Community Sequencing Program: **Project Proposal**

Project Title:

Sequencing *Solanum pennellii* - the wild parent of tomato introgression lines (ILs) that reveal a molecular and a system view of complex phenotypes

CSP Letter of Intent ID: 'CSP_LOI_783890'

Proposer's Name: Dani Zamir

Participants representing international sequencing groups of the tomato genome: James J. Giovannoni (US, Chromosome # 1, 10, 11), Doil Choi (Korea #2), Huang Sanwen (China #3), Graham Seymour (UK #4), Akhilesh K. Tyagi (India #5), Willem Stiekema (The Netherlands #6), Mondher Bouzayen (France #7), Satoshi Tabata (Japan #8), Antonio Granell Richart (Spain #9), Giovanni Giuliano (Italy #12).

List of Authors: Fernando Carrari, Roger Chetelat, Yuval Eshed, Hiroshi Ezura, Alisdair Fernie, Yossi Hirschberg, Jonathan Jones, Harry Klee, Sandra Knapp, Rob Last, Zachary Lippman, Leonie Moyle, Lukas Mueller, Eran Pichersky, Ramesh Sharma, Mark Stitt, Steve Tanksley, Joyce Van Eck.

A) BRIEF DESCRIPTION

Abstract

The resolution of functional genomic analyses is only as powerful as the diversity of phenotypes that are associated with genome sequences. The focus of this proposal is the sequencing of ***Solanum pennellii* (~1200 Mb)**, which is the wild species donor parent of the extensively phenotyped tomato introgression line (IL) population. The impetus for this project is to leverage the rich complex trait variation exposed in the *S. pennellii* ILs over the last 15 years, as well as the widespread research in international Solanaceae projects. The ILs comprise the most diverse inter-specific quantitative trait mapping population in any system, and are composed of 76 immortalized nearly isogenic lines, each carrying a single marker-defined 'exotic' chromosome segment, developed through a succession of backcrosses with a variety of tomato, *S. lycopersicum*. Solanaceae researchers worldwide have utilized the publicly available full genome coverage ILs to map more than 2500 quantitative trait loci (QTL) affecting plant biomass, yield, drought-tolerance, morphology, metabolism and gene expression. The success of the *S. pennellii* ILs in establishing new principles for plant breeding and for resolving the molecular basis of complex traits was demonstrated by the cloning of the first quantitative trait loci (QTL) ever: *FW2.2* (a fruit size gene) & *Brix9-2-5* (a sugar yield gene). The broad phenotypic diversity, integrated into a single map-based "system" framework, has also led to the identification of heterotic QTL for biomass production -- a major step towards isolating heterosis genes. The sequence of *S. pennellii* is needed to further expose, study, and utilize these and many other genes controlling quantitative phenotypes, particularly those that increase crop yield and biomass.

Solanum pennellii and *S. lycopersicum* are highly divergent species in the family Solanaceae, but are still related enough to produce viable progeny that expose the variation that has driven evolutionary change and provided the raw material for crop domestication and breeding. Presently, the euchromatic space of *S. lycopersicum* is being sequenced BAC by BAC by the International SOL Genome Project (SOL), involving collaborative research groups from 10 countries (~20% complete as of Feb, 2007; Expected completion 2009). SOL serves as a 'virtual umbrella' for many laboratories that will benefit from the sequence of *S. pennellii* as a bridge for linking genes and phenotypes of both fundamental and applied significance. The genome size of *S. pennellii* is similar to tomato (1.2×10^9 vs. 1×10^9 bp), and highly syntenic with potato, eggplant, pepper and other Solanaceae. Shotgun sequences of *S. pennellii* together with the BAC-based approach in tomato and potato will lead to a much better draft of the Solanaceae genome and to an in-depth molecular understanding of IL-based complex traits. This will promote further innovation in plant improvement that will impact the breeding of new energy crops.

Scope of work and relationship to the SOL project

We propose to sequence the genome of an inbred accession of *S. pennellii* (LA0716) to 8X coverage. Data assembly into contigs will be performed by JGI and integrated with the forthcoming sequence of *S. lycopersicum* through the bioinformatics engines of SOL. The partners for the proposal are representatives of the national sequencing programs who will ensure a collaborative and synergistic project for the benefit of a rapidly growing community.

B) TECHNICAL INFORMATION

The haploid **genome size** of *S. pennellii* has been estimated by flow cytometry to be **~1200 Mb**, which is very similar to the ~950 Mb estimate for *S. lycopersicum*[1]. The overall genome organization of *S. pennellii* is highly similar to that of *S. lycopersicum*, despite their ecologically and morphologically divergent phenotypes and their being distantly related taxa in the tomato clade[2]. Both species can inbreed, have the same chromosome number, and in pachytene spreads of *S. pennellii* x *S. lycopersicum* hybrids, species-specific chromosome sets are nearly indistinguishable[3]. Chromosome pairing and chiasma formation is completely normal in the F1 hybrid, and comparison of high density molecular maps (>4000 markers) between *S. pennellii* and *S. lycopersicum* show that the two genomes are nearly identical in gene order and content, with few examples of micro-rearrangements[4,5]. In fact, most genomes of the *Solanaceae* family [including the crop species potato, eggplant and pepper] are similar in size[6], with remarkable synteny throughout the *Solanaceae* as revealed by comparative mapping[7]. In contrast, nucleotide variation, or inter-specific polymorphism, between *S. pennellii* and *S. lycopersicum* is high, especially in non-coding regions. Levels of polymorphism in coding sequences range from 1-3%, but rise to 5-10% in introns and intergenic regions, including cis-regulatory domains where there is a high concentration of insertion/deletion (InDel) polymorphisms[8].

The macrostructure of tomato chromosomes consists of large tracts of pericentric heterochromatin flanked by distal euchromatin, which has been verified both cytologically and bioinformatically. Quantitative staining for DNA on pachytene chromosomes using the Feulgen technique has allowed measurements of the length, width and density of chromosomes showing that euchromatin comprises only 25% of the genome[9-11]. **Tranposable elements** and related heterochromatic repeats are a relatively large component of both the *S. pennellii* and *S. lycopersicum* genomes (~700 Mb), whereas the majority of the estimated 35,000 genes reside on ~220-250 Mb of DNA localized on the euchromatic arms of each of the twelve chromosomes[12]. These estimates are the basis of the sequencing strategy for the International SOL Genome Initiative[13]; analysis of BAC-end sequences and more than 20 Mb of non-redundant genomic DNA deposited to date supports the cytological estimation of the euchromatic space. From these data, extensive repeat libraries have been generated for *S. lycopersicum*. A repeat collection is available from SGN[13] that was generated from assembled BAC end sequences using the RepeatScout, and the repeats were annotated using homology searches. A repeat dataset is also available from TIGR[14]. Annotation of the BAC end sequences with the *de-novo* repeat database generated at SGN shows that 40% of the sequence matches transposon sequences, and another 30% matches other repeats. From heterochromatic BAC sequencing and analyses of the BAC end sequences, it is clear that a large fraction of transposon repeats locate to centromeric regions.

GC content for *S. pennellii* is similar to *S. lycopersicum*, for which extensive sequence resources are available. The *S. lycopersicum* genome sequencing project currently has more than 200 euchromatic BACs fully sequenced; the GC content overall in these BACs is 33.5%[15]. BAC end sequencing indicates that the GC content of the heterochromatin is not significantly different from the euchromatin, as indicated by the absence of bipartite GC content distribution in the BAC end sequence (excluding ribosomal RNA genes,

which have a higher than average GC content). The average GC content (including the ribosomal DNA) of the BAC ends is 36.1%. The GC content of the transcript sequence contigs, or unigenes, derived from the large-scale EST sequencing projects is somewhat higher at 40.5%. Transcript sequences from *S. pennellii* have a GC content that is very close to that number, at 40.8%.

