

Analysis of the transcriptome of the Fagaceae

Ali Barakat

The Penn State University

PENNSSTATE



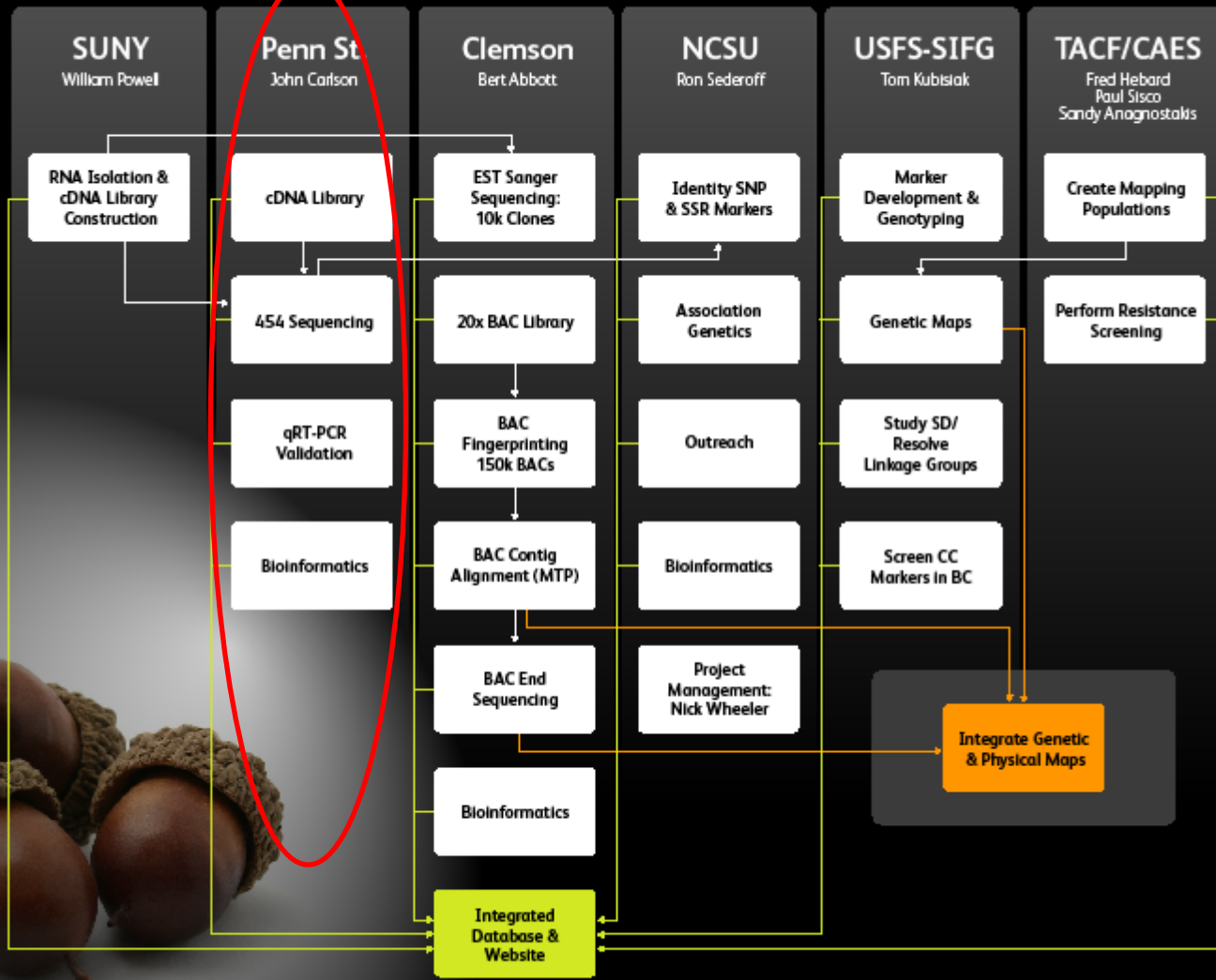
College of Agricultural Sciences

School of Forest Resources



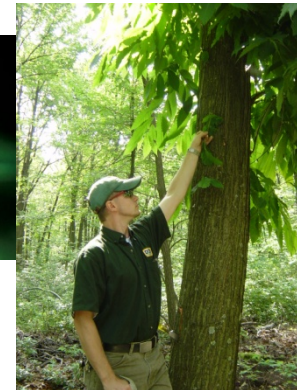
Penn State's Role in the Fagaceae Genome Project

FAGACEAE GENOMICS PROJECT



cDNA Libraries Constructed:

- (1) Infected stem tissue for Chinese chestnut cv. Nanking (TACF) (GS20)
- (2) Infected stem tissue for American chestnut cv BA69 (TACF) (GS20)
- (3) Uninfected stem tissue from Chinese chestnut cv Mahogany (TACF)
- (4) Uninfected stem tissue from Chinese chestnut cv Nanking (TACF)
- (5) Uninfected stem tissue from American chestnut Wisneiwski (CAES)
- (6) Uninfected stem tissue from American chestnut Watertown (CAES)
- (7) Whole plant tissues for Chinese chestnut Nanking GR119 (TACF)
- (8) Whole plant tissues for two Chinese chestnut Mahogany BX316 (TACF)
- (9) Whole plant tissue for American chestnut Wisneiwski (CAES)
- (10) Whole plant tissue for American chestnut Watertown (CAES)
- (11) Northern red oak tree – Above ground tissues (PSU)
- (12) Northern red oak seedlings –Roots (PSU)
- (13) White oak tree - Above ground tissues (PSU)
- (14) White oak tree seedlings – Roots (PSU)
- (15) Whole plant tissues for healthy American beech tree 1504 (USFS)
- (16) Whole plant tissues for infected American beech tree 1506 (USFS)



Status of the transcriptome sequencing of the Fagaceae species

Sample	System	# Plates	# of Reads	# of bp	AL of Reads	# Contigs	AL of All Contigs	# Large Contigs	AL of Large Contigs
ACCanker	GS20	1	129,508	13,080,308	101	7,171	168	247	689
ACHS1	FLX	3/4	222,939	29,828,910	247	11,496	276	885	817
ACHS2	FLX	3/4	254,810	38,165,054	246	9,431	271	691	816
ACWP1	FLX	1/4	47653	11380607	238.8	4723		239	780
ACWP2	FLX	1/4	33288	7364484	221	3793		239	758
CCCanker	GS20	1	235,635	23,799,135	101	14,308	168	436	773
CCMHS	FLX	3/4	228,594	56,051,191	246	21,828	344	3,074	851
CCNHS	FLX	3/4	259,859	64,271,926	247	28,784	339	4,451	848
CCWP1	FLX	1/4	60,445	14643040	242	4553		307	854
CCWP2	FLX	1/4	53,939	13249397	246	6348		485	827
ABWP1	FLX	1/4	35,424	8112694	174	4163	229	174	790
ABWP2	FLX	1/4	28,829	6599184	229	2467	229	83	824
WOA	GS20+flx	1/4	119,385	28669562	241	14674		1058	764
WOB	GS20+flx	1/2	83,821	19297054	231.5	8025		645	829
ROA	FLX	1/2	131,630	32092278	242	15494		1181	780
ROB	FLX	1/2	145,524	34237369	235	16138		1274	849
Total	both	8	2,184,941	428,799,020	198	93.018	261	9,784	799

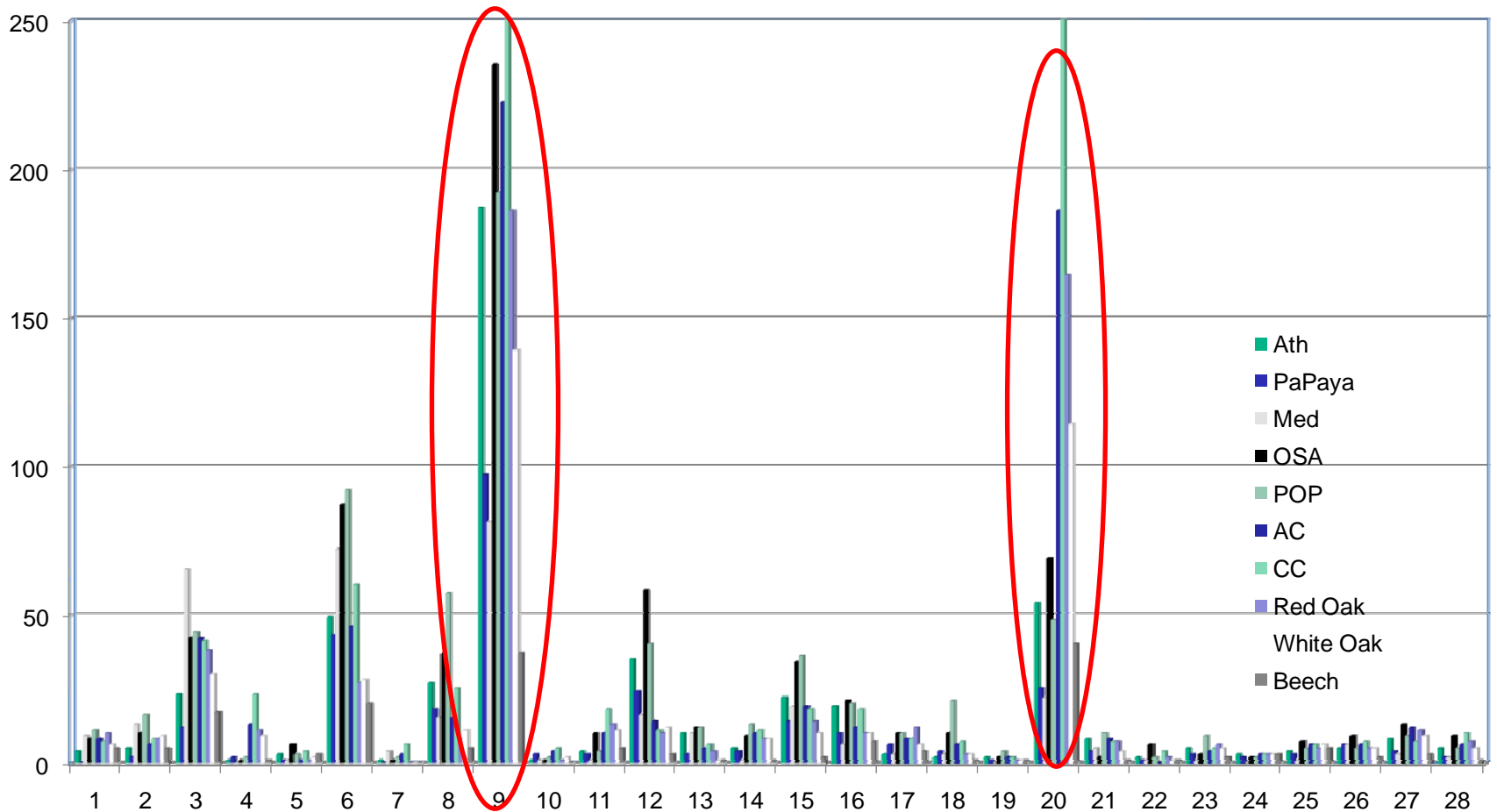
Mass assembly of unigenes from the five Fagaceae species studied

