Genome **Biology**

OPINION

The tomato genome: implications for plant breeding, genomics and evolution

Aashish Ranjan, Yasunori Ichihashi and Neelima R Sinha*

Abstract

The genome sequence of tomato (*Solanum lycopersicum*), one of the most important vegetable crops, has recently been decoded. We address implications of the tomato genome for plant breeding, genomics and evolutionary studies, and its potential to fuel future crop biology research.

Keywords Genome, *Arabidopsis*, Tomato, Solanaceae, Biodiversity, Genomics, QTL, Omics

Plant genomes: current status

Plants are indispensable for human life, as they not only provide food, fiber, and fuel but also are critical for provision of oxygen and adsorption of CO_2. Plant breeding and genetics are powerful tools for increasing plant productivity through development of improved varieties. The rapid progress of plant genomics in recent years has opened new possibilities in targeted breeding of specific traits, and provides a powerful approach to sustainable crop production [1]. Plant genomics, in combination with genetics and breeding, has a particularly crucial role to play in ensuring food security to the rapidly growing world population.

Many plant genomes are large and complex due to an abundance of transposable elements and a long history of repeated genome duplication, making genome sequencing a major challenge [2]. The era of plant genomics began with release of the *Arabidopsis* genome sequence in 2000 [3]. It was a milestone in plant biology and made *Arabidopsis* one of the most popular species for basic plant research. Rice, a staple food in most of the world, was the second available plant genome in 2002 [4,5]. Rapid progress in the development of new sequencing technology and bioinformatic tools in recent years has

allowed faster and more efficient sequencing, and assembly of genomes at lower cost. As many as 20 plant genomes have been sequenced and assembled in the last two years [6-25]. Genome sequences of plants belonging to different groups, such as two plants from the early land plant clades (a moss *Physcomitrella patens* and a spike-moss *Selaginella moellendorffii*) and numerous economically important monocots (Box 1), such as rice, maize, sorghum, and so on, have now been decoded [4,5,25-28] (Figure 1). Eudicots (Box 1), the largest group in flowering plants, are composed of two major clades, the Eurosids and Euasterids. Genome sequences of many members of group Eurosids, such as *Arabidopsis*, grape, poplar, medicago, cucumber, and so on, have already been published [3,13,29-31]. However, members of the group Euasterids, which has many plants of economic importance, were not represented in the list of known plant genome sequences until the release of the potato (*Solanum tuberosum*) genome belonging to the family Solanaceae [6]. The recently decoded genome sequences of tomato (*Solanum lycopersicum*) and its close wild relative (*Solanum pimpinellifolium*) are significant additions to published Euasterid genomes [7]. These genomes will not only promote plant genomics and breeding studies for crop improvement programs in the Euasterids, particularly in the family Solanaceae, but also provide an unprecedented opportunity for basic plant biological research in the area of development and evolution.

Arabidopsis as a model plant for research

Arabidopsis has served as a model plant for basic plant research due to its small size, self-pollination, short life cycle, ease of propagation and genetic transformation [32]. Additionally, its small genome and the availability of its genome sequence made it a favorite for genetic and molecular studies. The sheer volume and extent of *Arabidopsis*-related research and integration of genetic, molecular, biochemical, genomic and morphological data from *Arabidopsis* provided insights into many universal aspects of plant biology. Due to a high level of synteny (Box 1) between the genomes of various plant species, the *Arabidopsis* genome also provided information on the structure of other Eurosid genomes [33].

*Correspondence: nrsinha@ucdavis.edu
Department of Plant Biology, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA

Monocots: A group of flowering plants whose seeds typically have one cotyledon (embryonic-leaf).

Dicots: A group of flowering plants whose seeds typically have two cotyledons (embryonic leaves). However, the term dicots is paraphyletic. 'Eudicots' defines a monophyletic group that excludes monocots, their allies and more basal 'dicot' species.
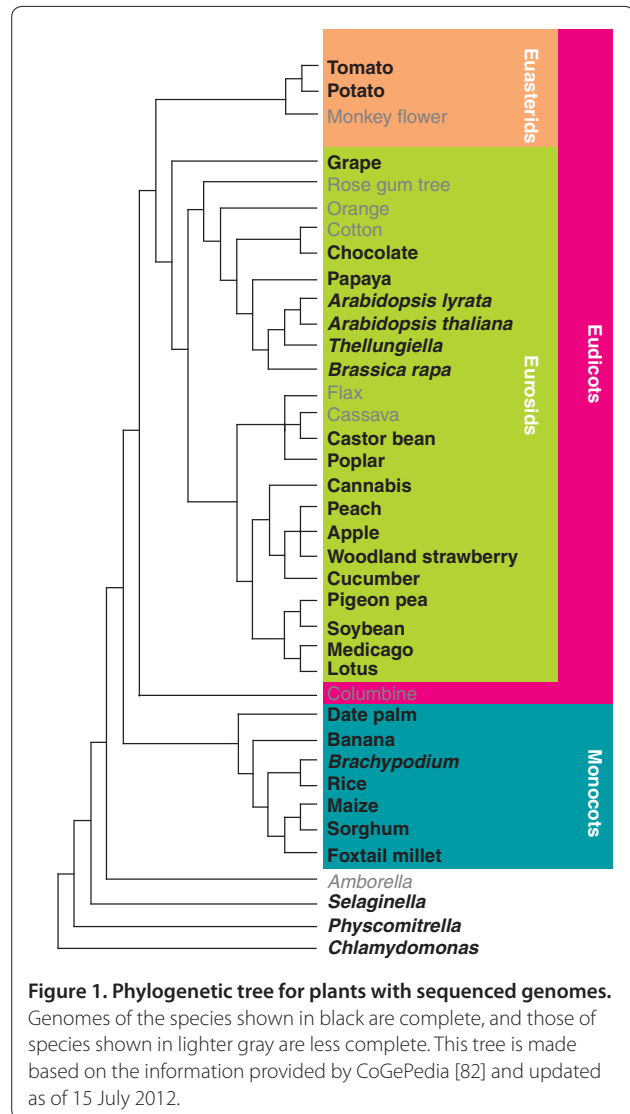
Synteny: Physical co-localization of genetic loci in the same relative positions in two or more genomes.