The **level of polymorphism** within *S. pennellii* (LA0716) is low, as expected for an inbreeding accession. To estimate the overall level of polymorphism, an EST collection was analyzed by assembling the ESTs and looking for mis-matches in the assembled sequence alignments. In total, 11188 mismatches were identified as putative SNPs in 596,445 base-pairs of contigged sequence, which corresponds to a polymorphism level of 1.8%. However, the actual level of polymorphism is likely much smaller when accounting for sequencing errors of 1% (phred scores of 20), leaving an upper boundary for the polymorphism rate of 0.8% in *S. pennellii* transcript sequence. This estimate for the low level of polymorphism in *S. pennellii* is supported by the extensive surveys of molecular markers conducted over the past 30 years.

In summary, such extraordinary synteny and similarity in genome size and structure between the species will facilitate alignment of *S. pennellii* whole genome sequencing contigs to the *S. lycopersicum* reference genome. In this respect, each genome will benefit the other by providing mutual anchors for physical assembly, resequencing of coding regions, comparative sequencing of inter-genic and heterochromatic regions, and a comprehensive annotation in a unified database.

Available Biological Resources

An ultra-high density molecular map is available between *S. pennellii* and *S. lycopersicum* having more than 4000 markers mapped on a population of 80 F2 individuals, which corresponds to a marker density of ~ 1 marker/0.35 cM. In addition, a population of 76 *S. pennellii* ILs is available. The two collections have been extensively phenotyped over many years, allowing the robust identification of thousands of morphological and chemical QTL.

The sequencing strategy for *S. lycopersicum* uses euchromatic markers to identify 'seed BACs' from which contigs are then extended. For this purpose, a BAC library of *S. lycopersicum* has been made available with 15X coverage, from which ~180,000 BAC end sequences have already been obtained, allowing for the generation of a first draft physical map. Four additional BAC libraries are available – two for *S. pennellii* and two for *S. lycopersicum*, each ranging from 5 – 8X coverage[16]. The *S. lycopersicum* BAC libraries have also been end sequenced adding further 150,000 sequences to the collection just mentioned. There is an extensive EST library for *S. lycopersicum* representing more than 34,000 unigenes, from which there are ~7000 ESTs from *S. pennellii*. Together, these resources will greatly enable a *S. pennellii* whole-genome shotgun sequencing approach with *S. lycopersicum* serving as a reference genome, both for assembly and annotation.

SOL bioinformatics: Data Access and Dissemination Strategy

As the SOL community has grown, so has the need for centralized databases for genetic, genomic, and phenotypic information. The SOL Genomics Network is one of the nodes of the *S. lycopersicum* genome sequencing project that has the tools in place to house and display DNA sequences and their annotations. For the analysis of the tomato genome sequence, the International Tomato Annotation Group (ITAG) was formed. The group has developed a rigorous annotation pipeline[17] in which several project partners share the workload of the annotation analyses and contribute their expertise; we propose to use the same pipeline for the 8X-derived contigs of the *S. pennellii* sequence.

To maximize the usefulness of the deliverables to the research community, the following data dissemination strategy will be implemented. In addition to submission to Genbank, the sequence and annotation data will be made available in different formats and modes of access on SGN. The goal is to make the information accessible to a large audience with diverse interests in a user-friendly manner. Use cases for access will include advanced query functions for annotations, graphical browsing of the data, comparative queries and visualization, homology searches, and bulk queries. These main modes of access will be implemented as follows: (1) For bulk download of the complete sequence and annotation data, files will be published on the SGN FTP site; (2) For interactive browsing of contigs and their annotations, the data will be loaded into the SGN genome viewer (Gbrowse) that currently displays the tomato genome project annotations; (3) Comparative information with *S. lycopersicum* will be displayed using the SynBrowse viewer; (4) Annotations will be searchable through the SGN database search functions; and (5) Sequences will be searchable through the SGN BLAST tool. In addition, SGN implements a prototype database of annotated genes, and interlinks them to other datatypes, such as sequences, markers and maps, images, loci, alleles, and germplasm. This prototype of the database is also used for representing simple and complex phenotypes of mutant and mapping populations in an easy to browse user-interface using controlled vocabularies. **According to the above strategy the SOL bioinformatics groups will cater the tomato species genome sequences to the community.**

Technical Challenges

The primary technical challenge will be dealing with assembly of repetitive heterochromatic regions of DNA, which is an issue with all large plant genomes. Tomato contains a relatively large proportion of repetitive DNA compared to *Arabidopsis thaliana* (70% vs. 15%, respectively), but assembly will only be hampered in highly repetitive regions, such as core centromeric regions and immediate flanking DNA, where tandem repeats may be abundant along with a high concentration of transposable elements. This issue is much less relevant in the gene rich regions because of the availability of anchored *S. lycopersicum* sequences. Beyond this expected challenge, no other major technical difficulties are foreseen.

Starting Materials

Stocks of inbred *S. pennellii* (LA0716) seed and DNA are readily available as are all mapping populations and BAC libraries.

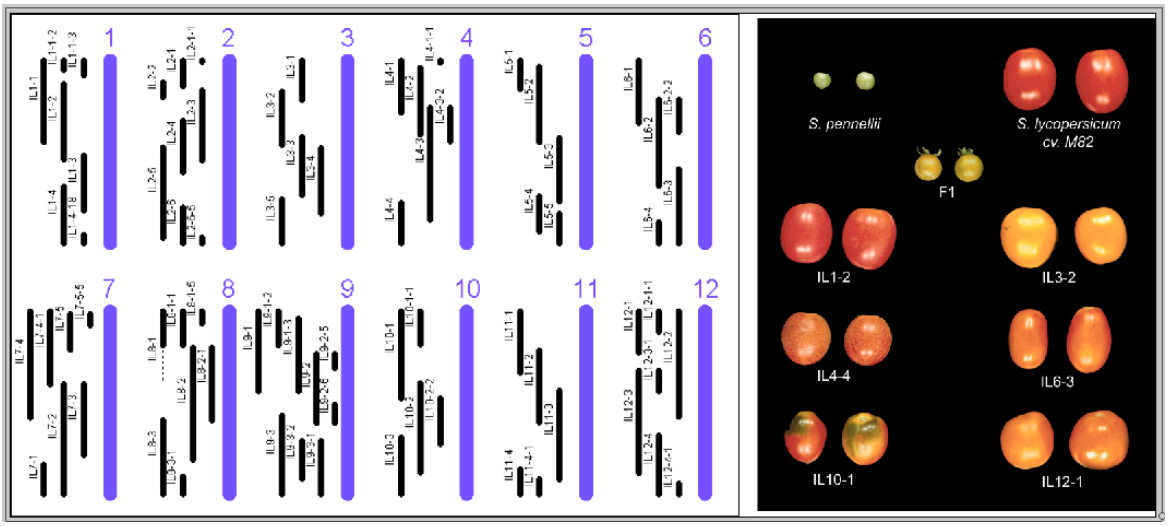
C) PROJECT DESCRIPTION

Tomato as a Model Species for the Genomic Resolution of Complex Traits

Variation in nature often takes the form of a quantitative phenotypic range, as opposed to qualitative phenotypes that fall into discrete Mendelian categories[18, 19]. The genetic variation underlying quantitative traits, such as human height or plant yield, is notoriously difficult to dissect due to the segregation of numerous genes, or quantitative trait loci (QTL), each explaining a portion of the total variation, and whose expression is modified by interactions with other genes and by the environment[20]. Even for the limited number of loci known to control quantitative variation (less than 20), the molecular mechanisms are poorly understood, making it difficult to exploit them for applied biology. **A major bottleneck in the functional annotation of natural genetic variation of both fundamental and applied importance is genetic resources representing wide-diversity, and the sensitivity and scale of phenotyping platforms**.

