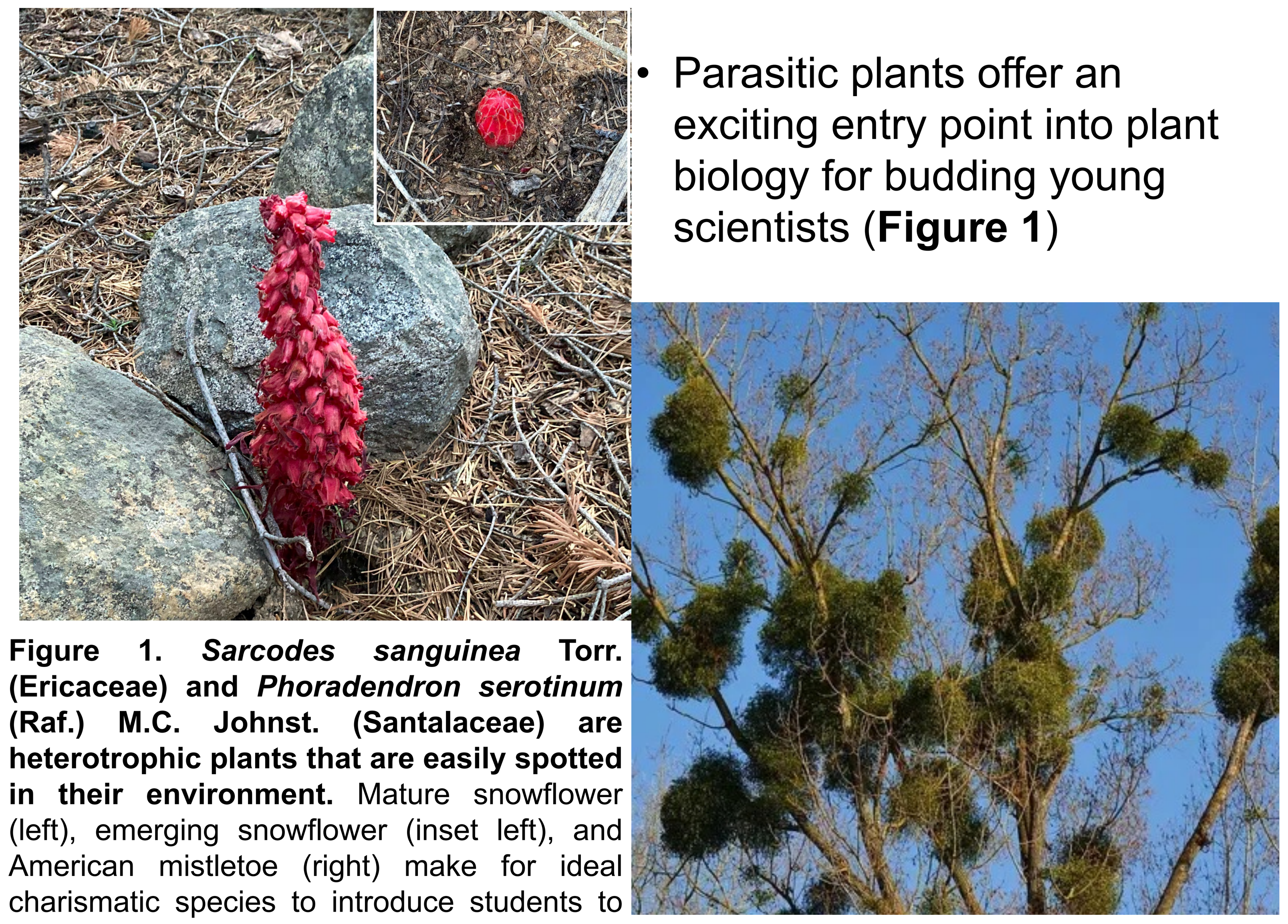
Don't know what you got 'til it's gone – plastid genome evolution in non-photosynthetic snowflower (Ericaceae)

Joel Sharbrough[1], 2022-2024 NTHS AP Biology Class[2], David L. Steakley[2]

[1] – New Mexico Tech, Socorro, NM; [2] – North Tahoe High School, Tahoe City, CA

Nanopore sequencing of charismatic plants to broaden participation in plant biology



- Parasitic plants offer an exciting entry point into plant biology for budding young scientists (**Figure 1**)

