Using the SEAN package to find SNPs in cultivated tomato

Angela Baldo[1],Derek Huntley[2],Joanne Labate[1]

[1]USDA ARS Plant Genetic Resources Unit, Geneva NY USA

[2]Department of Computing, Imperial College, London, UK

Visual screening & lab work: **Katie Timmer**[1], **Susan Sheffer**[1]
Primer3 Wrapper: **Mike D'Amico**[1]

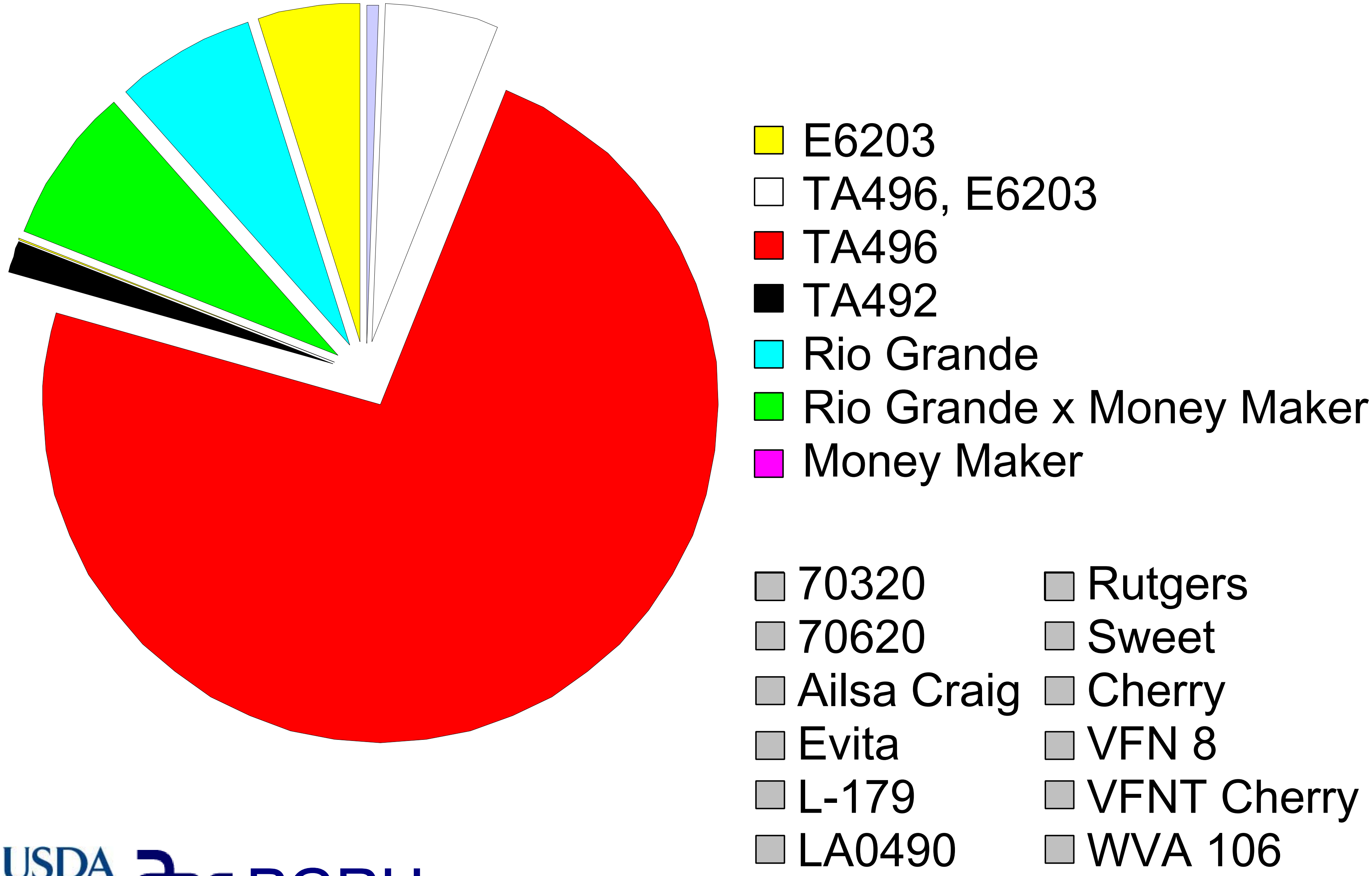
USDA ars PGRU

WHY?