Plants are the founding model systems of quantitative trait genetics because they are particularly amenable to controlled population development and phenotyping, thereby permitting the use of DNA markers to map complex phenotypes defining myriad aspects of multicellular growth and development. Among all model systems, the wild and domesticated species of the tomato clade in the family Solanaceae have pioneered novel population development to break-down complex traits into Mendelian components, which has facilitated robust assessments of mean phenotypic values - a challenging prerequisite for isolating complex trait genes[21, 22]. In particular, the last fifteen years of research on the tomato *S. pennellii* introgression line (IL) population (Fig. 1) has established the value of such a resource in fundamental biology and real-world applications[23].

Figure 1. The *S. pennellii* IL population. A) Genome introgressions in the 76 *S. pennellii* ILs, which are nearly isogenic to each other and to M82 and differ only for the marked introgressed chromosome segments. B) Green fruits of the wild species, *S. pennellii*, the lycopene rich red fruits of *S. lycopersicum*, their F1 hybrid and six introgression lines (ILs). Five genes affecting the IL color phenotypes were cloned; notably IL6-3 carries an allele of ***Beta(B)*** that is up regulated during ripening and thus beta-Carotene (pro-Vitamin-a) is increased by 30 fold[24]. This figure highlights one of the major characters of exotic variation breeding where **the progeny phenotypes can not be predicted based on those of their parents (transgressive variation).**



The phenotypic resolution provided by the *S. pennellii* ILs is much greater than other systems because of the wide-diversity introduced from the *S. pennellii* donor genome – a rare drought tolerant, green-fruited perennial species (see below). **We are proposing to sequence *S. pennellii*, which, when united with current sequencing efforts of the domesticated tomato, *S. lycopersicum*, will enable the Solanaceae community to decipher the molecular nature and interactions of complex traits as revealed in the comprehensive IL phenotypic database**.

The Value of 'Exotic Libraries' in Plant Breeding as Revealed by S. pennellii

From an applied standpoint, the ability to resolve complex variation into single gene components is perhaps most important in plant breeding, which is the art and science of the genetic improvement of crops for increased quality, productivity and environmental friendliness[23]. Ancient breeders who made the first dramatic advances in crop productivity worked 10,000 years ago in the Fertile Crescent of the Near East domesticating several cereal and pulse species[25]. Domestication of nearly all crop plants has been based on only a few founding genotypes, resulting in a 'founder effect' in crop evolution explaining why many species contain only a small fraction of the genetic variation that is present in their wild relatives. Such narrow variation has, in turn, limited the breeding potential of all major crop species[26, 27].

Tomato breeders have been at the forefront of establishing new principles for crop breeding over the past 50 years by using exotic germplasm to improve modern cultivars[28, 29]. Already in 1969, Prof. Charles Rick generated *S. pennellii* ILs for four regions of the genome[3]. In the past decade, the ever-growing tomato research community has invested in the development of permanent mapping populations using wild species introgression breeding in the genetic background of elite genotypes[30]. These exotic libraries have been and are being screened using cutting edge phenotyping platforms to identify novel alleles of economic and scientific importance. Among all populations, the *Solanum pennellii* ILs have been adopted as the standard for exploring and utilizing the hidden breeding potential of wild species to improve crop quality and nutrition, and increase biomass and yield, also in drought stress environments[31]. The *S. pennellii* introgression lines have demonstrated that future plant breeding challenges of producing better food and biofuel crops can be based on the wild ancestors of crop plants[32]. **Such genetic diversity is the engine that propels plant breeding to meet its goals, which will be expedited with a *S. pennellii* genome sequence.**

The Family Solanaceae as a Model for Exploring and Exploiting Natural Variation

Solanum pennellii is one of the most distant relatives of the cultivated tomato *S. lycopersicum* within the tomato clade, and both are members of the genus Solanum in the family Solanaceae in the derived Euasterid clade[33]. The Solanaceae family includes more than 3000 species, many of which evolved in the Andean/Amazonian regions of South America in extreme habitats that vary dramatically and include rain forests that receive more than 3 meters of rainfall annually, deserts with virtually no rainfall and high mountains with regular snowfall and sub freezing temperatures. Consequently, Solanaceae exhibit wide adaptation, form and chemistry. Representative species include -- potato, tomato, tobacco, petunia, pepper and related Euasterid species (e.g. Coffee, Mimulus and Antirrhinum).

The tomato clade is the most intensively studied of all Solanaceae groups and includes 13 core species[33]. *Solanum pennellii* is a basal species in the tomato clade, and has evolved unique adaptations in terms of morphology, mating system, chemistry (especially secondary compounds) and responses to biotic/abiotic stress. *Solanum pennellii* inhabits one of the driest deserts on Earth (the Atacama), where rainfall is scarce (0-10mm/year) and episodic. Endemic to Peru, *S. pennellii* occurs as scattered populations consisting of relatively few individuals – it is such a rare plant in the wild that even Prof. Charles Rick, who collected many of the wild tomato accessions available today, completely missed it during his first forays into the native region. Despite their drastic differences in ecology, *S. pennellii* is sexually compatible and produces fertile hybrids with *S. lycopersicum*, making it ideally suited for exploring the genetic and molecular basis of inter-specific variation, environmental adaptation, reproductive biology, and many other topics[2, 3].

Thus, populations derived from *S. pennellii* and *S. lycopersicum* are unique tools for exploring Solanaceae, which is an excellent taxon to address a central question in biology posed as a major research goal of the International SOL Genome Project[34]:

- How can a common set of genes/proteins give rise to such a wide range of morphologically and ecologically distinct organisms?

The corollary question of agricultural importance is:

- How can a deeper understanding of the genetic basis of plant diversity be harnessed to better meet the needs of society in an environmentally-friendly and sustainable manner?

A key attribute of Solanaceae that make them particularly suited to address the above questions is the high overall conservation of their genome structure and gene repertoire[35]. Whole genome comparative genetic maps of tomato, potato, eggplant and pepper reveal large tracts of collinear markers (e.g. five paracentric inversions differ between tomato and potato), which are consistent with their similar gene content[36-38]. Considerable synteny is also evident beyond the family Solanaceae family as was exemplified recently for Coffee (a member of the family Rubiaceae, also in the Euasterid clade), which is being genetically mapped using orthologous markers[39]. Single copy orthologs (COS)-based maps will be further used to link a whole host of related Solanaceous species to a common framework. Thus, a deeper understanding of the molecular basis of natural variation within the Solanaceae, beginning with the integration of genome sequence from *S. pennellii* and *S. lycopersicum*, will directly impact numerous other plant species, including many with worldwide agricultural importance.

The Solanum pennellii Introgression Lines (ILs)

Introgression lines (ILs) are a set of nearly isogenic lines (NILs) developed through a succession of backcrosses, where each line carries a single genetically defined chromosome segment from a divergent genome[23]. In many cases, QTL mapping studies involve whole genome segregating populations, but from a practical plant breeding perspective, epistatic interactions in segregating populations, whether F2 or RILs, make it difficult to fully define and characterize individual loci that control complex phenotypes. ILs are largely devoid of epistasis, because unlinked QTL from other regions of the

genome are absent. A complete IL population has enough members to reconstitute the donor parent in overlapping chromosomal segments and is immortal since it can be maintained by self-pollination. Consequently, these populations are very effective in identifying QTL, because any phenotypic difference between an IL and the recurrent parent is attributed to the introgressed chromosomal segment.

