

RNA-Seq sheds novel insight into the transcriptomic  
interplay between *Glycine max* and *Phakopsora  
pachyrhizi*  
MAPMBS 2012

Parsa Hosseini <sup>1</sup>    Arianne Tremblay <sup>2</sup>    Benjamin F. Matthews <sup>3</sup>

<sup>1</sup>Dept. Bioinformatics & Computational Biology, George Mason University, VA

<sup>2</sup>Dept. Biological Science, UMBC, MD,

<sup>3</sup>USDA-ARS Soybean Genomics & Improvement Lab., MD,

August 16<sup>th</sup>, 2012

## Covered Topics

- 1 Overview of Soybean Rust
- 2 Time-series soybean-inoculated RNA-Seq study
  - Dataset pre-processing
  - Transcript quantification & annotation
  - Novel SR transcript assembly
- 3 Systematic Transcription Factor (TF) Analysis
  - Modeling TFBSs
  - Quantifying TF abundance
- 4 Conclusions

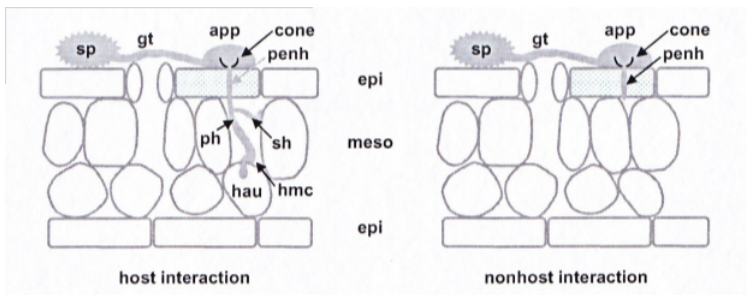
# Introduction

## Definition

**Soybean Rust** (SR): A major crop disease caused by the obligate parasite *Phakopsora pachyrhizi*

- *Phakopsora pachyrhizi* has wide host-range, encompassing 30+ legumous crops such as soybean, kidney bean, and cowpea.
- Being a foliar parasite, management of SR is difficult; spores are easily spread over long distances via. wind gusts.
- Magnitude of crop loss is dependent on temperature; losses in-upwards of 80% have been reported.

## Introduction – SR lifecycle



K. Goellner, et. al., Mol. Plant Path., 2010

# Experimental Overview

## RNA extraction

- **Tissues:** 9x leaves from *G. max* Williams 82 susceptible plant:
  - 1 0dai (uninoculated control; 21 day old plant)
  - 2 10 days after infection (dai) with SR
- **Sequencing:** Single-end recipe on the Genome Analyzer (GA) IIx yielded 70bp (0dai) and 40bp (10dai) fragments

Lane	cDNA conc. (pM)		# raw reads	
	0dai	10dai	0dai	10dai
1	-	1	-	1,724,879
2	-	1	-	2,178,249
3	-	4	-	5,574,892
4	-	4	-	4,997,270
5	PhiX	PhiX	-	-
6	-	2	-	3,060,798
7	-	2	-	3,319,178
8	8	2	5,940,995	3,461,920
	-	-	5,940,995	24,317,186

# Dataset processing & statistics

## Pre-processing & read-sanitization

**Quality filters:** 1) Subtract low-quality reads. 2) Trim low-quality 3'.

**Contamination filter:** Filter reads mapping to *H. sapiens* or microbe.

		0dai	10dai
<b>Pre-processing</b>	Passed filtering	$\approx 84.4\%$ (5,015,459)	$\approx 97.2\%$ (5,420,745)
	$\overline{Q_{score}}$	27	30
<b>Alignment</b>	% map to G. max	75.66% (3,795,091)	64.08% (3,473,688)
<b>Quantification</b>	RPKM range	0.0020 ... 6730	0.0021 ... 4195.3

## Induced DEGs 10dai

Transcript	RPKM		GO process
	0dai	10dai	
Glyma20g34820.1	0.43	1436.87	response to wounding
Glyma08g25520.1	0.36	1206	negative regulation of endopeptidase activity
Glyma10g32830.1	0.24	825.54	response to wounding
Glyma13g42210.1	0.27	248.31	cell wall macromolecule catabolic process
Glyma13g35320.2	0.15	132.94	defense response
Glyma07g32650.1	0.08	32.12	defense response to fungus, incompatible interaction
Glyma17g03360.1	0.14	53.01	response to biotic stimulus

Transcript	RPKM		GO function
	0dai	10dai	
Glyma10g33650.1	0.14	253.20	glutathione transferase activity
Glyma13g42210.1	0.27	248.31	chitinase activity
Glyma05g27960.2	0.05	41.09	O-methyltransferase activity
Glyma09g07330.1	0.07	53.31	alliin lyase activity
Glyma11g32520.1	0.04	20.31	protein serine/threonine kinase activity
Glyma20g38930.2	0.01	5.92	prenyltransferase activity