Sample	System	# Plates	# of Reads	# of bp	AL of Reads	# Contigs	AL of All Contigs	# Large Contigs	AL of Large Contigs	# Unigenes	Observed redundancy	# of full length cDNA	#cDNA with L>70%
AC	FLX/GS20	3	688198	146479551	200	46364	374	11431	879	46,364	13.8	344	2031
CC	FLX/GS20	3	838472	171849098	207	32738	501	10016	1038	40,039	19.94	874	5051
AB	FLX	1/2	64253	14705938	229	7123	256	423	787	7,123	8.02	50	308
WO	GS20+flx	1	203206	47966616	238	13626	364	2477	850	19,351	9.50	311	1488
RO	FLX	1	277154	66339647	241	17502	394	3693	889	24,846	10.15	361	1964
Total	both	8.5	2071283	447340850		117353		28040		137,723	14.04	1679	9662

The size of gene families in the Fagaceae species is proportional to ones from other plant species

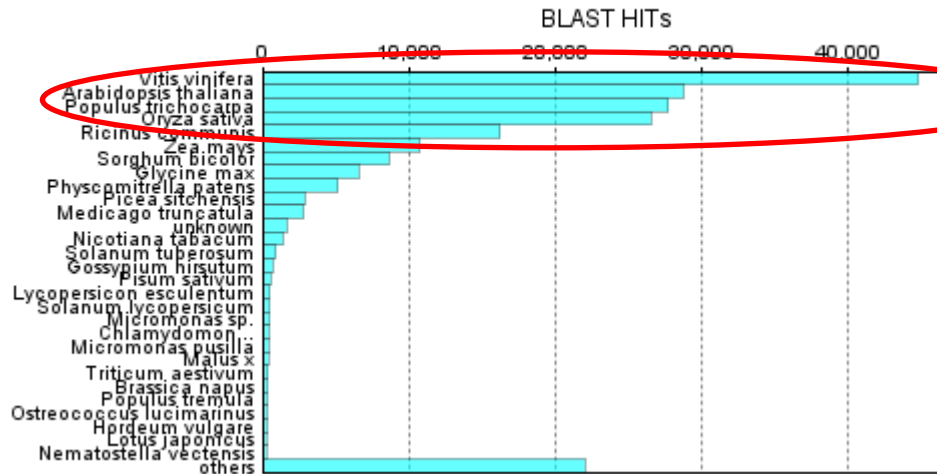
Accession	#AC	#CC	Annotation	Ath	PaPaya	Med	OSA	POP	AC	CC	Red Oak	White Oak	Beech
ATCG00120	5	3	Encodes the ATPase alpha subunit	4	0	9	8	11	8	7	10	6	5
ATCG00480	3	1	chloroplast-encoded gene for beta subunit of ATP synthase	5	2	13	10	16	6	8	8	9	5
AT1G09080	3	1	luminal binding protein 3 (BiP-3) (BP3)	23	12	65	42	44	42	41	38	30	17
AT1G09770	4	0	myb family transcription factor	1	2	0	1	2	13	23	11	9	1
AT1G09780	3	1	2,3-biphosphoglycerate-independent phosphoglycerate mutase	3	1	1	6	3	1	4	1	2	3
AT1G15520	6	1	PLEIOTROPIC DRUG RESISTANCE 12	49	43	72	87	92	46	60	27	28	20
AT1G17870	3	1	S2P-like putative metalloprotease	1	0	4	1	2	3	6	0	0	0
AT1G19320	3	0	pathogenesis-related thaumatin family protein	27	18	15	37	57	15	25	8	11	5
AT1G30270	5	1	similar to CBL-interacting protein kinase 9 (CIPK9)	187	97	81	235	192	222	250	186	139	37