In tomato, the *S. pennellii* ILs were the founding members of the first introgression line population used for QTL identification and plant breeding (Fig. 1). The ILs, representing whole-genome coverage of *S. pennellii* in overlapping segments in the genetic background of *S. lycopersicum* cv. M82, were first phenotyped in 1993, and presently this library consists of 76 genotypes. In the framework of a currently running EU project (EU-SOL)[40], 500 sub-ILs are being produced in a manner such that the mapping resolution of target traits will be markedly improved. Because the ILs differ by only a single defined chromosomal segment, the resulting lines generally resemble the cultivated parent, thus allowing the reproducible mapping of QTL for the most complex integrated traits in plants - yield and biomass production. The *S. pennellii* ILs have been publicly available for more than 15 years, and have been requested 265 times from the Tomato Genetics Resource Center -TGRC[41] resulting in the distribution of 6682 seed samples; LA0716 is the single most frequently requested item in TGRC, which maintains over 3600 lines (see R. Chetelat, support letter). IL seed were also made available to the Solanaceae community and distributed from Israel (a total of 3400 seed envelopes). Over the years the ILs have been phenotyped for hundreds of traits including repeated measurements of the same traits, thus allowing for the identification of 2795 QTL (Table 1). Importantly, the QTL data is most robust for the components of yield and biomass - 10-seasons of repeated measurements are available.

Table 1. The prevalence and diversity of QTL identified by the *S. pennellii* ILs.

QTL #	Trait #	Traits measured	Reference
36	6	Carotenoids content	McQuinn and Giovannoni, unpub
18	2	Fruit Phenotypes	White and Giovannoni, unpub
6	1	Ascorbate (Vitamin C)	McQuinn and Giovannoni, unpub
584	101	Primary metabolites	Semel Y, unpub
889	74	Primary metabolites	Schauer, et al. (2006)[42]
841	35	Morphology and yield	Semel, et al. (2006)[43]
88	23	Volatile compounds	Tieman et al. (2006)[44]
20	3	Nutritional and antioxidant	Rousseaux, et al. (2005)[45]
82	9	Sugars and acid content	Baxter et al. (2005[ab])[46, 47]
81	9	Metabolites, brix and fruit wt.	Causse et al. (2004)[48]
30	8	Leaf morphology	Holtan and Hake (2003)[49]
16	1	Intensity of red color of ripe fruit	Liu et al. (2003)[50]
104	6	Yield related traits	Eshed and Zamir (1995)[31]

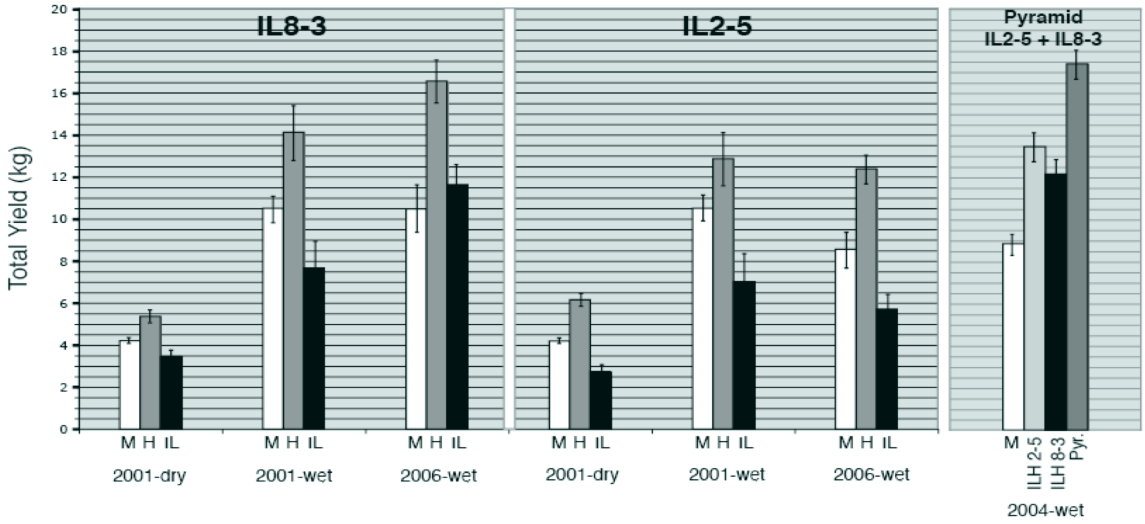
The benefits of the IL approach are now being realized in other major food and biofuel crops. IL population structures were developed for pepper[51, 52], rice[53, 54], barley[55], wheat[56], maize[57], soybean[58], *Arabidopsis*[59], and even mice[60]. Yet, the amount of polymorphism detected in the *S. pennellii* IL population is exceptionally high due to the divergent nature of its parents. Thus, while phenotyping of five traits in an intra-specific IL population of *Arabidopsis* uncovered, on average, 5 QTL per trait, analyzing the *S. pennellii* ILs for

comparable traits and statistical procedures (Dunnet 5%) revealed 17 QTL per trait. Given the high phenotypic variability and the accumulation of 15 years worth of communal data, the *S. pennellii* ILs will continue paving the way for the exploitation of biodiversity-based nearly isogenic populations. **To fully benefit from this permanent and dynamic resource we need the genome sequence of both IL parents.**

Leveraging Solanaceae to Unlock the Mystery of Heterosis for Yield and Biomass

Heterosis, or hybrid vigor, is a major contributor to agricultural production world-wide[61]. Such hybid vigor, or heterosis, was discovered by maize breeders nearly a century ago as a "miraculous" agricultural phenomenon that has subsequently been found to occur in many crop species[62]. Nearly 100 years of research suggests that the genetic basis of this hybrid vigor is determined by non-mutually exclusive mechanisms that include dominance complementation, overdominance and epistasis[63, 64]. The importance of heterosis in agriculture is evident from the dramatic increases in yield measured over the past 50 years following the influx of hybrids to crop production; yield advantages due to hybrids ranges between 15-50%, depending on the crop[65]. With such great benefits it is clear that the breeding of future food and biofuel crops will be based on principles that govern heterosis, yet those principles and their molecular determinants are still not known.

Figure 2. *Solanum pennellii* IL heterosis. IL8-3 and IL2-5 were crossed to M82 (M) and the hybrids (H) and parents were evaluated for total fruit yield per plant (per M^2) in three growing seasons. In 2001 the plants were grown in normally irrigated field (wet) and in dry conditions that received 10% of the water[66]. The pyramiding of the two introgressions was achieved by crossing the ILs to generate a hybrid that is heterozygous for both genomic segments (Pyr).



The *S. pennellii* ILs are being used to discover the genes that govern heterosis. In a large phenomic oriented field study, the 76 homozygous *S. pennellii* ILs were crossed to the recurrent parent M82 to test the effect of heterozygosity on 35 diverse phenotypes with the eventual identification of the chromosomal segments that contribute to heterosis[43]. The analysis revealed 841 QTL where heterotic QTL were much more prevalent for phenotypes associated with reproductive fitness. For example, the mean

fruit yield of plants heterozygous for IL2-5 and IL8-3 was significantly higher than both parents in three years of testing in irrigated fields and under drought conditions (Fig. 2). Moreover, the pyramiding of the two heterotic introgressions in the heterozygous condition further increased yield beyond the individual components.