Sympodial branching: In this type of branching the shoot apex terminates in a flower and an axillary bud or buds continues growth of the inflorescence. The process is reiterated many times.
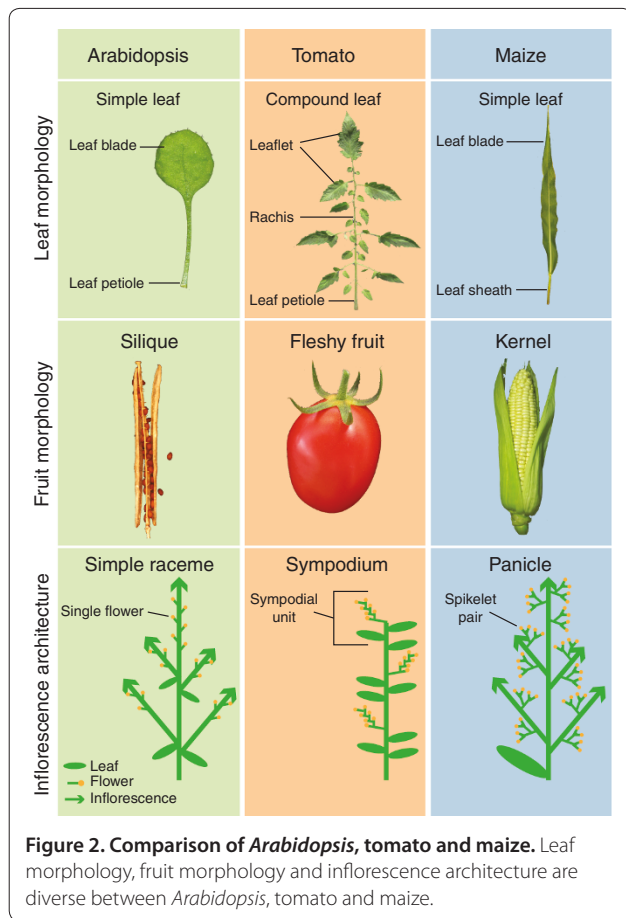
In spite of these advantages, *Arabidopsis* is an 'atypical' plant (Figure 2). Its small genome and features such as leaf morphology, fruit characteristics and plant architecture differ from most agriculturally important plants [34]. While *Arabidopsis* is excellent for research in basic plant biology, it is not amenable to investigations of domestication or crop improvement through selective breeding. The ongoing 1001 *Arabidopsis* genome project aims to look at natural selection and alleles underlying phenotypic diversity across the entire genome and the entire species [35]. Mostly the questions related to domestication and crop improvement have been investigated in monocot cereal crops such as maize and rice [36-38]. Genome sequences of other agriculturally and economically important plant species are needed to answer major questions about genome function and genome evolution, and application of genomic information to the practical problems of yield and quality enhancement. Consistent with this, the US National Plant Genome Initiative, established in 1998, also made a call for genome sequences of every major plant of economical importance.

Figure 1. Phylogenetic tree for plants with sequenced genomes. Genomes of the species shown in black are complete, and those of species shown in lighter gray are less complete. This tree is made based on the information provided by CoGePedia [82] and updated as of 15 July 2012.

Tomato genetics and genomics: past and present

The family Solanaceae, having many species of economic importance, such as tomato, potato, tobacco, pepper, eggplant, and so on, is the most extensively studied family among the Euasterids. Solanaceous crop genomics is in an exciting phase of development following the recent sequencing of the potato and tomato genomes [6,7]. Tomato (*Solanum lycopersicum* L.), originated in South America, and was spread around the world to become one of the most extensively used vegetable crops. Besides its economic value, it has interesting developmental features, such as compound leaves, fleshy fruits, and sympodial shoot branching (Box 1, Figure 2) [39-41]. Moreover, it has simple diploid genetics, short generation time, and routine transformation technology and is easy to maintain. Together these make tomato an excellent species for both basic and applied plant research.