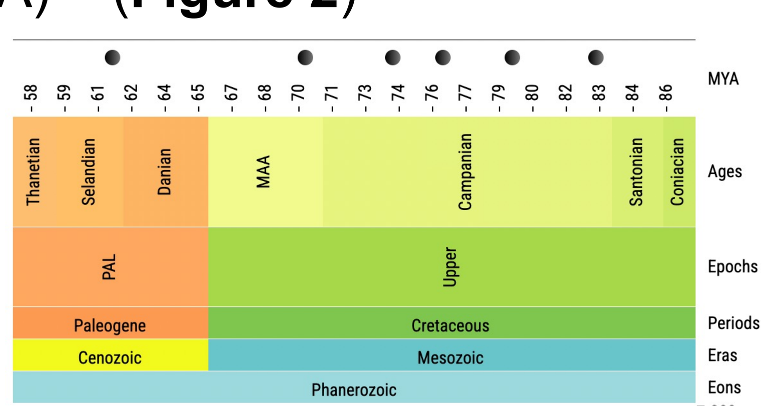
Figure 1. *Sarcodes sanguinea* Torr. (Ericaceae) and *Phoradendron serotinum* (Raf.) M.C. Johnst. (Santalaceae) are heterotrophic plants that are easily spotted in their environment. Mature snowflower (left), emerging snowflower (inset left), and American mistletoe (right) make for ideal charismatic species to introduce students to plant biology and the DNA Sequencing Revolution. Snowflower photo credits: J Sharbrough. Mistletoe photo credit: Pixabay.com

- We have developed workshops to introduce high-school students to plant biology and the DNA sequencing revolution centered on local, charismatic species
- Workshops are tailored to fit local organizations' needs, including a 2-hour sequencing demonstration (Albuquerque Botanic Gardens) and a week-long hands on wet- and dry-lab (North Tahoe High School)

Testing for degradation of the plastid genome following loss of photosynthesis in snowflower

- Snowflower is a mycotrophic plant nested within the entirely non-photosynthetic Monotropoideae, which is estimated to have diverged ~75 MYA (61-83 MYA)[1-6] (**Figure 2**)

Figure 2. Time-tree estimate of divergence between snowflower and it's closest photosynthetic relative. Black circles represent time estimates (in MYA) of divergence between *Sarcodes sanguinea* and *Arbutus unedo* from various studies[1-6].



- We therefore **hypothesized** **that after losing photosynthesis, the snowflower plastid genome has experienced extensive degradation and pseudogenization**

Sequencing snowflower genomes in the classroom

- In spring 2023, NTHS AP Biology students, extracted total cellular DNA from a snowflower collected from Burton Creek (Tahoe City, CA) using a Qiagen DNEasy Plant Mini kit
- Students then prepared barcoded Oxford Nanopore libraries in groups of four using the rapid barcoding library prep
- Libraries were sequenced on an Oxford Nanopore MinION with an R9.4.1 flow cell live-and-in-person at NTHS
- Additional HMW DNA was extracted in the Sharbrough Lab at NMT using the Circulomics Plant Nuclei Kit, and an additional ~3 Gbp of DNA (~0.1x nuclear coverage) was sequenced on a MinION on another R9.4.1 flow cell
- Sequencing data were basecalled using Guppy (v. 6.4.6+ae70e8f) with the "Super Accurate" base-calling model (dna_r9.4.1_450bps_sup.cfg), and assembled *de novo* using Flye (v. 2.9.4-b1799) using default parameters (**Figure 3**)

