Massive Transcriptome Changes during Leaf Senescence in Field Grown *Populus trichocarpa* Nisqually-1 using KBase Tools

Haiwei Lu and Steve Strauss

Oregon State University

BER Plant Genomic Science Workshop, PAG XXVII

January 14, 2019

Oregon State University

The concept of leaf senescence

- The final developmental stage before leaf death

- A slow process during which nutrients are remobilized into seeds of annual plants or bark and other tissues of perennial plants
  - An actively and highly regulated process
  - Essential to plant reproduction and survival

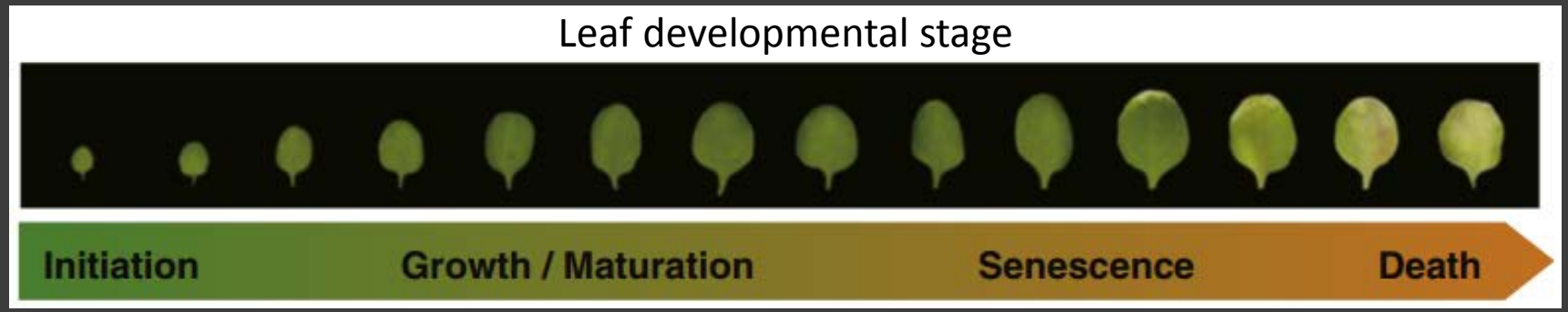
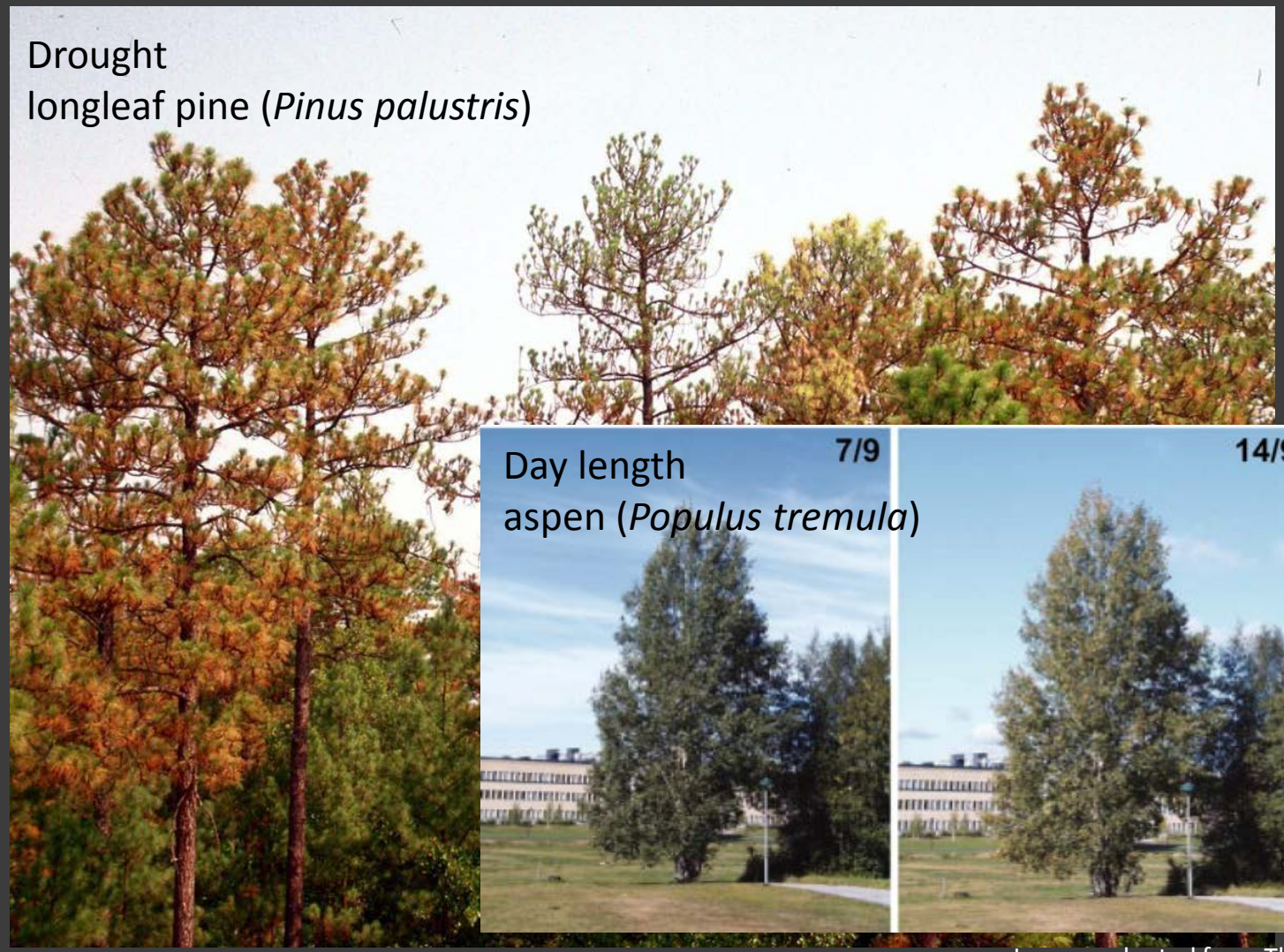


Image adapted from Kim et al. *Current Opinion in Plant Biology* (2016)

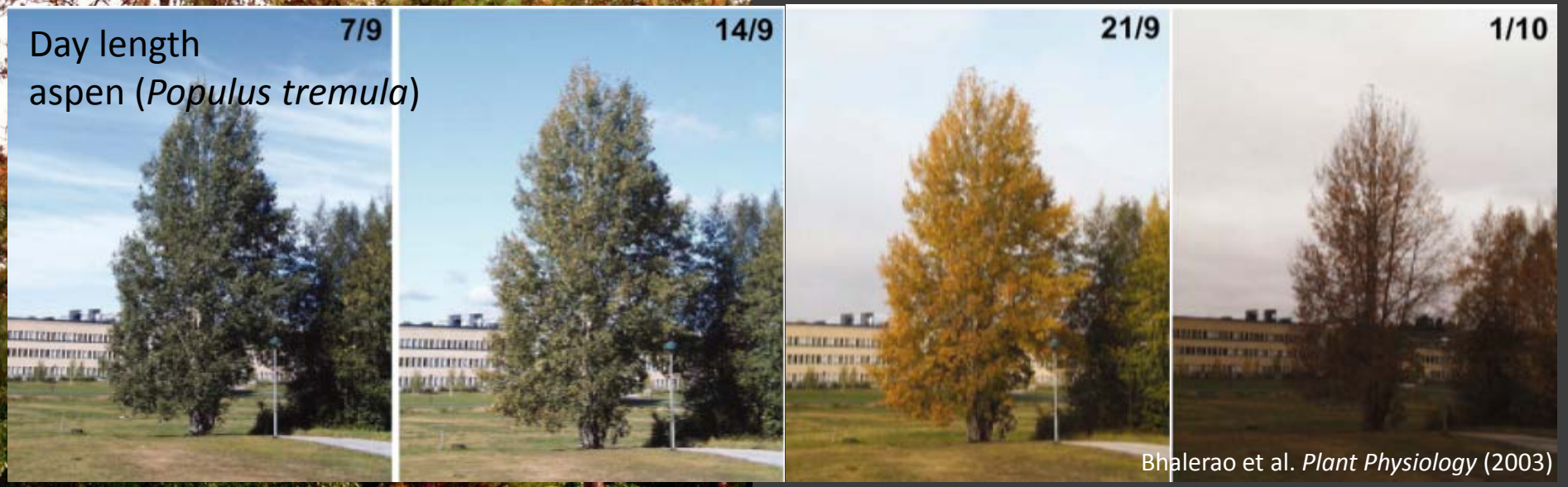
Senescence has many triggers

Drought
longleaf pine (*Pinus palustris*)