The increased confidence in the stability of heterosis effects resulting from multi-year phenotying has led to a major effort in fine mapping heterotic ILs to smaller intervals through recombination-mediated reduction of the QTL-carrying segment. A similar high resolution mapping approach was employed in past studies of the *S. pennellii* ILs leading to the isolation of the first two QTL ever cloned: *FW2.2*[21], which is one of 20 QTL that affect fruit weight and *Brix9-2-5*[22], which, together with at least 25 other genomic regions, increase sugar yield of the tomato fruit. Furthermore, the IL breeding concept has progressed beyond scientific publications into practical use in agriculture[66]: **Introgressions originating from *S. pennellii* were introduced into lines of processing tomato and the resulting hybrid, AB2, is presently the leading variety in California[67], which is the largest world producer of industrial processing tomatoes.**

IL Resolution of Integrated Developmental Networks – A Case Study

Of all model plant species, tomato is one of the few with "tree-like" characteristics. This feature is expressed both in the perennial habit of the plant and in its sympodial growth. The first, and so far most important agronomical and developmental switch to perennial habit in tomato lies in the *SELF PRUNING* (*SP*) gene[68]. Mutants in this gene exhibit compact "determinate" growth that deviate from the vine-like "indeterminate" habit of wild tomato species. As all *S. pennellii* ILs are in the background of the "determinate" line M82, the "indeterminate" line, IL6-3, immediately provided an isogenic perennial. Having this IL tool in hand facilitated the identification of *SP* by map based cloning and led to the discovery of its family member *SFT* - the long sought after Florigen[69] [70].

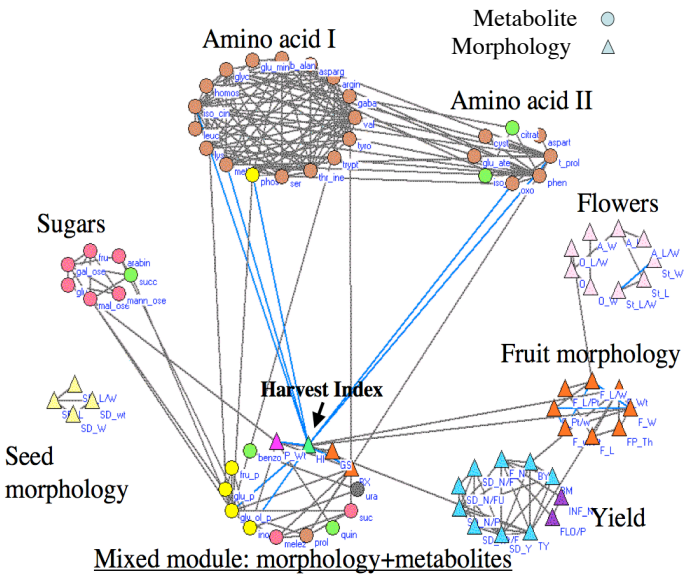


Figure 3. A system view of IL-born morphology and metabolism interplay. Cartographic representation of the combined metabolic and morphological networks of the tomato ILs. Each trait (node) is represented by a shape (morphological by a triangle and metabolic a circle). Correlation of all trait pairs was calculated using IL means (total of 76 lines); black lines represent positive correlations, blue lines represent negative correlations (Significance threshold of P<0.0001). The network algorithm partitions the nodes into modules, which are groups of vertices that are connected between themselves more than to nodes from other modules. Harvest index (HI), the ratio of fruit yield to total plant mass (plant weight + fruit yield)[72], is the central pleiotropic hub of the network.

Another feature of the tomato model is its fleshy fruit, and the numerous associated specialized metabolic pathways, including many important for human health[71]. A large-scale association study linking morphology and biochemistry was carried out by phenotyping the *S. pennellii* ILs for a wide range morphological and fruit metabolic profiles[42]. An integrated cartographical network revealed morphology-dependent and independent metabolic links and identified 889 QTL for fruit metabolism and 326 QTL that modified yield associated traits (among them 36 QTL that improve plant biomass). The analysis showed that Harvest Index[72] (HI; Fig. 3), which modulates the source-sink relationship between the vegetative to reproductive parts, is the central hub in fruit metabolism. The major pleiotropic gene and metabolic QTL was *SP*, which regulates plant architecture as described above[42]. Using this approach, multi-faceted phenotypic analyses of a wide biodiversity base permit a systems level perspective for classical biological questions. **This example illustrates that with the *S. pennellii* sequence in hand, more detailed systems-level analyses will be possible opening the door to discoveries through extending the phenomic network.**

Partitioning of Chemical Diversity

The rich chemical diversity in the Solanaceae is arrayed on a broad canvas of phenotypic variation. The area of metabolic research in the Solanaceae is very active as most of the structural and regulatory genes responsible for this diversity have not been discovered. Vast differences have been observed in the content of the energetically important polymeric carbohybdrate storage compounds, such as starch and cell wall components. There is also a high level of variance at the level of various primary metabolites, particulary with respect to sugar and amino acid content[42].

Comparative genomics approaches made use of several different Solanum species, including *S. pennellii* and its ILs, to characterize novel biosynthetic pathways and enzymes in the fruit and in specialized cells, the glandular trichomes. Metabolic profiling surveys have detected a large degree of change in the levels of various secondary metabolites including phenylpropanoids, pigments, phenolics, terpenoids, and fatty acid derivatives. For example, the ILs have already been used in combination with the reverse genetic approach to clarify the functional pathway of one of the most important fruit flavor volatiles[44].

The specialized metabolism in the glandular trichomes of *S. pennellii* is of particular interest since it secretes large amounts of fixed carbon as 'acyl sugars' - polyesters of glucose and sucrose decorated with fatty acids of various lengths[73, 74]. Moreover, volatile profiles are highly variable within the tomato clade, mostly in the content of sesquiterpenes[75, 76], and much of this biosynthesis occurs in the same glandular trichomes. Understanding the biosynthetic and secretion pathways for these polyesters, which have insecticidal properties, will provide insights about novel chemistries as well as elements of the interactions of plants with their biotic environment in nature and in cultivation[37].

Given that the differences between the alleles of the two species are both positive and negative with respect to fruit quality and human nutrition, knowledge of the overall levels of sequence differences between *S. pennellii* and *S. lycopersicum*, and of changes at

specific loci, will help in understanding the evolution of the various Solanum lineages and provide a rich resource for elucidating the flow of biochemical pathways.

Further Dimensions of the Solanum pennellii Sequence

The research activities described above focused on the demonstrated resolution power of the IL population. The briefly described topics below outline part of what could be explored once the genome sequence of both IL parents is available:

QTL epistasis: The advantage of ILs in resolving individual QTL is also a **drawback** - epistatic interactions between unlinked loci, which are a major component of the phenotypic variation, cannot be directly estimated. For this purpose, a new *S. pennellii* (LA0716) based population of 1000 backcrossed inbred lines (BIL; BC2 and BC3 selfed) is being constructed in the M82 background. Each genotype in BIL1000 carries multiple wild species introgressions permitting phenotypes to be associated with specific epistatically interacting QTL. Individual ILs can then be used to reconstruct any epistasis detected in BIL1000 to study the genetic and molecular components underlying specific interactions.

Genome evolution following introgression: If you were to make an IL population from scratch, would you get the same set of QTL/phenotypes? Did new phenotypes arise indepependently of the origin of the parental genome? Does the genome evolve upon introgression and recombination between diverse genomes? How does it evolve? Is genome evolution directional or random? Having the *S. pennellii* genome will expose structural changes in the ILs on a macro- and micro-scale, along with changes in transposon regulation, copy number, and distribution, as was found in wide crosses between synthetic polyploids[77].

