Herbaria provide a valuable resource for obtaining informative mRNA

Alexa S. Tyszka[1], Khong-Sam Chia[2], Eric C. Bretz[1], Linda Mansour[1], Drew A. Larson[3], Philip Carella[2] and Joseph F. Walker[1]

1. Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA
2. Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK
3. Department of Biology, Indiana University, Bloomington, IN, 47405, USA

Abstract

While DNA has built the framework for molecular insights from museum collections, the utility of archival RNA remains largely unexplored. Likely a consequence of the known instability of RNA relative to DNA, this has effectively nullified the use of herbaria for transcriptomics. Here, we challenge the assumption that herbaria cannot be used for transcriptomics by assembling transcriptomes from RNA extracted from herbarium specimens. Through systematic comparison of transcriptomes from fresh-collected, silica-dried, and archival specimens, we demonstrate the suitability of herbarium-derived RNA for transcriptomics. The practical applicability of archival mRNA was further illustrated by the functional validation of a plant immune receptor synthesized from a specimen collected in 1956. These results contradict the community consensus regarding archival RNA and open the door to subsequent transcriptomic explorations in rare and extinct species. Our findings highlight the importance of preserving and utilizing the diversity embedded within herbarium collections.

Main

The field of historical (or ancient) DNA has provided insights into human evolution(Green *et al.*, 2010), ancestral population structures of extinct organisms(Jónsson *et al.*, 2014) and selective sweeps that have shaped the genomes of crops(Da Fonseca *et al.*, 2015). Despite mRNA's potential as a historical molecule(Friedländer and Gilbert, 2024; Tyszka, Larson and Walker, 2024), few studies have examined the utility of RNA from archival plant specimens. To date, historical mRNA sequenced from museums has been limited to a single mammalian species(Mármol-Sánchez *et al.*, 2023). RNA from mammalian tissues preserved in permafrost has also proven useful for RNA-seq(Smith *et al.*, 2019). In plants, seed tissue stored at ambient temperatures for years has provided viable mRNA(Rollo, 1985; Fordyce *et al.*, 2013; Smith *et al.*, 2014). The Judean date palm was resurrected by germinating 2000 year old seeds(Sallon *et al.*, 2008; Gros-Balthazard *et al.*, 2021), indicating the presence of viable mRNA. Leaf tissue dried with silica gel, either frozen after drying(He *et al.*, 2022) or left at room temperature for six months(Ruiz-Vargas *et al.*, 2024), has provided transcriptome assemblies comparable to those preserved for mRNA extraction. Although these studies demonstrate that mRNA can be obtained from field-collected samples without standard preservation protocols

over a short timeframe, it remains unclear whether the vast wealth of tissue stored in herbaria are suitable for mRNA transcriptomics. In this Letter, we demonstrate that herbaria are a promising source for obtaining informative mRNA and that plant specimens dating back to 1950 can provide sufficient quality and quantity of RNA for next-generation sequencing and subsequent *in vivo* applications. New dimensions of utility for specimens reinforce the importance of maintaining the herbaria of the world(Davis, 2024).

In this study, we used samples stored under three preservation approaches: RNAlater followed by freezing (i.e., 'Fresh' samples), silica-dried tissue mimicking collection practices (i.e., 'Silica' samples), and herbarium tissue (i.e., 'Herbarium' samples). We used two approaches to explore the utility of RNA obtained from these preservation methods for transcriptomics: (1) quantifying differences in RNA quality between seven silica-dried leaves and corresponding fresh leaves, in six cases collected from the same individual plant, and in all cases at the same time and from the same tissue type, and (2) extracting and sequencing RNA from three herbarium specimens collected between 1995 to 1950 and three additional silica dried samples. We conducted a series of bioinformatics-based quality assessments on all assembled transcriptomes. To complement this and ensure the reliability of our data, we documented nucleotide-binding leucine-rich repeat (NLR) immune receptor genes in all samples and functionally validated an herbarium-derived NLR receptor by heterologous expression in *Nicotiana benthamiana*. For this study, we chose to sample members of the flowering plant order Ericales, a clade of over 13,000 species that contains important crops such as tea, kiwi, and blueberries, ornamentals, parasitic lineages, and carnivorous plants(Rose *et al.*, 2018; Larson *et al.*, 2020; Carruthers *et al.*, 2024).

The RNA integrity (RIN) scores of the silica and the herbarium samples were generally below the recommended level for sequencing (**Figure 1B; Supplementary Table 1**), unlike the fresh samples (**Supplementary Table 2).** Despite the low scores, libraries were successfully constructed, and we obtained the expected number of raw sequence reads. We compared the number of paired and unpaired reads as one proxy for RNA quality because there should be fewer paired reads as sequences degrade. When comparing results from samples collected from the same plant, the silica-dried samples had a higher proportion of unpaired sequence reads after trimming, ranging from 26.49% (*Navarretia*-Silica) to 43.69% (*Collomia*-Silica) compared to their respective fresh sample 15.52% (*Navarretia*-Fresh) to 15.79% (*Collomia*-Fresh) (**Supplementary Table 3**). This indicates a lower average quality of reverse reads for silica-dried samples. Herbarium samples had an even greater proportion of unpaired reads after trimming, with values between 70.08% (*Clavija*-Herbarium) and 79.95% (*Antistrophe*-Herbarium). The loss of the pairing reflects a decrease in RNA quality over time, which was also captured by the shifted read length distribution of the remaining reads that lost their pair (**Supplementary Figure 1, Supplementary Table 4**). This indicates that size selection and library type (paired vs unpaired) are worth accounting for when using older samples.