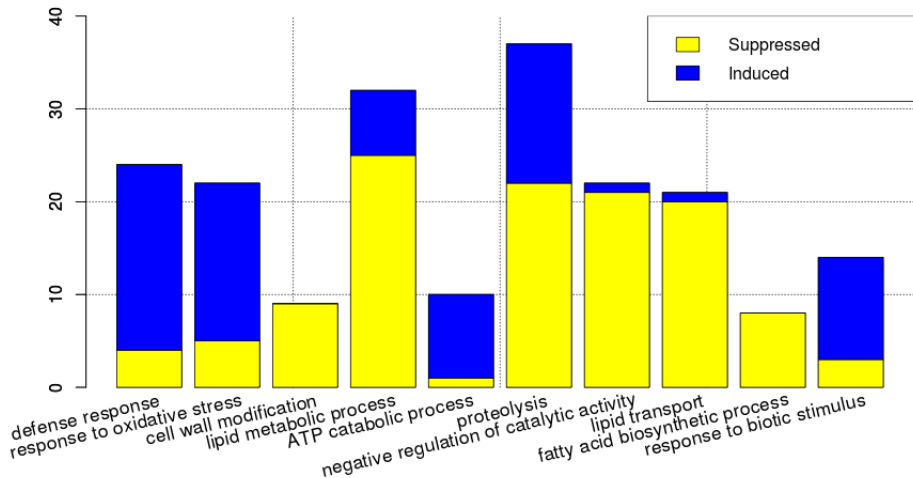
## Suppressed DEGs 10dai

Transcript	RPKM		GO process
	0dai	10dai	
Glyma14g14220.1	2039.93	0.16	lipid transport
Glyma10g00440.1	23.53	0.08	fatty acid biosynthetic process
Glyma03g41330.1	17.46	0.06	lipid metabolic process
Glyma18g47130.1	10.83	0.05	carbohydrate metabolic process
Glyma12g10960.1	14.15	0.07	cellular glucan metabolic process
Glyma16g28030.1	146	0.79	photosynthesis, light harvesting

Transcript	RPKM		GO function
	0dai	10dai	
Glyma03g24460.1	34.76	0.02	aldehyde decarboxylase activity
Glyma16g00980.2	183.33	0.14	nutrient reservoir activity
Glyma06g35110.1	86.70	0.13	anthocyanidin 3-O-glucosyltransferase activity
Glyma12g10960.1	14.15	0.07	xyloglucan:xyloglucosyl transferase activity



### GO Process abundance in transcripts 10dai



# SR transcript assembly

## Building an SR assembly

- Map all 10dai reads onto Phytozome & remove concordant reads
- 22,865,186 reads  $\xrightarrow{\text{no Phytozome hit}}$  8,258,851 reads
- Assemble and annotate de-novo transcripts from resultant reads

	Attribute	10dai
SR transcriptome assembly	# transcripts (count)	19,688
	Range (bp)	100 ... 1,905
	RPKM (min ... max)	6.37 ... 14,183
	N50 (bp)	132

## SR properties

- Min. transcript length  $\Rightarrow$  400bp, hence larger RPKM range.
- Top 875 DE transcripts were mapped onto EST databases
- Majority hit *P. pachyrhizi* dbEST entries (618 / 875)
- COGEME yielded far fewer annotations (25 / 875)
- Sequence-clustering created heterogenous clusters of un/annotated transcripts. Distinct unannotated clusters were also created

COGEME Function	Abundance
60S or 40S ribosomal protein	11
Histones	2
MAS protein	2

# Modeling TFs & resources

## Example

### Pattern-based TFBS

GACTAACGGTAAG

Family: MYB, Gene: AtWER

GACGTCCGTACAACAAG

Family: SBP, Gene: SPL7

### Position Weight Matrix (PWM)

A	4	13	0	2	1	0	13
C	4	0	14	0	0	15	0
G	0	1	0	13	0	0	1
T	7	1	1	0	14	0	1

Family: Leucine-Zipper, Gene: TGA1A

TFBS resource	Pattern TFBS?	PWMs?
AGRIS	✓	✗
AthaMap	✓	✓
AtProbe	✓	✗
JASPAR	✗	✓
TransFAC	✓	✓

# Extrapolating TFBSs onto promoter regions

## Finding novel TF patterns in DEGs

- TFs such as WRKY, bZIP & MYB are associated with defense <sup>1</sup>
  - Probabilistic models can infer TF profile “interestingness” given multiple conditions or timepoints.
  - *Marina* <sup>2</sup> was developed to aide plant-researchers find novel and surprising TF patterns.
- 
- Obtained promoter sequences for top-300 DEGs in 0dai and 10dai.
  - Using *Marina*, TFBS propensities & probabilities were quantified

---

<sup>1</sup>Soria-Guerra et. al., Theor. Appl. Genet., 2010

<sup>2</sup>Manuscript in-preparation

# Extrapolating TFBSs onto promoter regions

TF name	Abundance (10dai)	Abundance (0dai)
WRKY26/WRKY38/WRKY43	23	1
ATHB6	42	3
ABFS	17	1
GT-1	26	2
ATHB2	128	14
ARR1/ARR2	40	4
ATHB1/ATHB5	39	4
MYB2	22	2
MYB98	57	7
WRKY11	10	1
WRKY6	10	1

An ensemble of metrics are used to infer interestingness

## Conclusions & Considerations

- Distinct transcriptomic signatures are evident; indicative of systematic defense mechanisms
- Potentially novel TF-profiles and clusterings could be identified
- *P. pachyrhizi* informatics is challenging but ESTs assist greatly

# Acknowledgements

- **USDA-ARS**

- Members of the Matthews lab.
- Bret Cooper, Ph.D.

- **UMBC**

- Arianne Tremblay Ph.D.

- **NCBI**

- Members of the Ovcharenko lab.

Thank you