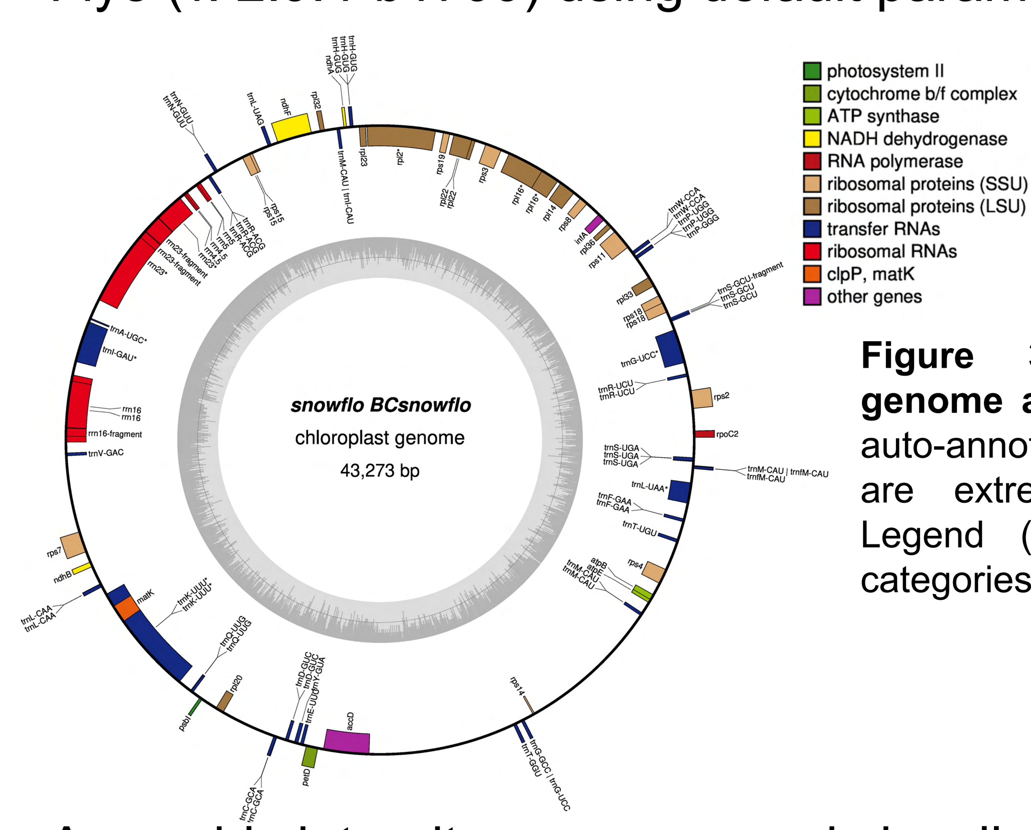
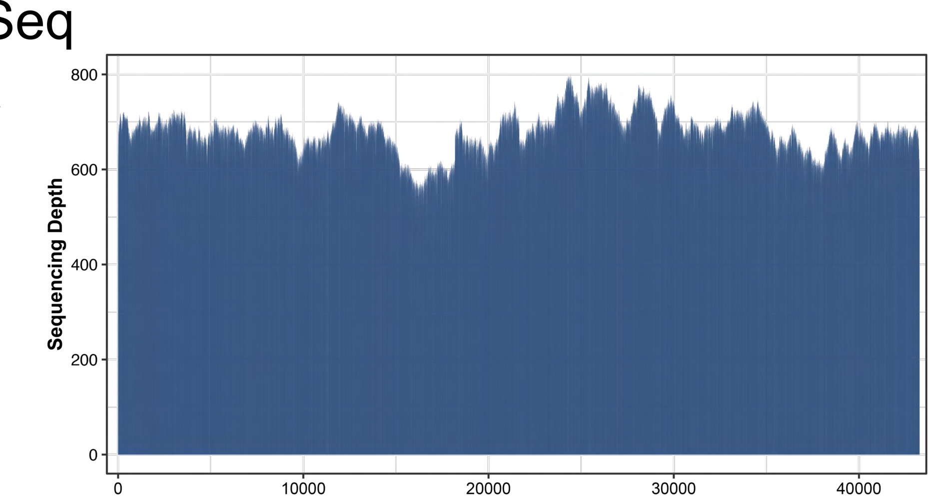


Figure 3. Snowflower plastid genome assembly. Boxes represent auto-annotated genes, many of which are extremely short (e.g., *psbI*). Legend (top right) lists all gene categories found.

- Assembly integrity was assessed visually using read coverage using mosdepth (v. 0.3.0) (**Figure 4**), and automatic annotations were performed in GeSeq

Figure 4. Coverage plot of nanopore reads across the assembled plastid genome of snowflower. Blue columns represent per-base coverage using all reads mapped against the plastid genome. **Mean depth was ~665x.**



- In spring 2024, NTHS AP Biology students manually annotated the assembled plastid genome, identified all missing genes and genes with premature stop codons present in their coding regions using tBLASTx against the *Vaccinium macrocarpa*, *Rhododendron henanense*, and *Arabidopsis thaliana* plastid-encoded proteins as queries

Snowflower plastid genome is massively reduced compared to photosynthetic relatives

- Snowflowert ptDNA is only ~43.3 kb in length (**-79.2% compared to Rhododendron**), with only 18 intact genes (**-79.1%**)
- Nearly all intact genes related to translation (e.g., ribosomal proteins, initiation factor) (**Table 1**)
- All photosynthesis genes have been lost or pseudogenized
- An additional 8 pseudogenes were found, each of which had many premature stop codons
- The inverted repeat has been entirely lost, and the genome is highly rearranged compared to Rhododendron (**Figure 5**)

Table 1. Plastid gene functional category predictions and presence/absence in snowflower ptDNA.

