Fagaceae Comparative Analysis

Meg Staton

BAC End Sequences

- Mimimum Tile: 15,802 BACs
- Randomly selected 6,048 BACs to sequence now
- Yielded 10,322 high quality sequences, average length of 554
- For 4,511 BACs we have both ends

Why sequence BAC ends?

- Scaffold of sequence to associate whole genome sequence to physical map.
- Longer sequences to help pull together short reads into a cohesive assembly
- Early glimpse of the genome for better planning
- Marker-rich resource

Repeats

- 20.4% of sequences show low complexity
- 12.3% of sequences match known plant repeats

Transposable elements	1995	
DNA transposon	233	11.7%
LTR Retrotransposon	1347	67.5%
Endogenous Retrovirus	34	1.7%
Non-LTR Retrotransposon	415	20.8%

BES repeats

Matching elements

ATLANTYS1_I	59
SIRE1_INT	59
SHALINE10_MT	57
Gypsy3-VV_I	49
COPIA3-I_LC	43
Gypsy16-VV_I	35
Gret1_I	34
SHACOP14_I_MT	33
VLINE6_VV	29
Copia-34-I_VV	27

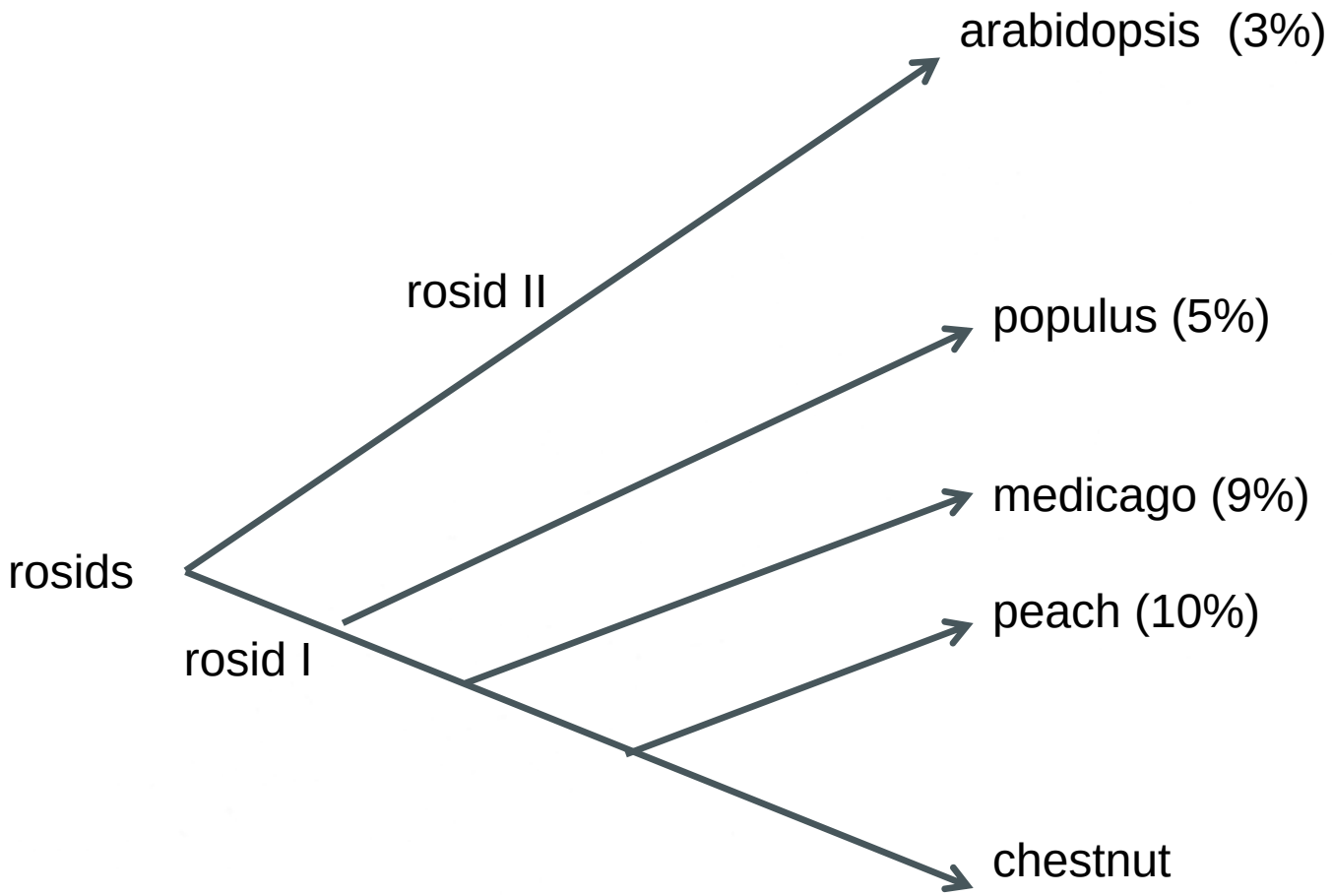
By organism

Eukaryota	1050	10.2%
Vitis vinifera	449	4.3%
Medicago truncatula	392	3.8%
Arabidopsis thaliana	300	2.9%
Populus trichocarpa	197	1.9%
Oryza sativa	173	1.7%
Zea mays	152	1.5%

Gene Coverage

- Augustus gene prediction
- BLAT of unigene as “evidence”
- 2214 coding
- 21% of genome is coding?

