RNA-seq and prospects for obtaining non-model plant transcriptomes

Alexa S. Tyszka[1], Drew A. Larson[2], Joseph F. Walker[1]

[1]University of Illinois at Chicago Biological Sciences [2]Indiana University Department of Biology

Background

Transcriptomes are more cost-effective than genomes and can serve the same purposes in many cases (Yang et al. 2017). However, there are many difficulties involved in obtaining and preserving tissue for RNA-seq. Most of these come from the need to immediately flash freeze tissue in liquid nitrogen and store it at low temperatures. This makes remote field work more arduous and expensive, and excludes many areas where liquid nitrogen is not available. Technological advances have extended the time that samples can stay at room temperature, but many protocols still require freezer storage. It is yet unclear how RNA degrades in tissue that is not preserved using common methods. Multiple studies (most recently He et al. 2022 and Ruiz-Vargas et al. 2023) have shown the viability of silica-dried plant tissue for RNA extraction. Some studies have detected RNA survival under typical herbarium preservation methods and the survival of seed RNA for centuries (see lower middle panel).

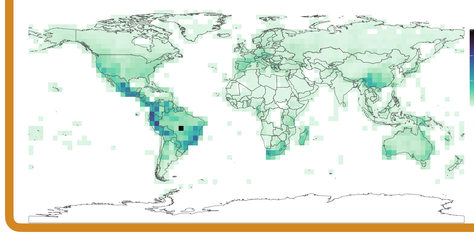
Research Questions

- **How well is global plant diversity captured by existing RNA-seq data and are there major gaps in sampling?**
- **How can we leverage silica-dried plant material for a more complete transcriptome tree of plant life?**

How do RNA-seq records reflect global vascular plant species richness?

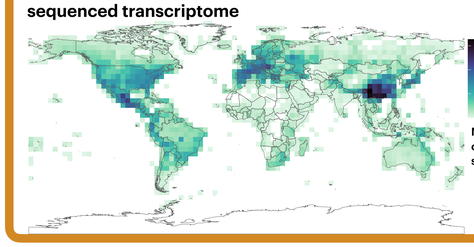
Interactive figures available here:

1a) Total global vascular plant species richness



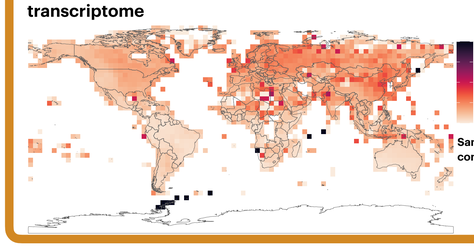
1a) Global species occurrence data were collected from the RBIEN database. Each species in the Kew World Checklist of Vascular Plants (WCVP) was searched in RBIEN (WCVP n=356,618, within RBIEN n= 174,106). Lat/long pairs (n=22,881,401) were downloaded and rasterized over a 4x4 latitude by longitude grid.

1b) Richness for vascular plant species with a sequenced transcriptome



1b) The records for transcriptome sequenced species which occurred in the SRA (n= 4,653 valid species were present in RBIEN, out of 6,966) were collected. This returned 6,441,980 lat/long pairs.

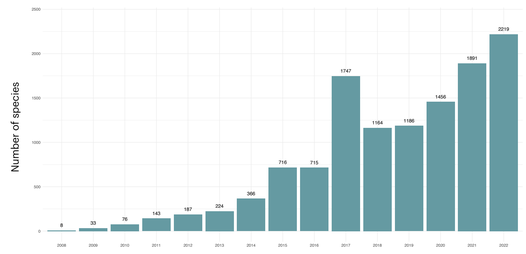
1c) Proportion of species with a sequenced transcriptome



1c) An approximate measure of completeness was calculated by taking the number of species with transcriptomes sequenced, compared to the predicted number of species in a grid cell.