The quality of the produced transcripts was assessed using the TransRate algorithm(Smith-Unna *et al.*, 2016) as implemented in TransRate2 (**Figure 1C**). Between 65.8% (*Leucothoe*-Fresh) and 79.9% (*Pyrola*-Fresh) of the transcripts from the fresh tissue were considered good quality by this metric. For the silica-dried samples, these numbers varied from 46.4% (*Ipomopsis*-Silica) to 93.7% (*Collomia*-Silica), and for the herbarium samples, they

ranged from 66.3% (*Antistrophe*-Herbarium) to 94.1% (*Clavija*-Herbarium) (**Supplementary Table 5**). Five of the seven fresh samples produced a greater proportion of good quality transcripts than their silica-dried counterpart. The wide range of transcript scores for each treatment indicates that producing good quality transcripts from all three storage methods is possible based on the TransRate metrics.

We examined the inferred transcriptome completeness using Benchmarking Universal Single-Copy Orthologs (BUSCOs). In all comparisons between silica-dried and fresh samples, the silica-dried samples had fewer missing BUSCOs (**Figure 1D; Supplementary Table 6**). The sequencing depth for the silica-dried samples was, on average, much deeper than for fresh-collected samples (**Supplementary Table 3**). Thus, sequencing depth likely explains the greater number of complete BUSCOs recovered from silica-dried samples. These results indicate that single copy orthologs are still identifiable in silica-dried samples, despite being stored at room temperature for five years. For transcriptomes derived from herbarium specimens, we found that the 1950s samples had fewer complete BUSCOs despite similar sequencing depth to the silica-dried samples and higher sequencing depth than the fresh samples. However, the total missing BUSCOs for herbarium samples were still comparable to those of the fresh-collected samples, underlining the importance of sequencing depth when assembling transcriptomes from herbarium specimens.

Next, we used a phylotranscriptomic approach to estimate gene trees that included all samples. We estimated a coalescent based species tree with the branch lengths converted to molecular rates (see methods). The overall topology of the inferred species tree matched the current consensus relationships among species(Larson *et al.*, 2020; Zuntini *et al.*, 2024), and samples collected from the same plant were always sister in the tree (**Figure 1E**). We found no evidence of longer branches subtending silica-dried specimens or herbarium specimens, as is often found for degraded museum samples due to damage to nucleic acids(Derkarabetian, Benavides and Giribet, 2019). Although the assemblies from the same individual were not identical, this is not unexpected because stochastic differences introduced during library prep, sequencing, and assembly processes can result in differences in the final assembled transcript, especially for transcripts containing heterozygous sites. Importantly, this demonstrates that mRNA sequencing of herbarium samples can be used to infer evolutionary relationships among species.

After we confirmed the quality of the assembled transcriptomes, we examined the NLR (Nucleotide-binding and leucine-rich repeat) immune receptor repertoires in transcriptomes from each of the three collection types (i.e., fresh, silica-dried, and herbarium samples). We selected this gene family since (i) NLRs are fast evolving and complex in architecture, (ii) their expression often correlates with function, and (iii) herbarium samples represent an untapped resource to interrogate NLR diversity. Using the NLRtracker algorithm (Kourelis et al., 2021), we first defined the total set of expressed NLRs in transcriptomes from fresh, silica-dried, and herbarium samples (Supporting Dataset), which revealed an overall increase in detectable NLRs in silica-dried samples relative to freshly collected controls for each species (**Figure 2A, Supplementary Table 7**), potentially due to differences in sequencing depth (**Supplementary Table 3**). By comparison, fewer NLR-related genes were detected for archival RNA from the 1950s, though a full-length receptor was recovered (**Figure 2A**). Next, we generated a maximum

likelihood phylogeny using the conserved NB-ARC domain to examine NLR relationships among all samples relative to a reference set of functionally validated receptors(Kourelis *et al.*, 2021). Consistent with known NLR phylogenetic relationships(Chia *et al.*, 2024), we observed three distinct clades corresponding to NLR subtype as defined by their N-terminal RPW8, TIR, or CC domains (**Figure 2B**). Importantly, these analyses revealed overlap between NLRs from each of the three collection types, highlighting the diversity of NLRs present within archival RNA and providing new candidates for functional investigation.

To determine whether NLRs from archival RNA retain functions in immunity, we synthesized two full-length CC-NLRs tagged with 3xFLAG epitopes for comparison against the well-established NbZAR1 CC-NLR from *Nicotiana benthamiana*. The CC-NLRs were DcCNL from a *Darlingtonia californica* sample dried with silica in 2018, and CcCNL from a *Clavija costaricana* herbarium specimen pressed in 1956. To assess their function, we synthesized wild-type versions of each receptor and compared them against variants carrying 'autoactive' D (aspartic acid) to V (valine) mutations in the conserved 'MHD' motif that is known to sensitize CC-NLRs and produce immune related cell death(Adachi *et al.*, 2019). When transiently expressed in *N. benthamiana* leaves, autoactivated NbZAR1^{DV} and CcCNL^{DV} produced appreciable immune-related cell death phenotypes observable under typical light and UV illumination, whereas DcCNL^{DV} and all wild-type (WT) controls were inactive (**Figure 2CD**). Importantly, immunoblotting demonstrated sufficient accumulation of each construct in *N. benthamiana* (**Figure 2E**). Our results demonstrate that the CcCNL receptor sequence, collected in 1956 and archived at the University of Michigan Herbarium (MICH), is indeed viable and supports the idea that herbaria can be used to explore functionally relevant NLR immune receptors.