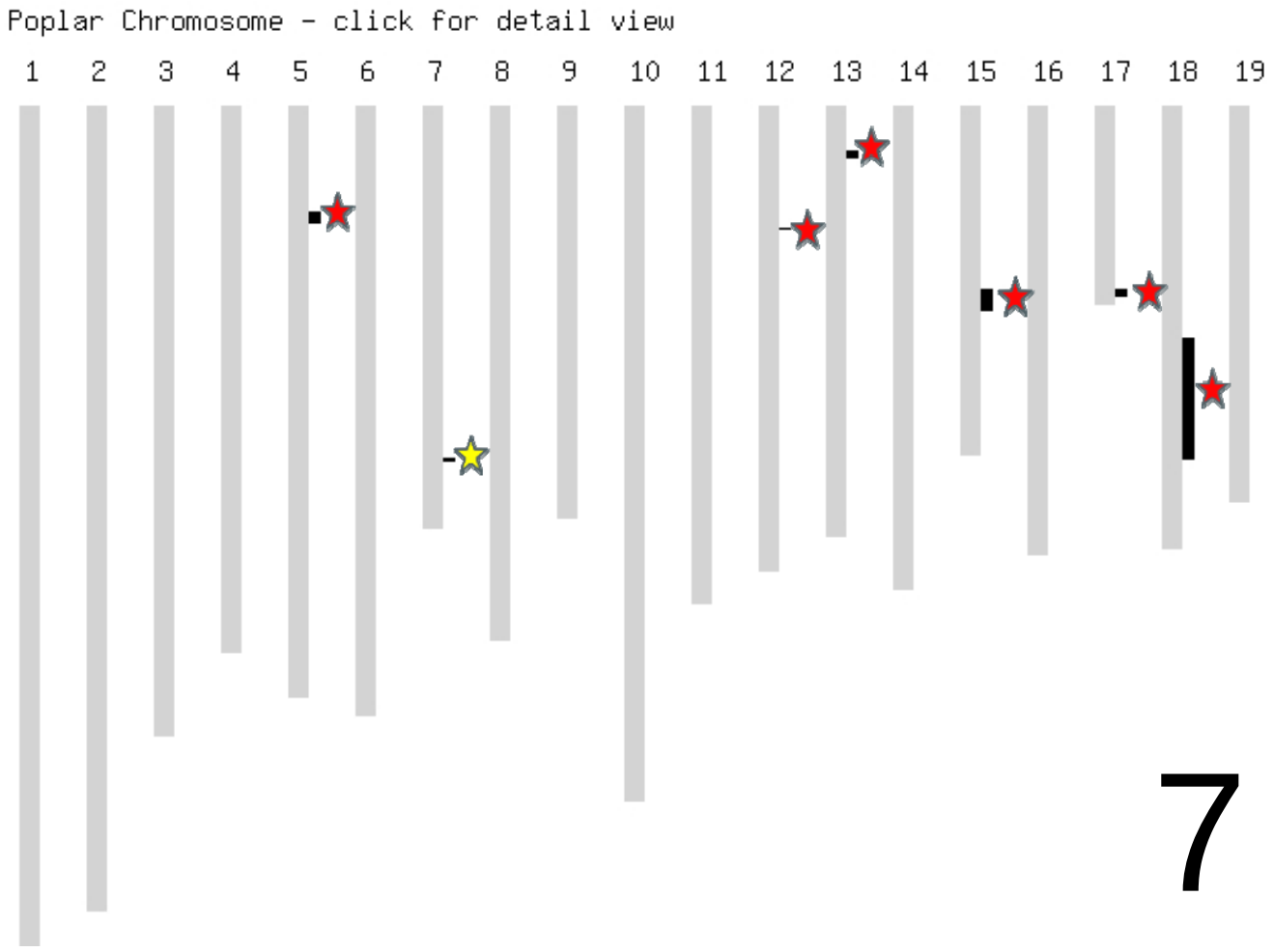
Similarity to other genomes



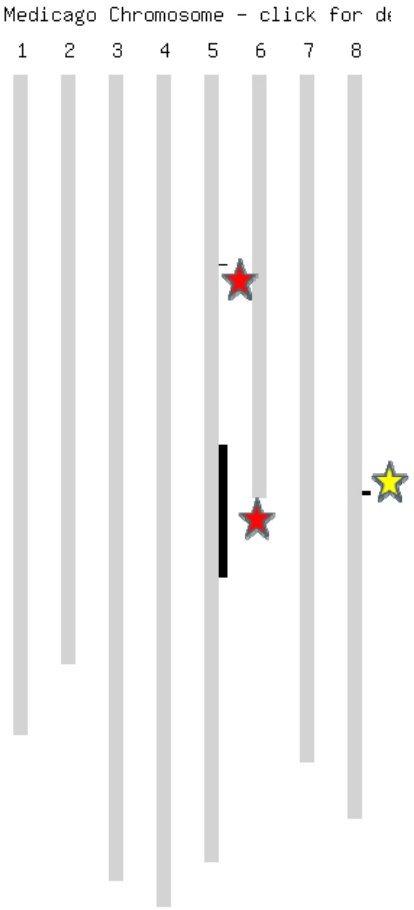
SyMap

- From Soderlund et al, 2006
- FPC to Genome
  - BAC end sequences
  - Hyb marker sequences
- FPC to FPC
  - by fingerprints
  - by shared markers

Chestnut to Poplar



Chestnut to Medicago



Orthologs

- Comparative mapping
- Evolutionary markers
  - study fundamental processes of genome evolution
- Allow scientists to move between species, translating information from one to another

KAAS Results

- KEGG Automatic Annotation Server
- Orthologs are detected by bi-directional best hit information.
- Most accurate when full genomic sequence is known.
- KEGG Orthology (KO) system is a pathway-based definition of orthologous genes
- Allows us to link our sequence data to other biological systems and molecular interaction pathways in a systematized manner