Several genetic and genomic resources were available for tomato before the inception of the tomato genome sequencing project. Large germplasm collections consisting of numerous accessions of landraces of tomato (*S. lycopersicum*) and its wild relatives (Box 2) [42,43] had been established, many of which are sexually compatible with tomato and are also are a source of valuable disease resistance and other genes that had been exploited by breeders to develop modern cultivated tomato varieties. Tomato geneticists had used a number of morphological and isozyme markers to construct a genetic map of tomato and identified the 12 linkage groups corresponding to the cytologically visible chromosomes. This aided the construction of an RFLP (restriction fragment length polymorphism) linkage map [44,45]. The resulting comprehensive molecular linkage map enabled breeders to identify quantitative trait loci (QTLs) leading to an

Figure 2. Comparison of *Arabidopsis*, tomato and maize. Leaf morphology, fruit morphology and inflorescence architecture are diverse between *Arabidopsis*, tomato and maize.

Box 2: Tomato wild relatives

Tomato wild relatives fall into two groups: red- or orange-fruited species such as *S. pimpinellifolium*, *S. galapagense*, and *S. cheesmaniae*, and the green-fruited species such as *S. neorickii*, *S. chmielewskii*, *S. chilense*, *S. arcanum*, *S. corneliomulleri*, *S. huaylasense*, *S. peruvianum*, *S. habrochaites*, *S. pennellii*, *S. lycopersicoides*, *S. sitiens*, *S. ochranthum* and *S. juglandifolium*. Red-fruited species accumulate glucose and fructose, while the green-fruited species accumulate sucrose [42]. Traditionally, wild and cultivated tomatoes were grouped within the genus *Lycopersicon* in the Solanaceae. However, molecular phylogenetic studies support placement of tomatoes in the genus *Solanum* [43].

understanding of the genetic basis of numerous quantitative traits [46]. The Solanaceae Genomics Network website provides extensive information on the available tomato genetic and genomic resources [47].

Domesticated tomato and related wild species (Box 2) exhibit tremendous genetic and trait biodiversity, making the group highly suitable for evolutionary and domestication studies [48,49]. Sequence diversity analysis of extensive expressed sequence tags from domesticated and wild tomato species identified numerous inter- and intra-specific polymorphisms, many of which could be important for domestication [50]. In order to exploit the rich trait reservoir of domesticated and wild tomato species, tomato breeders developed advanced backcross mapping populations for identifying and transferring favorable QTLs from wild to cultivated germplasm. This subsequently led to development of permanent mapping populations in the form of introgression lines (ILs) where, by repeated backcrossing, a segment of a wild species genome is introduced into a cultivated tomato background [51,52]. A set of 76 ILs, ensuring complete genome coverage of *S. pennellii* introgressed into the cultivated tomato M82 variety, have been extensively phenotyped for numerous traits such as morphology,

yield, fruit quality, and fruit primary and secondary metabolites for the identification of QTLs [53]. The high-resolution mapping approach applied to *S. pennellii* ILs has led to the map-based cloning of the sugar yield QTL *Brix9-2-5*, and the fruit weight QTL *fw2.2* [54,55]. *Brix9-2-5* was delimited to an invertase gene, which is expressed early in fruit development, whereas *fw2.2* was delimited to the gene *ORFX*, which is expressed early in floral development. Classical breeding and marker analysis has also made remarkable contributions to improve various yield traits of tomato. For example, the fruit size QTL *fasciated*, initially identified using a cross between *S. lycopersicum* and *S. pimpenellifolium*, has recently been characterized [56]. The first example of yield improvement by a single overdominant gene (*SINGLE FLOWER TRUSS*) through heterosis has been demonstrated in tomato [57]. Furthermore, tomato and its wild relatives have also been used as a model for self/hybrid incompatibility studies [58,59].

Tomato genome in brief

Sequencing of the tomato genome was initiated in 2005 as a multinational effort between 14 countries. The genome of the domesticated tomato Heinz 1706 was sequenced using a combination of longer Sanger and 454/Roche GS FLX reads and high-coverage, shorter SOLiD and Illumina GAIIx reads. The sequences were assembled into 91 scaffolds, covering 760 Mb of the approximately 900 Mb of genome, aligned to the 12 tomato chromosomes with 34,727 predicted protein-coding genes [7]. Most of the gaps were restricted to repeat-rich pericentromeric regions. Additionally, a draft sequence of the closest wild relative *S. pimpinellifolium* was compared to the Heinz sequence. The two genomes are highly similar showing only 0.6% nucleotide divergence. Sixty percent of the genes are identical or with only synonymous changes between domesticated and wild tomato, while the remaining 40% have non-synonymous changes, including alterations of stop codons with

potential consequences for gene function. Compared to the potato genome, the tomato and *S. pimpinellifolium* genomes show more than 8% nucleotide divergence. Moreover, the tomato genome is highly syntenic with the genomes of other economically important members of the family Solanaceae, such as eggplant and pepper. Comparative genome analysis identified two consecutive triplication events in the *Solanum* lineage. Interestingly, these genome triplications added new gene family members such as transcription factors and enzymes necessary for ethylene biosynthesis and perception, which mediate important fruit-specific functions.