Gene family size is similar in the Fagaceae and several model species

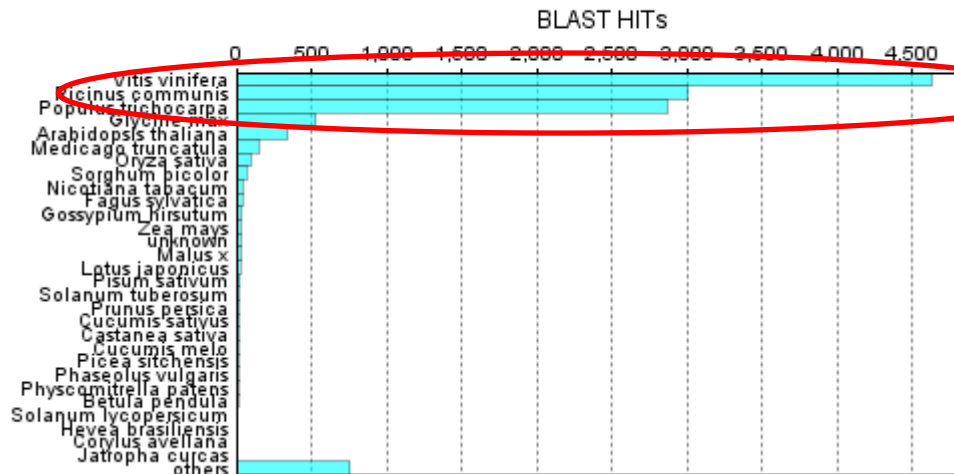


Top-hit species are from Eurosid I

Species distribution

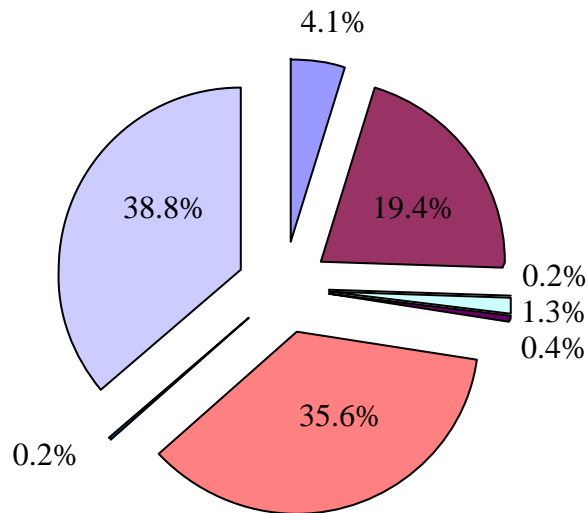


Top-Hit species distribution

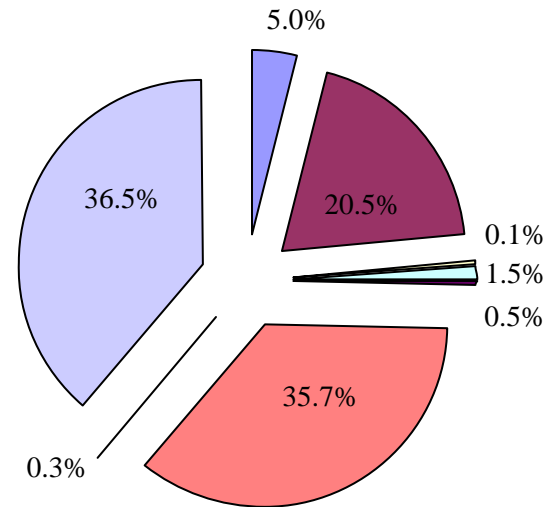


A large fraction of the Fagaceae unigenes has a best hit on tree sequences

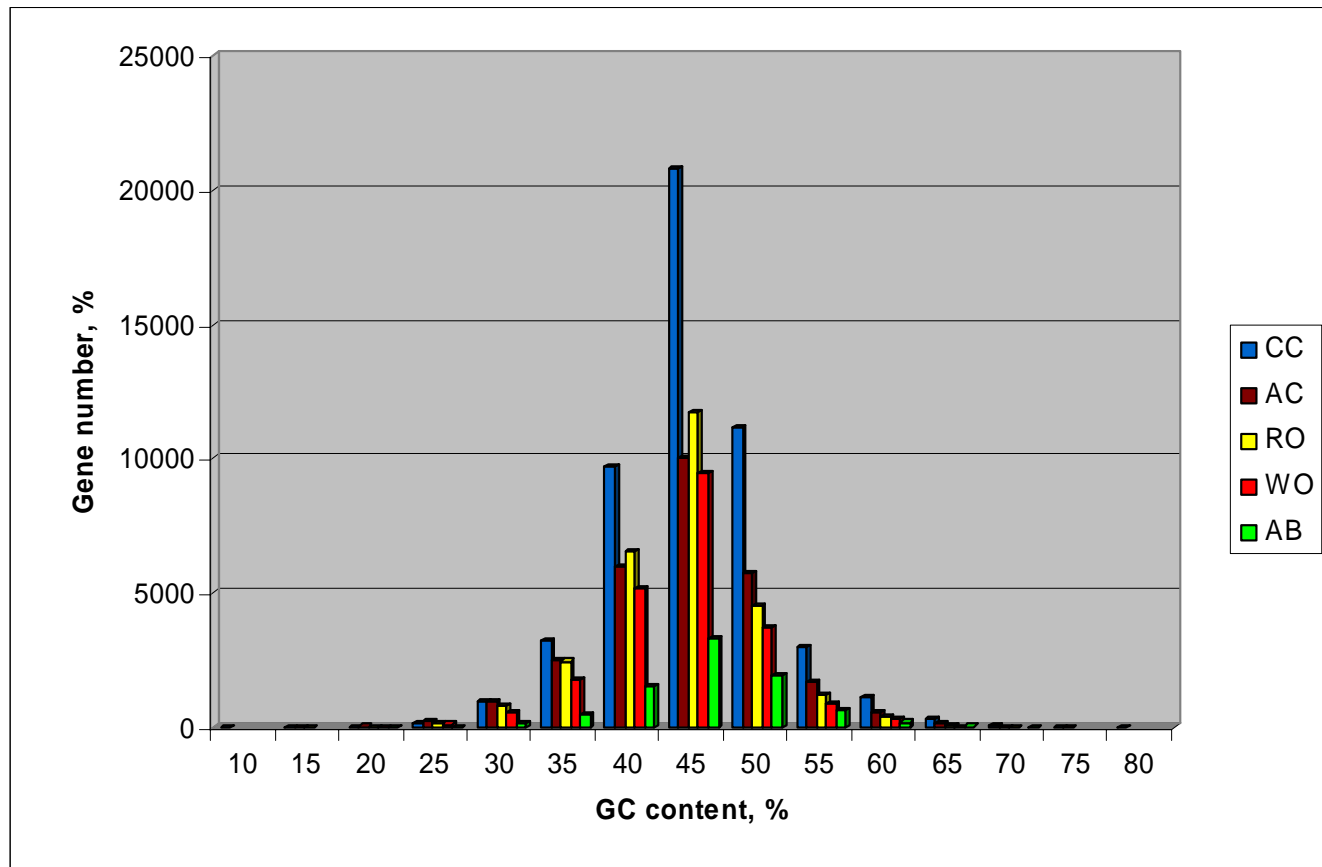
American chestnut



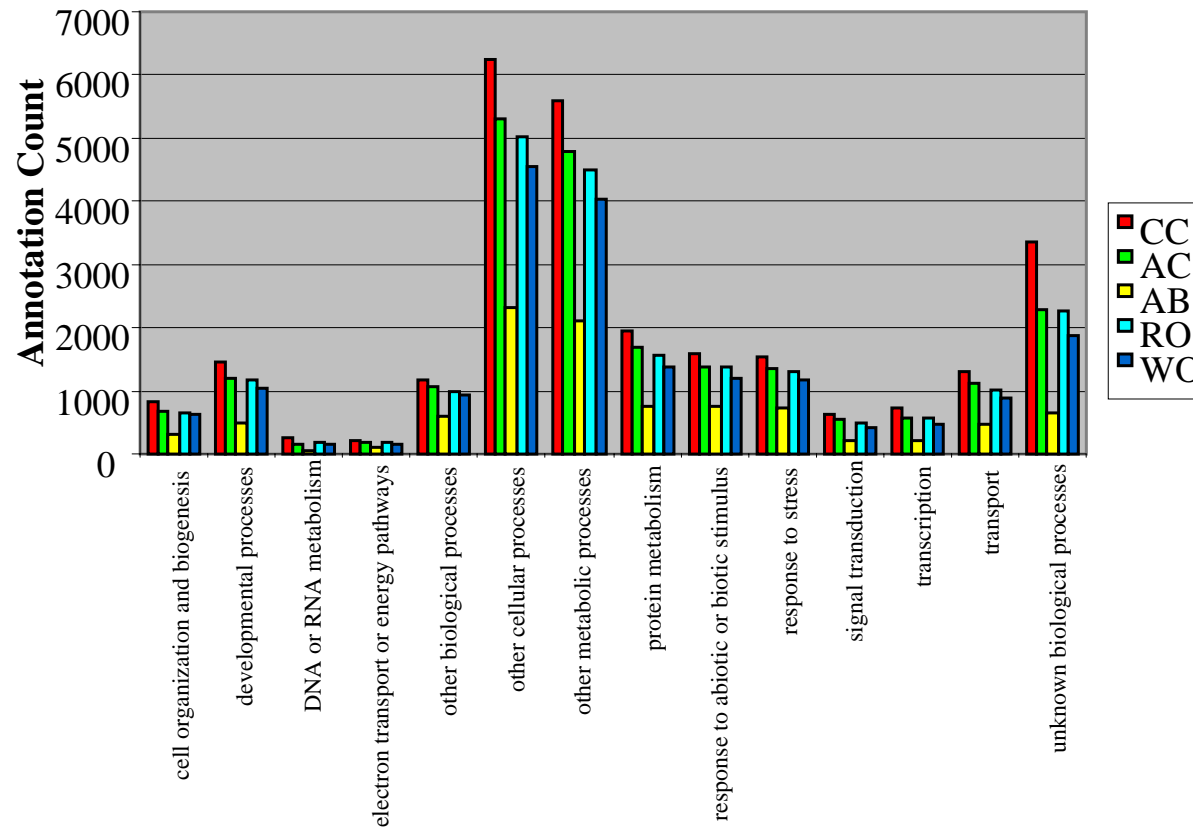
Chinese chestnut



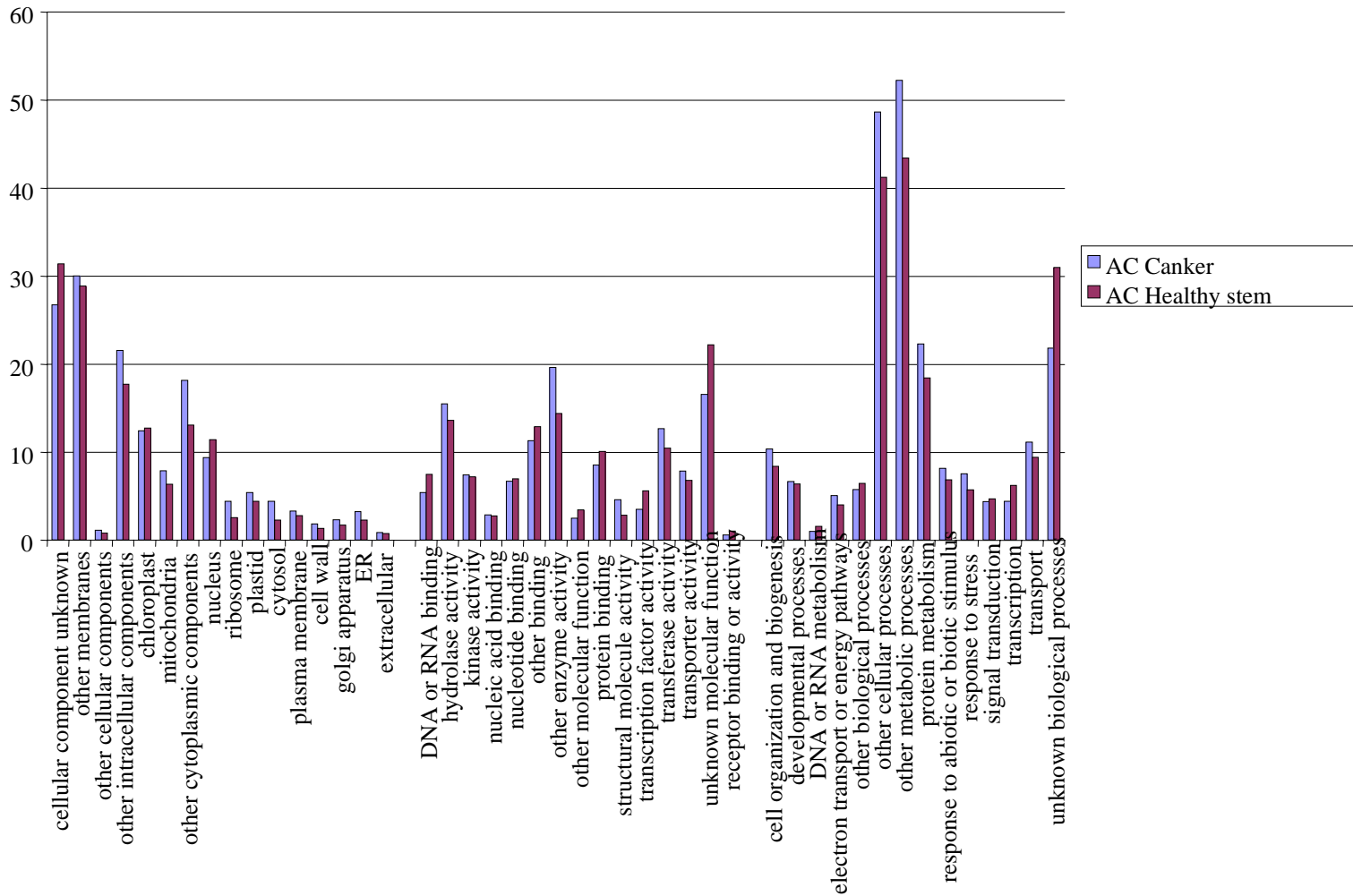
The Fagaceae species present similar GC content distribution as *Arabidopsis* and *Populus*



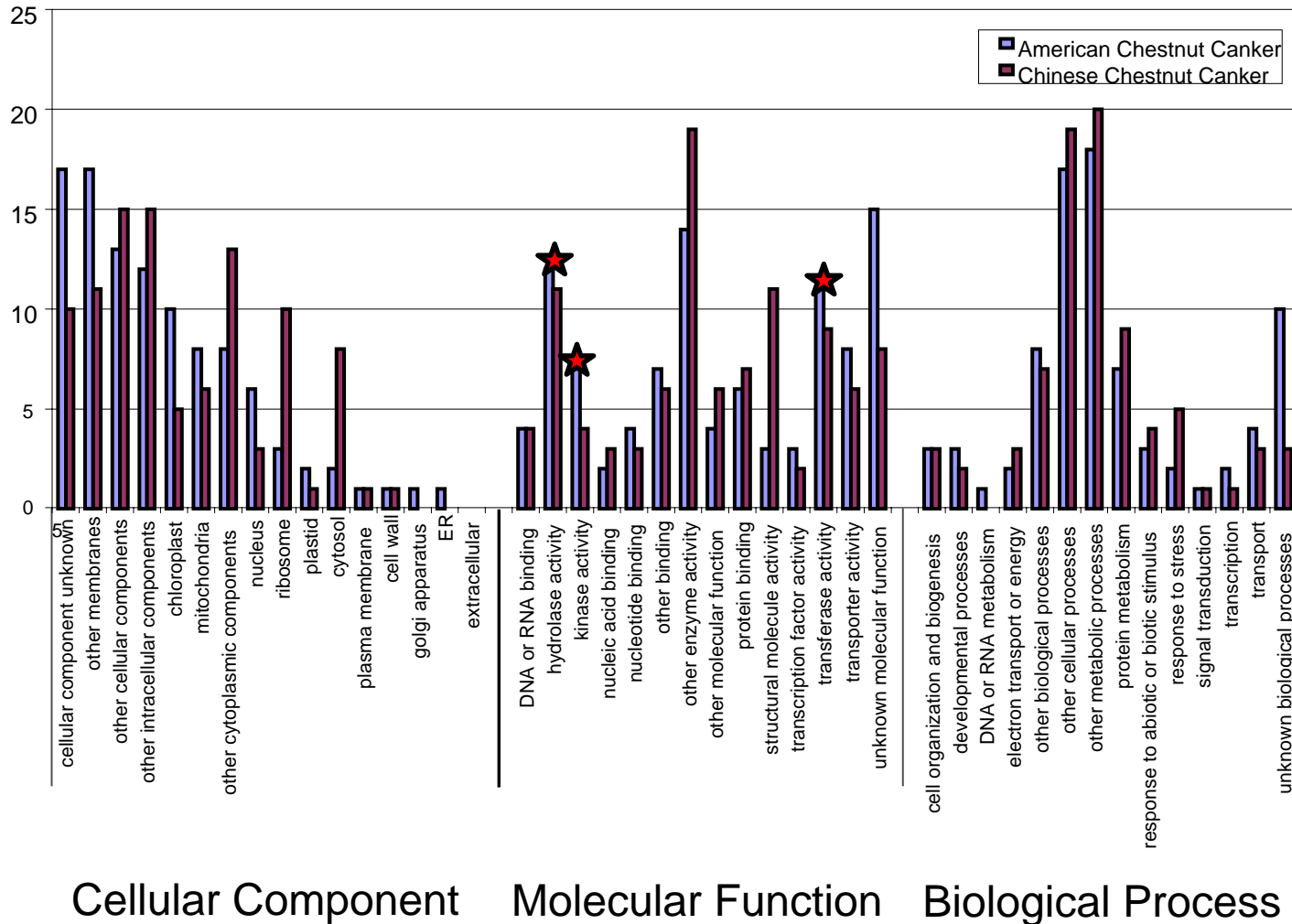
Unigenes from Fagaceae species cover various biological processes



Similar pattern of GO-annotation function distribution for ACC and ACHS



Hydrolase and transferase functions are highly represented



Transcriptome comparison

Goals: Identify genes induced in ACC/CCC, ACC/ACHS, and CCC/CCHS.