In summary, our work (1) demonstrates that mRNA may be obtained from pressed herbarium specimens and years-old silica-dried samples; (2) this mRNA may be assembled into reliable transcriptomes that can be used for transcriptomic studies focused on leveraging the sequence data; and (3) using transcriptomes from herbarium specimens, we can predict NLR genes, indicating the continued presence of genetic information including essential variation underlying disease resistance. To date, work in historical mRNA remains almost non-existent across the tree of eukaryotes, with great potential for insights across broad disciplines in biology(Friedländer and Gilbert, 2024; Tyszka, Larson and Walker, 2024). This work further highlights the value of the world's herbaria by explicitly combining the historical collections with the insights of next generation sequencing. Our results emphasize the importance of protecting these invaluable resources as novel uses for them continue to arise.

Materials and Methods

RNA extraction

We obtained mRNA from three herbarium samples (named Genus-Herbarium) and 10 silica-dried samples (named Genus-Silica) packed with an abundance of silica gel and then stored at room temperature for ca. 5 years. Tissue grinding, RNA extractions, and quality assessment were

performed using the same procedure as Ruiz-Vargas *et al.*(Ruiz-Vargas *et al.*, 2024). For two of the samples, the RNA concentrations assessed with a Qubit fluorometer (Thermofisher) were found to be below the minimum concentration needed for analysis on TapeStation (**Supplementary Table 1**). Ribosomal RNAs were removed using the QIAseq FastSelect kit (Qiagen), and the RNAseq libraries were prepared with the Kapa Hyper Stranded mRNA library kit (Roche). Paired-end 150 bp was carried out on a single lane of an Illumina NovaSeq X. Sequencing efforts for these samples were successful despite low RNA integrity (RIN) scores for many silica-dried and herbarium samples (**Figure 1B**) with undetectable mRNA via tape station in two cases (**Supplementary Table 1**). Seven of the silica gel samples had a matching fresh sample (named Genus-Fresh) collected from the same tissue type and species simultaneously. Six of these pairs were from the same individual (**Supplementary Table 1**). The fresh samples were collected in RNAlater followed by storage at -80C; these data were published by Carruthers *et al.*(Carruthers *et al.*, 2024).

Transcriptome Assembly and Quality Assessment

Transcriptomes for all samples, including the fresh samples from Carruthers *et al.*(Carruthers *et al.*, 2024), were assembled using the same procedure. The quality of the raw sequence reads was initially assessed using the program fastqc (**https://www.bioinformatics.babraham.ac.uk/projects/fastqc/**). Single base pair errors were corrected using the program Rcorrector v1.0.4(Song and Florea, 2015), with reads deemed unfixable removed using the unfixable_filter.py script from Morales-Briones *et al.*(Morales-Briones *et al.*, 2021). Trimmomatic v0.39(Bolger, Lohse and Usadel, 2014) was used to remove adapter sequences and trim low-quality reads.

The program Kraken2 v2.1.2(Wood, Lu and Langmead, 2019) was used to classify any material matching the standard protist and fungal database, and any classified material was removed as a contaminant. The same procedure was then used to classify reads that matched the chloroplast and then the mitochondrion database from the program Kakapo(Ramanauskas and Igić, 2023). The remaining reads were classified as nuclear reads.

The program Trinity v2.1.1(Haas *et al.*, 2013) was used to assemble the nuclear sequence reads into transcripts. Chimeric transcripts were detected using the procedure of Yang and Smith(Yang and Smith, 2013) with a database composed of the peptide sequences for *Arabidopsis thaliana* (TAIR10) and *Actinidia chinensis* (Red5_PS1_1.69.0). The abundances of the nuclear reads were then quantified using Salmon v1.3.0(Patro *et al.*, 2017) and clustered into the highest supported transcript using Corset v1.09(Davidson and Oshlack, 2014). The open reading frames were predicted using Transdecoder v5.5.0 **(https://github.com/TransDecoder/TransDecoder)**, with the best supported ORF being retained.

The completeness of the assemblies was assessed using BUSCO v.5.8.2(Manni *et al.*, 2021) using "—mode transcriptomes" and the viridiplantae_odb10 database. The quality assessment of the transcripts was done using TransRate2 (https://github.com/ericbretz/transrate2), an implementation of the TransRate algorithm(Smith-Unna *et al.*, 2016) that utilizes HiSat v2.2.1(Kim *et al.*, 2019). We used HiSat as the alignment program as it is specifically designed to be splice aware and capable of handling complex eukaryotic genomes.

Phylogenetic Analysis

We conducted a phylogenomic analysis by identifying homologous gene clusters from the assembled transcriptomes of fresh, silica-dried, and herbarium samples. The phylogenomic analysis also included a previously assembled transcriptome of *Cornus florida* (NCBI RefSeq assembly: GCF_030987335.1) Cd-hit v4.8.1(Fu *et al.*, 2012) with a sequence identity threshold of 0.99 was used to reduce redundancy in the transcriptomes. An all-by-all blast(Altschul, *et al.*, 1990) search on the transcriptomes was conducted for initial homology detection, with an e-value of 10 and maximum target sequences set to 1000. The homolog clusters, as detected by blast, were then further refined using mcl v14-137(Van Dongen, 2008), with an inflation value of 1.4. All remaining homolog clusters with at least 10 taxa were retained for phylogenomic analysis.