Plastid Enzyme Complex	Biological Function	Prediction?	Present?	Functional?
ACCase	Fatty acid biosynthesis in chloroplast	Maintained	1/1	Appears to be functional
ATP Synthase	Generate ATP from ADP	Not maintained	3/6	No – deleted/ pseudogenes
Chloroplast Envelope Membrane Protein	Contributes to proton efflux to modulate photosynthesis	Not maintained	0/1	No – deleted
CLP Protease	Protein homeostasis in chloroplast	Maintained	1/1	Appears to be functional
Cytochrome b6-f complex	Electron transport during photosynthesis	Not maintained	1/6	No – deleted/ pseudogene
Cytochrome c Biogenesis Protein	Cytochrome c biogenesis	Not maintained	0/1	No – deleted
Maturase K	RNA Maturation	Maintained	1/1	Appears to be functional
NADH Dehydrogenase	Electron transport, redox homeostasis	Maintained	3/11	No – deleted/ pseudogene
Photosystem I	Photosynthesis	Not maintained	0/5	No – deleted
Photosystem II	Photosynthesis	Not maintained	0/15	No – deleted
Ribosome; Large Subunit	Translation	Maintained	8/9	Appears to be functional
Ribosome; Small Subunit	Translation	Maintained	1/1	Appears to be functional
RNA Polymerase	Transcription	Maintained	1/4	No – deleted/ pseudogene
Rubisco	Carbon fixation	not maintained	0/1	No – deleted
Translation Initiation Factor	Translation	maintained	1/1	Appears to be functional
ycf genes	Unknown	maintained	0/5	No – deleted

Green shading – some/most genes present and intact
Yellow shading – genes can be found, but none are functional and if present have premature stop codons
Red shading – no clear evidence of gene relics remains present

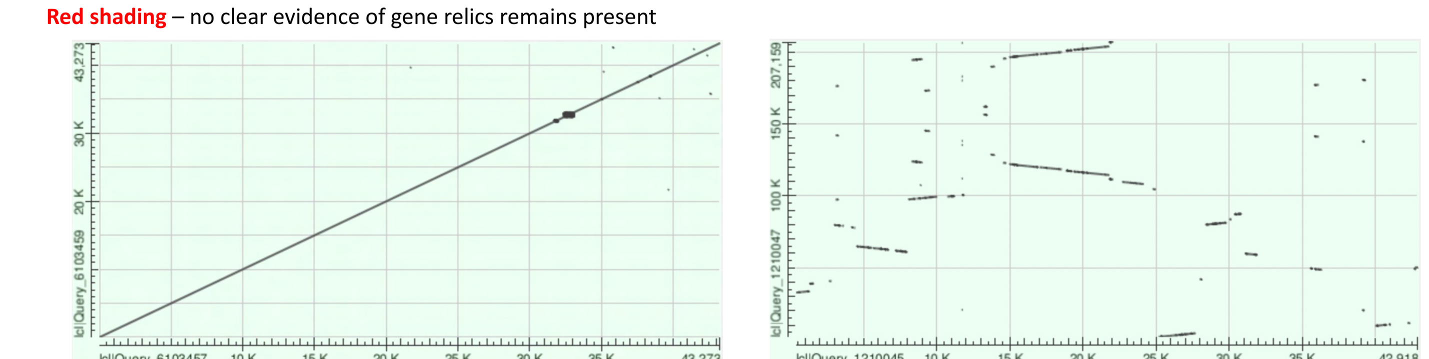


Figure 5. Dotplot of snowflower and Rhododendron plastomes. Left: Snowflower aligned to itself. No evidence of an inverted repeat was detected. Right: Snowflower (x-axis) aligned to Rhododendron (y-axis). The snowflower plastome is highly rearranged and reduced.

Summary & Conclusions

- The loss of photosynthesis has dramatically affected snowflower ptDNA evolution
- Translation genes are nearly all that remains, but pseudogenes persist
- Plastid genome maintained at extremely high copy number (**~10-12k copies per cell!**)
- Successful deployment of plant genome sequencing module has permeated into the school's art class (right image)

Photo credit: NTHS Art Class, Ms. Jo Anna Battaglia

Acknowledgements: This work was supported by funding from the National Science Foundation (IOS-2145811). Some data presented here utilized the RMACC Alpine supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University.

References:
1. Elliott & Davies. 2019. *Biodiv Conserv*, *28*: 711-728.
2. Ramírez-Barahona et al. 2020. *Nat Ecol Evol*, *4*: 1232-1238.
3. Li et al. 2019. *Nat Plants*, *5*: 461-470.
4. Rose et al. 2018. *Mol Phylogenet Evol*, *122*: 59-79.
5. Liu et al. 2014. *Ann Bot*, *114*: 1701-1709.
6. Bouchenak-Khelladi et al. 2015. *New Phytol*, *207*: 313-326.