Future directions of tomato genetics and genomics

Modern tomato genetics had already used molecular markers and functional analysis to identify a handful of genes underlying developmental or yield traits, but the availability of the tomato genome sequence will further revolutionize tomato genetics and breeding. However, since the domesticated tomato varieties show limited genetic diversity, the wild tomato relatives provide a rich source of useful allelic variation. The 150 tomato genome resequencing project was recently initiated with an objective to reveal and explore extant genetic variation in tomato, and will provide a major boost to identification of valuable alleles. The project aims to sequence 83 genotypes, including 30 wild accessions, 43 land races and 10 old varieties [60]. This will not only help identify useful SNPs from the wild accessions but also rare SNPs within domesticated varieties. Tomato breeders can then target gene variants (SNPs) in the wild species associated with desirable traits such as disease or pest resistance or growth in extreme environmental conditions and introduce them into cultivars in order to exploit the rich tomato germplasm for breeding purposes. More genome sequences will facilitate QTL identification, mapping and cloning of underlying genes, and provide new SNP markers for marker-assisted breeding. For example, genome-wide association studies (GWAS) will allow detection and fine mapping of QTLs in the post-genome era, given the high phenotypic diversity among various tomato wild relatives [61,62]. QTL analyses will also help to investigate the process of domestication and associated yield increase [56,63]. Additionally, millions of informative markers (SNPs/InDels) and structural variations, such as duplications, inversions, transpositions, and so on, identified through comparison of genome sequences of domesticated and wild tomatoes will promote investigations into the genetic and molecular basis of the process of domestication and crop improvement.
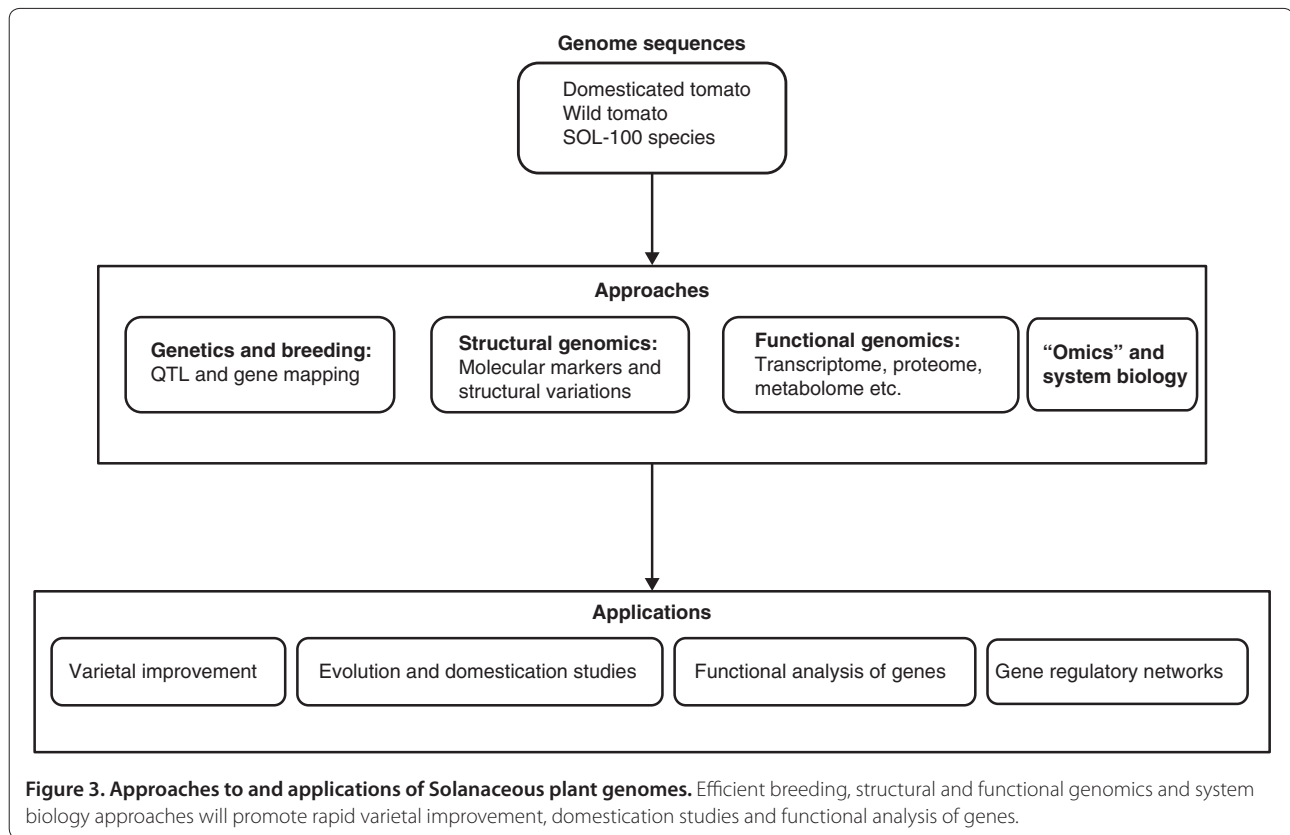
Identification of introgressions of segments of the *S. pimpinellifolium* genome into the Heinz genome already suggests that introgression through conventional (rather than marker assisted) breeding has been a significant factor in crop improvement/domestication in tomato [7]. These wild-species introgressions have provided disease resistance and others have been associated with small fruit size (cherry tomatoes). ILs in the background of cultivated tomato exist for many wild tomato species [51,64-68]. The tomato ILs are an excellent tool for functional genomics studies to investigate genetic and environmental interactions. Expression QTLs (eQTLs), as identified by large-scale transcriptome profiling of the ILs, will be useful in connecting phenotypic variation to genotypic diversity, thus leading to a hypothetical regulatory network based on location of eQTLs and phenotypic QTLs [69]. Integrating additional functional genomics approaches such as metabolomics and proteomics can significantly reduce the number of candidate genes for a given QTL [70,71]. One of the major thrusts of functional genomics in future will be RNA-seq enabled transcriptome profiling. For example, comparison of transcriptome profiles from domesticated and wild tomato species will give us insights into the gene expression differences associated with the process of domestication and trait diversity. The tomato functional genomics database (TFGD), which includes microarray, metabolite and small RNA data, has already been established as a comprehensive resource even before the complete tomato genome sequence was released [72].