Changes in gene expression networks: Having in hand the allelic variation that make the *S. pennellii* ILs, an allele-specific set of probes can be designed on a genomic scale. Using such expression arrays will allow for the comparison of transcriptional networks in ILs versus their parental backgrounds. Furthermore, *S. pennellii* sequence will allow for exploring the evolution of the transcriptome following diverse genome hybridization and its relationship to heterosis. This includes the role of small RNA molecules in evolution and plant breeding, and the contribution of specific alleles to overall gene expression.

Epigenetics and complex trait variation: An unresolved question in complex trait biology is whether epigenetic variation, such as variation in DNA methylation patterns, has a major role in phenotypic variation[78]. Current populations to address this question are limited because they are based on intra-specific crosses. Epigenomic profiling in *S. pennellii* ILs has great potential to explore epigenome evolution and its phenotypic consequences[79].

Conservation of QTL across the tomato clade: IL populations are now available for a number of additional tomato species *S. habrochaites*[80-82], *S. peruvianum*[83], *S. pimpinellifolium*[84], *S. lycopersicoides*[85], *S. neorickii*[86], *S. chmielewskii*[87], *S. chilense*[88], and a different accession of *S. pennellii* (LA1657)[89]. These populations are being phenotyped widely but are also being used for focusing on specific traits, such as the genetic basis of interspecific crossability barriers and the development of strategies to overcome them in applied breeding[90]. Recently, the quantification of recombination rates in different tomato IL crosses revealed an efficient and elegant approach to enhance crossing over thus

improving the efficiency of elimination of linkage drag as well as providing a tool to explore the molecular biology of recombination rates in plants[91].

Further association of phenotypes with sequences: The above experiments rely on the extensive allelic variation between cultivated and exotic germplasm. However, the golden standard for understanding biological processes is still knock-out alleles. Thus, beyond introgression populations many other resources are available and being developed in tomato for establishing causal links between DNA and traits. These include: 1) A phenotyped core collection of 7000 accessions representing heirloom varieties and modern inbreds as well as ancient varieties from Mexico and wild species[40]. 2) A saturated mutagenesis population (13,000 M2 families and 3800 individual mutants) in the M82 background that is amenable to TILLING aimed at isolation variation in specific gene sequences[92, 93]. 3) The MICROTOM dwarf variety which is suitable for high throughput experimentation in limited space[94-96].

Integration of IL Informatics and Beyond

A current bioinformatic challenge is the creation of frameworks that describe the genetic components of subtle QTL variation in relation to genome sequence[97]. One objective is to include detailed information in genomic databases about IL phenotypic differences in a form of statistical and graphical outputs that describe the components of the genetic variation and their inter-relationships. A first-generation *S. pennellii* IL based working model that enables mapping QTL, viewing basic trends, associations between traits, and finding specific combinations of phenotypes was implemented on SGN and is called 'Real Time QTL'[98, 99]. Another database that exhibits *S. pennellii* IL data is the Tomato Metabolite Database (TOMET[100]) , which includes the expression profile data of more than 10,000 unique genes and composition data of approximately 60 metabolites that contribute to fruit flavor and human nutrition. Data retrieval and analysis functions have been developed that allow users to identify gene expression ratios that correlate with accumulation of a selected metabolite in all or a subset of ILs for a given trial.

The goal of the SOL community is to apply a common system to integrate and mine QTL information across the IL experiments[101-103]. Such a system should combine and organize information from public databases that integrates QTL, genetic and physical maps across the species. In order to make progress towards this goal a task force was established to combine existing tools with various R statistical packages[104] or AnimalQTLdb[105], and adapt them for our purposes that aim to use ILs as a basis for a system view of complex phenotypes. This engine will continually be expanded to incorporate QTL information from additional tomato-based crosses as well as from other Euasterid taxa.

A proof of concept for the beyond the Solanaceae complex trait view has been the identification an orthologous yield associated QTL common to potato and tomato: The invertase gene *invGE* co-localizes with cold-sweetening and starch content QTL from potato[106]. The potato invertase is orthologous to the tomato *Lin5*, which is the fruit-sugar-yield QTL *Brix9-2-5*[107], indicating that natural variation of sugar yield in tomato fruits and potato tubers is controlled by functional variants of orthologous genes. **This example illustrates the potential of the Solanaceae family to link new dimensions of a biology that is moving from the 'vertical' age of focusing on a single species to a 'horizontal' view comparing traits across species both functionally and phylogenetically.**

The SOL Community

A meeting was held in Washington DC on Nov 3, 2003 where representatives of the grassroots Solanaceae community established the International Solanaceae Genome Initiative (SOL) with the aim of creating a coordinated network of knowledge about the Solanaceae family in order to address key questions in biology and agriculture (see the SOL white paper at[17]). SOL functions as a 'virtual umbrella' for this community-led initiative by promoting, coordinating and actively seeking additional scientists, countries and funding agencies to participate in the ten-year expedition to develop, understanding and utilize natural biodiversity. Presently SOL includes scientists from 28 countries, 10 of which are involved in sequencing the cultivated tomato genome. Since 2004 in the Netherlands, the SOL community has been conducting yearly meeting; more than 500 scientists attended SOL2006 at Madison, WI, which included physiologists, geneticists, taxonomists, breeders, and computational biologists – groups of scientists that infrequently mingle. In addition SOL country workshops have also been held during the past years in the UK, India, Japan, Spain and Italy. Thus over the past three years a diverse SOL community has become united around common biological questions that relate to biodiversity and its exploration.

The use of natural variation is now recognized as a key strategy in functional and applied genomics as this diversity represents the fruits of millions of years of evolution as compared to man-made mutagenesis. In this light, the long-term objective of the SOL community is to create a common Solanaceae-based genomic framework that includes sequences, phenotypes, and functional characterization of 100 genomes. The present proposal will provide the methodological and technological foundations for the 'wider view' of SOL100 by focusing first on *S. pennellii*.

The BAC-by-BAC sequence developed by the international community for *S. lycopersicum* will insure accurate localization, organization and assembly of *S. pennellii* shotgun sequence across the gene-rich euchromatin, while the *S. pennellii* shotgun sequence will provide a view of the heterochromatin sequence organization and gene content that will be absent from the *S. lycopersicum* project. The major international effort to sequence the potato genome on a BAC by BAC basis relies on the most marker populated map of any Solanaceae species[108, 109] and will undoubtedly contribute in constructing a draft of the SOL genome. Infrastructure for annotation and display of SOL genome sequence is in place and will be used to integrate and disseminate the proposed sequence for *S. pennellii*. To date it is estimated that the SOL initiative has helped in raising funds which amount to ~ 60 M $ (including 40 M $ from EU-SOL). Such a large worldwide endowment is already attracting young scientists who intend to look at nature through the eyes of the Solanaceae.

Relevance of the Project to the DOE Mission

The DOE goal of alternative energy crops are linked at its base to advancing concepts, techniques and tools for plant breeding. In recent years it became clear that naturally occurring genetic variation can act as reagents for discovery and characterization of new alleles that underlie traits of agricultural value[27, 110]. This is true for wild relatives of crop plants as well as new species that could be domesticated and bred for their biofuel potential.

Solanum pennellii created the source of the most advanced complex trait mapping population that proved the utility of the 'exotic library' breeding principles and is setting the road map for the use of such resources in other organisms. More than a decade of phenotypic measurements has yielded thousands of QTL, which are waiting to be localized to their individual gene components. Specifically, *S. pennellii* is a resource to identify, manipulate, and incorporate genes controlling plant growth and biomass production, and responses to drought and salt stress, which are key to DOE goals. *S. pennellii* research also sheds a renewed light on hererosis, or hybrid vigor, which is a major genetic force that contributes to world food and biomass production. Further understanding of the molecular basis of complex traits derived from wild germplasm will reveal ways to overcome conserved bottlenecks for yield and introduce a higher level of predictability to the domestication and breeding of biofuel crops.

