expressed above 25 CPM

What is a Gene Atlas?

- Curated database for high quality, large-scale gene expression data
- Represents different tissues, developmental stages, or treatment conditions

EMBL-EBI

Why Use a Gene Atlas?

- To accelerate genetics research in cacao
- To speed up functional genomics work and gene discovery
    - Individual gene expression
    - Gene expression changes
    - Dissecting multigene families
    - Mining for molecular markers

Cacao Betulia

Project Workflow

Data Collection and Analysis

Gene Discovery Resources

Tissue Harvest:
3-5 replicates per tissue type
402 RNA libraries

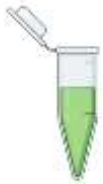
RNA Extraction

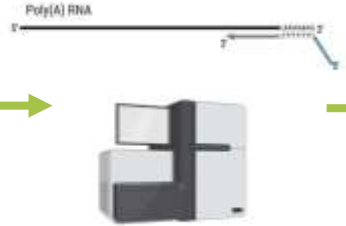
Library construction
Lexogen QuantSeq 3’ and Illumina Sequencing

Bioinformatics:
Read mapping
Annotation

Transcriptome Data Matrix

Visualization Browser





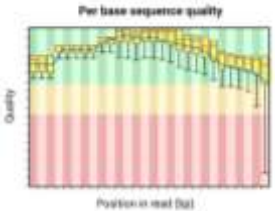
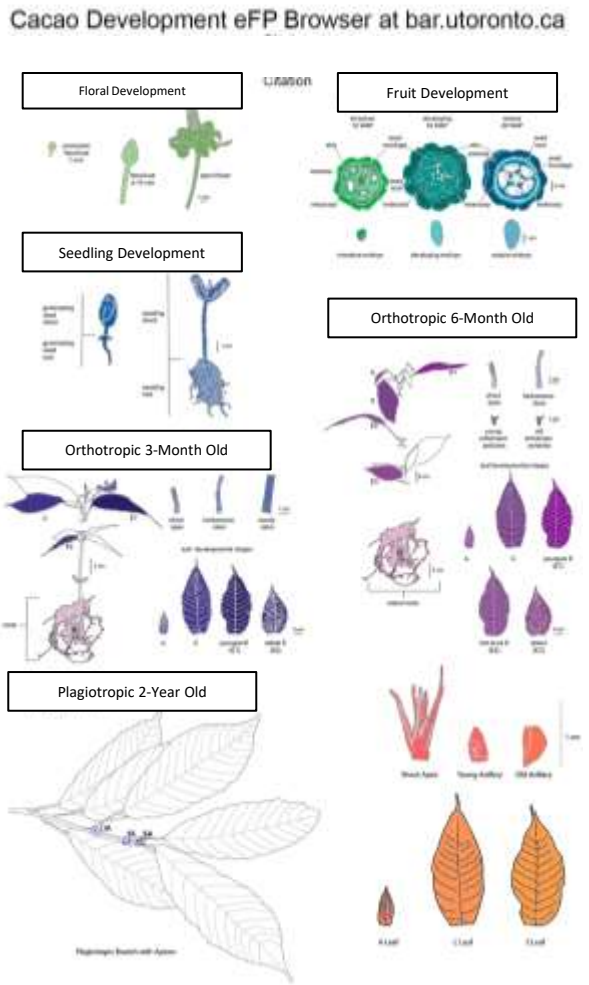
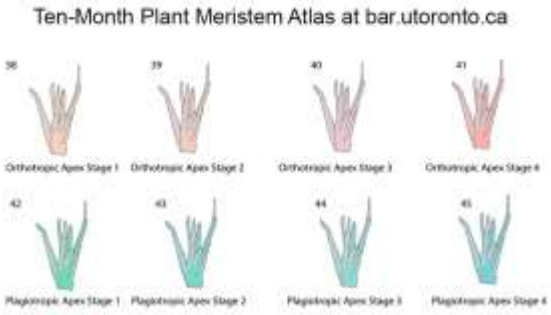
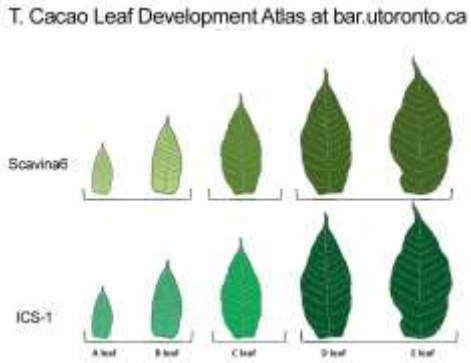