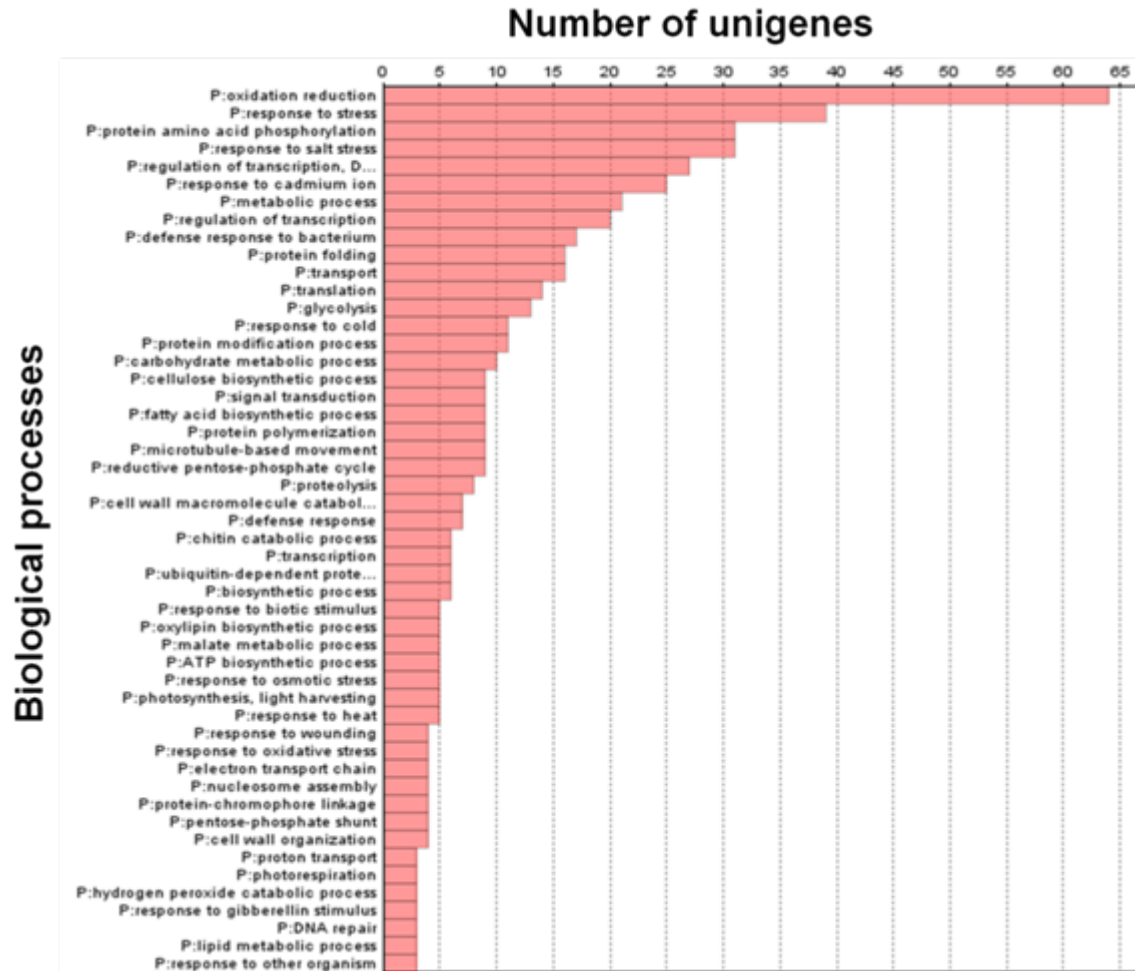
Approach:

- Blast contigs on Arabidopsis
- Identify defense-related genes
- Compare the number of reads from “ortholog” genes
- Identify metabolic pathways
- Confirm the expression of candidate genes

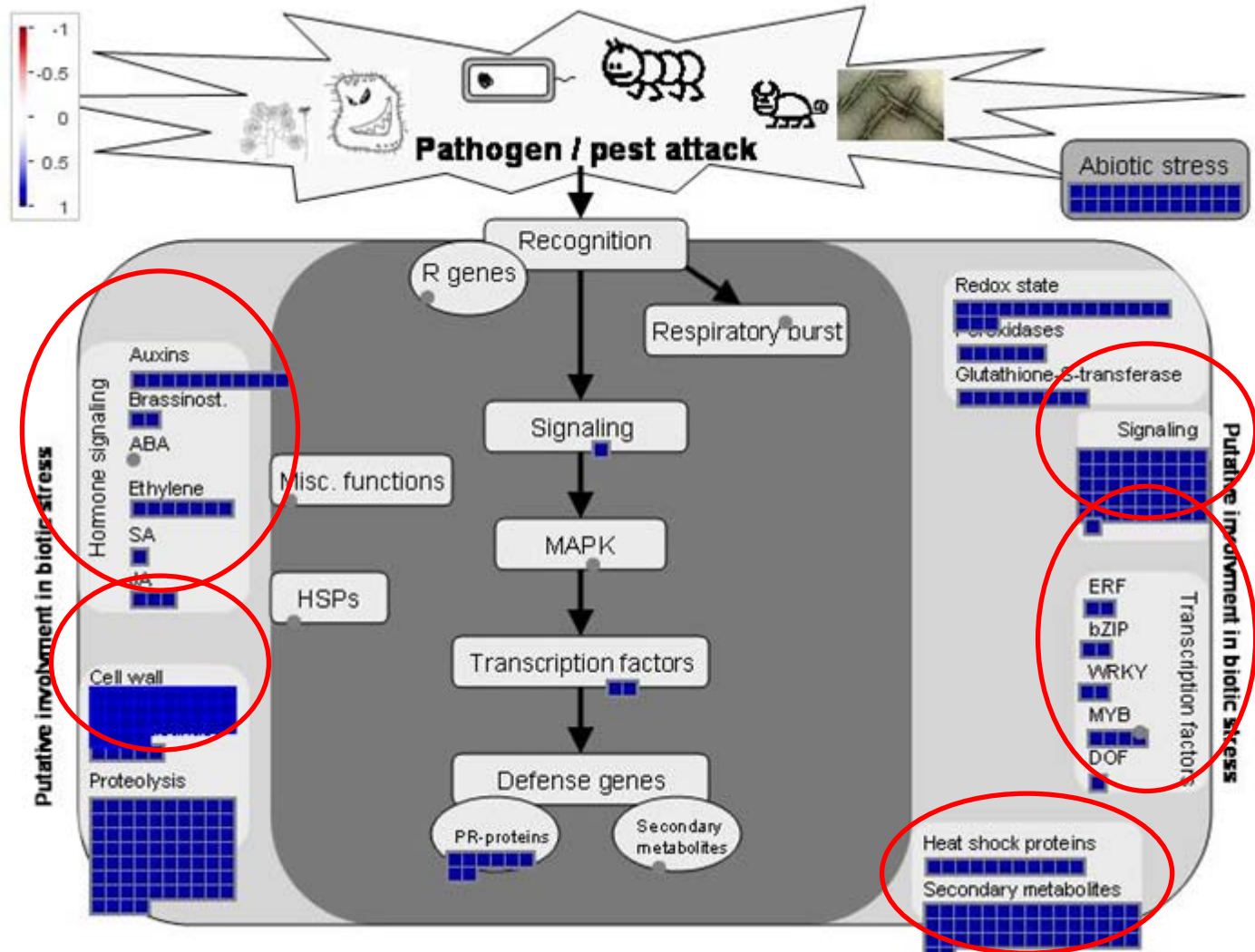
Disease and Defense Response Genes More Highly Expressed in Infected Tissues of Chinese Chestnut (CC) than in American Chestnut (AC). (*) significant at 95% conf' level. Barakat et al. (2009)