The advent of next generation sequencing and available genome sequence should make characterization of large collections of tomato mutants even more rapid and robust through sequencing of phenotyped sub-pools from F2 populations and subsequent mapping using methods such as SHOREmap and next generation mapping (NGM) [73,74]. Availability of the tomato genome sequence will speed up the understanding of gene function in developmental and metabolic pathways and identify key steps in co-regulation mechanisms by mapping relevant tomato mutants. Additionally, multiple TILLING (Targeting Induced Local Lesions IN Genomes) resources in different backgrounds have already been developed for tomato functional genomics [75,76]. These TILLING resources, in combination with the tomato genome sequence, should be useful for both forward and reverse genetics in tomato for both basic science and/or crop improvement.

Beyond tomato to the family Solanaceae

Besides the genus *Solanum*, the family Solanaceae has more than 3,000 species that exhibit diversity in development, organ morphology, metabolism and geographic distribution. Many of these species have high economic, nutritional and agricultural importance. The SOL-100 sequencing project, with an objective to centralize sequences and phenotypes of 100 different species across the phylogenetic diversity of the group, will facilitate the

Figure 3. Approaches to and applications of Solanaceous plant genomes. Efficient breeding, structural and functional genomics and system biology approaches will promote rapid varietal improvement, domestication studies and functional analysis of genes.

genetic mapping of simple and complex phenotypes affecting the numerous diverse traits in SOL-100 species [77]. It has long been known that gene duplication events followed by sub-functionalization and neo-functionalization have fueled the process of evolution. Genome sequences of SOL-100 species will promote comparative genetics efforts in the family Solanaceae, and will provide important insights into the evolution of gene families [7]. This knowledge will help identify genes for traits that may be useful in tomato breeding - either through introgression from sexually compatible species or by moving them into tomato via a transgenic route.

In the post-genomic era, an overwhelming amount of data from different 'omics' approaches is being generated and utilized for genomics research. It is becoming increasingly clear that with the availability of new plant genome sequences from both the family Solanacae and more crops of agricultural interest coupled with cheap next generation sequencing technologies, conversion of raw data into biologically meaningful information will require better and easily accessible bioinformatics tools [78]. Progress in plant research in general will depend on our ability to tie together independent components such as genotypic information, phenotypic data, expression profiles and so on into higher order complexity with multiple dimensions. The availability of genome sequences

and the ability to handle large data sets will promote system biology approaches for crop plant research to build higher order gene regulatory networks for understanding plant developmental and metabolic pathways [79]. For example, laser capture microdissection (LCM) or fluorescence-activated cell sorting (FACS) in combination with RNA-seq have enabled us to generate transcriptome profiles of specific tissues/cell types, such as leaf, inflorescence, fruit, and so on, at high spatial and temporal resolution. The resulting large-scale data generated can be integrated using bioinformatic tools to understand how cell/tissue-specific gene expression leads to the production of whole organ phenotypes and will address the developmental changes associated with environmental adaptation and diversification of crop plants [80,81]. Integration of other 'omics' data from proteomics, metabolomics, epigenomics and other studies will further allow crop biology to be approached from a systems perspective.

Together with genome sequences of related tomato wild species and other SOL-100 species, tomato sequence information will not only accelerate the elucidation of evolutionary relationships within the Solanaceae, but also aid in the discovery of new genes, allele-mining, and large-scale SNP genotyping. More efficient use of genetics and breeding methods and large scale structural and

Table 1. Genome sequencing of Euasterids in progress

Species	Current status	Sequencing initiative/url
Nicotiana benthamiana	Draft sequence available	ftp://ftp.solgenomics.net/genomes/Nicotiana_benthamiana/
Petunia	In progress	Joint effort of Northern Illinois University, The Petunia Platform and BGI
Sunflower	In progress	http://www.genomebc.ca/portfolio/projects/bioenergy-projects/genomics-of-sunflower/
Lettuce	In progress	http://compgenomics.ucdavis.edu/
Coffee	In progress	http://www.coffeegenome.org/

functional genomics studies in combination with system biology approaches will promote rapid varietal improvement, evolution and domestication studies and functional analysis of genes (Figure 3). Considering the highly diverse developmental and metabolic behavior of different plants and differences in their human usage, it is evident that a single model plant will not able to answer multiple intriguing questions in the area of basic and applied plant biology. With the shifting focus of plant biology from a single model plant, *Arabidopsis*, to multiple model plants of economic importance, tomato genomics studies will serve as a model for other crop plants, with particular focus in the area of exotic germplasm genomics for trait improvement and use of system biology approaches. For example, crop plants within the family Solanaceae, such as eggplant, pepper and potato, have wild relatives that have been utilized for crop improvement and will benefit from incorporation of genome scale information into breeding efforts.

