Techniques & Applications

Cell PRESS

Resolution by recombination: breaking up *Solanum pennellii* introgressions

Saleh Alseekh[1,2,3], Itai Ofner[2], Tzili Pleban[2], Pasquale Tripodi[4,5], Francesco Di Dato[4,5], Maria Cammareri[4], Ayed Mohammad[3], Silvana Grandillo[4], Alisdair R. Fernie[1], and Dani Zamir[2]

[1] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany
[2] The Institute of Plant Sciences and Genetics in Agriculture, Faculty of Agriculture, The Hebrew University of Jerusalem, P.O. Box 12, 76100 Rehovot, Israel
[3] Hebron University, P.O. Box 40, Hebron, Palestine
[4] Italian National Research Council – Institute of Plant Genetics (CNR – IGV), Research Division Portici, Via Università 133, 80055 Portici, Italy
[5] Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Research Centre for Vegetable Crops (CRA-ORT), Pontecagnano (SA), Italy

Quantitative trait locus (QTL) genetics retains an important role in the study of biological and agronomic processes; however, its genetic resolution is often comparatively low. Community-based strategies are thus required to address this issue. Here we detail such a strategy wherein the widely used *Solanum pennellii* introgression lines (ILs) in the genetic background of the cultivated tomato (*Solanum lycopersicum*) are broken up into molecular marker-defined sublines as a community resource for map-based cloning.

Genome-wide association studies (GWAS)

The emergence of GWAS and the adoption of genotyping by sequencing [1] has allowed the construction of ultrahigh-resolution haplotype maps of our major crops [2,3] and facilitated the identification of causative variation in important agronomic characteristics to the resolution of a few genes [3]. Despite the resolving power of GWAS, pitfalls do exist [4] and these can be overcome by going back to the traditional experimental populations derived from biparental crosses. A recurring theme in QTL analysis is the lack of sufficient mapping resolution. Although examples exist in which high-resolution mapping has been utilized to identify single-nucleotide changes [5] as causal for phenotypic variation, most studies merely provide relatively coarse genome-wide surveys of the genetic architecture of the traits under study. One of the reasons behind the lack of resolution is the extreme time investment required to generate new subpopulations of specific regions to zoom in on the genetic factors underlying traits of interest. One solution to address this is to create genome-wide subpopulations harboring smaller, introgressed segments from the 'donor parent' (Box 1). Here we illustrate this approach by describing two publically available genetic resources derived from the extensively phenotyped *S. pennellii* (LA0716) ILs in the genetic background of *S. lycopersicum* cv M82: (i) a set of sublines with marker-defined introgressions smaller than the original set of ILs; and (ii) large quantities of F2 seeds derived from selfing of the hybrids of the ILs with the recurrent parent M82.

Corresponding authors: Fernie, A.R. (fernie@mpimp-golm.mpg.de); Zamir, D. (dani.zamir@mail.huji.ac.il).
Keywords: genome sequence; map-based cloning; interspecific cross; recombinant; sublines; tomato.

S. pennellii ILs

The *S. pennellii* ILs have been publically available since 1995 [6]. *S. pennellii*, a distant relative of the cultivated tomato *S. lycopersicum*, has evolved highly specific morphology, mating system, and phytochemical diversity and is particularly important given its desert habit responses to abiotic stress. Despite these specificities, it is sexually compatible with *S. lycopersicum*, producing fertile hybrids. For these reasons it was chosen as the founding donor parent of the first IL population used for interspecific QTL identification, cloning, and plant breeding [7]. The ILs represent whole-genome coverage of *S. pennellii* in overlapping segments in the background of M82, presently comprising a core set of 76 genotypes (Figure S1 in the supplementary material online).

A unique feature of the tomato IL phenotypes is that the raw data of the replicated measurements have been deposited in the phenotype warehouse of Phenom Networks (http://phnserver.phenome-networks.com/). The data can be browsed and statistically analyzed online or downloaded from the site for further analysis using other statistical software. Phenom Networks harbors data from 45 IL experiments in which 355 traits were scored in replicated measurements. These studies were conducted by multiple laboratories and allowed the identification of diverse phenotypes underlying QTLs, such as whole-plant morphology and yield (including heterosis), metabolic composition, fruit color, enzyme activities, leaf, fruit, and root morphology, cellular development, and biotic and abiotic stress tolerance (see, for example, [7–10]). Over 3069 QTLs have been identified in this population to date (Table S1 in the supplementary material online).