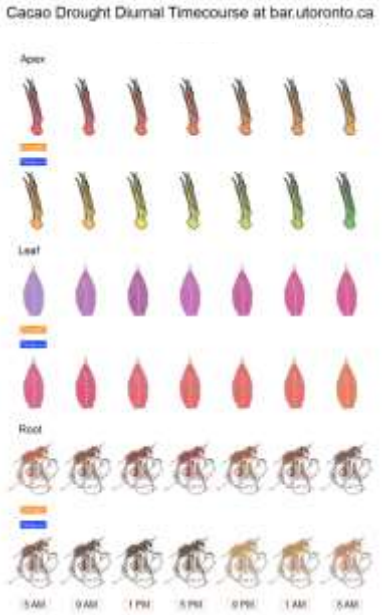
Figure made using Biorender

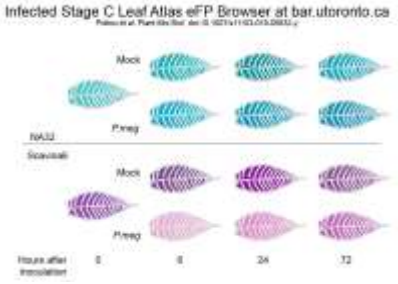
Atlas Overview





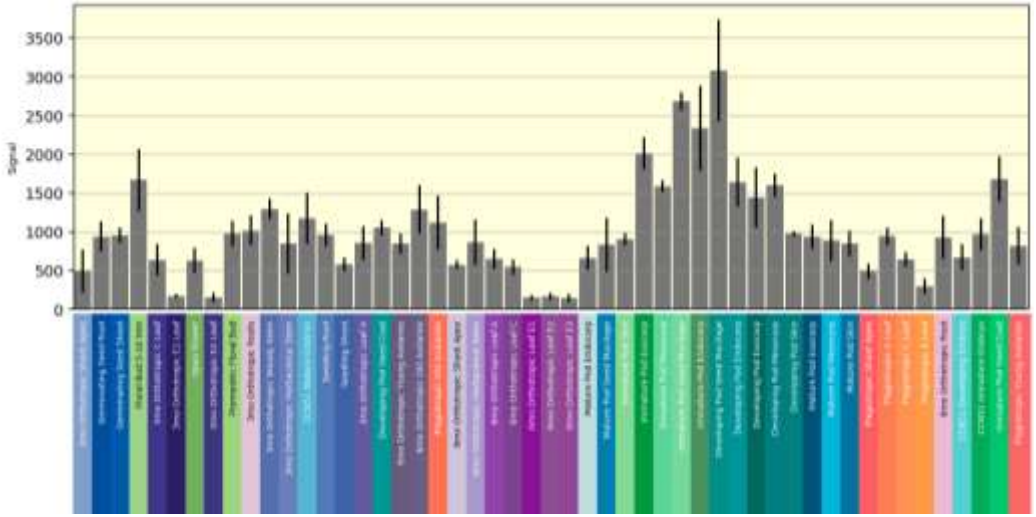


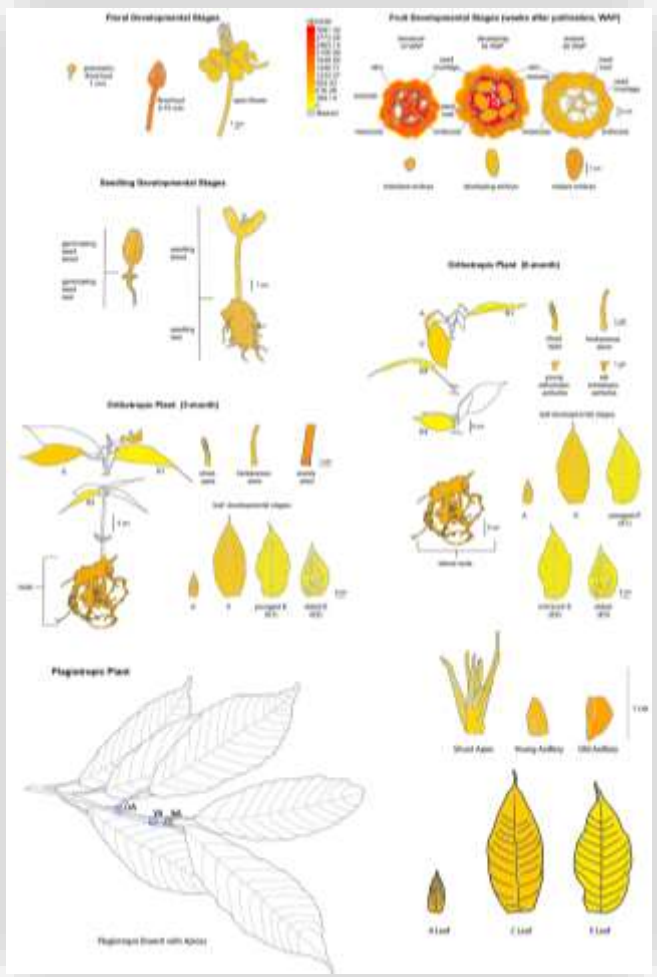




Atlas Examples: *ACT7*

Actin: Housekeeping Gene

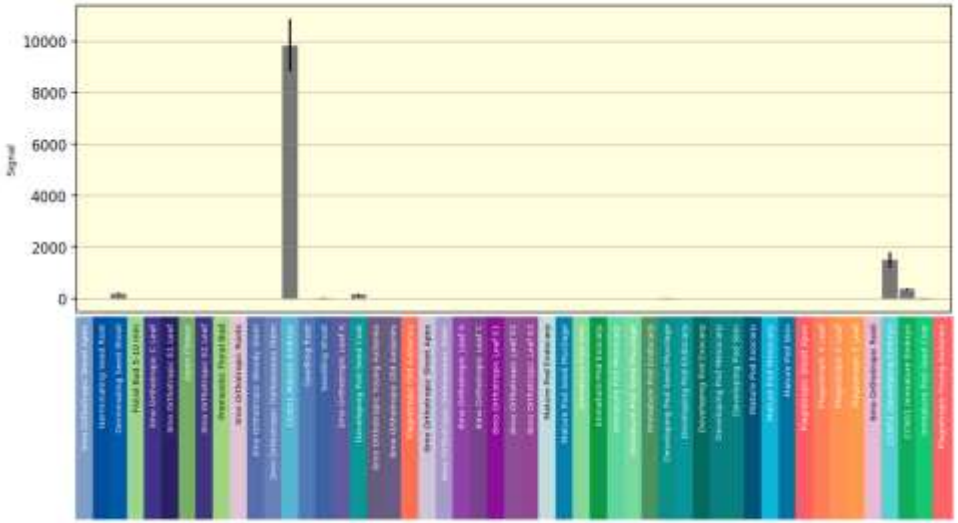


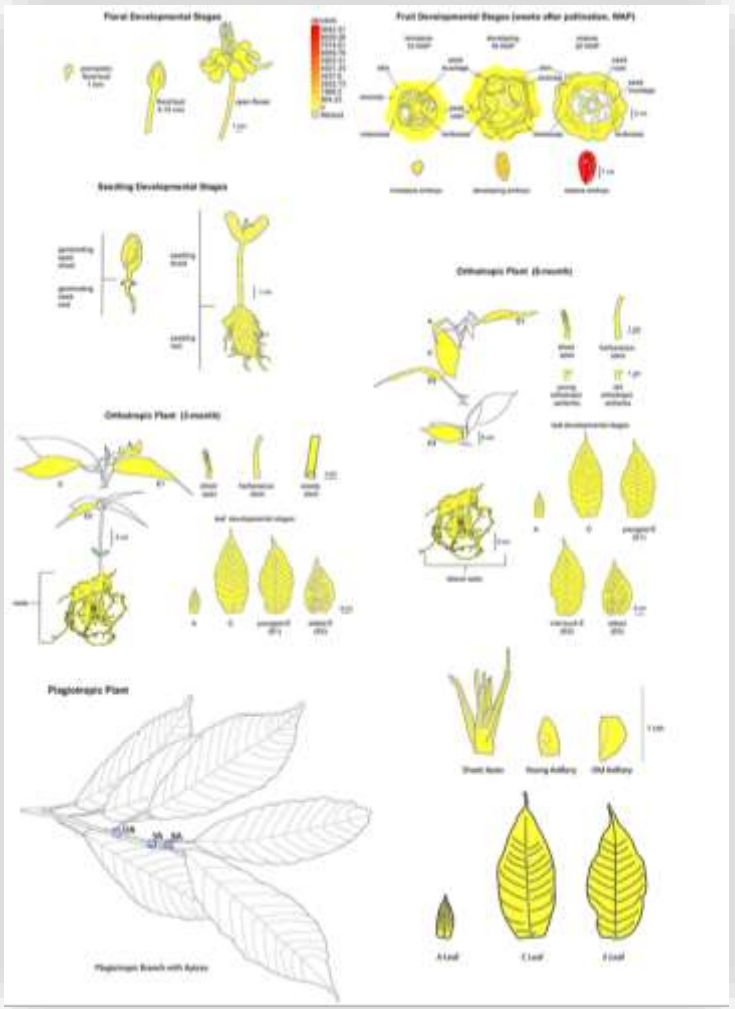


- Actin gene is expressed in all tissues
- Expression levels vary 100 fold
- Highest expression in fruit tissues

Atlas Examples: *LEA19*

Late embryogenesis abundant protein D-19: Plant growth and development





- Highly expressed in mature embryos
- Extremely tissue specific

Resources and Conclusions

Data Matrix

- Expression data