Arabidopsis Accession #	# CC Canker reads	% CC Canker transcriptome	# AC Canker reads	% AC Canker transcriptome	Description (Annotation in Arabidopsis)
AT5G36880.2	12	0.51%	2	0.15%	AMP binding / acetate-CoA ligase/ catalytic
*AT5G18650.1	12	0.51%	0	0.00%	zinc finger (C3HC4-type RING finger) family protein
*AT4G37640.1	11	0.47%	0	0.00%	calcium-transporting ATPase 2
ATCG00190.1	10	0.42%	4	0.31%	Chloroplast DNA-dependent RNA polymerase B subunit
*AT3G42050.1	9	0.38%	0	0.00%	vacuolar ATP synthase subunit H family protein
*AT5G07390.1	8	0.34%	0	0.00%	respiratory burst oxidase protein A (RbohA) / NADPH oxidase
AT1G63770.1	7	0.30%	2	0.15%	peptidase M1 family protein
ATCG01010.1	7	0.30%	2	0.15%	Chloroplast encoded NADH dehydrogenase unit
AT2G20650.2	7	0.30%	0	0.00%	zinc finger (C3HC4-type RING finger) family protein
AT5G14450.1	7	0.30%	0	0.00%	GDSL-motif lipase/hydrolase family protein
AT5G58860.1	7	0.30%	0	0.00%	cytochrome P450 86A1 (CYP86) (CYP86A1)
AT2G43790.1	7	0.30%	0	0.00%	mitogen-activated protein kinase
AT2G01060.1	7	0.30%	0	0.00%	myb family transcription factor
AT3G01420.1	7	0.30%	0	0.00%	pathogen-responsive alpha-dioxygenase
AT5G01220.1	6	0.25%	3	0.23%	DAG sulfoquinovosyltransferase
ATCG00180.1	6	0.25%	2	0.15%	RNA polymerase beta' subunit-1
AT1G14810.1	6	0.25%	1	0.08%	semialdehyde dehydrogenase family protein
AT1G63660.2	6	0.25%	1	0.08%	GMP synthase (glutamine-hydrolyzing)
AT3G08590.2	6	0.25%	1	0.08%	2,3-biphosphoglycerate-independent phosphoglycerate mutase
AT3G60750.1	6	0.25%	1	0.08%	transketolase, putative
AT1G53510.1	6	0.25%	0	0.00%	similar to mitogen-activated protein kinase
AT5G60020.1	6	0.25%	0	0.00%	laccase
ATCG00120.1	5	0.21%	3	0.23%	Encodes the ATPase alpha subunit
AT2G26080.1	5	0.21%	2	0.15%	glycine dehydrogenase (decarboxylating)
AT4G33010.1	5	0.21%	2	0.15%	glycine dehydrogenase (decarboxylating)
AT5G10330.2	5	0.21%	2	0.15%	EMB2196 (EMBRYO DEFECTIVE 2196)
AT5G42740.1	5	0.21%	2	0.15%	glucose-6-phosphate isomerase
AT2G22250.3	5	0.21%	1	0.08%	AAT/ATAAT/MEE17 (maternal effect embryo arrest 17)
AT2G38290.1	5	0.21%	1	0.08%	ammonium transporter
AT3G10050.1	5	0.21%	1	0.08%	OMR1 (L-O-METHYLTHREONINE RESISTANT 1)
AT3G22200.1	5	0.21%	1	0.08%	POP2 (POLLEN-PISTIL INCOMPATIBILITY 2)
AT4G24830.1	5	0.21%	1	0.08%	arginosuccinate synthase family
AT5G08530.1	5	0.21%	1	0.08%	NADH-ubiquinone oxidoreductase 51 kDa subunit
AT5G11770.1	5	0.21%	1	0.08%	NADH-ubiquinone oxidoreductase 20 kDa subunit
AT5G37510.2	5	0.21%	1	0.08%	EMB1467 (EMBRYO DEFECTIVE 1467); NADH dehydrogenase
AT5G63890.2	5	0.21%	1	0.08%	ATHDH (HISTIDINOL DEHYDROGENASE)
AT1G01190.1	5	0.21%	0	0.00%	cytochrome P450
AT5G04885.1	5	0.21%	0	0.00%	glycosyl hydrolase family 3 protein
AT5G13410.1	5	0.21%	0	0.00%	immunophilin / FKBP-type peptidyl-prolyl cis-trans isomerase family protein
AT1G48090.1	5	0.21%	0	0.00%	C2 domain-containing protein

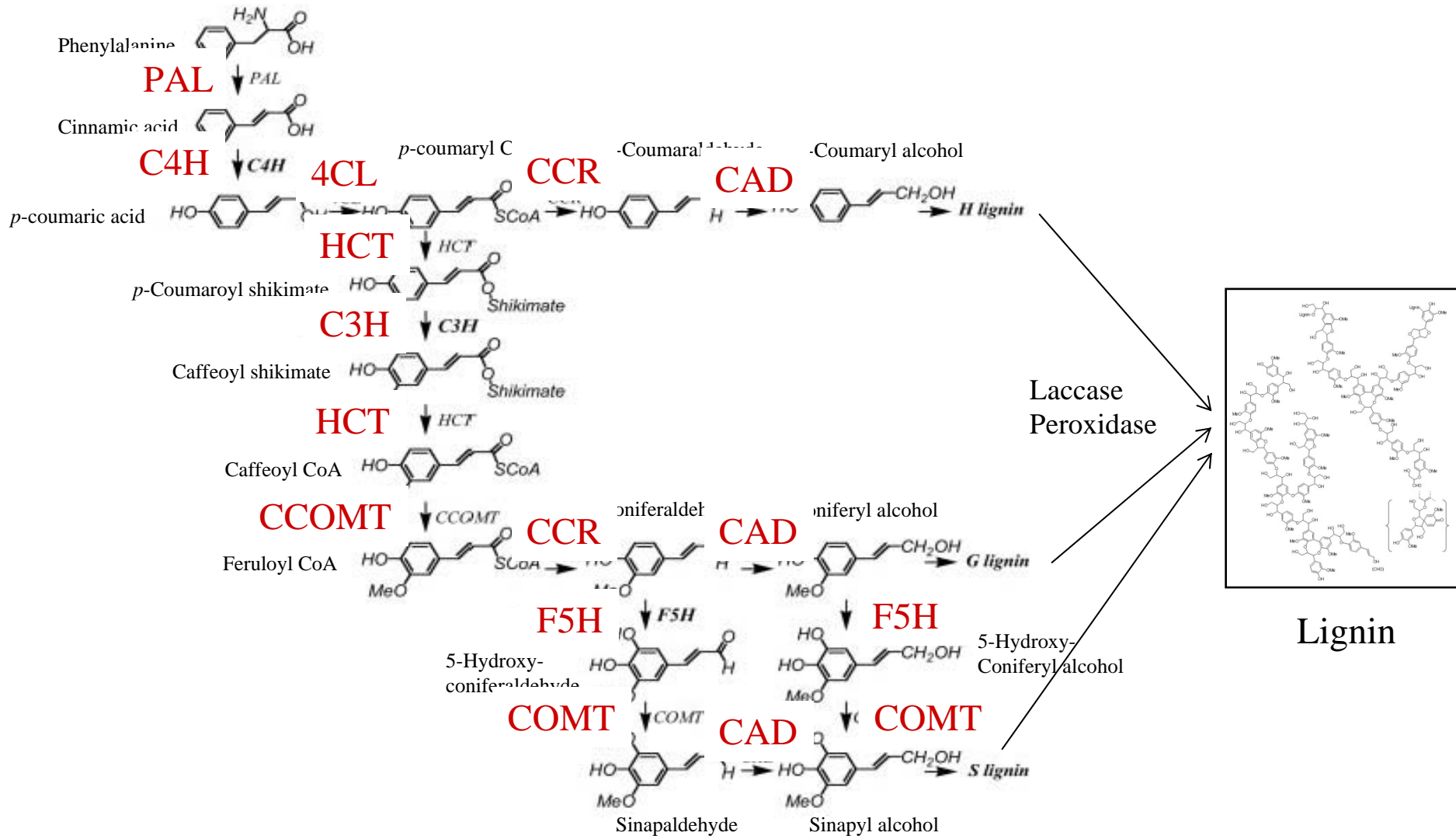
Defense related genes induced following *Cryphonectria* infection belong to various biological processes



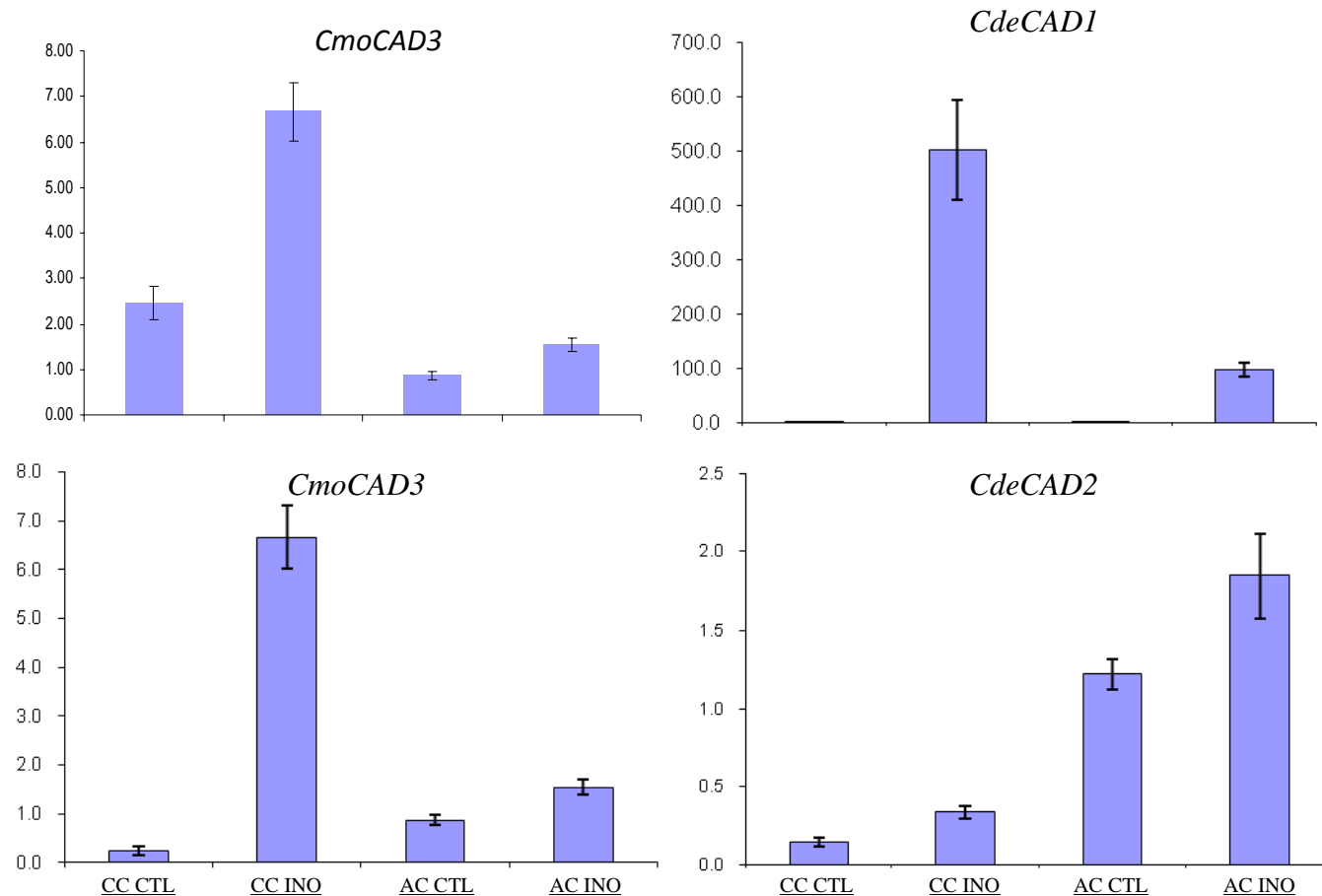
Blight disease related defense genes in American and Chinese chestnut