KAAS results

- KEGG Automatic Annotation Server

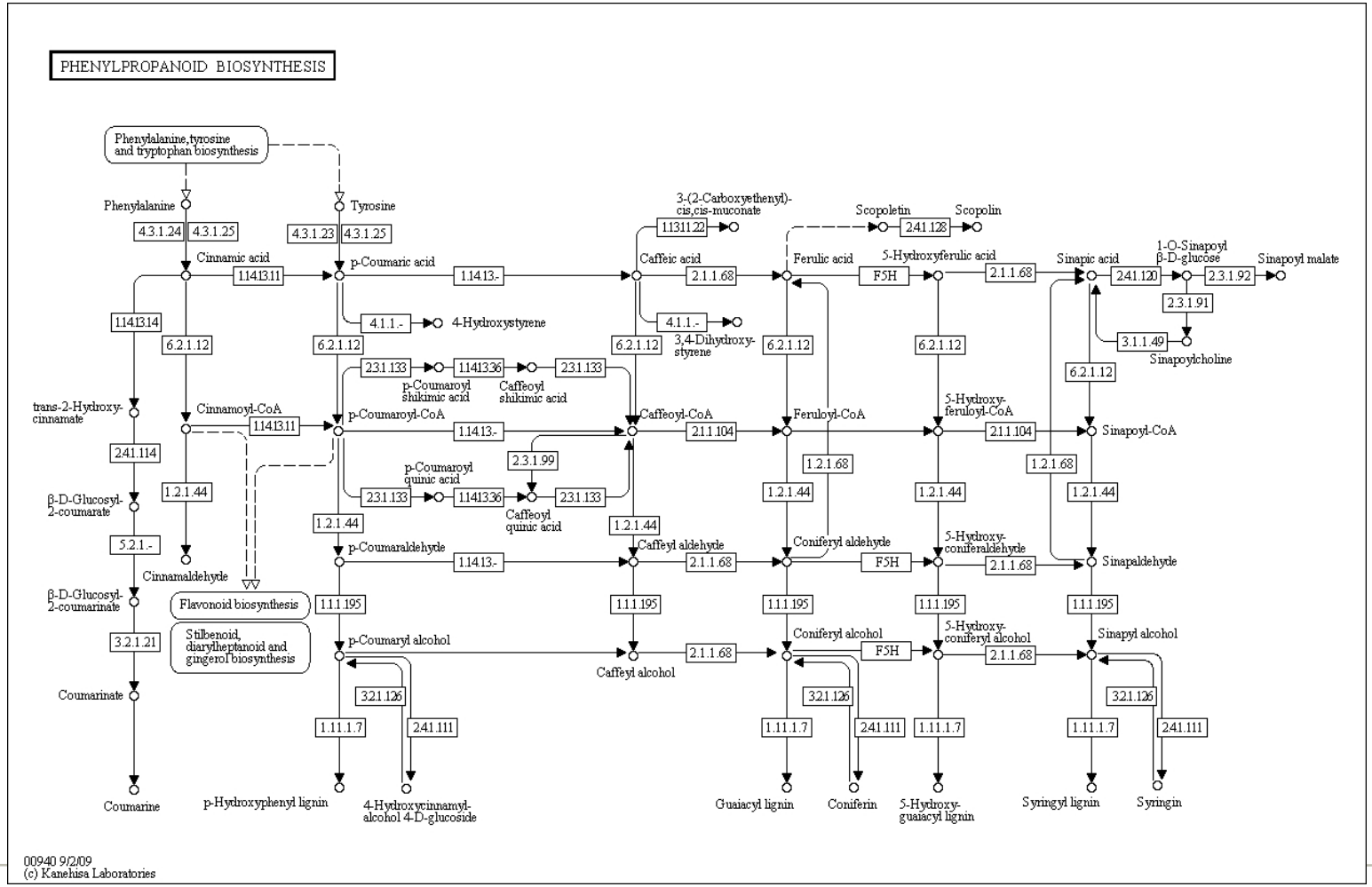
ABall V2	1000
AC454 V3	2006
CCall V2	2157
RO454 V2	1781
WO454 V2	1682

- 591 KOG groups with members in all 5 fagaceae species, populus, and prunus

Linking 3 tree genomes

- Found Prunus, populus, and chinese chestnut sequences linked to the same KOG
- 591 KOG groups with members in all 5 fagaceae species, populus, and prunus
- Developed overgos from 250
- First 125 completed, next 125 underway

Phenylpropanoid Pathway



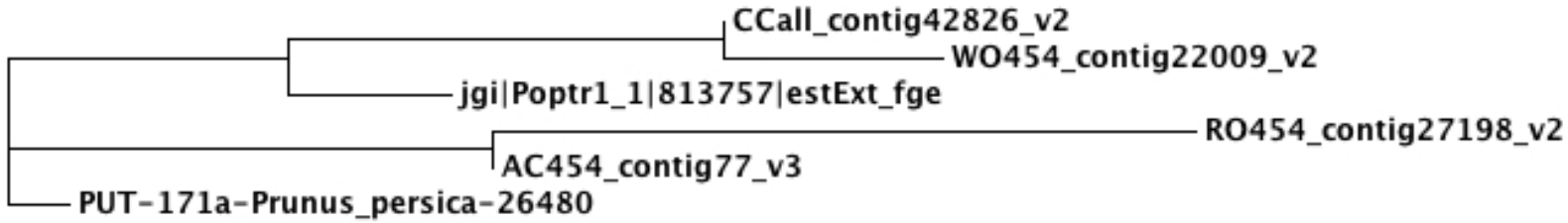
Phenylpropanoid Pathway

- **ko01061 Biosynthesis of phenylpropanoids** (41)