The putative homologs were aligned using MAFFT v7.490(Katoh and Standley, 2013) with the –auto and –maxiterate 1000 settings. The phyx program pxclsq v1.2(Brown, Walker and Smith, 2017) was used to remove all alignment columns with less than 10% occupancy. Maximum likelihood, as implemented in IQ-TREE v1.6.12(Nguyen *et al.*, 2015), with the GTR model of evolution and gamma rate variation, was used to infer a phylogenetic tree for each homolog. From the inferred trees, all tips that had branch lengths of 0.2 subs/bp or were 10x longer than their sister or 0.3 subs/bp absolute value were removed, and all clades in trees that consisted of gene sequences from a single sample were collapsed to retain only the one with the most information in the alignment. The refined homologs were then run through the same procedure three more times to further refine the phylogenetic dataset.

Orthologous groups were extracted from the homolog trees using the maximum inclusion method(Yang and Smith, 2013), requiring a minimum of 10 taxa and removing any branch lengths with a relative value of 0.2 subs/bp or an absolute value of 0.3 subs/bp. The sequences corresponding to the extracted ortholog trees were then aligned using MAFFT, cleaned using pxclsq, and the gene trees were inferred using IQ-TREE with the same settings used in the homology procedure. A species tree was then inferred using the maximum quartet support species tree method implemented in ASTRAL v5.7.8(Zhang *et al.*, 2018). The coalescent units were converted to subs/bp using the median branch length value of all concordant branches in the gene trees as implemented in the program BES(Walker *et al.*, 2022).

NLR identification and phylogenetic analysis

NLR identification was performed using the default settings of the NLRtracker pipeline, which is available online (https://github.com/slt666666/NLRtracker). Phylogenetic analysis was performed by comparing the central NB-ARC domains of fresh, silica-dried, and herbarium samples against a reference set of flowering plant NLRs(Kourelis *et al.*, 2021). Isolated domains were aligned using MAFFT and trimmed using trimAl(Capella-Gutiérrez, Silla-Martínez and Gabaldón, 2009). The resulting alignments were used for maximum likelihood phylogenetic

analysis via IQtree v2.0.3 with the "JTT+F+G4" model selected through ModelFinder(Kalyaanamoorthy *et al.*, 2017) and with 1000 ultrafast bootstrap replicates(Hoang *et al.*, 2018). The phylogenetic tree was subsequently rendered using TVbot(Xie *et al.*, 2023).

Plant growth conditions

N. benthamiana plants were grown in soil under controlled conditions, with a long-day photoperiod (16 h light; 160–200 µE, using Sylvania F58W/GRO fluorescent lighting) at 22°C.

Cloning, immunoblotting, and cell death assays

Constructs for cell death assays were assembled using the Modular Cloning (MoClo) system with the following components: pICH85281 (mannopine synthase promoter + W [MasWpro], Addgene #50272), pICSL50007 (AddGene #50308), pICSL60008 (Arabidopsis heat shock protein terminator [HSPter], TSL SynBio), binary vector pICH47742 (Addgene #48001) and the appropriate synthesized NLR gene or variant. All resulting constructs were transformed into *Agrobacterium tumefaciens* GV3101 (pMP90) by electroporation. Cell death assays were performed by Agrobacterium-mediated transient expression, where fully expanded leaves of 4-week-old *N. benthamiana* plants were infiltrated with *A. tumefaciens* strains containing a binary expression plasmid mixed at a 1:1 ratio with a suspension of *A. tumefaciens* containing the p19 silencing suppressor. Bacterial cells were resuspended in a fresh infiltration buffer (10 mM MES-KOH, 10 mM $MgCl_2$, and 150 mM acetosyringone, pH 5.6) and adjusted to an OD600 of 0.3. Immune-related cell death phenotypes were assessed using an HR index ranging from 0 (no visible symptoms) to 7 (fully confluent cell death). For immunoblotting, protein samples were prepared from six *N. benthamiana* leaf discs (8mm diameters) at 2 d after agroinfiltration. Leaf samples were then ground to a fine powder and homogenized in 150 ul of 2x LDS sample buffer (NuPAGE LDS Sample Buffer). Immunoblotting was performed with anti-flag antibody coupled with HRP (Sigma-Aldrich A8592, 1:5,000 dilution), and total protein loading was visualized with Ponceau S solution (Sigma-Aldrich, P7170).

Acknowledgements

The authors thank the UIC Sequencing Core, Zarema Arbieva and Nina S. Los, for library prep and quality assessment. Tom Carruthers for assistance organizing the sequence data for the fresh tissue samples. Jim Belsher-Howe and Plumas National Forest staff for logistical support and permission to collect specimens. Steve Kelly about the function of TransRate and Nathanael Walker-Hale for helpful discussion. A.S.T. is supported by the National Science Foundation (NSF) grant GRFP-2236870. D.A.L. is supported by NSF IOS-2109716. J.F.W. acknowledges start-up funding support from the University of Illinois Chicago. P.C. is supported by UKRI (UK Research and Innovation) Biotechnology and Biological Sciences Research Council (BBSRC) core funding [BB/X010996/1]. All authors thank the University of Michigan Herbarium for providing access to archival specimens.

