

## Symposium: Functional genomics and bioinformatics an overview

Orlando Borrás - Hidalgo

Plant Division. Center for Genetic Engineering and Biotechnology.  
P.O. Box 6162, Havana, 10600, Cuba. E-mail: orlando.borras@cigb.edu.cu

### Introduction

A primary challenge in plant biology is the identification of genes that regulate the unique features that distinguish major crop plants from botanical models. Although research on the model organism *Arabidopsis* has advanced rapidly, our understanding of most commercially important plants has not developed properly. However, recent landmark achievements in plant genomics have provided opportunities to efficiently apply the knowledge of model organisms to expedite future research in all plants [1].

The best example of such a landmark is the publication of the genomic sequence of *Arabidopsis*, which has made it possible to identify the entire array of genes in the plant [2]. In association with these sequencing efforts, technological advances have allowed researchers to survey the genome-wide distribution and expression of genes. This progress in genomics research has fostered new approaches to address traditional problems in plant biology, one of which is gene identification.

On completely sequencing the *Arabidopsis* genome, the scientific community has reached a status, at least for this model plant, in which the application of the high-throughput technologies will dramatically increase the knowledge on complex biological networks. The sequence information in itself cannot provide a significant knowledge of the biology of the organisms. But it is a sound basis and framework for further research investigations. In general, biology depends on the selective readout of individual genes from the genome. These are converted into primary transcripts, processed into mRNA, translated into a protein sequence and ultimately, post-translational protein modifications are carried out, such as glycosylation and phosphorylation (Figure 1).

Knowing when and where a gene product (RNA and / or protein) is expressed can provide important clues to its biological function. The ease with which a high throughput approach might be used to study gene expression depends largely on the level of regulation that is being addressed.

On the other hand, bioinformatic analysis allows the theoretical discrimination of open reading frames and genetic control elements; however, such studies are inaccurate by nature [3]. Even where our knowledge on genomic structure allows us to infer the presence of genes, we can say little about the functioning or control of these transcriptional units. While 69% of the publicly annotated *Arabidopsis* genes have some sequence similarity to those with known functions in other organisms, only 9% of genes have been experimentally characterized [2].

Information on both the physical and functional annotation of the genome can be gained through transcript profiling [4]. In recent years, transcript profiling has become synonymous with gene expression analysis, largely because of the technical difficulties

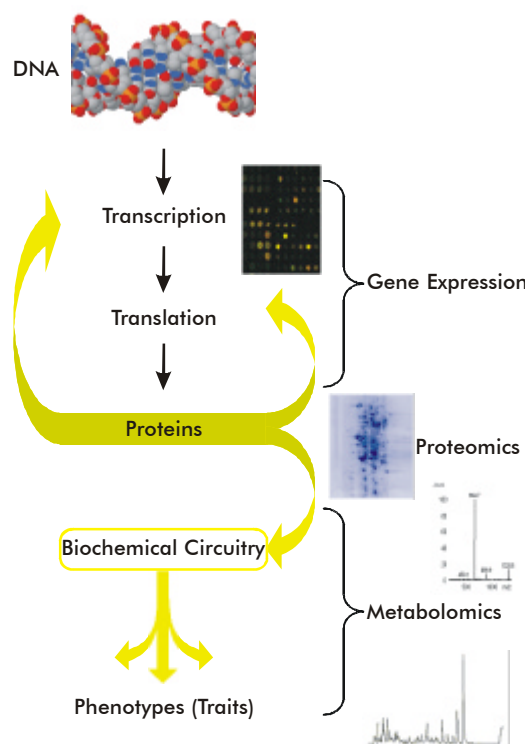


Figure 1. General description of biological pathway in plants.

and greater molecular complexity of proteomics and metabolomics [5]. These correlations are acceptable in many cases [6], even though the term 'gene expression' is often used to refer more directly to the compendium of gene products that ultimately cause cellular responses, and that are more frequently proteins.