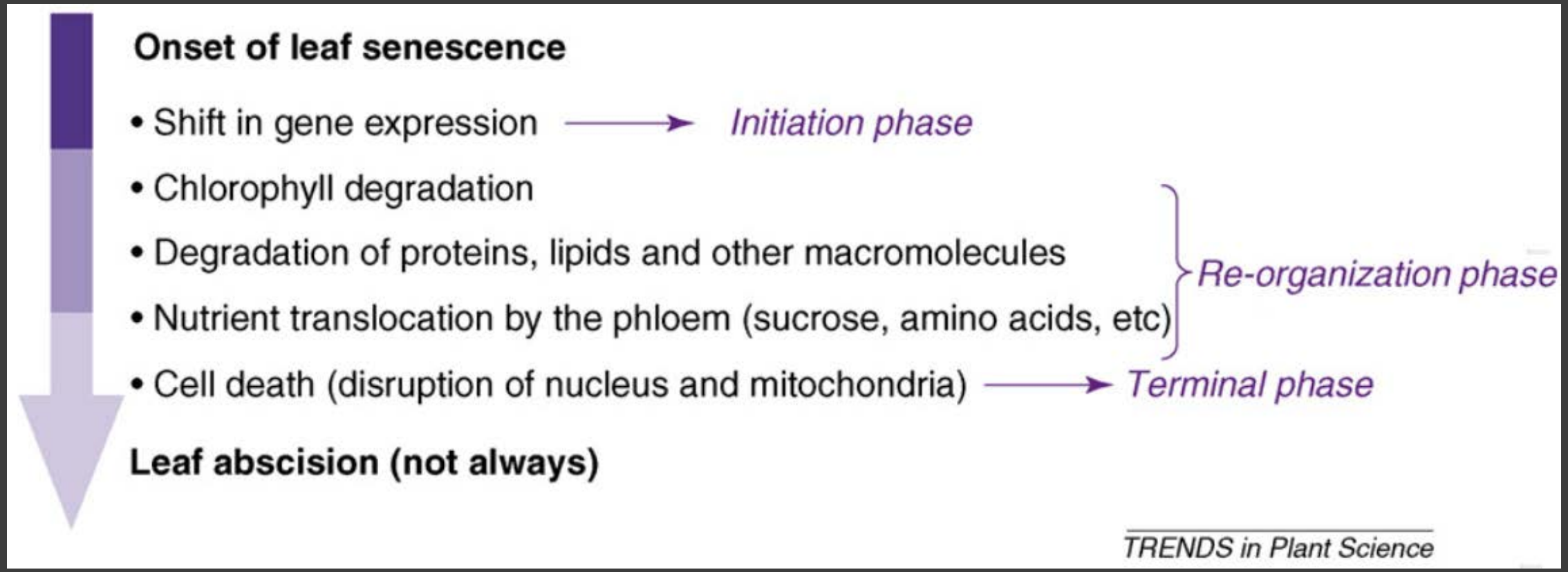


Development of seeds
rice (*Oryza sativa*)

Day length
aspen (*Populus tremula*)



Coordinated changes occur in cell structure, metabolism pathways, and gene expression



Munné-Bosch. *Trends in Plant Science* (2008)

Previous studies focused on transcriptome changes during senescence

The Plant Cell, Vol. 23: 873–894, March 2011, www.plantcell.org © 2011 American Society of Plant Biologists

LARGE-SCALE BIOLOGY ARTICLE

High-Resolution Temporal Profiling of Transcripts du
***Arabidopsis* Leaf Senescence Reveals a Distinct Chron**
Processes and Regulation C W OA

Emily Breeze,[a,1] Elizabeth Harrison,[a,1] Stuart McHattie,[a,b,1] Linda Hughes,[a,2] Richard Hickman,[a]
Steven Kiddle,[a,b] Youn-sung Kim,[a,3] Christopher A. Penfold,[b] Dafyd Jenkins,[b] Cunjin Zhang,[a] Karl Mo
Carol Jenner,[a] Stephen Jackson,[a] Brian Thomas,[a] Alexandra Tabrett,[a] Roxane Legaie,[b] Jonathan D.

Lin *et al. BMC Plant Biology* (2015) 15:43
DOI 10.1186/s12870-015-0433-5

BMC Plant Biolog

RESEARCH ARTICLE Open Acces

Global analysis of the *Gossypium hirsutum* Transcriptome during leaf senescence by

Min Lin, Chaoyou Pang, Shuli Fan, Meizhen Song, Hengling Wei and Shuxun Yu*

Plant Biotechnology Journal (2007) **5**, pp. 192–206

Transcriptome analysis of senescence in th wheat (*Triticum aestivum* L.)

Per L. Gregersen* and Preben Bach Holm

Danish Institute of Agricultural Sciences, Department of Genetics and Biotechnology, Research Centre Flakkebjerg, D

Genome Analysis

Gene Expression in Autumn Leaves[1]

Rupali Bhalerao, Johanna Keskitalo, Fredrik Sterky, Rikard Erlandsson, Harry Björkbacka[2], Simon Jonsson Birve, Jan Karlsson, Per Gardeström, Petter Gustafsson, Joakim Lundeberg, and Stefan Jansson*

Umea Plant Science Center, Department of Plant Physiology, Umea University, 901 87 Umea, Sweden (R.B., Jo.K., H.B., S.J.B., Ja.K., Per G., Pet. G., S.J.); and Department of Biotechnology, Kungliga Tekniska Högsl
106 9

Research Open Access

A transcriptional timetable of autumn senescence

Anders Andersson*, Johanna Keskitalo†, Andreas Sjödin†, Rupali Bhalerao*†, Fredrik Sterky*, Kirsten Wissel†, Karolina Tandre‡, Henrik Aspeborg*, Richard Moyle‡, Yasunori Ohmiya‡, Rishikesh Bhalerao‡, Amy Brunner§, Petter Gustafsson†, Jan Karlsson†, Joakim Lundeberg*, Ove Nilsson‡, Göran Sandberg‡, Steven Strauss§, Björn Sundberg‡,

Transcriptome Analysis of a Subtropical Deciduous Tree: Autumn Leaf Senescence Gene Expression Profile of Formosan Gum

Chi-Hsiang Wen[1], Shih-Shun Lin[2,3,4] and Fang-Hua Chu[1,5,*]

[1]School of Forestry and Resource Conservation, National Taiwan University, Taipei, Taiwan
[2]Institute of Biotechnology, National Taiwan University, Taipei, Taiwan
[3]Genome and Systems Biology Degree Program, National Taiwan University, Taipei, Taiwan
[4]Agriculture Biotechnology Research Center, Academia Sinica, Taipei, Taiwan
[5]Experimental Forest, National Taiwan University, Nan-Tou, Taiwan
*Corresponding author: E-mail, fhchu@ntu.edu.tw; Fax, +886-2-23654520.
(Received September 2, 2014; Accepted November 3, 2014)

PCP PLANT & CELL PHYSIOLOGY

Limitations of previous work

- Focused on annual species, therefore provided limited insights to questions related to dormancy and perennial growth