Perspective

With the recently decoded tomato and potato genomes (both belonging to the family Solanaceae), plant biologists have just started to explore the genome sequences of Euasterids for both applied and basic research [6,7]. To promote large-scale genomic studies in Euasterids, more species belonging to this group need to be sequenced to match the plethora of sequence information from Eurosids. The family Solanaceae itself has many other plants of economic importance, such as eggplant, pepper, tobacco, petunia, and so on. Furthermore, Euasterids include numerous other plants of economic importance: the beverage-producing plants tea and coffee; oil-producing plants such as sunflower, safflower and olive; vegetable plants like carrot, parsley, lettuce and so on; and aromatic plants like mint, basil, rosemary and so on. Genome sequencing of a number of these species is already in progress (Table 1). Genome information for these plants will not only help to enhance the economic value of these crops but also answer specific questions related to their basic development and the regulation of secondary metabolism in these species. Orobanchaceae, the largest family of parasitic flowering plants, also belongs to the Euasetrids. Deciphering the genomes of parasitic plants such as *Striga* and their close relatives will help in saving millions of hectares of crop plants worldwide from destruction by these parasites. Moreover, availability of genome sequences from Euasterids and basal eudicot (for example, Columbine; Figure 1), along with the numerous sequenced Eurosid genomes, will promote large-scale evolutionary and biodiversity studies across all the flowering plants.

References
1. Bevan M, Waugh R: **Applying plant genomics to crop improvement.** *Genome Biol* 2007, **8**:302.
2. Schatz MC, Witkowski J, McCombie WR: **Current challenges in de novo plant genome sequencing and assembly.** *Genome Biol* 2012, **13**:243.
3. The *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**:796-815.
4. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*).** *Science* 2002, **296**:92-100.
5. Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*).** *Science* 2002, **296**:79-92.
6. Potato Genome Sequencing Consortium, Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, Orjeda G, Guzman F, Torres M, Lozano R, Ponce O, Martinez D, De la Cruz G, Chakrabarti SK, Patil VU, Skryabin KG, Kuznetsov BB, Ravin NV, Kolganova TV, Beletsky AV, Mardanov AV, Di Genova A, Bolser DM, Martin DM, Li G, *et al.*: **Genome sequence and analysis of the tuber crop potato.** *Nature* 2011, **475**:189-195.
7. The Tomato Genome Consortium: **The tomato genome sequence provides insights into fleshy fruit evolution.** *Nature* 2012, **485**:635-641.
8. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E, Alioto T, Capella-Gutiérrez S, Blanca J, Cañizares J, Ziarsolo P, Gonzalez-Ibeas D, Rodríguez-Moreno L, Droege M, Du L, Alvarez-Tejado M, Lorente-Galdos B, Melé M, Yang L, Weng Y, Navarro A, Marques-Bonet T, Aranda MA, Nuez F, Picó B, Gabaldón T, *et al.*: **The genome of melon (*Cucumis melo* L.).** *Proc Natl Acad Sci U S A* 2012, **109**:11872-11877.
9. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton JM, Rees DJ, Williams KP, Holt SH, Ruiz Rojas

JJ, Chatterjee M, *et al.:* The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* 2011, **43:**109-116.

10. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Ravel J, Rabinowicz PD: **Draft genome sequence of the oilseed species *Ricinus communis*.** *Nat Biotechnol* 2010, **28:**951-956.

11. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, *et al.:* **The genome of the domesticated apple (*Malus x domestica* Borkh.).** *Nat Genet* 2010, **42:**833-839.

12. van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE: **The draft genome and transcriptome of *Cannabis sativa*.** *Genome Biol* 2011, **12:**R102.

13. Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KA, Tang H, Rombauts S, Zhao PX, Zhou P, *et al.:* **The Medicago genome provides insight into the evolution of rhizobial symbioses.** *Nature* 2011, **480:**520-524.

14. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al.:* **Genome sequence of the palaeopolyploid soybean.** *Nature* 2010, **463:**178-183.

15. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, *et al.:* **Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers.** *Nat Biotechnol* 2012, **30:**83-89.

16. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, *et al.:* **The genome of *Theobroma cacao*.** *Nat Genet* 2011, **43:**101-108.

17. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, *et al.:* **The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change.** *Nat Genet* 2011, **43:**476-481.

18. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, *et al.:* **The genome of the mesopolyploid crop species *Brassica rapa*.** *Nat Genet* 2011, **43:**1035-1039.

19. Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu JK, Bohnert HJ, Cheeseman JM: **The genome of the extremophile crucifer *Thellungiella parvula*.** *Nat Genet* 2011, **43:**913-918.

20. Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J, *et al.:* **De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*).** *Nat Biotechnol* 2011, **29:**521-527.

21. The International Brachypodium Initiative: **Genome sequencing and analysis of the model grass *Brachypodium distachyon*.** *Nature* 2010, **463:**763-768.

22. Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, *et al.:* **Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential.** *Nat Biotechnol* 2012, **30:**549-554.

23. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, *et al.:* **Reference genome sequence of the model plant Setaria.** *Nat Biotechnol* 2012, **30:**555-561.

24. D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, *et al.:* **The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants.** *Nature* 2012, in press.

25. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, *et al.:* **The Selaginella genome identifies genetic changes associated with the evolution of vascular plants.** *Science* 2011, **332:**960-963.

26. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, *et al.:* **The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants.** *Science* 2008, **319:**64-69.

27. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, *et al.:* **The B73 maize genome: complexity, diversity, and dynamics.** *Science* 2009, **326:**1112-1115.

28. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, *et al.:* **The *Sorghum bicolor* genome and the diversification of grasses.** *Nature* 2009, **457:**551-556.

29. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, *et al.:* **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313:**1596-1604.

30. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, *et al.:* **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla.** *Nature* 2007, **449:**463-467.

31. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, *et al.:* **The genome of the cucumber, *Cucumis sativus* L.** *Nat Genet* 2009, **41:**1275-1281.

32. Koornneef M, Meinke D: **The development of *Arabidopsis* as a model plant.** *Plant J* 2010, **61:**909-921.

33. Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D, Freeling M: **Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids.** *Plant Physiol* 2008, **148:**1772-1781.

34. Barakat A, Matassi G, Bernardi G: **Distribution of genes in the genome of *Arabidopsis* thaliana and its implications for the genome organization of plants.** *Proc Natl Acad Sci U S A* 1998, **95:**10044-10049.

35. Weigel D, Mott R: **The 1001 genomes project for *Arabidopsis thaliana*.** *Genome Biol* 2009, **10:**107.

36. Doebley JF, Gaut BS, Smith BD: **The molecular genetics of crop domestication.** *Cell* 2006, **127:**1309-1321.

37. Lin Z, Li X, Shannon LM, Yeh CT, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, *et al.:* **Parallel domestication of the *Shattering1* genes in cereals.** *Nat Genet* 2012, **44:**720-724.

38. Ross-Ibarra J, Morrell PL, Gaut BS: **Plant domestication, a unique opportunity to identify the genetic basis of adaptation.** *Proc Natl Acad Sci U S A* 2007, **104 Suppl 1:**8641-8648.

39. Townsley BT, Sinha NR: **A new development: evolving concepts in leaf ontogeny.** *Annu Rev Plant Biol* 2012, **63:**535-562.

40. Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB: **A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening.** *Nat Genet* 2006, **38:**948-952.

41. Schmitz G, Theres K: **Genetic control of branching in *Arabidopsis* and tomato.** *Curr Opin Plant Biol* 1999, **2:**51-55.

42. Schauer N, Zamir D, Fernie AR: **Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex.** *J Exp Bot* 2005, **56:**297-307.

43. Labate JA, Grandillo S, Fulton T, Munos S, Caicedo AL, Peralta I, Ji Y, Chetelat RT, Scott JW, Gonzalo MJ, *et al.:* **Tomato.** *Genome Mapping Mol Breeding Plants* 2007, **5:**1-125.

44. Linkage Committee, "Linkage summary,". *Tomato Genetics Cooperative* 1973, **23:**9-11.

45. Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, *et al.:* **High density molecular linkage maps of the tomato and potato genomes.** *Genetics* 1992, **132:**1141-1160.

46. Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD: **Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms.** *Nature* 1988, **335:**721-726.

47. Sol Genomics Network [http://solgenomics.net/]

48. Larry R, Joanne L: **Genetic resources of tomato.** In *Genetic Improvement of Solanaceous Crops. Volume 2.* Edited by Razdan MK, Mattoo AK. Enfield, NH: Science Publishers; 2007.

49. Nakazato T, Warren DL, Moyle LC: **Ecological and geographic modes of species divergence in wild tomatoes.** *Am J Bot* 2010, **97:**680-693.

50. Jimenez-Gomez JM, Maloof JN: **Sequence diversity in three tomato species: SNPs, markers, and molecular evolution.** *BMC Plant Biol* 2009, **9:**85.

51. Eshed Y, Zamir D: **An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL.** *Genetics* 1995, **141:**1147-1162.

52. Tanksley SD, Nelson JC: **Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines.** *Theor Appl Genet* 1996, **92:**191-203.

53. Gur A, Semel Y, Cahaner A, Zamir D: **Real Time QTL of complex phenotypes in tomato interspecific introgression lines.** *Trends Plant Sci* 2004, **9:**107-109.

54. Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert

KB, Tanksley SD: *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 2000, **289**:85-88.

55. Fridman E, Pleban T, Zamir D: **A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene.** *Proc Natl Acad Sci U S A* 2000, **97**:4718-4723.