Data Availability

All data involved in the conclusions of this Letter are present in the paper, supplementary material, or available publicly. The raw sequence reads for samples sequenced in this study are available on the NCBI Sequence Read Archive (PRJNA1200749). Voucher information is available in Supplementary Table 8. Analysis outputs for BUSCO and transRate2 are available as Supplementary Information Tables 4-5. Assembled sequences, trees, and NLRtracker outputs are available via https://zenodo.org/records/14720388.

Citations

Adachi, H. *et al.* (2019) 'An N-terminal motif in NLR immune receptors is functionally conserved across distantly related plant species', *eLife*, 8, p. e49956. Available at: https://doi.org/10.7554/eLife.49956.

Altschul, S. *et al.* (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, (3), pp. 403–410. Available at: https://doi.org/10.1016/S0022-2836(05)80360-2.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. Available at: https://doi.org/10.1093/bioinformatics/btu170.

Brown, J.W., Walker, J.F. and Smith, S.A. (2017) 'Phyx: phylogenetic tools for unix', *Bioinformatics*. Edited by J. Kelso, 33(12), pp. 1886–1888. Available at: https://doi.org/10.1093/bioinformatics/btx063.

Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. (2009) 'trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses', *Bioinformatics*, 25(15), pp. 1972–1973. Available at: https://doi.org/10.1093/bioinformatics/btp348.

Carruthers, T. *et al.* (2024) 'Repeated shifts out of tropical climates preceded by whole genome duplication', *New Phytologist*, 244(6), pp. 2561–2575. Available at: https://doi.org/10.1111/nph.20200.

Chia, K.-S. *et al.* (2024) 'The N-terminal domains of NLR immune receptors exhibit structural and functional similarities across divergent plant lineages', *The Plant Cell*, 36(7), pp. 2491–2511. Available at: https://doi.org/10.1093/plcell/koae113.

Da Fonseca, R.R. *et al.* (2015) 'The origin and evolution of maize in the Southwestern United States', *Nature Plants*, 1(1), p. 14003. Available at: https://doi.org/10.1038/nplants.2014.3.

Davidson, N.M. and Oshlack, A. (2014) 'Corset: enabling differential gene expression analysis for de novoassembled transcriptomes', *Genome Biology*, 15(7), p. 410. Available at: https://doi.org/10.1186/s13059-014-0410-6.

Davis, C.C. (2024) ‘Collections are truly priceless’, *Science*, 383(6687), pp. 1035–1035. Available at: https://doi.org/10.1126/science.ado9732.

Derkarabetian, S., Benavides, L.R. and Giribet, G. (2019) ‘Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault’, *Molecular Ecology Resources*, 19(6), pp. 1531–1544. Available at: https://doi.org/10.1111/1755-0998.13072.

Fordyce, S.L. *et al.* (2013) ‘Deep Sequencing of RNA from Ancient Maize Kernels’, *PLoS ONE*. Edited by D.Q. Fuller, 8(1), p. e50961. Available at: https://doi.org/10.1371/journal.pone.0050961.

Friedländer, M.R. and Gilbert, M.T.P. (2024) ‘How ancient RNA survives and what we can learn from it’, *Nature Reviews Molecular Cell Biology*, 25(6), pp. 417–418. Available at: https://doi.org/10.1038/s41580-024-00726-y.

Fu, L. *et al.* (2012) ‘CD-HIT: accelerated for clustering the next-generation sequencing data’, *Bioinformatics*, 28(23), pp. 3150–3152. Available at: https://doi.org/10.1093/bioinformatics/bts565.

Green, R.E. *et al.* (2010) ‘A Draft Sequence of the Neandertal Genome’, *Science*, 328(5979), pp. 710–722. Available at: https://doi.org/10.1126/science.1188021.

Gros-Balthazard, M. *et al.* (2021) ‘The genomes of ancient date palms germinated from 2,000 y old seeds’, *Proceedings of the National Academy of Sciences*, 118(19), p. e2025337118. Available at: https://doi.org/10.1073/pnas.2025337118.

Haas, B.J. *et al.* (2013) ‘De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis’, *Nature Protocols*, 8(8), pp. 1494–1512. Available at: https://doi.org/10.1038/nprot.2013.084.

He, J. *et al.* (2022) ‘A phylotranscriptome study using silica gel-dried leaf tissues produces an updated robust phylogeny of Ranunculaceae’, *Molecular Phylogenetics and Evolution*, 174, p. 107545. Available at: https://doi.org/10.1016/j.ympev.2022.107545.

Hoang, D.T. *et al.* (2018) ‘UFBoot2: Improving the Ultrafast Bootstrap Approximation’, *Molecular Biology and Evolution*, 35(2), pp. 518–522. Available at: https://doi.org/10.1093/molbev/msx281.

Jónsson, H. *et al.* (2014) ‘Speciation with gene flow in equids despite extensive chromosomal plasticity’, *Proceedings of the National Academy of Sciences*, 111(52), pp. 18655–18660. Available at: https://doi.org/10.1073/pnas.1412627111.

Kalyaanamoorthy, S. *et al.* (2017) ‘ModelFinder: fast model selection for accurate phylogenetic estimates’, *Nature Methods*, 14(6), pp. 587–589. Available at: https://doi.org/10.1038/nmeth.4285.

Katoh, K. and Standley, D.M. (2013) ‘MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability’, *Molecular Biology and Evolution*, 30(4), pp. 772–780. Available at: https://doi.org/10.1093/molbev/mst010.