- Used microarray or Sanger-sequencing based methods which are low throughput and inefficient at detecting low abundance transcripts

Goals of the project

- Examine gene expression patterns in depth during temperature and photoperiod-induced natural senescence in *Populus*

- Develop a toolkit useful for metabolic engineering of senescence-related traits, such as knowledge of key transcription factors, regulatory networks, and promoters useful for developmentally timed activities

We use RNA-seq to examine gene expression pattern during photoperiod-induced natural senescence in *Populus trichocarpa*

NATURE REVIEWS | GENETICS VOLUME 10 | JANUARY 2009 | 57

INNOVATION

RNA-Seq: a revolutionary tool for transcriptomics

Zhong Wang, Mark Gerstein and Michael Snyder

Abstract | RNA-Seq is a recently developed approach to transcriptome profi
that uses deep-sequencing technologies. Studies using this method h

RESEARCH ARTICLES

The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray)

G. A. Tuskan,[1,3*] S. DiFazio,[1,4†] S. Jansson,[5†] J. Bohlmann,[6†] I. Grigoriev,[9†] U. Hellsten,[9†] N. Putnam,[9†] S. Ralph,[6†] S. Rombauts,[10†] A. Salamov,[9†] J. Schein,[11†] L. Sterck,[10†] A. Aerts,[9] R. R. Bhalerao,[5] R. P. Bhalerao,[12] D. Blaudez,[13] W. Boerjan,[10] A. Brun,[13] A. Brunner,[14] V. Busov,[15] M. Campbell,[16] J. Carlson,[17] M. Chalot,[13] J. Chapman,[9] G.-L. Chen,[2] D. Cooper,[6] P. M. Coutinho,[19] J. Couturier,[13] S. Covert,[20] Q. Cronk,[7] R. Cunningham,[1] J. Davis,[22] S. Degroeve,[10] A. Déjardin,[23] C. dePamphilis,[18] J. Detter,[9] B. Dirks,[24] I. Dubchak,[9,25] S. Duplessis,[13] J. Ehlting,[7] B. Ellis,[6] K. Gendler,[26] D. Goodstein,[9] M. Gribskov,[27] J. Grimwood,[28] A. Groover,[29] L. Gunter,[1] B. Hamberger,[7] B. Heinze,[30] Y. Helariutta,[12,31,33] B. Henrissat,[19] D. Holligan,[21] R. Holt,[11] W. Huang,[9] N. Islam-Faridi,[34] S. Jones,[11] M. Jones-Rhoades,[35] R. Jorgensen,[26] C. Joshi,[15] J. Kangasjärvi,[32] J. Karlsson,[5] C. Kelleher,[6] R. Kirkpatrick,[11] M. Kirst,[22] A. Kohler,[13] U. Kalluri,[1] F. Larimer,[2] J. Leebens-Mack,[21] J.-C. Leplé,[23] P. Locascio,[2] Y. Lou,[9] S. Lucas,[9] F. Martin,[13] B. Montanini,[13] C. Napoli,[26] D. R. Nelson,[36] C. Nelson,[37] K. Nieminen,[31] O. Nilsson,[12] V. Pereda,[13] G. Peter,[22] R. Philippe,[6] G. Pilate,[23] A. Poliakov,[25] J. Razumovskaya,[2] P. Richardson,[9] C. Rinaldi,[13] K. Ritland,[8] P. Rouzé,[10] D. Ryaboy,[25] J. Schmutz,[28] J. Schrader,[38] B. Segerman,[5] H. Shin,[11] A. Siddiqui,[11] F. Sterky,[39] A. Terry,[9] C.-J. Tsai,[15] E. Uberbacher,[2] P. Unneberg,[39] J. Vahala,[32] K. Wall,[18] S. Wessler,[21] G. Yang,[21] T. Yin,[1] C. Douglas,[7‡] M. Marra,[11‡] G. Sandberg,[12‡] Y. Van de Peer,[10‡] D. Rokhsar[9,24‡]

Experimental strategy

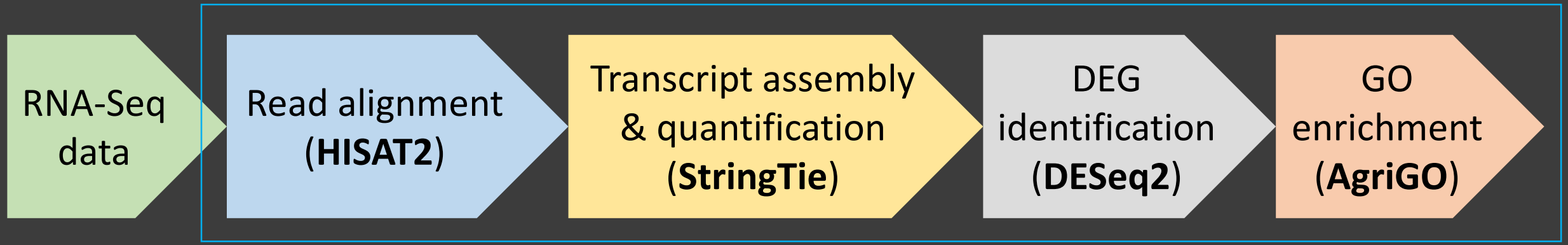
- Collected leaf samples (*P. trichocarpa*) at the end of each month from May to October in 2012, 2015, and 2016
- Built and sequenced a total of 54 RNA-seq libraries
  - three biological replicates for each collection timepoint
- Identified differentially expressed genes (DEGs, FDR < 0.05) and gene ontologies (GOs, log FC = 1.5; FDR < 0.05)
  - primary analysis focused on 2015 and 2016 data

Replicate trees of *Populus trichocarpa* Nisqually-1 used for RNA-seq near OSU (Corvallis, OR)

Hypotheses to test

- Samples collected from different months differ in expression patterns, and cluster according to chronology and physiological state
- Genes related to chlorophyll biosynthesis, photosynthesis, protein synthesis and other energy-requiring activities, are down-regulated as leaves senesce
- Genes related to reactive oxygen species (ROS) and catalytic activity are up-regulated as leaves senesce
- Due to the depth of sampling in our RNAseq study compared to prior work, we will detect large numbers of genes not previously associated with senescence in perennial plants

Workflow of RNA-seq analysis



KBase

Generate lists of DEGs and GOs

R

AGRIGO v2

Visualize results

KBase:
An open environment for systems biology

RNA Seq
Create RNA-seq Sample Set
Provide RNA-seq reads and the metadata to create an RNA-seq Sample Set

HISAT-2
Align Reads using HISAT2 - v2.1.0
Align sequencing reads to long reference sequences using HISAT2

String Tie
Assemble Transcripts using StringTie - v1.3.3b
Assemble the transcripts from RNA-seq reads using StringTie

DESeq2
Create Differential Expression Matrix using DESeq2 - v1.20.0
Create differential expression matrix based on a given threshold cutoff