56. Cong B, Barrero LS, Tanksley SD: **Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication.** *Nat Genet* 2008, **40**:800-804.

57. Krieger U, Lippman ZB, Zamir D: **The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato.** *Nat Genet* 2010, **42**:459-463.

58. Li W, Chetelat RT: **A pollen factor linking inter- and intraspecific pollen rejection in tomato.** *Science* 2010, **330**:1827-1830.

59. Moyle LC, Nakazato T: **Hybrid incompatibility "snowballs" between Solanum species.** *Science* 2010, **329**:1521-1523.

60. **150 Tomato Genome ReSequencing project** [http://www.tomatogenome.net/]

61. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, *et al*.: **Genome-wide association studies of 14 agronomic traits in rice landraces.** *Nat Genet* 2010, **42**:961-967.

62. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES: **Genome-wide association study of leaf architecture in the maize nested association mapping population.** *Nat Genet* 2011, **43**:159-162.

63. Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WT, Macaulay M, MacKenzie K, Simpson C, Fuller J, Bonar N, *et al*.: *INTERMEDIUM-C*, **a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene** *TEOSINTE BRANCHED 1*. *Nat Genet* 2011, **43**:169-172.

64. Monforte AJ, Tanksley SD: **Development of a set of near isogenic and backcross recombinant inbred lines containing most of the** *Lycopersicon hirsutum* **genome in a** *L. esculentum* **genetic background: a tool for gene mapping and gene discovery.** *Genome* 2000, **43**:803-813.

65. Finkers R, van Heusden AW, Meijer-Dekens F, van Kan JA, Maris P, Lindhout P: **The construction of a** *Solanum habrochaites* **LYC4 introgression line population and the identification of QTLs for resistance to** *Botrytis cinerea*. *Theor Appl Genet* 2007, **114**:1071-1080.

66. Doganlar S, Frary A, Ku HM, Tanksley SD: **Mapping quantitative trait loci in inbred backcross lines of** *Lycopersicon pimpinellifolium* (LA1589). *Genome* 2002, **45**:1189-1202.

67. Canady MA, Meglic V, Chetelat RT: **A library of** *Solanum lycopersicoides* **introgression lines in cultivated tomato.** *Genome* 2005, **48**:685-697.

68. Peleman JD, van der Voort JR: **Breeding by design.** *Trends Plant Sci* 2003, **8**:330-334.

69. Joosen RV, Ligterink W, Hilhorst HW, Keurentjes JJ: **Advances in genetical genomics of plants.** *Curr Genomics* 2009, **10**:540-549.

70. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE: **Genomic and metabolic prediction of complex heterotic traits in hybrid maize.** *Nat Genet* 2012, **44**:217-220.

71. Toubiana D, Semel Y, Tohge T, Beleggia R, Cattivelli L, Rosental L, Nikoloski Z, Zamir D, Fernie AR, Fait A: **Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations.** *PLoS Genet* 2012, **8**:e1002612.

72. Fei Z, Joung JG, Tang X, Zheng Y, Huang M, Lee JM, McQuinn R, Tieman DM, Alba R, Klee HJ, Giovannoni JJ: **Tomato Functional Genomics Database: a comprehensive resource and analysis package for tomato functional genomics.** *Nucleic Acids Res* 2011, **39**:D1156-1163.

73. Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, *et al*.: **Next-generation mapping of** *Arabidopsis* **genes.** *Plant J* 2011, **67**:715-725.

74. Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jorgensen JE, Weigel D, Andersen SU: **SHOREmap: simultaneous mapping and mutation identification by deep sequencing.** *Nat Methods* 2009, **6**:550-551.

75. Minoia S, Petrozza A, D'Onofrio O, Piron F, Mosca G, Sozio G, Cellini F, Bendahmane A, Carriero F: **A new mutant genetic resource for tomato crop improvement by TILLING technology.** *BMC Res Notes* 2010, **3**:69.

76. Okabe Y, Asamizu E, Saito T, Matsukura C, Ariizumi T, Bres C, Rothan C, Mizoguchi T, Ezura H: **Tomato TILLING technology: development of a reverse genetics tool for the efficient isolation of mutants from Micro-Tom mutant libraries.** *Plant Cell Physiol* 2011, **52**:1994-2005.

77. **The SOL-100 sequencing project** [http://solgenomics.net/organism/sol100/view]

78. Khatri P, Sirota M, Butte AJ: **Ten years of pathway analysis: current approaches and outstanding challenges.** *PLoS Comput Biol* 2012, **8**:e1002375.

79. Pu L, Brady S: **Systems biology update: cell type-specific transcriptional regulatory networks.** *Plant Physiol* 2010, **152**:411-419.

80. Park SJ, Jiang K, Schatz MC, Lippman ZB: **Rate of meristem maturation determines inflorescence architecture in tomato.** *Proc Natl Acad Sci U S A* 2012, **109**:639-644.

81. Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, *et al*.: **The developmental dynamics of the maize leaf transcriptome.** *Nat Genet* 2010, **42**:1060-1067.

82. **CoGePedia** [http://genomevolution.org/wiki/index.php/Sequenced_plant_genomes]