Box 1. Generation of *S. pennellii* sublines

To improve resolution, we took the strategy of backcrossing the initial 76 ILs with the recipient parent (*S. lycoperscium*) (Figure I). These crosses, together with marker-assisted selection, were based mainly on the use of conserved orthologous sequence (COSII) markers [15], but additionally relying on simple sequence repeats (SSRs) and other anchor markers. Via this strategy, 285 smaller ILs have been obtained to date, which break up the 37 largest ILs of the initial population – corresponding to approximately 75% of the genome. Progress on the remaining 25% of the genome is ongoing and seeds from the sublines resulting from the current work will eventually also be available as part of the same resource. Two hundred and eighty-five smaller introgression lines are available. Only 108 can be unambiguously be defined as unique on the basis of the markers used (Figure S1 in the supplementary material online); however, in instances where potentially identical ILs span a genome region of interest for a particular trait these could be rapidly re-genotyped using a higher density of markers to aid the mapping process. Although this population, on its own, will not allow the same level of resolution that was ultimately achieved for the cell-wall invertase example, it will greatly aid in the identification of the genes responsible for the variation in the many QTLs described to date. Furthermore, these new lines, like the original 76, will be maintained indefinitely as a permanent resource for tomato quantitative genetics. It is our belief that this novel population will greatly accelerate the discovery of the genes that influence important traits and thereby enhance our fundamental and agronomic understanding of crop species.

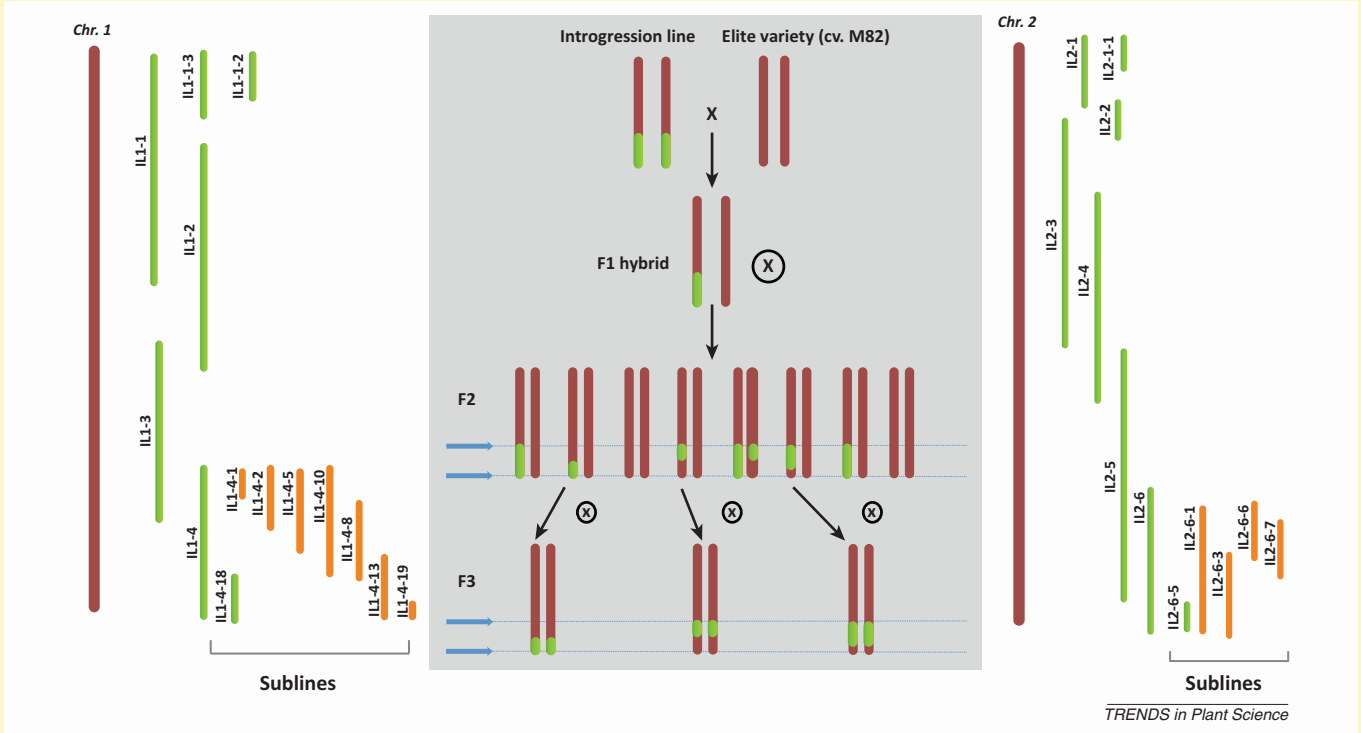

Figure I. A single chromosome is shown where the introgression lines (ILs) exhibit introgression from the green fruited species *Solanum pennellii* with the recurrent parent *Solanum lycoperscium* lines. IL chromosomes are shown in red (*S. lycoperscium*) and green (*S. pennellii*). Subline chromosome segments are shown in orange. The gray panel shows the breeding strategies; initially the introgression line is crossed to the recipient parent (*S. lycoperscium*), the F1 hybrid is selfed, and F2 plants are genotyped to identify recombinants. The lines with the best resolution and coverage of the original line are selected and selfed. From the F3 plants, those homozygous for the introgression are selected as a new set of sublines.