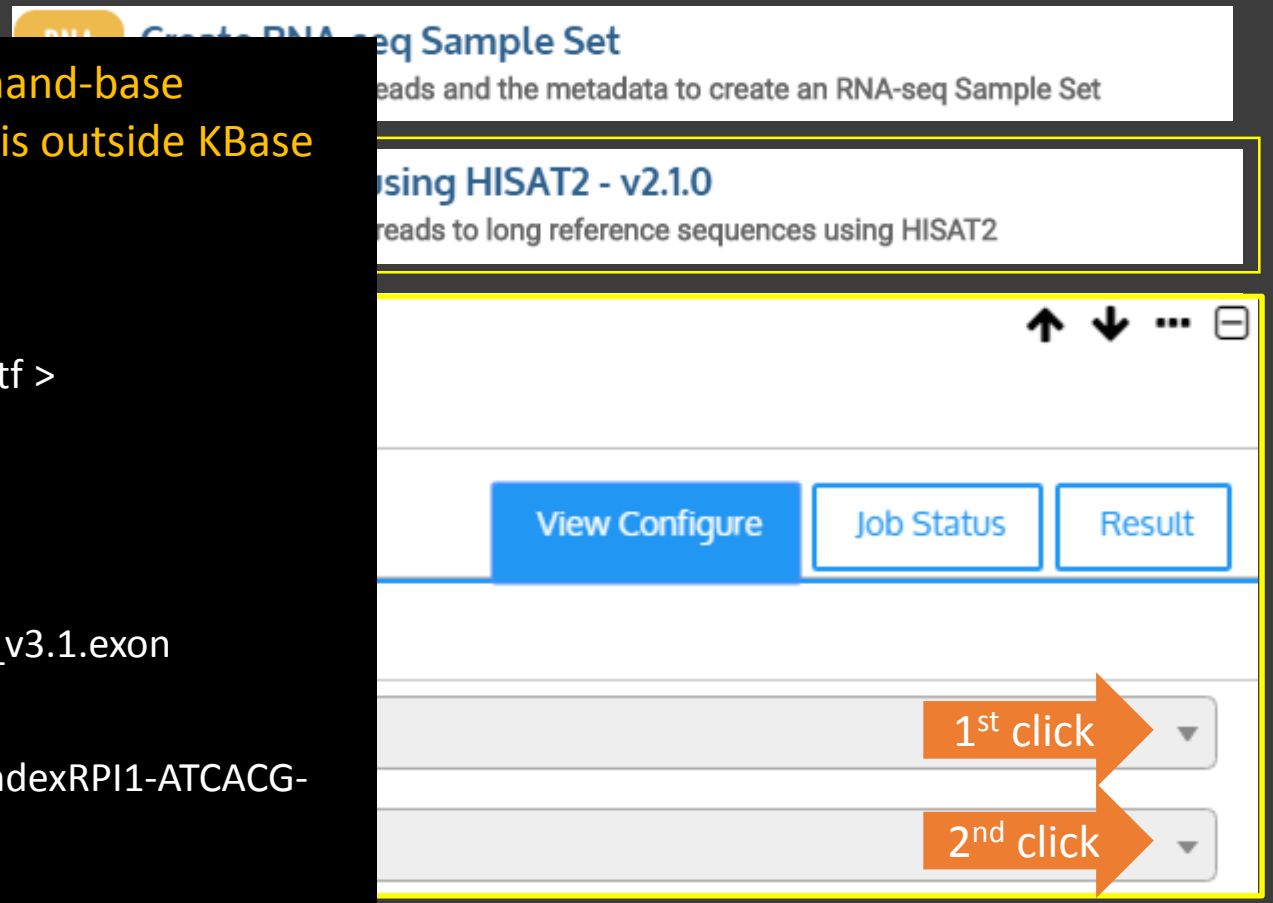
DATA
Create Feature Set/Filtered Expression Matrix From Differential
Create FeatureSet/Filtered Expression Matrix based on given threshold cutoffs

KB
Functional Enrichment for GO Terms - v1.0.8
Compute GO term enrichment for genomic features

KBase provided an easier and faster way for analyzing our large dataset

Command-base analysis outside KBase

```
$ gunzip Ptrichocarpa_444_v3.0.fa.gz

$ gunzip Ptrichocarpa_444_v3.1.gene_exons.gff3.gz

$ gffread -E Ptrichocarpa_444_v3.1.gene_exons.gff3 -T -o- > Ptrichocarpa_444_v3.1.gene_exons.gtf

$ hisat2_extract_splice_sites.py Ptrichocarpa_444_v3.1.gene_exons.gtf > Ptrichocarpa_444_v3.1.ss

$ hisat2_extract_exons.py Ptrichocarpa_444_v3.1.gene_exons.gtf > Ptrichocarpa_444_v3.1.exon

$ hisat2-build --ss Ptrichocarpa_444_v3.1.ss --exon Ptrichocarpa_444_v3.1.exon Ptrichocarpa_444_v3.0.fa Ptrichocarpa_444_v3.1._tran

$ hisat2 -p 8 --dta -x Populus_trichocarpa_v3.40_tran -U lane5-s003-indexRPI1-ATCACG-15MT1_S3_L005_R1_001.fastq -S 15May_T1.sam
…
$ samtools sort -@ 8 -o 15May_T1.bam 15May_T1.sam
…
```



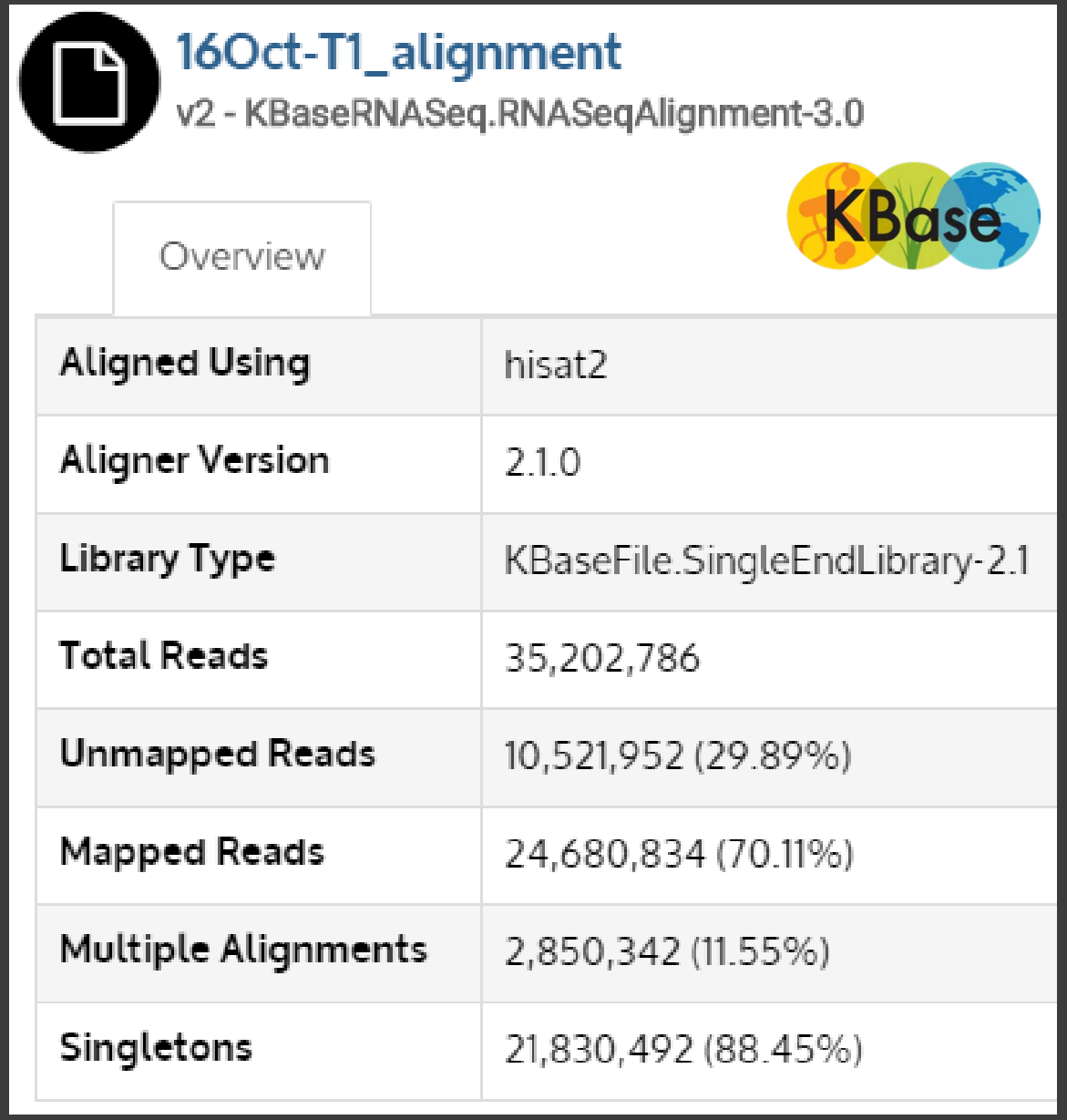
Each library had 18 to 30 million reads mapped to the *P. trichocarpa* genome