ko:K00026 E1.1.1.37B, mdh; malate dehydrogenase [EC:1.1.1.37]
ko:K00031 IDH1, IDH2, icd; isocitrate dehydrogenase [EC:1.1.1.42]
ko:K00083 E1.1.1.195; cinnamyl-alcohol dehydrogenase [EC:1.1.1.195]
ko:K00134 GAPDH, gapA; glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]
ko:K00161 PDHA, pdhA; pyruvate dehydrogenase E1 component subunit alpha [EC:1.2.4.1]
ko:K00162 PDHB, pdhB; pyruvate dehydrogenase E1 component subunit beta [EC:1.2.4.1]
ko:K00164 OGDH, sucA; 2-oxoglutarate dehydrogenase E1 component [EC:1.2.4.2]
ko:K00234 SDHA, SDH1; succinate dehydrogenase (ubiquinone) flavoprotein subunit [EC:1.3.5.1]
ko:K00235 SDHB, SDH2; succinate dehydrogenase (ubiquinone) iron-sulfur protein [EC:1.3.5.1]
ko:K00382 DLD, lpd, pdhD; dihydrolipoamide dehydrogenase [EC:1.8.1.4]
ko:K00430 E1.11.1.7; peroxidase [EC:1.11.1.7]
ko:K00475 E1.14.11.9; naringenin 3-dioxygenase [EC:1.14.11.9]
ko:K00487 CYP73A; trans-cinnamate 4-monooxygenase [EC:1.14.13.11]
ko:K00588 E2.1.1.104; caffeoyl-CoA O-methyltransferase [EC:2.1.1.104]
ko:K00615 E2.2.1.1, tktA, tktB; transketolase [EC:2.2.1.1]
ko:K00616 E2.2.1.2, talA, talB; transaldolase [EC:2.2.1.2]
ko:K00627 DLAT, aceF, pdhC; pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase) [EC:2.3.1.12]
ko:K00658 DLST, sucB; 2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase) [EC:2.3.1.61]
ko:K00660 E2.3.1.74, bcsA; chalcone synthase [EC:2.3.1.74]
ko:K00800 E2.5.1.19, aroA; 3-phosphoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19]
ko:K00813 E2.6.1.1B, aspC; aspartate aminotransferase [EC:2.6.1.1]
ko:K00815 E2.6.1.5; tyrosine aminotransferase [EC:2.6.1.5]
ko:K00817 hisC; histidinol-phosphate aminotransferase [EC:2.6.1.9]
ko:K00850 PFK, pfk; 6-phosphofructokinase [EC:2.7.1.11]
ko:K00873 PK, pyk; pyruvate kinase [EC:2.7.1.40]
ko:K00891 E2.7.1.71, aroK; shikimate kinase [EC:2.7.1.71]
ko:K00927 PGK, pgk; phosphoglycerate kinase [EC:2.7.2.3]
ko:K01623 ALDO, fbaB; fructose-bisphosphate aldolase, class I [EC:4.1.2.13]
ko:K01626 E2.5.1.54, aroF, aroG, aroH; 3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54]
ko:K01648 ACLY; ATP citrate (pro-S)-lyase [EC:2.3.3.8]
ko:K01681 ACO, acnA; aconitate hydratase 1 [EC:4.2.1.3]
ko:K01689 ENO, eno; enolase [EC:4.2.1.11]
ko:K01736 E4.2.3.5, aroC; chorismate synthase [EC:4.2.3.5]
ko:K01834 PGAM, gpm; phosphoglycerate mutase [EC:5.4.2.1]
ko:K01850 E5.4.99.5; chorismate mutase [EC:5.4.99.5]
ko:K01859 E5.5.1.6; chalcone isomerase [EC:5.5.1.6]
ko:K01899 LSC1; succinyl-CoA synthetase alpha subunit [EC:6.2.1.4 6.2.1.5]
ko:K01904 E6.2.1.12; 4-coumarate--CoA ligase [EC:6.2.1.12]
ko:K03841 FBP, fbp; fructose-1,6-bisphosphatase I [EC:3.1.3.11]
ko:K05298 E1.2.1.13, gap2; glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) [EC:1.2.1.13]
ko:K09753 CCR; cinnamoyl-CoA reductase [EC:1.2.1.44]