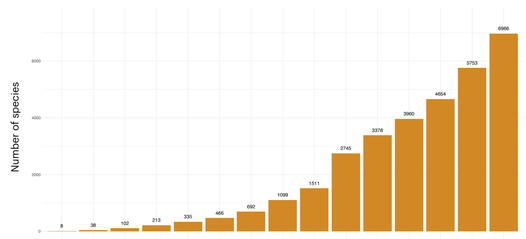
How has vascular plant diversity in the NCBI SRA increased over time?

2a) Number of species with sequenced transcriptomes uploaded to the SRA per year



2a) The increase in RNA-seq species diversity within vascular plants over time. The NCBI SRA reports statistics from uploaded RNA-seq runs, which were filtered by name to species level and referenced to currently accepted scientific names using the *rgbif* taxonomic backbone (https://github.com/ropensci/rgbif).

2b) Number of species with a sequenced transcriptome on the SRA by year



2b) The cumulative summation of unique vascular plant species that have been sequenced by year. Each species is only added to the total on the year it is first sequenced. A total of 6,966 unique and valid species within vascular plants have been sampled and uploaded to the NCBI SRA from 2008 to 2022, representing 406,177 records.

Selected studies inspiring this research

Study	Organism	Organism type	Preservation type	Sample age or year of origin	RNA detection method
Ruiz-Vargas et al. 2023	*Pitcairnia spp.*	Plant	Silica gel at RT	3-6 months old	RNA sequencing; assembly; RIN; total RNA
He et al. 2022	Multiple	Plant	Silica dried then frozen	0 to 2 years old	RNA sequencing; assembly; total RNA concentration; total RNA mass; r26S/18S; RNA integrity numbers (RIN); OD
Hamim 2022	Multiple	Plant virus	RNAlater at changing temperatures	2 to 2.5 months old	cDNA synthesis; RNA sequencing, assembly; PCR; gel electrophoresis; spectrophotometry
Mark et al. 2022	Multiple	Plant virus	Herbarium dry at RT	Up to 56 months old	Gel DNA/RNA separation; cDNA synthesis; PCR
Rieux et al. 2021	*African cassava mosaic virus*	Plant virus	Herbarium dry at RT	Originally from 1928	sRNA sequencing
Jimenez et al. 2021	Multiple	Plant virus	Silica gel at RT or cold storage	1.5 to 9 months	RNA concentration; OD260nm / OD280nm; OD260nm /OD230nm; total nucleic acid yield
De Wever et al. 2020	*Theobroma cacao L.*	Plant	Seed at multiple temperatures	0 to 5 weeks old	Qubit RNA HS assay kit; Qubit 2.0 Fluorometer; A260/280; A260/230; fragment analyzer; RQN; RT qPCR
Fleming et al. 2019	Multiple; >40 species	Plant	Seed at multiple temperatures	Originally from 1959 to 2017	Germination assay; RIN; RNA concentration
Fleming et al. 2017	*Glycine max, cv 'Williams 82'*	Plant	Seed	Originally from 1989 to 2015	Germination assay; RIN; fragment size analysis; RNA yield per seed
Mangeot-Peter et al. 2016	*Cannabis sativa*	Plant	Ethanol	1-8 days old	RIN; PCR; Bioanalyzer
Hartung et al. 2015	*Citrus leprosis virus*	Plant virus	Herbarium dry at RT	Originally from 1932 to 1967	sRNA sequencing; assembly; Qubit 2.0 Fluorometer; Bioanalyzer
García-Baldenegro et al. 2015	*Vitis vinifera L. cv. 'Flame Seedless'*	Plant	Freezing then lyophilization	6 months at -80C, then 0-6 weeks of lyophilization	Semi-quantitative and real-time RT-PCR; A260/A280; A260/A230; electrophoresis; Bioanalyzer; cDNA
Smith et al. 2014	*Barley Stripe Mosaic Virus*	Plant virus	Seed	1264 ± 150 CE	sRNA sequencing; assembly
Fordyce et al. 2013	*Zea mays spp. mays)*	Plant	Seed	1290 ± 23 CE	PCR; long and short read sequencing of cDNA; assembly
Sallon et al. 2008	*Phoenix dactylifera*	Plant	Seed	295 ± 47 CE	RAPD
Natarajan et al. 2000	*Solanum tuberosum*	Plant	FTA® Card	5 days old	RT-PCR
Rollo 1985	*Lepidium sativum L.*	Plant	Seed	1400 BCE	spot hybridization