  - 402 RNA libraries
  - ~11.2 million gene expression datapoints
- Available for further bioinformatic analysis
  - Differential gene expression analysis
  - Gene ontology (GO) analysis

Visualization Browser

- Housed at University of Toronto BAR
- Explore gene expression changes with minimal bioinformatic skills needed
- Graphical User Interface representing the data matrix
- Simplifies large data sets into easily interpreted images
- Resource for science communication

Implications of this research on the income of cacao farmers

- Accelerate and reduce cost of genomics research and breeding
- Elite varieties will provide better yields with less inputs
- Sustainable intensification will increase farmers income

Data Access and Collaboration

- Open data access to the public
  - Publication in preparation
  - Website will be open after publication

Early access for collaborations

Scan QR Code

OR

Contact

Evelyn Kulesza:
evk5405@psu.edu

Acknowledgments

Cacao Gene Atlas Consortium

- Penn State University
  - Patrick Thomas
  - Mark Guiltinan
  - Siela Maximova
  - Sarah Prewitt
  - Akiva Shalit-Kaneh
  - Eric Wafula
  - Benjamin Knollenburg
  - Noah Winters
  - Claude Depamphilis
  - Melanie Perryman
  - Lena Sheaffer
  - Paula Ralph
- U Toronto BAR
  - Nicholas Provart
  - Asher Pasha
- Penn State Genomics Core
- Fernando Crespo (Ecuador)