- Germplasm Collection Management
- Relationships among accessions
- Identifying duplicates
- Association Mapping for Traits
- Population Genetics

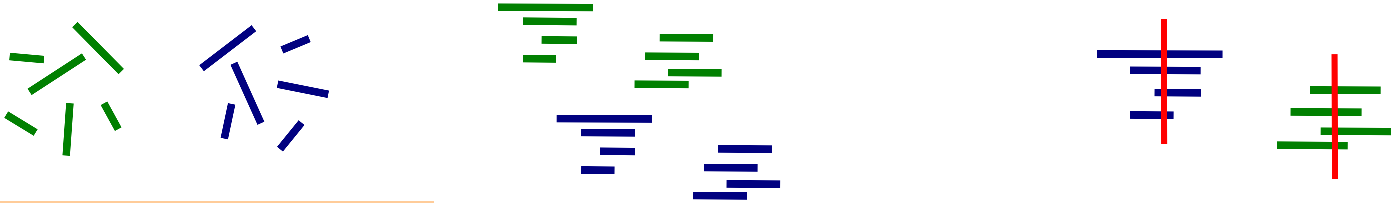
→Low diversity between tomato varieties (~1 SNP per 7 KB)

150,000 publically available expressed tomato sequences



USDA ars PGRU

Wholesale SNP Discovery



NCBI Unigene Set
- Sequence similarity
- cDNA clone origin

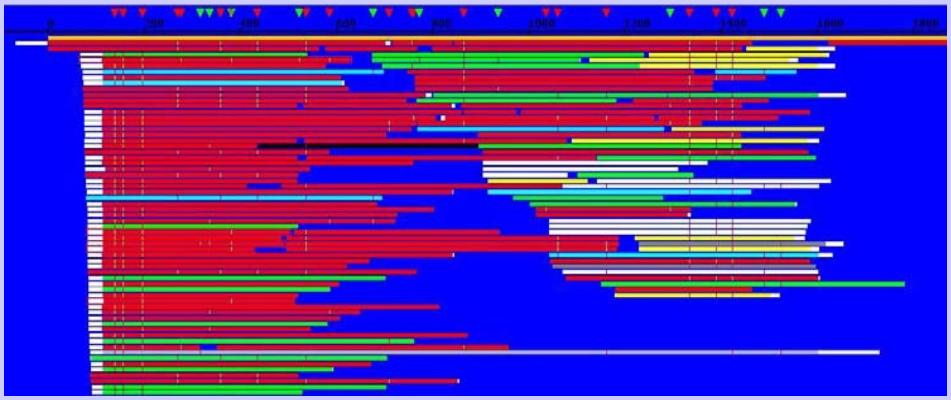
Cluster, Align
- Phrap

Identify potential SNPs
- Modified Picoult-Newberg (1999)

Generate correct consensi and BLAST against Swissprot & mapped tomato markers

Visually Examine Alignments

Design Primers
- Primer3



▼ SNP confirmed in more than one clone from the same cultivar
▼ SNP not confirmed in more than one clone from same cultivar

🟪 SEAN package
🟩 Phrap
🟧 PGRU scripts
🟦 Primer3

USDA ars PGRU

73 Predictions Tested:

10 No Amplification

40 Observed = Expected Size

23 Observed > Expected Size

50-1300 BP Introns?

25 Prediction was wrong

10 Looks heterozygous - Multigene family?

12 Poor Quality Sequence

1 Doesn't match consensus

3 Too Large to sequence

11 Confirmed

1 Unexpected SNP

~ 1 per 1066 BP tested

(Six times better than random)
-if you only count the useful ones!

USDA ars PGRU

Resources
(Everything GPL)

ISMB Poster A-9

SEAN

(perl & java available now):
http://zebrafish.doc.ic.ac.uk/SEAN

PGRU Website

(perl wrappers available soon):
http://www.ars-grin.gov/gen