Kim, D. *et al.* (2019) ‘Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype’, *Nature Biotechnology*, 37(8), pp. 907–915. Available at: https://doi.org/10.1038/s41587-019-0201-4.

Kourelis, J. *et al.* (2021) ‘RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family’, *PLOS Biology*. Edited by X. Dong, 19(10), p. e3001124. Available at: https://doi.org/10.1371/journal.pbio.3001124.

Larson, D.A. *et al.* (2020) ‘A consensus phylogenomic approach highlights paleopolyploid and rapid radiation in the history of Ericales’, *American Journal of Botany*, 107(5), pp. 773–789. Available at: https://doi.org/10.1002/ajb2.1469.

Manni, M. *et al.* (2021) ‘BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes’, *Molecular Biology and Evolution*. Edited by J. Kelley, 38(10), pp. 4647–4654. Available at: https://doi.org/10.1093/molbev/msab199.

Mármol-Sánchez, E. *et al.* (2023) ‘Historical RNA expression profiles from the extinct Tasmanian tiger’, *Genome Research*, 33(8), pp. 1299–1316. Available at: https://doi.org/10.1101/gr.277663.123.

Morales-Briones, D.F. *et al.* (2021) ‘Disentangling Sources of Gene Tree Discordance in Phylogenomic Data Sets: Testing Ancient Hybridizations in Amaranthaceae s.l’, *Systematic Biology*. Edited by R. Ree, 70(2), pp. 219–235. Available at: https://doi.org/10.1093/sysbio/syaa066.

Nguyen, L.-T. *et al.* (2015) ‘IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies’, *Molecular Biology and Evolution*, 32(1), pp. 268–274. Available at: https://doi.org/10.1093/molbev/msu300.

Patro, R. *et al.* (2017) ‘Salmon provides fast and bias-aware quantification of transcript expression’, *Nature Methods*, 14(4), pp. 417–419. Available at: https://doi.org/10.1038/nmeth.4197.

Ramanauskas, K. and Igić, B. (2023) ‘kakapo : easy extraction and annotation of genes from raw RNA-seq reads’, *PeerJ*, 11, p. e16456. Available at: https://doi.org/10.7717/peerj.16456.

Rollo, F. (1985) ‘Characterisation by molecular hybridization of RNA fragments isolated from ancient (1400 B.C.) seeds’, *Theoretical and Applied Genetics*, 71(2), pp. 330–333. Available at: https://doi.org/10.1007/BF00252076.

Rose, J.P. *et al.* (2018) ‘Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections’, *Molecular*

Phylogenetics and Evolution, 122, pp. 59–79. Available at: https://doi.org/10.1016/j.ympev.2018.01.014.

Ruiz-Vargas, N. *et al.* (2024) 'Transcriptome data from silica-preserved leaf tissue reveal gene flow patterns in a Caribbean bromeliad', *Annals of Botany*, 133(3), pp. 459–472. Available at: https://doi.org/10.1093/aob/mcae002.

Sallon, S. *et al.* (2008) 'Germination, Genetics, and Growth of an Ancient Date Seed', *Science*, 320(5882), pp. 1464–1464. Available at: https://doi.org/10.1126/science.1153600.

Smith, O. *et al.* (2014) 'Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process', *Scientific Reports*, 4(1), p. 5559. Available at: https://doi.org/10.1038/srep05559.

Smith, O. *et al.* (2019) 'Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival', *PLOS Biology*. Edited by C. Tyler-Smith, 17(7), p. e3000166. Available at: https://doi.org/10.1371/journal.pbio.3000166.

Smith-Unna, R. *et al.* (2016) 'TransRate: reference-free quality assessment of de novo transcriptome assemblies', *Genome Research*, 26(8), pp. 1134–1144. Available at: https://doi.org/10.1101/gr.196469.115.

Song, L. and Florea, L. (2015) 'Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads', *Gigascience*, 4(1), pp. s13742-015-0089-y. Available at: https://doi.org/10.1186/s13742-015-0089-y.

Tyszka, A.S., Larson, D.A. and Walker, J.F. (2024) 'Sequencing historical RNA: unrealized potential to increase understanding of the plant tree of life', *Trends in Plant Science*, p. S1360138524003054. Available at: https://doi.org/10.1016/j.tplants.2024.11.004.

Van Dongen, S. (2008) 'Graph Clustering Via a Discrete Uncoupling Process', *SIAM Journal on Matrix Analysis and Applications*, 30(1), pp. 121–141. Available at: https://doi.org/10.1137/040608635.

Walker, J.F. *et al.* (2022) 'Concordance-Based Approaches for the Inference of Relationships and Molecular Rates with Phylogenomic Data Sets', *Systematic Biology*. Edited by R. Thomson, 71(4), pp. 943–958. Available at: https://doi.org/10.1093/sysbio/syab052.

Wood, D.E., Lu, J. and Langmead, B. (2019) 'Improved metagenomic analysis with Kraken 2', *Genome Biology*, 20(1), p. 257. Available at: https://doi.org/10.1186/s13059-019-1891-0.

Xie, J. *et al.* (2023) 'Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees', *Nucleic Acids Research*, 51(W1), pp. W587–W592. Available at: https://doi.org/10.1093/nar/gkad359.