In summary, the sequence of *S. pennellii*, when united with *S. lycopersicum* will pave the way for a molecular understanding of complex traits in plants with real-world applications unrivaled by other model systems.

D) REFERENCES

1. Arumuganathan, K. & Earle, E. Estimation of nuclear DNA content of plants by flow cytometry. Plant. Molec. Biol. Rep. 9, 208-218 (1991).
2. Rick, C. M. Hybridization between *Lycopersicon esculentum* and *Solanum pennellii*: phylogenetic and cytogenetic significance. Proc Natl Acad Sci U S A 46, 78-82 (1960).
3. Rick, C. M. Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum*: segregation and recombination. Genetics 62, 753-768 (1969).
4. Barrero, L. S., Cong, B., Wu, F. & Tanksley, S. D. Developmental characterization of the fasciated locus and mapping of Arabidopsis candidate genes involved in the control of floral meristem size and carpel number in tomato. Genome 49, 991-1006 (2006).
5. van der Knaap, E., Sanyal, A., Jackson, S. A. & Tanksley, S. D. High-resolution fine mapping and fluorescence in situ hybridization analysis of sun, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics 168, 2127-40 (2004).
6. http://www.rbgkew.org.uk/cval/.
7. Paterson, A. H. et al. Comparative genomics of plant chromosomes. Plant Cell 12, 1523-40 (2000).
8. Eshed, Y., Lippman, Z. & Mueller, L. Pers. Comm.
9. Wang, Y. et al. Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. Genetics 172, 2529-40 (2006).
10. Peterson, D. G., Price, H. J., Johnson, J. S. & Stack, S. M. DNA content of heterochromatin and euchromatin in tomato (*Lycopersicum esculentum*) pachytene chromosomes. Genome 39, 77-82 (1996).
11. Sherman, J. D., Herickhoff, L. A. & Stack, S. M. Silver staining two types of meiotic nodules. Genome 35, 907-15 (1992).

12. Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G. & Tanksley, S. Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell 14, 1441-56 (2002).
13. http://www.sgn.cornell.edu/.
14. http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=tomato.
15. Mueller, L. Pers. Comm.
16. Giovannoni, J. Pers. Comm.
17. http://www.sgn.cornell.edu/solanaceae-project/index.pl.
18. Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu Rev Plant Biol 55, 141-72 (2004).
19. Mackay, T. F. The genetic architecture of quantitative traits. Annu Rev Genet 35, 303-39 (2001).
20. Salvi, S. & Tuberosa, R. To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10, 297-304 (2005).
21. Frary, A. et al. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289, 85-8 (2000).
22. 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 97, 4718-23 (2000).
23. Zamir, D. Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2, 983-9 (2001).
24. Ronen, G., Carmel-Goren, L., Zamir, D. & Hirschberg, J. An alternative pathway to beta -carotene formation in plant chromoplasts discovered by map-based cloning of beta and old-gold color mutations in tomato. Proc Natl Acad Sci U S A 97, 11102-7 (2000).
25. Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. Cell 127, 1309-21 (2006).
26. Tanksley, S. D. & McCouch, S. R. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277, 1063-6 (1997).
27. McCouch, S. Diversifying selection in plant breeding. PLoS Biol 2, e347 (2004).
28. Powers, L. Inheritance of quantitative characters in crosses involving two species of *Lycopersicon*. Jour. Agri. Res. 63, 149-175 (1941).
29. Rick, C. M. High soluble-solids content in large-fruited tomato lines derived from a wild green-fruited species. Hilgardia 42, 493-510 (1974).
30. Mueller, L. A. et al. The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol 138, 1310-7 (2005).
31. 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 141, 1147-62 (1995).
32. Varshney, R. K., Graner, A. & Sorrells, M. E. Genomics-assisted breeding for crop improvement. Trends Plant Sci 10, 621-30 (2005).
33. Knaap, E., Bohs, L., Nee, M. & Spooner, D. M. Solanaceae - a model for linking genomics with biodiversity. Comp. Funct. Genom. 5, 285-291 (2004).
34. Mueller, L. et al. The Tomato Sequencing Project, the first cornerstone of the International Solanaceae Project (SOL). Comp. Funct. Genom. 6, 153-158 (2005).

35. Rensink, W. A. et al. Comparative analyses of six solanaceous transcriptomes reveal a high degree of sequence conservation and species-specific transcripts. BMC Genomics 6, 124 (2005).
36. Livingstone, K. D., Lackney, V. K., Blauth, J. R., van Wijk, R. & Jahn, M. K. Genome mapping in Capsicum and the evolution of genome structure in the Solanaceae. Genetics 152, 1183-202 (1999).
37. Gebhardt, C. & Valkonen, J. P. Organization of genes controlling disease resistance in the potato genome. Annu Rev Phytopathol 39, 79-102 (2001).
38. Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N. & Tanksley, S. D. A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the solanaceae. Genetics 161, 1697-711 (2002).
39. Wu, F., Mueller, L. A., Crouzillat, D., Petiard, V. & Tanksley, S. D. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the Euasterid plant clade. Genetics 174, 1407-20 (2006).
40. http://www.eu-sol.net/.
41. http://tgrc.ucdavis.edu/.
42. Schauer, N. et al. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol 24, 447-54 (2006).
43. Semel, Y. et al. Overdominant quantitative trait loci for yield and fitness in tomato. Proc Natl Acad Sci U S A 103, 12981-6 (2006).
44. Tieman, D. et al. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. Proc Natl Acad Sci U S A 103, 8287-92 (2006).
45. Rousseaux, M. C. et al. QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. Theor Appl Genet 111, 1396-408 (2005).
46. Baxter, C. J. et al. Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. Plant Cell Physiol 46, 425-37 (2005).
47. Baxter, C. J., Sabar, M., Quick, W. P. & Sweetlove, L. J. Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. J Exp Bot 56, 1591-604 (2005).
48. Causse, M. et al. A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot 55, 1671-85 (2004).
49. Holtan, H. E. & Hake, S. Quantitative trait locus analysis of leaf dissection in tomato using *Lycopersicon pennellii* segmental introgression lines. Genetics 165, 1541-50 (2003).
50. Liu, Y. S. et al. There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnol J 1, 195-207 (2003).
51. Ben Chaim, A., Borovsky, Y., Rao, G. U., Tanyolac, B. & Paran, I. *fs3.1*: a major fruit shape QTL conserved in Capsicum. Genome 46, 1-9 (2003).