Sample	Total Reads	Mapped	Multiple Alignments	Singletons
2015MayT1	39,028,092	27,546,804	5,242,879	22,303,925
2015MayT2	37,618,748	25,994,451	3,787,680	22,206,771
2015MayT3	41,864,782	27,022,013	7,136,967	19,885,046
2015JuneT1	38,383,412	17,662,780	1,669,728	15,993,052
2015JuneT2	36,034,085	23,808,026	2,515,832	21,292,194
2015JuneT3	40,129,485	21,895,438	1,932,833	19,962,605
2015JulyT1	35,998,575	24,708,472	2,367,903	22,340,569
2015JulyT2	35,816,142	22,099,437	2,431,281	19,668,156
2015JulyT3	38,873,070	28,341,217	2,851,425	25,489,792
2015AugT1	44,057,801	29,755,263	4,205,508	25,549,755
2015AugT2	35,787,635	18,744,647	2,388,791	16,355,856
2015AugT3	35,853,042	18,914,492	1,587,492	17,327,000
2015SeptT1	43,618,559	29,698,837	3,552,201	26,146,636
2015SeptT2	36,097,415	22,270,423	4,248,765	18,021,658
2015SeptT3	39,873,012	30,127,028	3,266,626	26,860,402
2015OctT1	48,464,438	31,575,757	5,916,244	25,659,513
2015OctT2	37,174,028	23,159,745	5,664,273	17,495,472
2015OctT3	37,428,947	18,307,836	2,386,960	15,920,876

Sample	Total Reads	Mapped	Multiple Alignments	Singletons
2016MayT1	45,063,842	30,586,150	2,633,596	27,952,554
2016MayT2	40,719,864	30,478,767	2,214,183	28,264,584
2016MayT3	36,905,845	26,138,307	1,964,928	24,173,379
2016JuneT1	40,347,881	23,337,970	1,880,171	21,457,799
2016JuneT2	34,019,874	18,115,784	1,449,646	16,666,138
2016JuneT3	43,456,828	27,814,089	1,889,700	25,924,389
2016JulyT1	41,609,097	30,508,159	2,918,727	27,589,432
2016JulyT2	39,097,800	21,968,256	4,861,689	17,106,567
2016JulyT3	40,315,148	28,000,800	3,158,295	24,842,505
2016AugT1	39,008,336	26,108,091	5,318,932	20,789,159
2016AugT2	46,381,838	30,411,124	5,456,267	24,954,857
2016AugT3	42,925,471	18,003,108	2,573,392	15,429,716
2016SeptT1	34,782,308	21,455,809	1,991,959	19,463,850
2016SeptT2	44,279,522	33,809,765	3,929,300	29,880,465
2016SeptT3	33,957,595	20,948,479	4,374,480	16,573,999
2016OctT1	35,202,786	24,680,834	2,850,342	21,830,492
2016OctT2	35,163,735	17,697,293	3,105,998	14,591,295
2016OctT3	33,293,602	21,551,049	1,439,902	20,111,147

On average, there were 24 million mapped reads per library

- 21 million reads mapped once
- 3 million reads mapped more than once



16Oct-T1_alignment

v2 - KBaseRNASeq.RNASeqAlignment-3.0

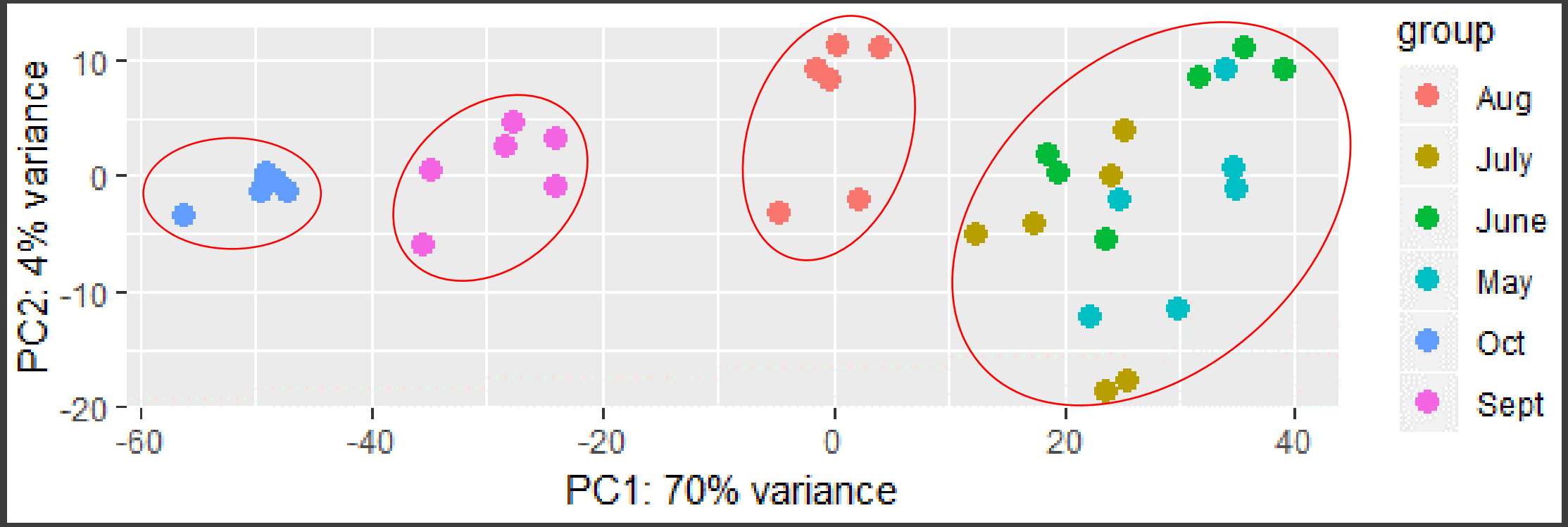
KBase

Overview

Aligned Using	hisat2
Aligner Version	2.1.0
Library Type	KBaseFile.SingleEndLibrary-2.1
Total Reads	35,202,786
Unmapped Reads	10,521,952 (29.89%)
Mapped Reads	24,680,834 (70.11%)
Multiple Alignments	2,850,342 (11.55%)
Singletons	21,830,492 (88.45%)

Collections were clustered into four major groups in the principal component analysis (PCA) plot

Collections were clustered into four major groups (simplified version)



Number (%) of DEGs identified when comparing the same month from two different years (FDR < 0.05)

Month - year	May2015	June2015	July2015	Aug2015	Sept2015	Oct2015
May2016	5,196 (16.1%) • 2,775 (8.6%) • 2,421 (7.5%)					
June2016		2,779 (8.6%) • 1,077 (3.3%) • 1,702 (5.3%)				
July2016			420 (1.3%) • 155 (0.5%) • 265 (0.8%)			
Aug2016				208 (0.6%) • 82 (0.3%) • 126 (0.4%)		
Sept2016					703 (2.2%) • 299 (0.9%) • 404 (1.2%)	
Oct2016						2,111 (6.5%) • 1,170 (3.6%) • 941 (2.9%)

Total
- Down regulated
- Up regulated

Number (%) of DEGs identified during pair-wise comparison of 2015 collection (FDR < 0.05)

