The tomato genome sequence: an enabling platform for tomato biology

The International Tomato Genome Sequencing Consortium

Jim Giovannoni, USDA-ARS Robert W. Holly Center, Cornell University
Why sequence the tomato genome?

.....as a reference genome for the Solanaceae
The need and goal is a high quality reference for the Solanaceae

Key Sequencing and Analysis Groups within the International Consortium:

Japan: Satoshi Tabata, Shusei Sato, Kazuza Inst.

S. Korea: Doil Choi, Seoul National

China: Sanwen Huang (CAAS), Chuanyou Li (CAS)

India: Akilesh Tyagi, ICPG, U. of Delhi

Italy: Giovanni Giuliano, ENEA

Spain: Antonio Granell, U. Politécnica de Valencia

The Netherlands: Rene Klein Lankhorst, Roeland VanHam, WU

France: Mondher Bouzayen, INRA

UK: Graham Seymour (U. of Nottingham), Gerard Bishop, Jane Rogers

US: L. Mueller, S. Tanksley, J. Giovannoni, S. Stack, J. VanEck, Z. Fei, B. Roe

Argentina: Fernando Carrari, U. of Buenos Aires

Syngenta, 454/Roche, Keygene, ABI

Stephane Rhambauts, Ghent

Andy Patterson, University of Georgia

Z. Lippman, R. McCombie, D. Ware, CSH

Dani Zamir, Hebrew University
Assumptions and proposed strategy (2004)

- 12 chromosomes
- 950MB of total DNA
- 220MB contiguous, gene rich euchromatin
- Sequence only gene-rich euchromatin (>90% all genes)
- Approx. 650 seed BACs available (one every 3.1 cM)
- Tiling path method preferred
As typical of genome sequencing projects, the landscape has changed rapidly......

- **454 (31x)**
  - 15x shotgun
  - 8x 3kb matepair
  - 4x 8kb matepair
  - 3x 20kb matepair

- **Sanger (3.6x)**
  - 134 Mb BACs
  - 3.3x 3kb paired-end
  - 0.1x 40kb fosmid ends

- **Illumina (82x)**
  - 70x ~300 bp paired-end
  - 3x 2kb matepair
  - 3x 3kb matepair
  - 0.2x 120kb BAC ends
  - 3x 5kb matepair

- **SOLiD (140x)**
  - 22x shotgun
  - 21x 1kb matepair
  - 31x 4kb matepair
  - 66x 8kb matepair
Improving the tomato genome......

Are we done??

The International Tomato Genome Sequencing Consortium is committed to developing a “gold Standard” genome sequence that will rival Arabidopsis, rice and mammalian genomes in terms of quality and utility.
Exploiting the tomato genome for sequence-enabled discovery........
Genetic regulation of the development and maturation of fleshy fruit

Chlorophyll degradation

non-photosynthetic pigments

7DPA  17DPA  27DPA  MG  Breaker
+5

ethylene

S. Shong, Z. Fei
Transcription Factors

- ripening inhibitor (rin)
- non-ripening (nor)
- Colorless non-ripening (Cnr)

Signal Transduction Components

- Green-ripe (Gr)
- Never-ripe (Nr)
- high-pigment 1 (hp-1)
- Never-ripe 2 (Nr-2)

Plastid Development

- green-flesh (gf)
- lutescent 2 (l2)
Decode the fruit transcriptome using large-scale strand-specific RNA-sequencing

27 Gb of data (from > 100 RNA-Seq libraries).

Yield comprehensive transcriptome profiles at single nucleotide resolution for *S. lycopersicum* cv AC and 4 ripening mutants (6 fruit developmental stages, 2-6 biological replicates).

S. Zhong, Z. Fei
Antisense RNA associate with intronic region

LEFL2040O15

Antisense RNA associated with exon

LEFL2002DC06

S. Shong, Z. Fei
Whole flesh – D. Zamir
uniform (u)

- Recessive, spontaneous mutation
- Described as early as 1930s.
- Another mutant called \textit{ug} has similar phenotype, but not allelic
- Consumer trait
Characterization of the tomato epigenome

- bisulfite sequencing to determine 5mC
- ChIP-seq to determine histone methylation
- RNA-seq transcriptome analysis
- RNA-seq sRNA analysis