Yang, Y. and Smith, S.A. (2013) 'Optimizing de novo assembly of short-read RNA-seq data for phylogenomics', *BMC Genomics*, 14(1), p. 328. Available at: https://doi.org/10.1186/1471-2164-14-328.

Zhang, C. *et al.* (2018) ‘ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees’, *BMC Bioinformatics*, 19(S6), p. 153. Available at: https://doi.org/10.1186/s12859-018-2129-y.

Zuntini, A.R. *et al.* (2024) ‘Phylogenomics and the rise of the angiosperms’, *Nature*, 629(8013), pp. 843–850. Available at: https://doi.org/10.1038/s41586-024-07324-0.

Figures

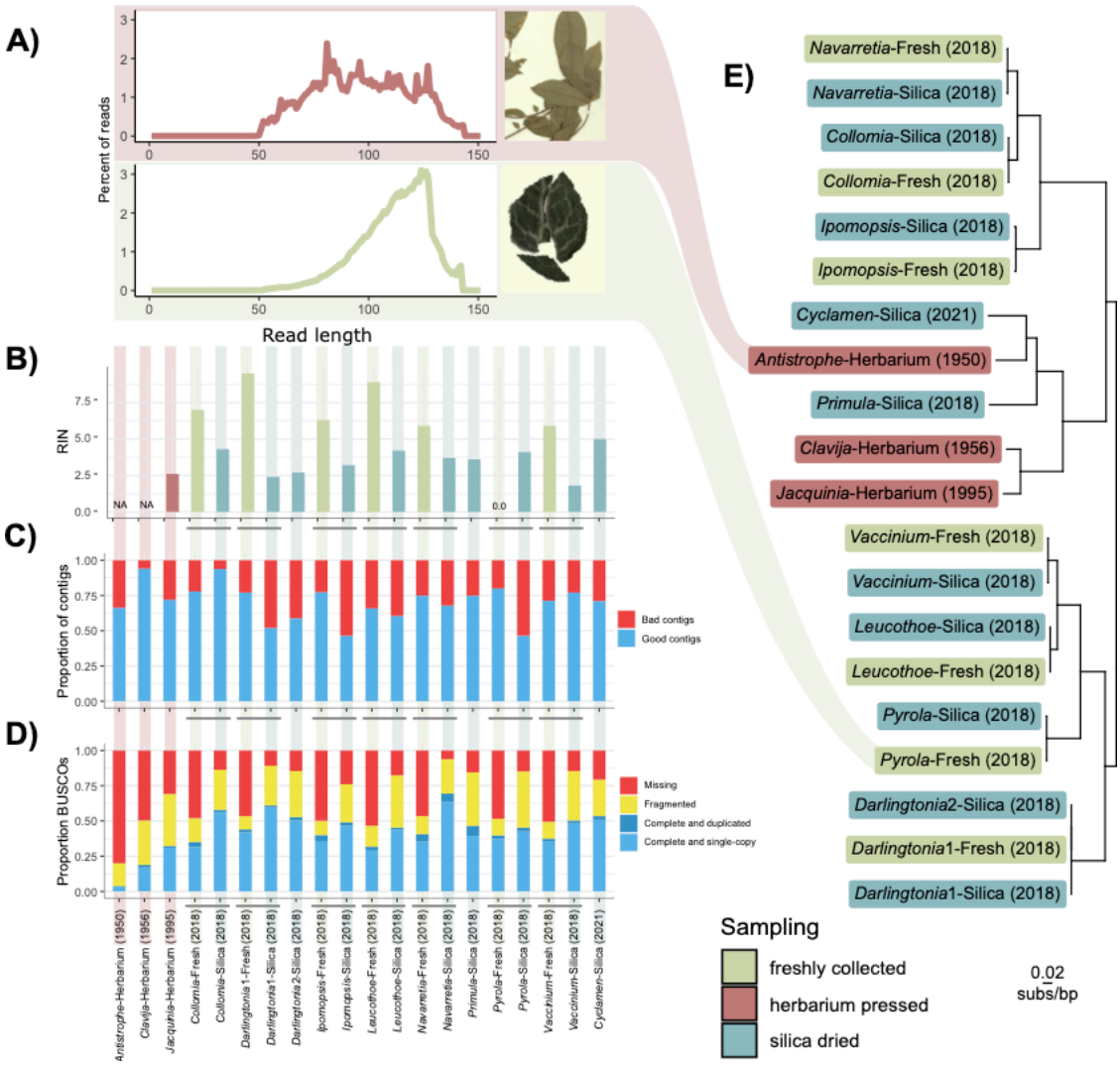

Figure 1. Extracted RNA from fresh, silica-dried, and herbarium specimens provide usable transcriptomes based on bioinformatic metrics. A) The distribution of forward only sequence read length for Antistrophe-Herbarium (1950) and Pyrola-Fresh (typical storage conditions) after Trimmomatic quality filtering. B) The RNA integrity (RIN) scores for samples used in this study. C) The proportion of contigs inferred to be "good contigs" compared to the proportion inferred as "bad contigs" based upon the TransRate metrics. D) The inferred completeness of the transcriptomes based upon sequenced single copy orthologs. E) The coalescent based inferred phylogenetic relationships and sample source for transcriptomes in this study. Branch lengths converted to molecular rates (subs/bp).