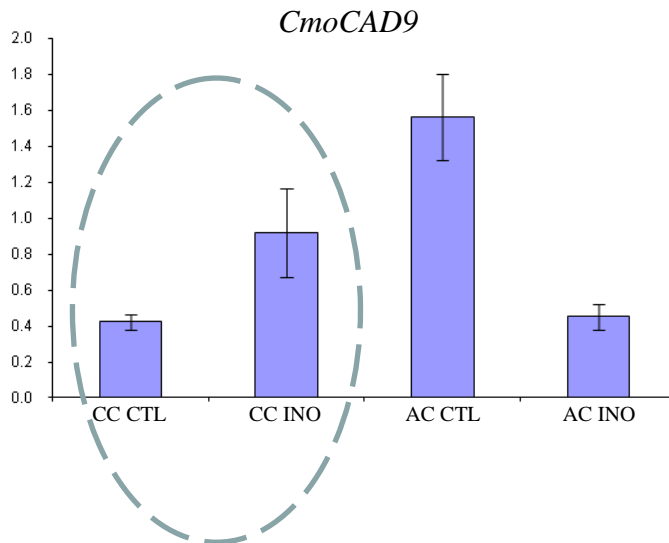
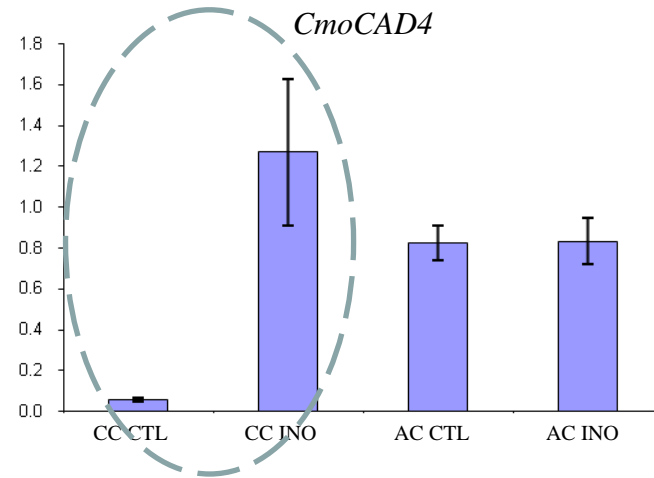
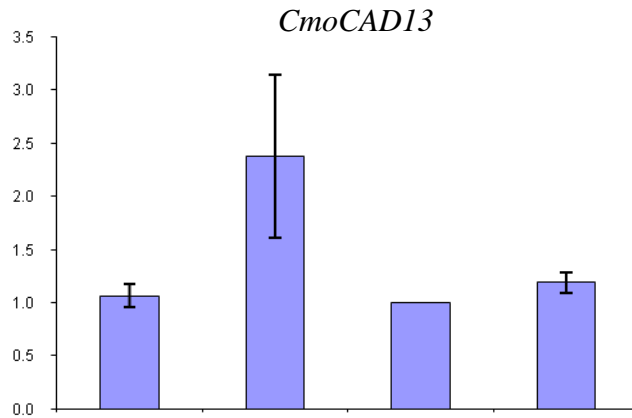
Lignin biosynthesis related genes were highly



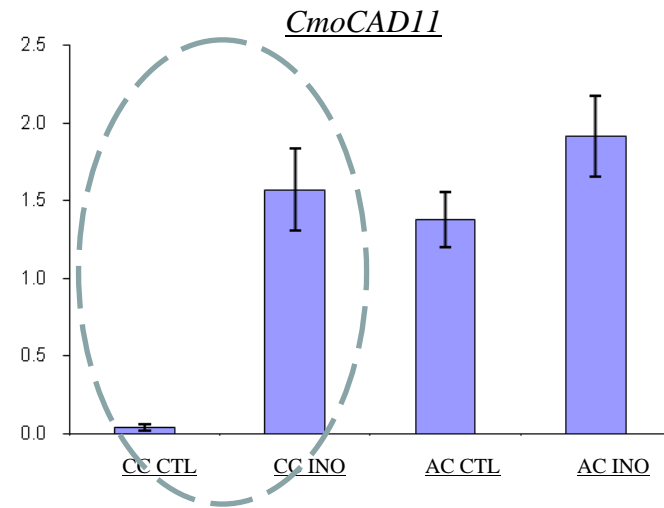
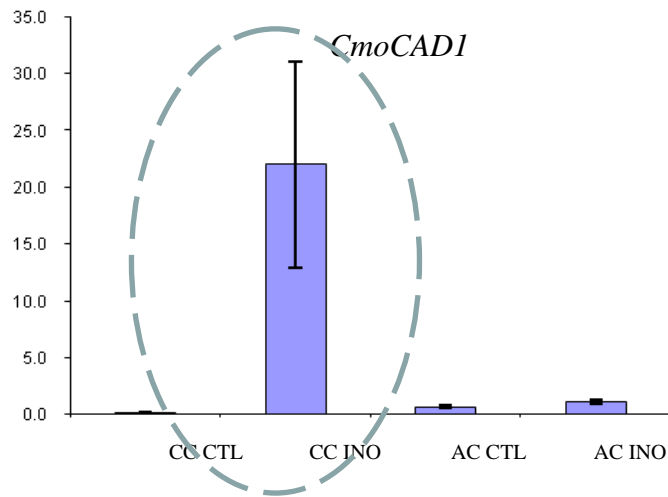
Several CAD genes were induced following *Cryphonectria* infection



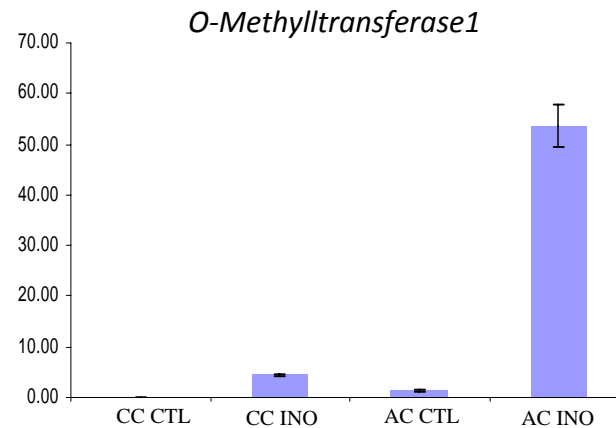
Several CAD genes were induced following *Cryphonectria* infection

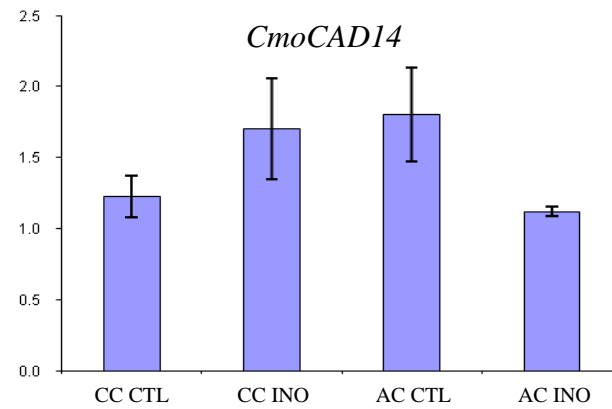
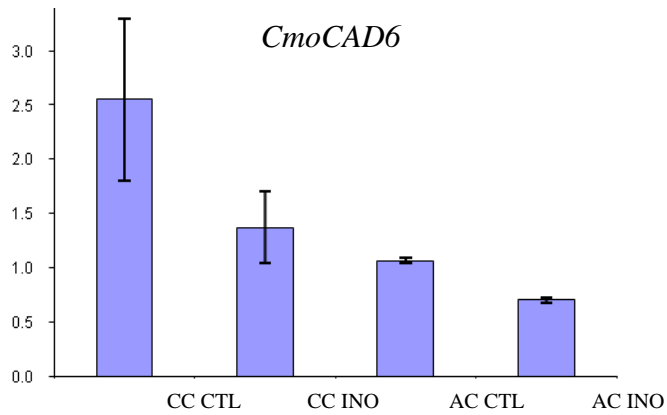
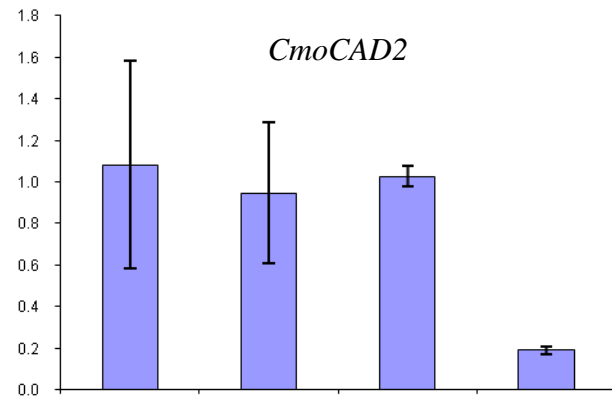
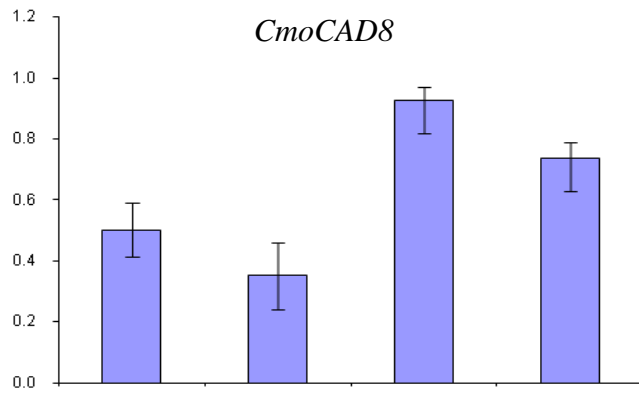
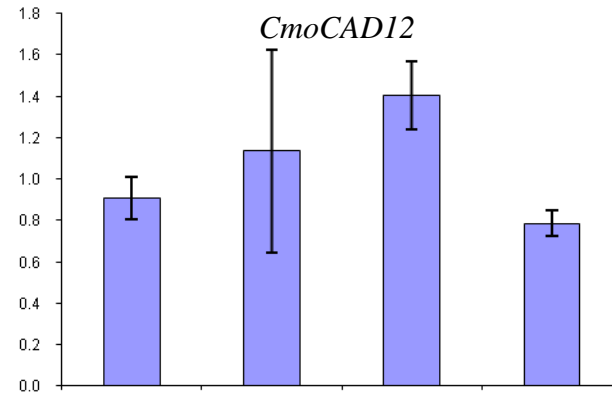
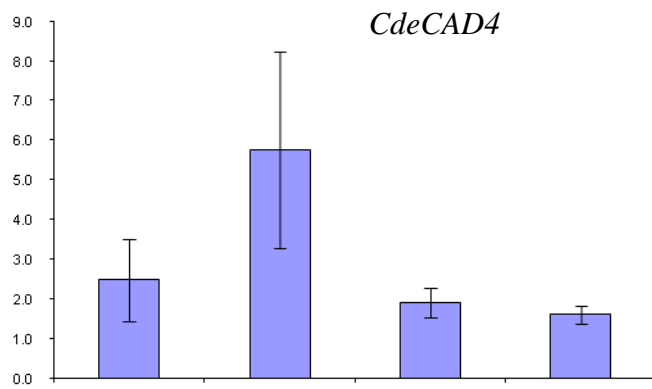