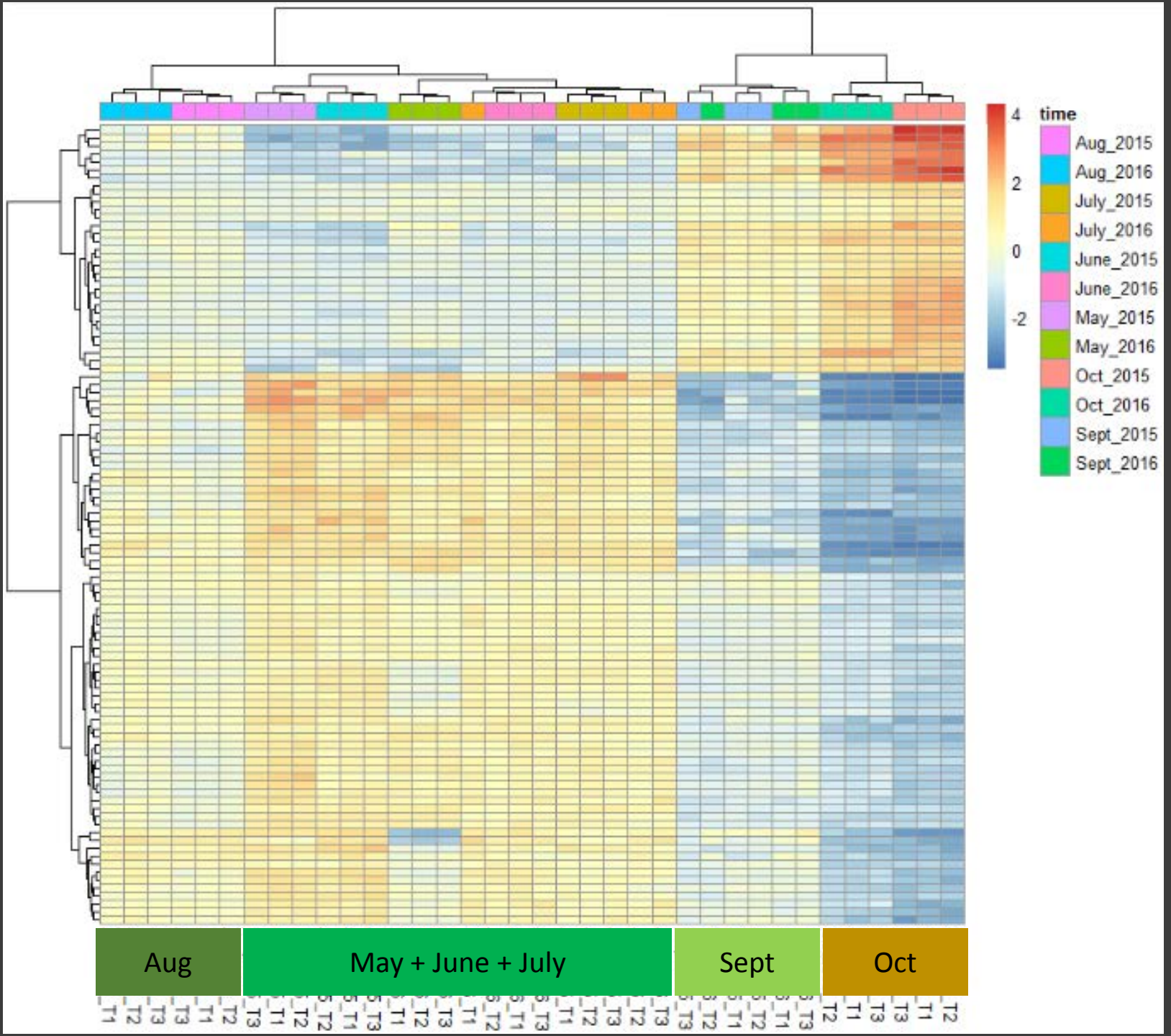
Month - year	May2015	June2015	July2015	Aug2015	Sept2015	Oct2015
May2015		1,739 (5.4%) • 988 (3.1%) • 751 (2.3%)	3,033 (9.4%) • 1,521 (4.7%) • 1,512 (4.7%)	5,679 (17.6%) • 2,945 (9.1%) • 2,734 (8.5%)	10,636 (32.9%) • 5,406 (16.7%) • 5,230 (16.2%)	14,962 (46.3%) • 7,549 (23.4%) • 7,413 (22.9%)
June2015			2,276 (7.0%) • 901 (2.8%) • 1,375 (4.3%)	5,956 (18.4%) • 2,915 (9.0%) • 3,041 (9.4%)	12,049 (37.3%) • 6,203 (19.2%) • 5,846 (18.1%)	15,434 (47.7%) • 7,842 (24.3%) • 7,592 (23.5%)
July2015				3,875 (12.0%) • 1,985 (6.1%) • 1,890 (5.8%)	9,202 (28.5%) • 4,738 (14.7%) • 4,464 (13.8%)	13,877 (42.9%) • 7,153 (22.1%) • 6,724 (20.8%)
Aug2015					4,128 (12.8%) • 2,027 (6.3%) • 2,101 (6.5%)	10,897 (33.7%) • 5,632 (17.4%) • 5,265 (16.3%)
Sept2015						7,226 (22.4%) • 3,922 (12.1%) • 3,304 (10.2%)
Oct2015						

Total
- Down regulated
- Up regulated

Number (%) of DEGs identified during pair-wise comparison of 2016 collection (FDR < 0.05)

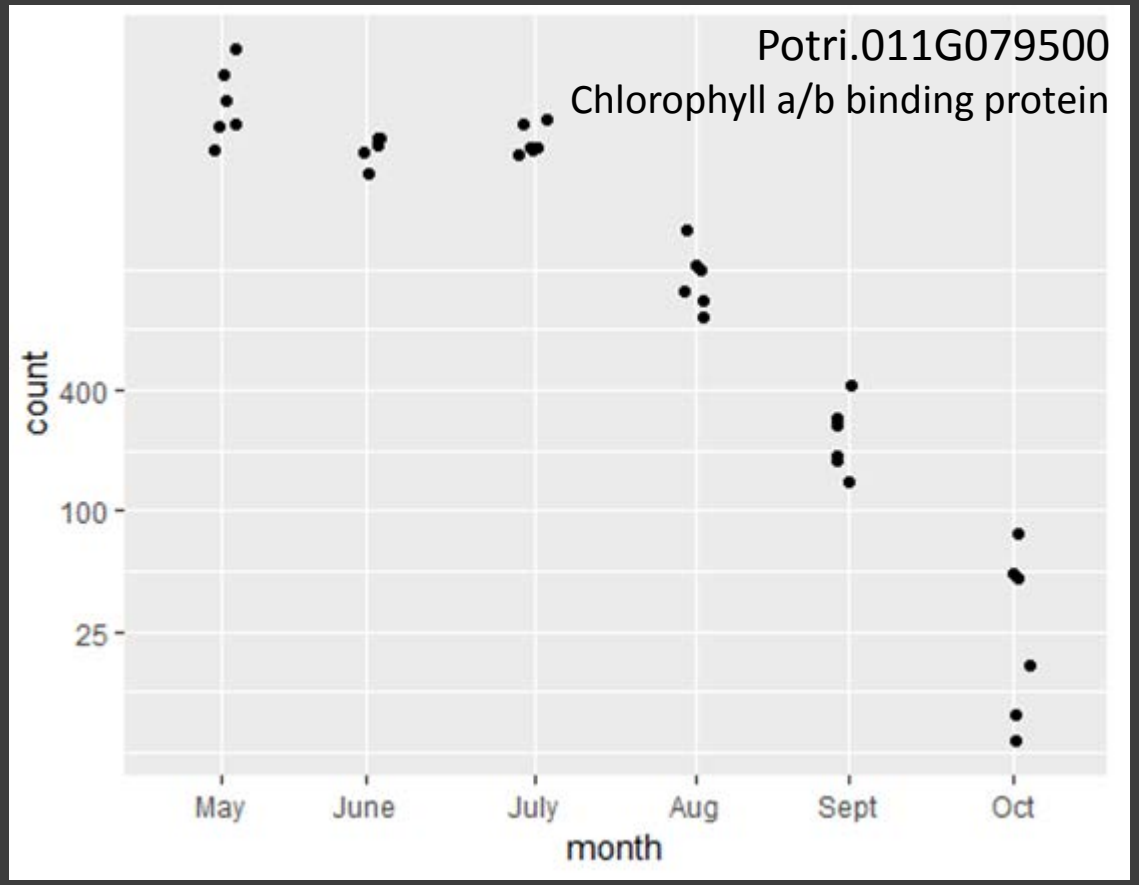
Month - year	May2016	June2016	July2016	Aug2016	Sept2016	Oct2016
May2016		4,619 (14.3%) • 2,097 (6.5%) • 2,522 (7.8%)	5,033 (15.6%) • 2,167 (6.7%) • 2,866 (8.9%)	7,882 (24.4%) • 3,659 (11.3%) • 4,223 (13.1%)	11,419 (35.3%) • 5,484 (17.0%) • 5,935 (18.4%)	14,242 (44.1%) • 7,113 (22.0%) • 7,119 (22.0%)
June2016			176 (0.5%) • 90 (0.3%) • 86 (0.3%)	1,582 (4.9%) • 729 (2.3%) • 853 (2.6%)	7,348 (22.7%) • 3,655 (11.3%) • 3,693 (11.4%)	11,695 36.2%) • 6,132 (19.0%) • 5,563 (17.2%)
July2016				1,199 (3.7%) • 622 (1.9%) • 577 (1.8%)	7,422 (23.0%) • 3,890 (12.0%) • 3,532 (10.9%)	11,661 36.1%) • 6,275 (19.4%) • 5,386 (16.7%)
Aug2016					3,701 (11.4%) • 1,999 (6.2%) • 1,702 (5.3%)	9,427 29.2%) • 5,179 (16.0%) • 4,248 (13.1%)
Sept2016						3,436 (10.6%) • 2,226 (6.9%) • 1,210 (3.7%)
Oct2016						