RNA can be obtained from six-month old silica-dried tissue

3a) *Pitcairnia jimenezii*

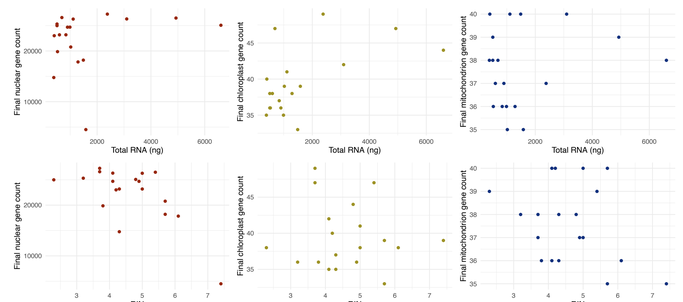
3b) From Ruiz-Vargas et al. 2023 (bioRxiv). Leaf tissue from 19 samples from the genus *Pitcairnia* (family Bromeliaceae) was stored in silica gel for three to six months at room temperature, then RNA was extracted and sequenced.

There is no strong trend associating total RNA (ng) or RIN and final predicted gene count.

Check out the preprint here:

Primary author Natalia Ruiz-Vargas is presenting poster **#162** on Monday, July 24.

3a) The number of recovered genes compared to RNA quality assessment metrics



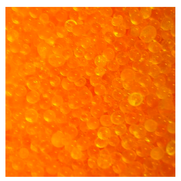
Future Directions

The observed transcriptomic sampling gap shows geographic bias, potentially due to the expenses involved in flash-freezing and long-term cold storage. The use of dry, room temperature plant samples for transcriptomics may greatly expand the number of plant species that can be sequenced.

We may be able to extend our transcriptomic sampling to the wealth of plant samples preserved in herbaria. The concept of herbariomics has thus far only encompassed DNA studies, but RNA from many herbarium samples may still be viable for many uses.

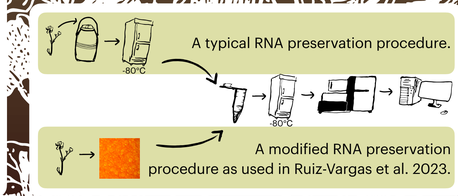
Further, the global climate cost of long-term cold storage should not be understated. Preserving, transporting, and storing plant samples at -80°C is costly for universities and for the environment. Each -80°C freezer consumes about 20 kWh/day, approximately as much energy as an average U.S. home (U.S. Federal Energy Management Program). Even small temperature adjustments from -80°C to -70°C for less sensitive samples can reduce cost and carbon footprint. Given this, it is important to understand the impact preservation methods have on the utility of plant samples.

A thorough investigation of the viability of plant RNA from dried samples has not been performed, although the ability to sample such specimens could greatly increase the number of plants that could be studied using transcriptomics. Silica drying can lower the burden of preservation and may still allow the production of usable transcriptome assemblies (Fig. **3a**; Ruiz-Vargas et al. 2023; He et al. 2022). In the future, we hope to examine the utility of multiple preservation methods for transcriptomics and the effect of the preservation method on gene recovery in an experimental fashion.



4a) Indicator colored silica gel, used for drying plant samples

4b) A gas-powered herbarium drying apparatus. Design of the Herbier de l'Université de Ouagadougou, Burkina Faso.



Contact information

@AlexaTyszka
atyszka2@uic.edu
atyszka.org

Acknowledgements

THE UNIVERSITY OF ILLINOIS AT CHICAGO

This work was supported by startup funds from the University of Illinois at Chicago to JW, as well as National Science Foundation GRFP award 2236870 to AT and IOS 2109716 to DL.

Citations and image credits