Several CAD genes were induced following *Cryphonectria* infection

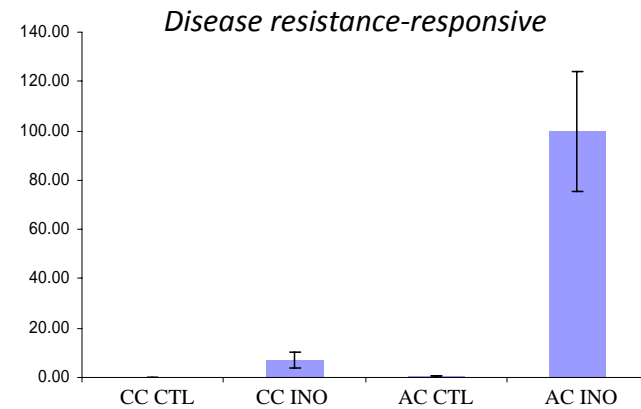
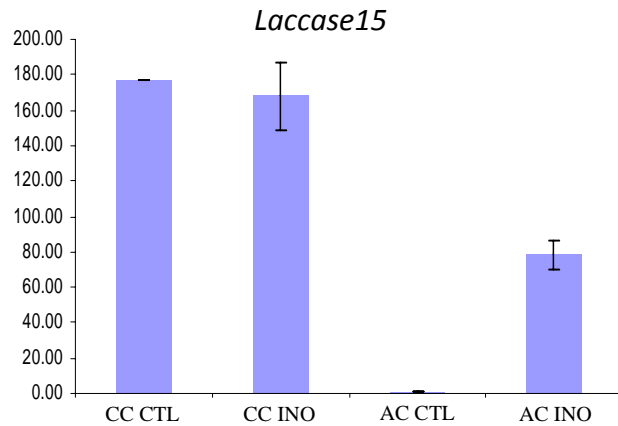


Other lignin biosynthesis related genes were induced following *Cryphonectria* infection

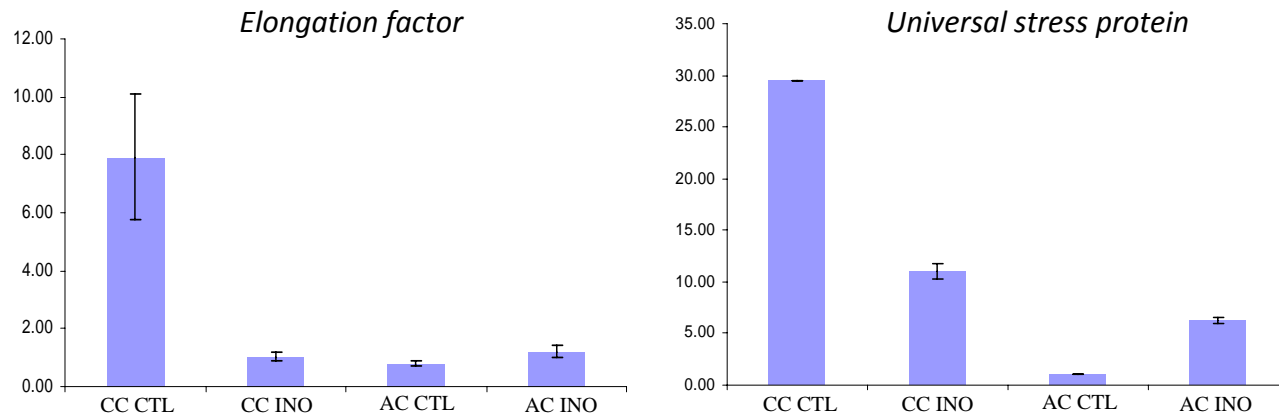




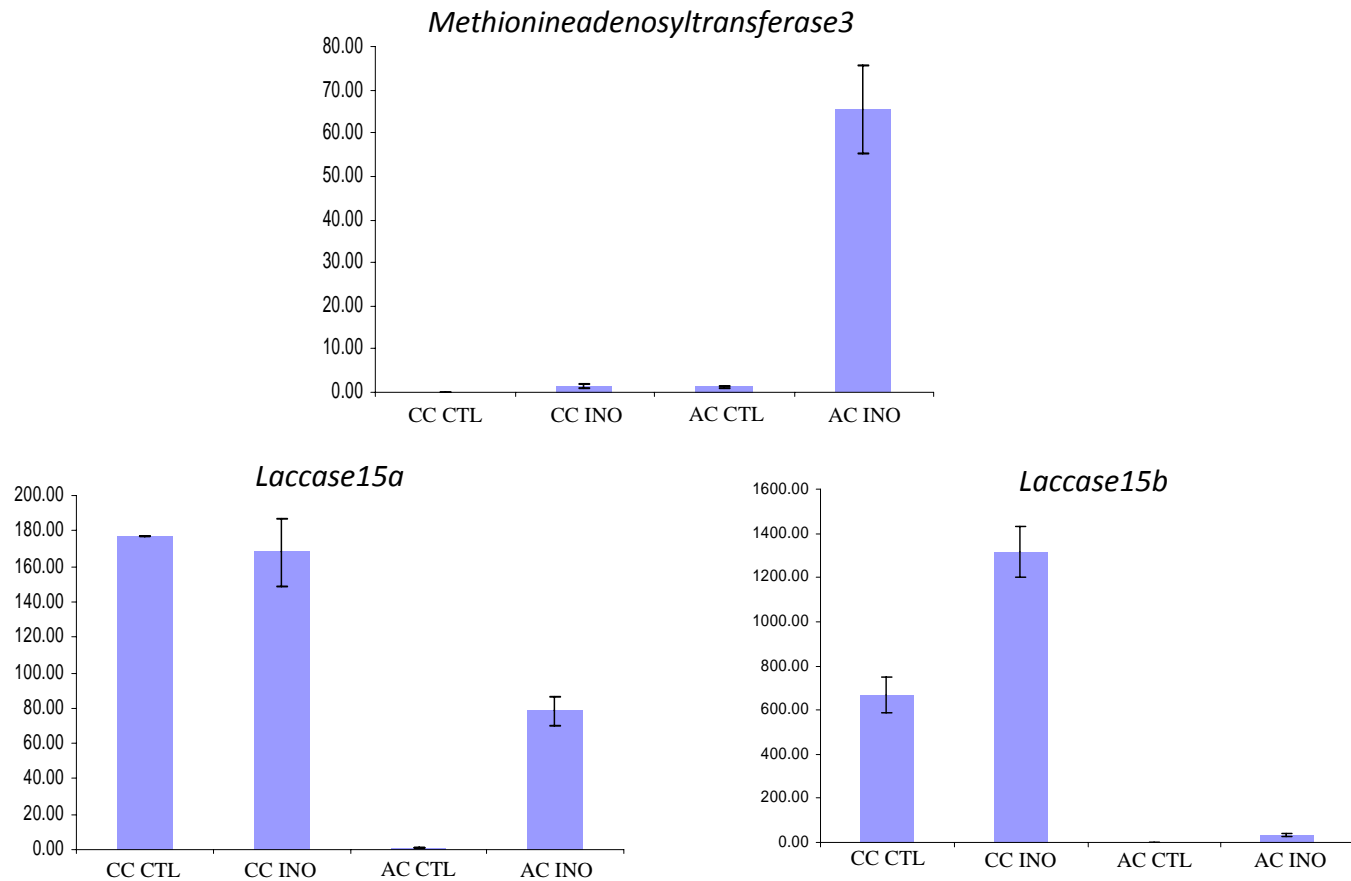
Several defense related genes were induced following *Cryphonectria* infection



Several defense related genes were induced following *Cryphonectria* infection



Several defense related genes were induced following *Cryphonectria* infection



Analyses of miRNAs in cDNAs from American and Chinese chestnut

• In silico analysis

- ✓ Search for conserved miRNAs in cDNA from American and Chinese chestnut
- ✓ Annotation of miRNAs : Check biogenesis and expression criteria
- ✓ Search for targets in American and Chinese chestnut

• Small RNA sequencing

- ✓ Sequence of small RNA libraries

Conserved miRNAs in American and Chinese chestnut

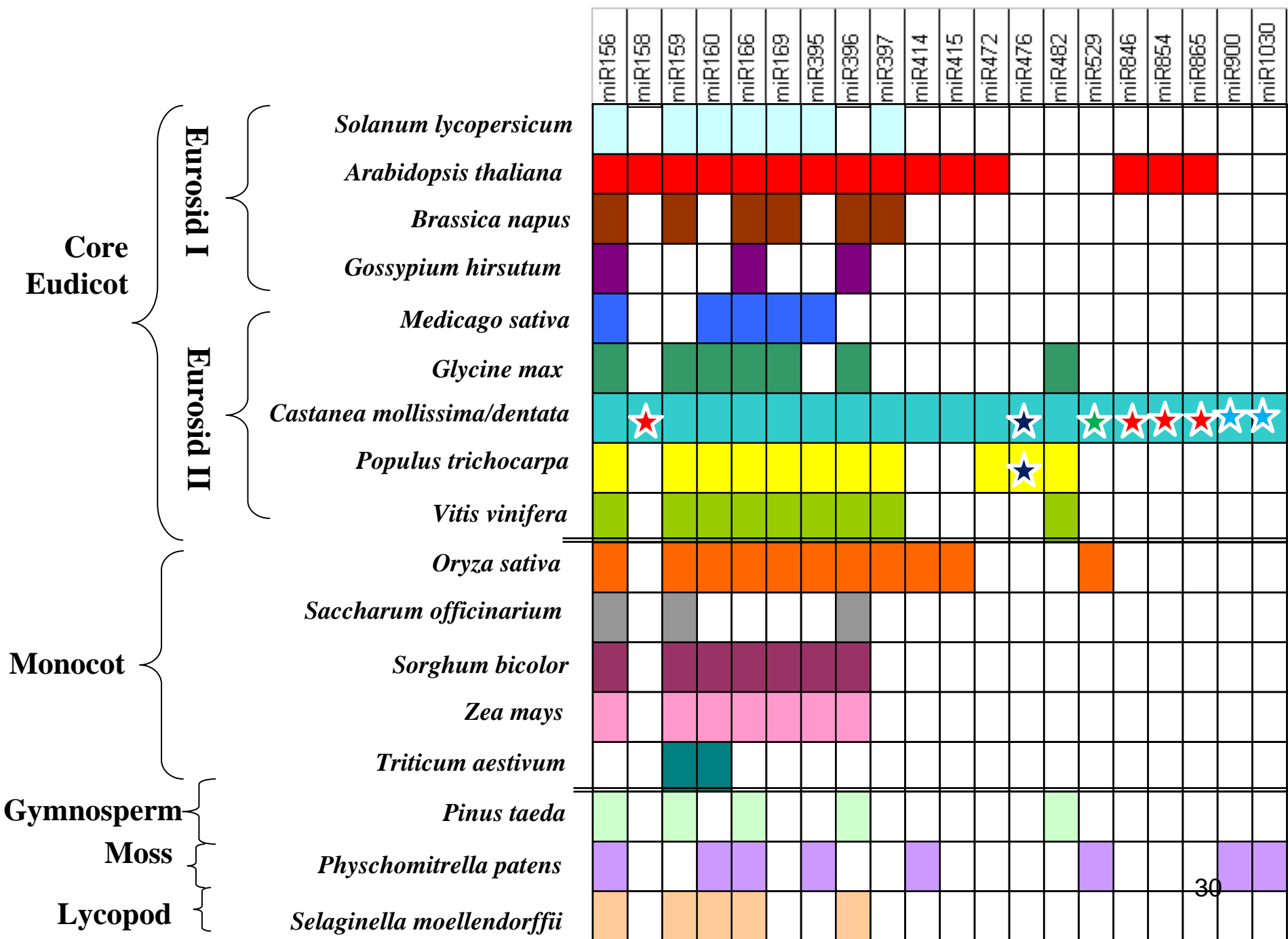
✓ Chinese chestnut

- 172 unique sequences
- 19 conserved miRNA families identified

✓ American Chestnut

- 114 unique miRNA sequences identified
- 9 non-conserved miRNA families identified