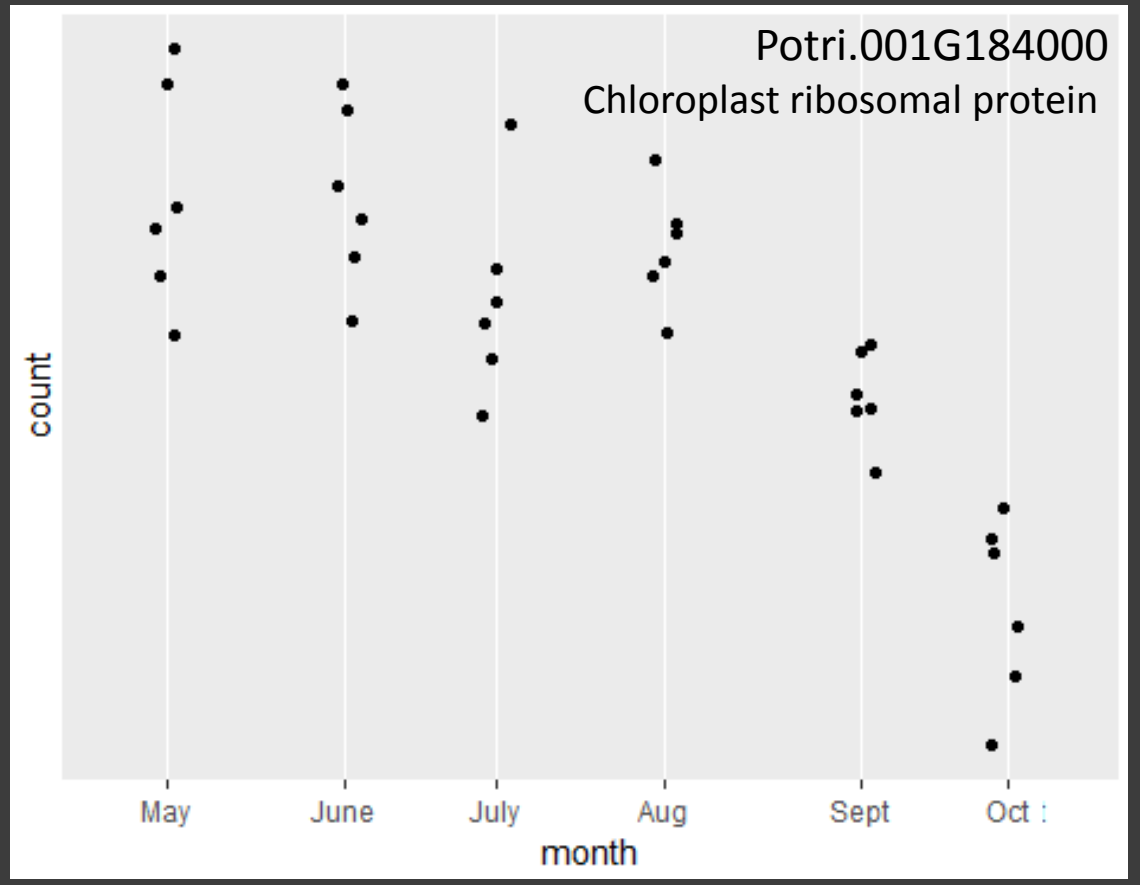
Total
- Down regulated
- Up regulated



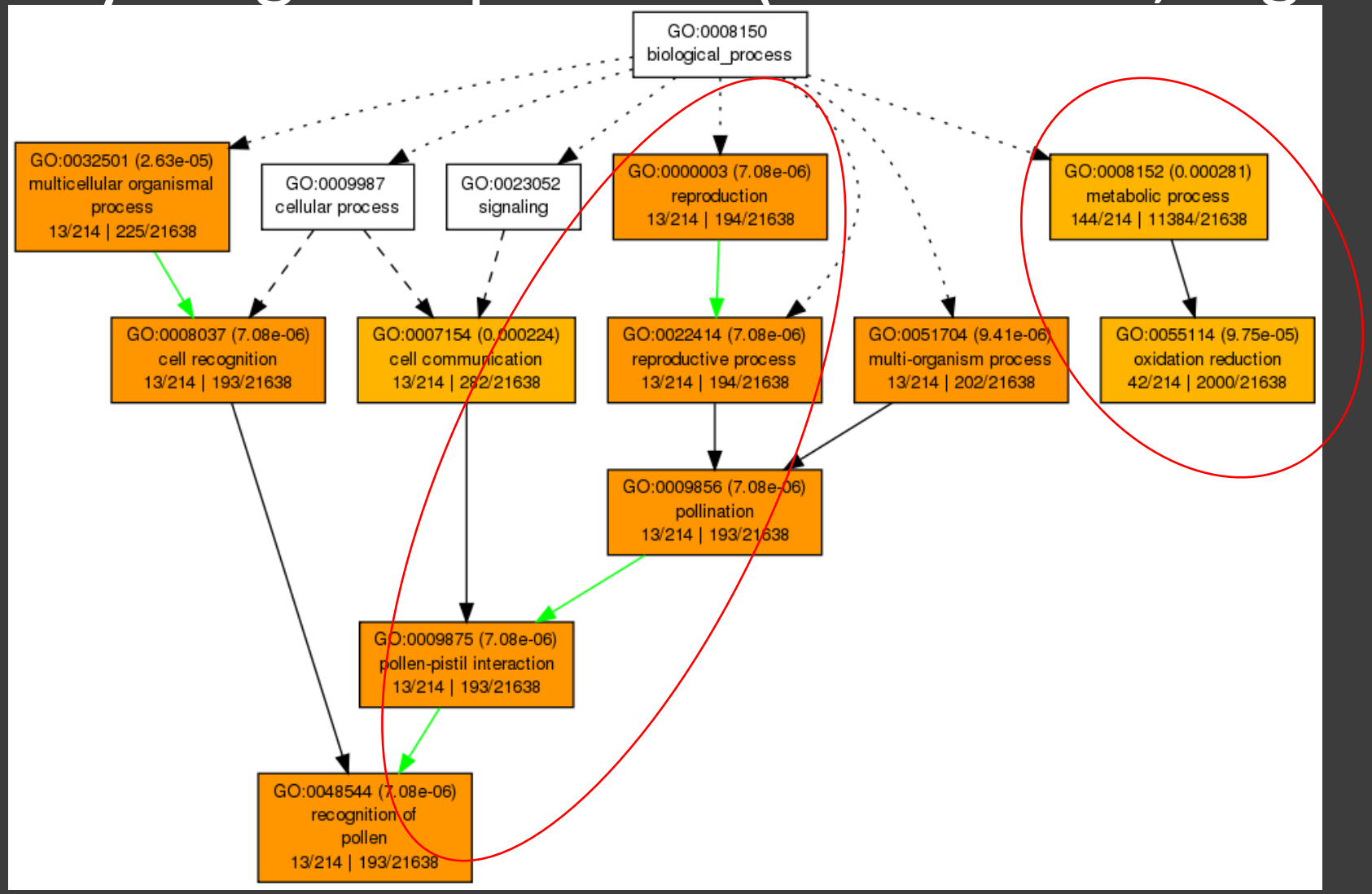
Samples were clustered into four groups in the heatmap based on top 100 DEGs between May and Oct from 2015