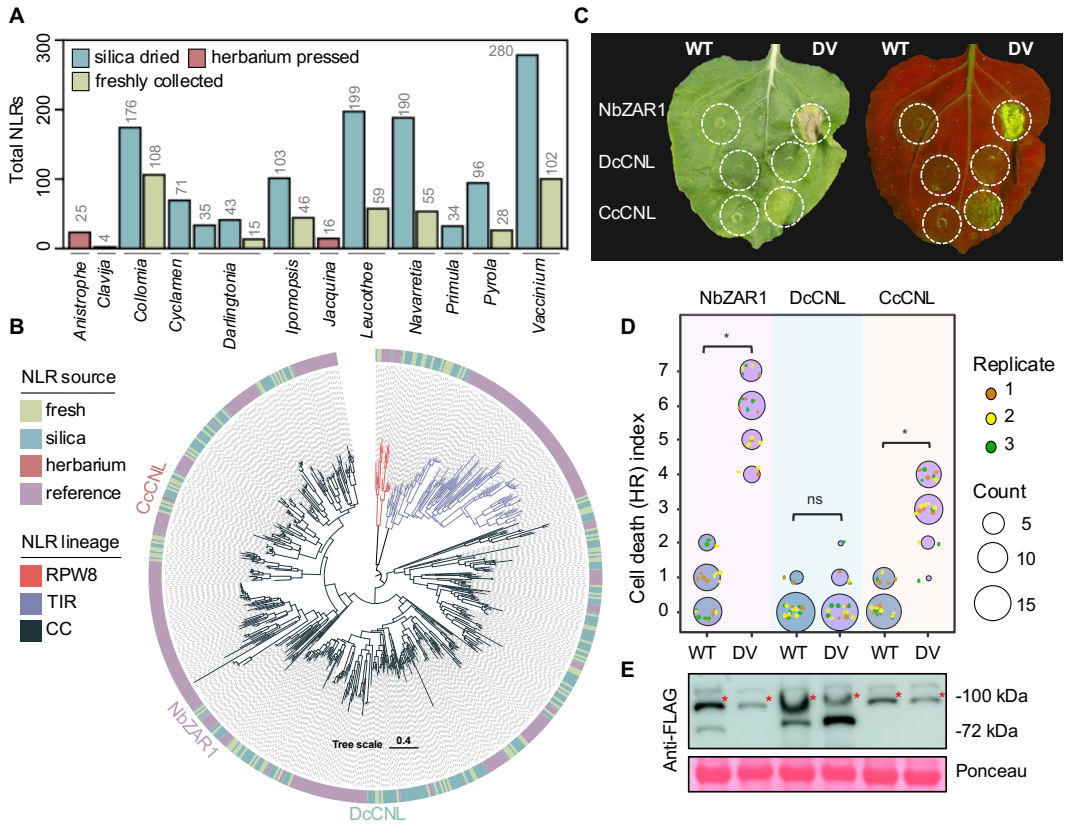

Figure 2. An NLR immune receptor sourced from a 68-year-old herbarium specimen of *Clavija costaricana* is functionally transferable to *Nicotiana benthamiana*. (A) Total number of NLRs identified in each transcriptome sample. (B) Maximum likelihood phylogeny of NLR immune receptors based on the central NB-ARC regulatory domain. The outer ring indicates whether each NLR was sourced from herbaria samples, silica samples, fresh samples, or belongs to a reference set of functional NLRs from flowering plants. Branch color represents NLR lineage; RPW8-NLRs (RPW8), TIR-NLRs (TIR), or CC-NLRs (CC). The three tested CC-NLRs (CNLs) are indicated on the tree. Tree scale = substitutions/site. (C) HR cell death phenotypes induced by wild-type (WT) or autoactive (DV) variants of DcCNL-3xFLAG, CcCNL-3xFLAG, or the NbZAR1-3xFLAG control in *Nicotiana benthamiana*. Leaves were imaged 7 days post agroinfiltration under bright light as well as UV light, where cell death responses appear as green fluorescence. (D) Quantification of HR cell death performed 7 d post agroinfiltration. Data from 3 independent experimental replicates are shown ($n \geq 5$ infiltrations per replicate). Statistically significant differences between the means of WT and DV treatments are indicated with an asterisk * (Kruskal Wallis test, $p < 0.01$; ns = not significant). (E) Immunoblots of all tested NLRs 2 d post agroinfiltration. Total protein levels were visualized with Ponceau staining. Immunoblotting was repeated twice with similar results. Red asterisks (*) indicate proteins of interest

Supplementary Figures and Tables

Supplementary Figure 1. The post quality control with Trimmomatic read length distribution for the forward only reads.

Supplementary Table 1. Table of newly sequenced samples (silica-dried and herbarium) and their respective quality scores as measured by NanoDrop, Qubit and TapeStation.

Supplementary Table 2. Table of samples used in this study previously published in Carruthers *et al.*(Carruthers *et al.*, 2024) (fresh), with their respective quality scores as measured by NanoDrop, Qubit and TapeStation.

Supplementary Table 3. The number of raw reads remaining and classified at each step of the assembly pipeline.

Supplementary Table 4. The percentage of each read length remaining for the forward only reads (those that lost their pair) after the Trimmomatic quality control step.

Supplementary Table 5. The TransRate Scores as calculated by TransRate2 for all transcriptomes assembled in the study.

Supplementary Table 6. The Benchmarking Universal Single Copy Orthologs (BUSCO) scores for each sample in the study.

Supplementary Table 7. The NLR's identified using NLR tracker on each of the assemblies.

Supplementary Table 8. The voucher information for each sample included as part of this study.