Family	# of Members	# of miRcheck Pass
miR1030	4	1
miR156	6	2
miR158	3	0
miR159	5	1
miR160	2	0
miR166	4	2
miR169	5	1
miR396	5	0
miR397	2	0
miR414	101	16
miR415	3	0
miR476	4	0
miR482	8	2
miR529	8	1
miR846	2	0
miR854	2	0
miR865	5	1
miR900	3	0



miRNA targets include several genes involved in resistance to biotic and abiotic stresses

miR414

DEAD box RNA helicase	At1g20920
F-box protein family	At1g15670
SNF2 domain/helicase domain-containing protein	At3g42670
Nucleosome assembly protein	At4g26110

miR415

PPR

miR476

PPR

miR482, miR472

Disease resistance gene

miR166

HD-ZIPIII

miRNA targets include several genes involved in resistance to biotic and abiotic stresses

miR395

DNA mismatch repair protein	At4g02460
DEAD box RNA helicase	At3g06480
SNF8 like protein	At4g27040
ATP sulfurylases	At3g22890

miR397

Laccase	At2g38080
ABC transporter family protein	At3g53480
PPR	At3g59040

Acknowledgements

Penn State:

John Carlson

Scott DiLoreto

Zchicho Sun

Norzawani Buang

Joseph Park

Haiying Liang

Kerr Wall/Claude depamphilis

Stephan Schuster

SUNY ESF, Syracuse:

William Powell

Kathleen Baier

Clemson University:

Meg Staton

Bert Abbott

North Carolina State University, Bioinformatics Research Center

Ron Sederoff, Nick Wheeler, Chris Smith

The American Chestnut Foundation:

Fred Hebard

Paul Sisco

The Connecticut Agricultural

Experiment Station:

Sandy Anagnostakis

This project is supported by grants from the NSF Plant Genome Research Program (DBI-PGRP-TRPGR 0605135) , The American Chestnut Foundation , and The Schatz Center for Molecular Genetics

On-going research activities

- Gene network analyses in American and Chinese chestnut
- Validation of candidate genes associated with resistance using real time RT-PCR
- Sequencing and isolation of miRNAs
- Expression profiling of miRNAs targeting defense related genes

Summary of SSR and SNP markers selected from the Fagaceae sequences

	total SSRs	SSRs with primers	SSRs with EP*	SSRs with primers and EP*	SNPs above 70% likelihood	SNPs above 70% likelihood with primers
AB454_v2	1319	442	37	13	624	536
ABall_v2	1345	451	39	11	607	524
AC454_v2	6605	2050	329	69	5962	5369
CC454_v2	12116	3196	835	114	15680	14412
CCall_v2	12539	3226	899	128	25904	23892
Oall_v2	10298	1638	1024	75	31345	28731
RO454_v2	6187	1358	419	56	12152	11289
WO454_v2	4350	931	293	28	8589	7918

Candidate Chestnut genes for plant resistance

- Hypersensitivity (control of the cell death)
 - *ABC transporter, C2-domain-containing gene, methylenetetrahydrofolate reductase, elongation factor-1 alpha, ATPase, and peroxidase*
- Systemic Resistance
 - *omega-3 fatty acid desaturase, suppressor of fatty acid desaturase deficiency (SFD1 and SFD2), Ras-related GTP-binding*

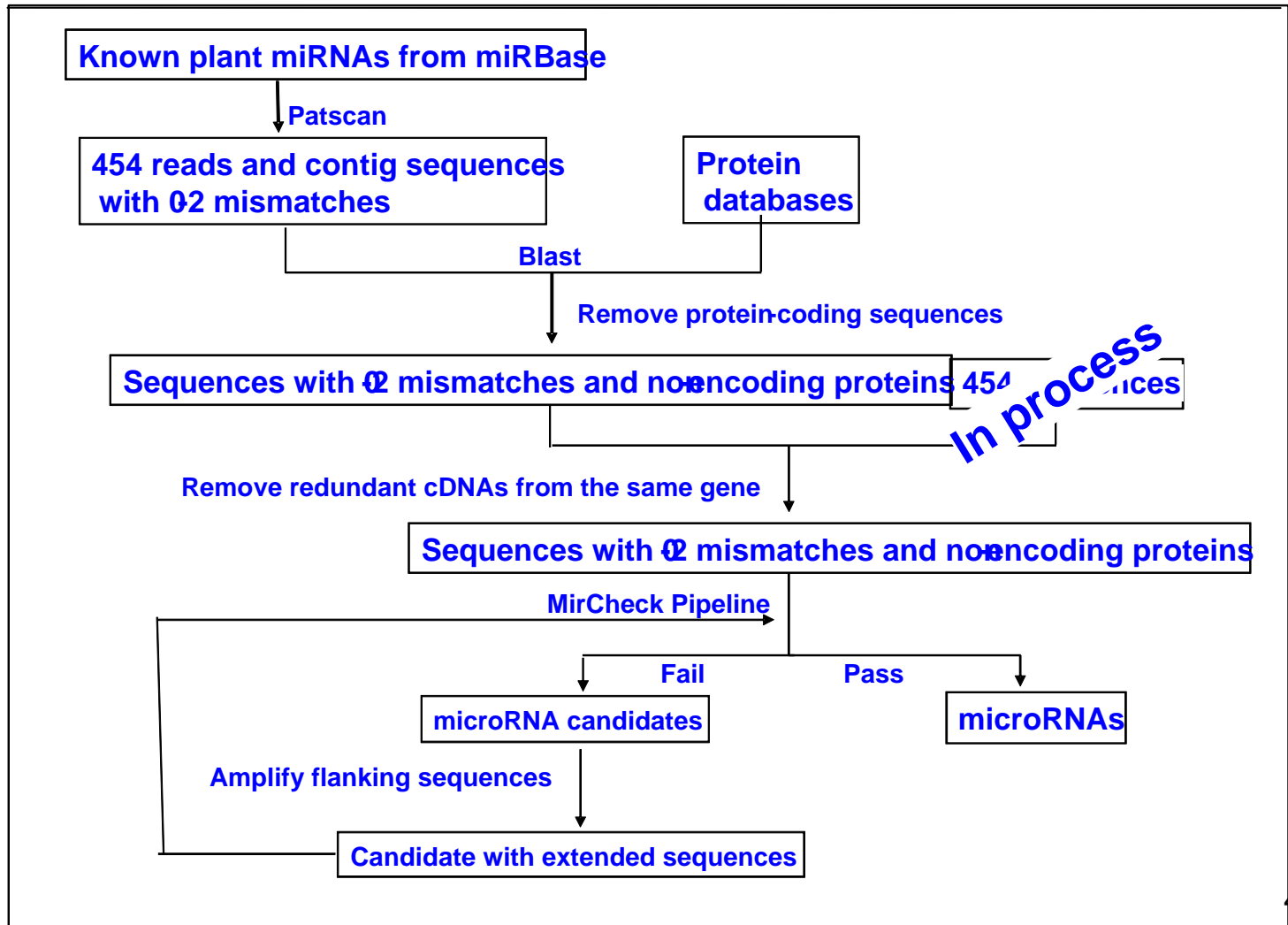
Other Chestnut induced defense-related genes

- Constitutive and Jasmonic-acid dependent defense
 - *Myb*, *ATP-binding cassette transport*, *ABC transporter family*, and *AP2*
- Signaling
 - Mitogen activated protein
- Others
 - *Acetyl co-enzyme A carboxyltransferase (CAC3)*, *SNF*, and several transcription factors such as *WRKY*, *Zinc finger*, *Myb*, etc were identified.

Combination of 454 pyro-sequencing and capillary sequencing: Benefit for cDNA contig construction

Library	Reads included in contigs	Reads not included in a contig	# Contigs	Average Contig Length
AB 454	40532	23721	8226	363.9
AB Hybrid	40243	24360	8319	365.4
WO_454	161920	41286	22102	433.4
RO 454	230723	46431	28041	454.9
Oak Hybrid	424611	65169	40272	512.9
AC 454	322652	76131	34800	412.3
CC 454	733010	105462	48501	525.9
CC Hybrid	743413	104539	48335	542.5

Pipeline for searching for miRNAs in cDNA sequences



*cDNAs prepared
for 454 sequencing:*



1. American Chestnut canker
2. Chinese Chestnut canker
3. Chinese Chestnut 'Mahogany' healthy stem cDNA
4. Chinese Chestnut 'Nanking' healthy stem cDNA
5. American Chestnut healthy stem tree 1 (AC 4T7 stem 1)
6. American Chestnut healthy stem tree 2 (AC 1T6 stem 2)
7. American Chestnut whole plant pool tree 1 (AC 477)
8. American Chestnut whole plant pool tree 2 (AC 1T6)
9. Chinese Chestnut whole plant pool tree 1 (CC BX316)
10. Chinese Chestnut whole plant pool tree 2 (CC GT119)
11. Red Oak-above ground cDNA, trees 1 and 2
12. Red Oak-below ground (root) cDNA, trees 1 and 2
13. Red Oak-above ground cDNA, trees 1 and 2
12. Red Oak-below ground (root) cDNAs trees 1 and 2
13. American beech, whole plant pool tree 1, healthy
14. American beech, whole plant pool tree 2, diseased