Phenylpropanoid Pathway

CCR



4-coumarate-CoA ligase

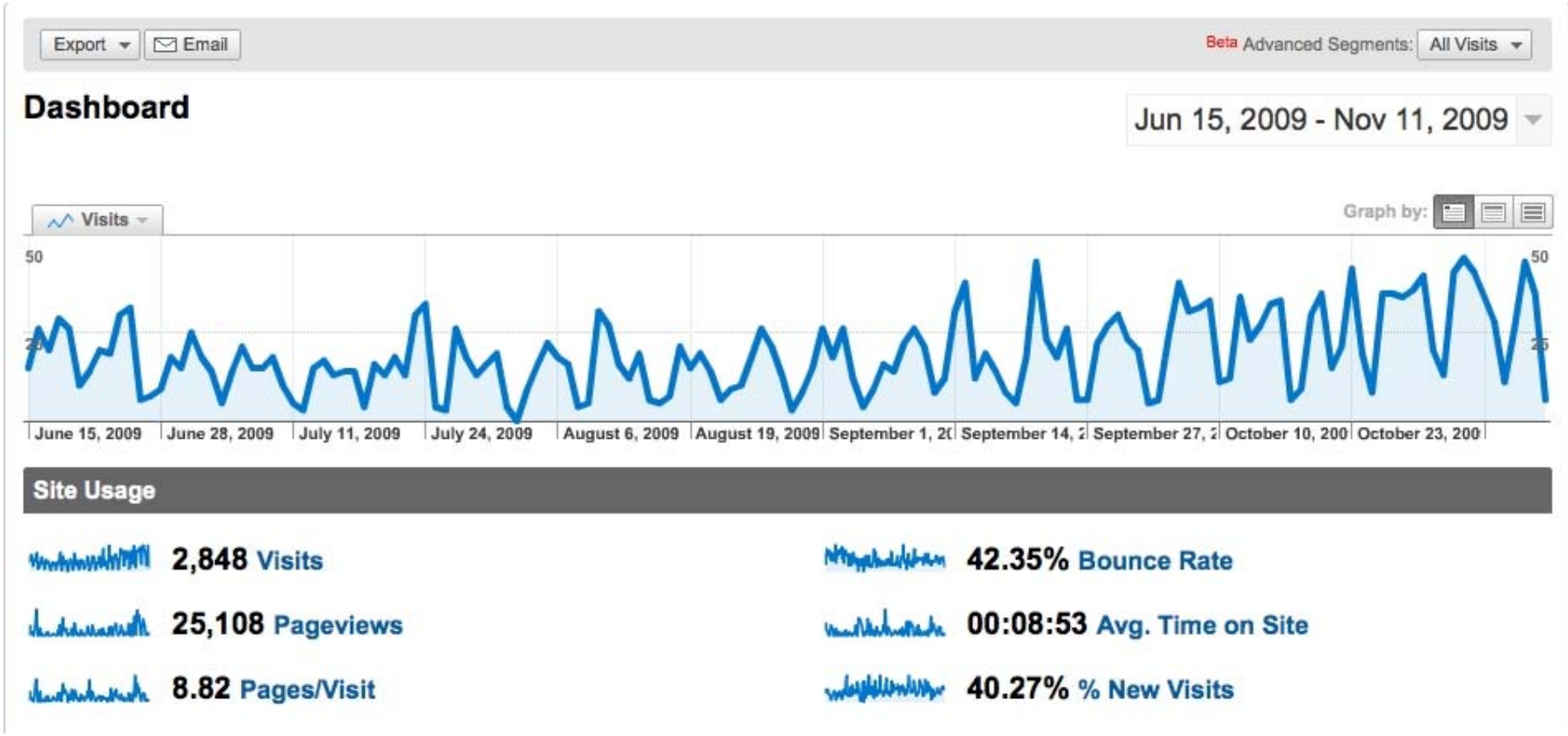
RBH method

- Reciprocal best hit method (similar to COSII from Solanaceae)

1. Group is initiated with Chinese Chestnut and Arabidopsis have a mutual best hit
2. Group is extended with the Arabidopsis and Chinese Chestnut sequences have the same mutual best hit in another Fagaceae species

4688 groups with 3 or more members

fagaceae.org



Conclusions

- Comparative work is just beginning
- New resources will open a whole new world of tools and knowledge for all trees
- the knowledge has to be shared and translated between trees and projects

Acknowledgements

- Bert Abbott, PI
- Eric Fang
- Clemson University Genomics Institute
  - Chun-Huai Cheng
  - Stephen Ficklin
- NSF Project