In addition, following on from the release of the genome sequence for tomato (*S. lycopersicum* cv Heinz) [11], genome sequences for the M82 cultivar and *S. pennellii* will shortly be released [12]. This will in principle allow traits, including yield-associated and metabolic traits, to be mapped to known sequence variation. It is worth noting that the observed sequence difference from the currently released *S. pennellii* genome sequence and that of *S. lycopersicum* is a considerable 1.6% [13], rendering sequence variance likely in every gene and thus making the association of phenotypes to functional single-nucleotide polymorphisms (SNPs) highly difficult. However, this is likely to be commonplace in all material derived from interspecific crosses and thus considerable research effort will be required to create genetic material of higher resolving power. This can perhaps be best illustrated in the *S. pennellii* population by mapping the *Brix*9-2-5 QTL that enhances fruit sugar content to a defined portion of the gene encoding a cell-wall invertase. Ultimately a single-nucleotide change was identified that led to a single amino-acid change in the corresponding protein in an area very close to the substrate-binding site of the enzyme [5]. Although a highly rewarding exercise, this was arduous and labor intensive; furthermore, each research group using this population has thus far been generating their own sublines, which will in the long term lead to redundancy. With this in mind, we here describe a research community resource of seeds that are available on request from the corresponding laboratories of 285 sublines currently covering 75% of the *S. pennellii* genome as well as many F2 seeds, which will save 1 year (two generations) for anyone who wants to fine map their favorite genes based on the available genome sequences. Seeds (~1000 for the F2 and 10–20 per subline) are

available by email request to either of the corresponding authors.

Concluding remarks

The take-home message of this techniques and applications article extends beyond the tomato population. In plant genetic research, scientists and breeders often collect data on simple and complex phenotypes in multiple locations. This data is being analyzed for the purpose of identifying genes, genomic regions, or genetic mechanisms that regulate the particular traits studied. Often the phenotypic data is superimposed on the genome of the studied organism or the sequence, haplotypes, or gene-expression patterns of the surveyed accessions. On the publication of the work, sequence and expression data need to be deposited in appropriate databases; however, the raw data of replicated phenotypic measurements are usually not deposited in any public repository and thus are lost. A major bioinformatic challenge facing the research community is to develop web-based resources to display the details of complex phenotypes, which are usually more expensive to collect than the genomic information [14]. The *S. pennellii* ILs provide such an example where researchers can share data and compare their phenotypes and analyses with those that have already been deposited in the Phenom Networks database, to identify wider pleiotropic links. Uniting data from multiple syntenic crops on a common framework will enable the identification of common and unique bottlenecks for crop productivity and the formulation of rational strategies for genomic assisted breeding. The future sharing of phenomic data is the key for continuity of collaborative projects, in both academic research groups and the seed industry, aimed at achieving superior crop varieties.

Acknowledgments

Work in the Zamir and Fernie laboratories is funded in the framework of Deutsche Israeli Project FE 552/12-1, which is administered by the Deutsche Forschungsgemeinschaft. The Grandillo laboratory is funded by the European Union SOL Integrated Project FOOD-CT-2006-016214

Disclaimer statement

D.Z. is a cofounder of the company Phenom Networks, which develops phenotype bioinformatics tools.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tplants.2013.08.003.

References

1 James, G.V. *et al.* (2013) User guide for mapping-by-sequencing in *Arabidopsis*. *Genome Biol.* 14, R61
2 Huang, X. *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967
3 Gore, M.A. (2009) A First-generation haplotype map of maize. *Science* 326, 1115–1117
4 Larsson, S.J. (2013) Lessons from *Dwarf8* on the strengths and weaknesses of structured association mapping. *PLoS Genet.* 9, e1003246
5 Fridman, E. *et al.* (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 305, 1786–1789
6 Eshed, Y. and Zamir, D. (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141, 1147–1162
7 Lippman, Z. *et al.* (2007) An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Curr. Opin. Genet. Dev.* 17, 545–552
8 Semel, Y. *et al.* (2006) Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12981–12986
9 Schauer, N. *et al.* (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* 24, 447–454
10 Ron, M. *et al.* (2013) Identification of novel loci regulating inter-specific variation in root morphology and cellular development in tomato. *Plant Physiol.* 162, 755–768
11 The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641
12 Shivaprasad, P.V. *et al.* (2013) Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J.* 31, 257–266
13 Kamenetzky, L. *et al.* (2010) Genomic analysis of wild tomato (*Solanum pennellii*) introgressions determining metabolic- and yield-associated traits. *Plant Physiol.* 152, 1772–1786
14 Zamir, D. (2013) Where have all the crop phenotypes gone? *PLoS Biol.* 11, e1001595
15 Wu, F. *et al.* (2006) 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, 1407–1420

http://dx.doi.org/10.1016/j.tplants.2013.08.003 Trends in Plant Science xx (2013) 1–3