Plots of individual genes indicate reduced chloroplast/chlorophyll in August



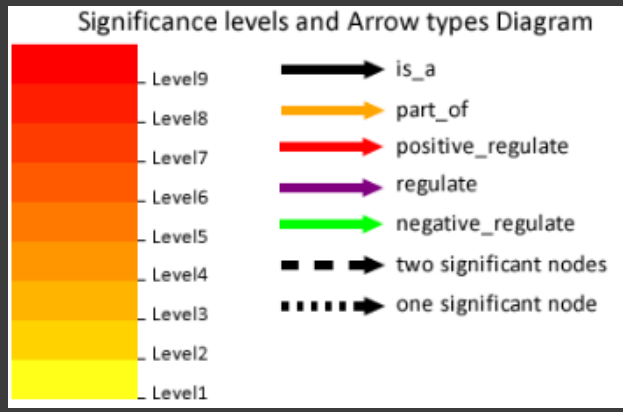


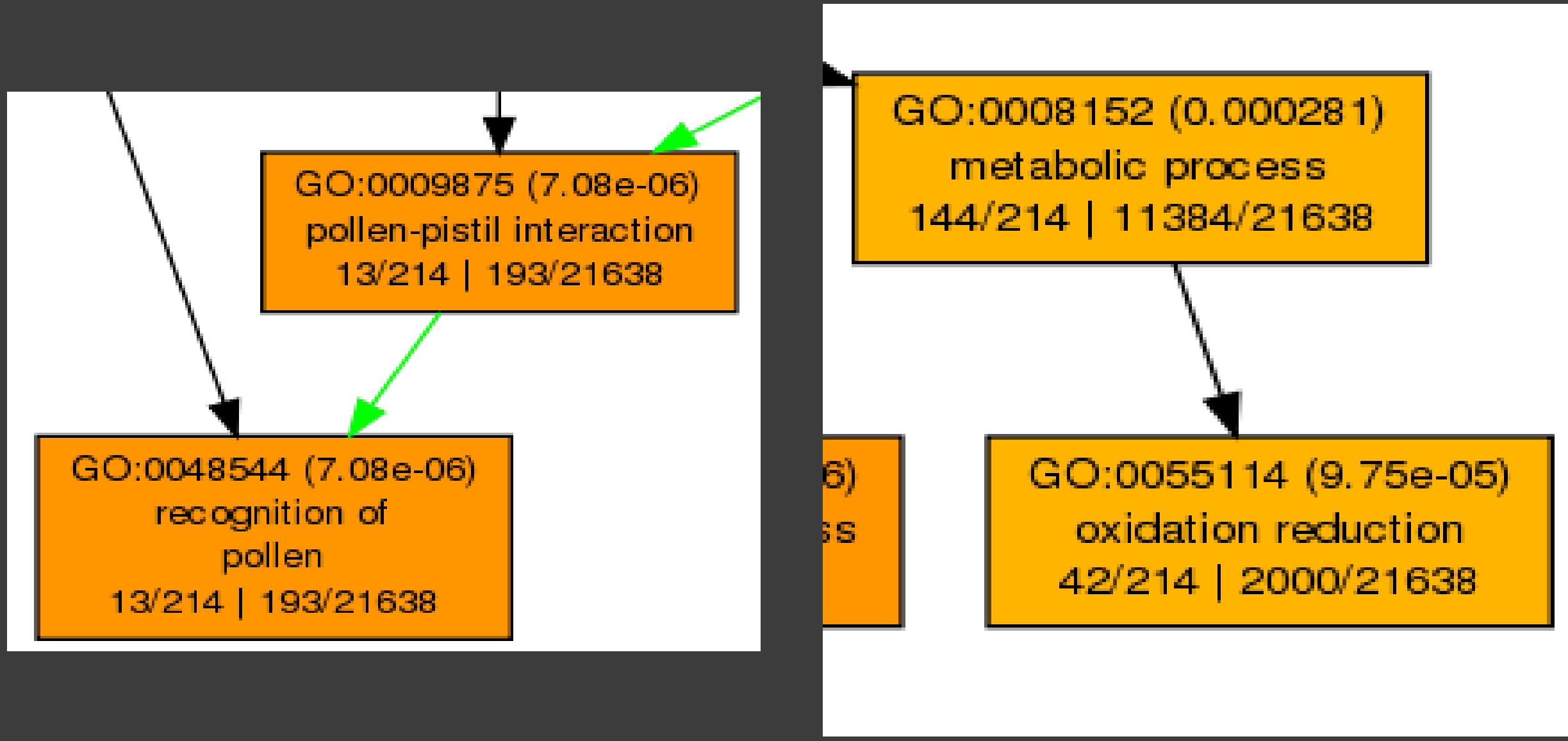
GOs related to cell recognition and communication, oxidation reduction, and reproduction were enriched in July-Aug comparison (FDR < 0.05; log FC= -1.5)



Legend:

GO term (adj p-value)
GO description
Annotated/Total number in query|
Annotated/Total number in reference




GO:0009875 (7.08e-06)
pollen-pistil interaction
13/214 | 193/21638
GO:0048544 (7.08e-06)
recognition of
pollen
13/214 | 193/21638
GO:0008152 (0.000281)
metabolic process
144/214 | 11384/21638
GO:0055114 (9.75e-05)
oxidation reduction
42/214 | 2000/21638

Conclusions

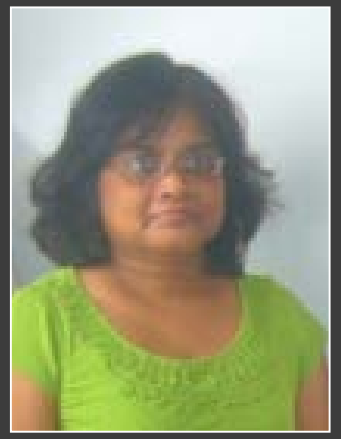
- Pair-wise comparison of different collection timepoints revealed extensive changes in gene expression during the collection period
- Collections (from 2015 and 2016) were clustered into four groups in PCA plot and heatmap
  - May, June, and July samples formed one cluster, Aug, Sept, and Oct samples each formed one cluster
- Reduced chlorophyll biosynthesis occurred in August, indicating the initiation of senescence
- Several GO terms, such as those related to metabolic process, reproduction, signaling, and cellular process, were enriched in early senescence (*i.e.*, July-Aug comparison)

Future directions

- Perform clustering analysis to classify collections into different developmental states (*e.g.*, growth, early senescence, and late senescence)
- Refine DEGs and GOs (after assigning collections into development states)
- Group DEGs into contrasting gene expression groups
- Identify novel isoforms associated with senescence (enabled by StringTie, the assembler used in the analysis pipeline)
- Identify sequence motifs associated with each gene expression group
- Validate gene expression using 2012 data and qRT-PCR

Key collaborators and funding sources

Priya Ranjan



Vindhya Amarasinghe