52. Chaim, A. B., Borovsky, Y., De Jong, W. & Paran, I. Linkage of the *A* locus for the presence of anthocyanin and *fs10.1*, a major fruit-shape QTL in pepper. Theor Appl Genet 106, 889-94 (2003).
53. Ashikari, M. & Matsuoka, M. Identification, isolation and pyramiding of quantitative trait loci for rice breeding. Trends Plant Sci 11, 344-50 (2006).
54. Xie, X. et al. Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon*. Theor Appl Genet 113, 885-94 (2006).
55. von Korff, M., Wang, H., Leon, J. & Pillen, K. Development of candidate introgression lines using an exotic barley accession (*Hordeum vulgare* ssp. *spontaneum)* as donor. Theor Appl Genet 109, 1736-45 (2004).
56. Liu, S., Zhou, R., Dong, Y., Li, P. & Jia, J. Development, utilization of introgression lines using a synthetic wheat as donor. Theor Appl Genet 112, 1360-73 (2006).
57. Szalma, S. J., Hostert, B. M., Ledeaux, J. R., Stuber, C. W. & Holland, J. B. QTL mapping with near-isogenic lines in maize. Theor Appl Genet (2007).
58. Concibido, V. C. et al. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor Appl Genet 106, 575-82 (2003).
59. Keurentjes, J. J. et al. Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics 175, 891-905 (2007).
60. Singer, J. B. et al. Genetic dissection of complex traits with chromosome substitution strains of mice. Science 304, 445-8 (2004).
61. Birchler, J. A., Auger, D. L. & Riddle, N. C. In search of the molecular basis of heterosis. Plant Cell 15, 2236-9 (2003).
62. Duvick, D. N. Biotechnology in the 1930s: the development of hybrid maize. Nat Rev Genet 2, 69-74 (2001).
63. Lippman, Z. B. & Zamir, D. Heterosis: revisiting the magic. Trends Genet 23, 60-6 (2007).
64. Springer, N. M. & Stupar, R. M. Allelic variation and heterosis in maize: How do two halves make more than a whole? Genome Res 17, 264-75 (2007).
65. Duvick, D. N. in The genetics and exploitation of heterosis in crops (eds. Coors, J. G. & Pandey, S.) 19-29 (ASSA, CSSA, SSSA, Madison, 1999).
66. Gur, A. & Zamir, D. Unused natural variation can lift yield barriers in plant breeding. PLoS Biol 2, e245 (2004).
67. http://www.ptab.org/ranking9.htm.
68. Pnueli, L. et al. The *SELF-PRUNING* gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFL1*. Development 125, 1979-89 (1998).
69. Lifschitz, E. et al. The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. Proc Natl Acad Sci U S A 103, 6398-403 (2006).
70. Lifschitz, E. & Eshed, Y. Universal florigenic signals triggered by *FT* homologues regulate growth and flowering cycles in perennial day-neutral tomato. J Exp Bot 57, 3405-14 (2006).

71. Ellinger, S., Ellinger, J. & Stehle, P. Tomatoes, tomato products and lycopene in the prevention and treatment of prostate cancer: do we have the evidence from intervention studies? Curr Opin Clin Nutr Metab Care 9, 722-7 (2006).
72. Yang, J. & Zhang, J. Grain filling of cereals under soil drying. New Phytol 169, 223-36 (2006).
73. Mutschler, M. A. et al. QTL analysis of pest resistance in the wild tomato *Lycopersicon pennellii*: QTLs controlling acylsugar level and composition. Theor Appl Genet 92, 709-718 (1996).
74. Hartman, J. B. & St Clair, D. A. Variation for aphid resistance and insecticidal acyl sugar expression among and within *Lycopersicon pennellii*-derived inbred backcross lines of tomato and their F1 progeny. Plant Breeding 118, 531-536 (1999).
75. van Der Hoeven, R. S., Monforte, A. J., Breeden, D., Tanksley, S. D. & Steffens, J. C. Genetic control and evolution of sesquiterpene biosynthesis in *Lycopersicon esculentum* and *L. hirsutum*. Plant Cell 12, 2283-94 (2000).
76. Fridman, E. et al. Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. Plant Cell 17, 1252-67 (2005).
77. Osborn, T. C. et al. Understanding mechanisms of novel gene expression in polyploids. Trends Genet 19, 141-7 (2003).
78. Zhang, X., Richards, E. J. & Borevitz, J. O. Genetic and epigenetic dissection of cis regulatory variation. Curr Opin Plant Biol 10, 142-8 (2007).
79. Manning, K. et al. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat Genet 38, 948-52 (2006).
80. Monforte, A. J. & Tanksley, S. D. 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 43, 803-13 (2000).
81. Brouwer, D. J. & St Clair, D. A. Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. Theor Appl Genet 108, 628-38 (2004).
82. Finkers, R. et al. The construction of a *Solanum habrochaites* LYC4 introgression line population and the identification of QTLs for resistance to *Botrytis cinerea*. Theor Appl Genet, in press (2007).
83. Fulton, T. M., Van der Hoeven, R., Eannetta, N. T. & Tanksley, S. D. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14, 1457-67 (2002).
84. Doganlar, S., Frary, A., Ku, H. M. & Tanksley, S. D. Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). Genome 45, 1189-202 (2002).
85. Canady, M. A., Meglic, V. & Chetelat, R. T. A library of *Solanum lycopersicoides* introgression lines in cultivated tomato. Genome 48, 685-97 (2005).

86. Fulton, T. M. et al. Advanced backcross QTL analysis of a *Lycopersicon esculentum* x *Lycopersicon parviflorum* cross. Theor Appl Genet 100, 1025-1042 (2000).
87. Peleman, J. D. & van der Voort, J. R. Breeding by design. Trends Plant Sci 8, 330-4 (2003).
88. Tanksley, S. unpublished.
89. Frary, A., Fulton, T. M., Zamir, D. & Tanksley, S. D. QTL analysis of an advanced backcross of *Lycopersicum peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. Theor Appl Genet 95, 881-894 (1997).
90. Moyle, L. C. & Graham, E. B. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Genome-wide associations between hybrid sterility QTL and marker transmission ratio distortion. Mol Biol Evol 23, 973-80 (2006).
91. Canady, M. A., Ji, Y. & Chetelat, R. T. Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics 174, 1775-88 (2006).
92. Menda, N., Semel, Y., Peled, D., Eshed, Y. & Zamir, D. In silico screening of a saturated mutation library of tomato. Plant J 38, 861-72 (2004).
93. Till, B. J. et al. High-throughput TILLING for Arabidopsis. Methods Mol Biol 323, 127-35 (2006).
94. Meissner, R. et al. Technical advance: a high throughput system for transposon tagging and promoter trapping in tomato. Plant J 22, 265-74 (2000).
95. Yamamoto, N. et al. Expressed sequence tags from the laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars. Gene 356, 127-34 (2005).
96. http://www.kazusa.or.jp/jsol/microtom/index.html.
97. Swertz, M. A. & Jansen, R. C. Beyond standardization: dynamic software infrastructures for systems biology. Nat Rev Genet 8, 235-43 (2007).
98. Gur, A., Semel, Y., Cahaner, A. & Zamit, D. Real Time QTL of complex phenotypes in tomato interspecific introgression lines. Trends Plant Sci 9, 107-9 (2004).
99. http://zamir.sgn.cornell.edu/Qtl/Html/home.htm.
100. http://ted.bti.cornell.edu/.
101. Labate, J. A. & Baldo, A. M. Tomato SNP discovery by EST mining and resequencing. Mol. Breed. 16, 343-349 (2005).
102. Urbanczyk-Wochniak, E. et al. Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. Plant Mol Biol 60, 773-92 (2006).
103. Wu, R. & Lin, M. Functional mapping - how to map and study the genetic architecture of dynamic complex traits. Nat Rev Genet 7, 229-37 (2006).
104. http://www.r-project.org/.
105. http://www.animalgenome.org/QTLdb/.

106. Li, L. et al. DNA variation at the invertase locus *invGE/GF* is associated with tuber quality traits in populations of potato breeding clones. Genetics 170, 813-21 (2005).
107. Fridman, E., Carrari, F., Liu, Y. S., Fernie, A. R. & Zamir, D. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305, 1786-9 (2004).
108. van Os, H. et al. Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173, 1075-87 (2006).
109. https://gabi.rzpd.de/projects/Pomamo/. http://www.potatogenome.net/
110. (EPSO), E. P. S. O. European plant science: a field of opportunities. J Exp Bot 56, 1699-1709 (2005).