Analogous to genomics, the systematic analysis and documentation of all protein species of an organism or tissue is termed proteomics. This concept implies the use of high-throughput analytical techniques. Proteomics addresses analytical questions about the amount and distribution of proteins in the organism, the expression profiles of different tissues, and the identification and localization of individual proteins of interest. These issues are closely connected to more functional ones, which are aimed to elucidate interactions between different proteins, or among proteins and other molecules, and may reveal the functional role of proteins.

Finally, our aim here is to offer an overview on the recent status in functional plant genomics and bioinformatics through important scientific papers presented at the Biotechnology Havana 2002 meeting. Functional plant genomics and bioinformatics applied to biotic and abiotic stress, secondary metabolism, signal transduction, proteomics and gene regulation were envisaged.

1. Mark T O, Paterson A H. Applied plant genomics: the secret is integration. *Current Opinion in Plant Biology* 2002; 5: 141-145.

2. The *Arabidopsis* Genome Initiative: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000; 408: 796 - 815.

3. Cho Y, Walbot V. Computational methods for gene annotation: the *Arabidopsis* genome. *Current Opinion Biotechnology* 2001; 12: 126 - 130.

4. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett - Engle P, McDonagh PD *et al*. Experimental annotation of the human genome using microarray technology. *Nature* 2001; 409: 922 - 927.

5. Smith RD. Probing proteomes: seeing the whole picture? *Nature Biotechnology* 2000; 18: 1041 - 1042.

6. Celis JE, Kruhoffer M, Gromova I, Frederiksen C, Ostergaard M, Thykjaer T *et al*. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Letter* 2000; 480: 2 - 16.

## Mechanisms of disease resistance and possibilities for engineering disease resistance

Pierre JGM de Wit

Wageningen University, Graduate School Experimental Plant Sciences,  
Laboratory of Phytopathology, Binnenhaven 5, 6709 PD Wageningen, The Netherlands.

The fungal pathogen *Cladosporium fulvum* secretes many stable cysteine-rich peptide elicitors to facilitate infection of tomato. Tomato resistant to *C. fulvum* has an efficient surveillance system to recognise these peptides and subsequently mount a hypersensitive response (HR). The HR is the most efficient defence mechanism of plants against biotrophic and obligate pathogens as it renders the plant immune for these pathogens. Our research group has cloned eight genes of *C. fulvum* encoding such peptide elicitors and has studied their structure and function. Also the role of the Cf (for *C. fulvum*) resistance genes in tomato encoding Cf proteins required for perception of these peptide elicitors and subsequent triggering of signal transduction pathways leading to plant resistance, are studied. Presently we are also unraveling in detail the signal transduction pathways leading to HR in tomato and other plant species by cDNA-AFLP analysis and subsequent gene silencing of cDNAs. We have discovered that induction of HR in tomato by some peptide elicitors is temperature-sensitive. Tomato seedlings expressing both the Avr4 gene of *C. fulvum* and the

matching Cf-4 resistance gene, quickly develop systemic HR at 23 °C, but grow normally at 33 °C. Thus, when the seedlings are grown at 33 °C and are subsequently incubated at 23 °C they synchronously start a cell death program within minutes after the temperature shift. We have identified mRNAs, either up- or down-regulated during the cell death program, by cDNA-AFLP analysis using KeyGene technology. This approach resulted in the identification of 420 cDNAs of which 200 have no homology to known genes and 220 have homology to genes or ESTs from tomato and Arabidopsis. The cDNA collection contains many genes with a presumed role in signal transduction leading to HR. We now try to uncover the function of the unique set of 420 tomato genes. For functional analysis of these 420 genes we use high throughput, virus-induced gene silencing (VIGS) technology employing tobacco rattle virus (TRV). VIGS will easily identify genes affecting HR and/or resistance. Genes crucial in HR induction can be exploited in molecular resistance breeding of plants to achieve resistance against biotrophic and obligate fungal pathogens.

## TraitMill™: Closing the application gap in functional genomics

WV Camp, C Reuzeau, V Frankard, Y Hatzfeld, A Sanz, AM Droual,  
P Lejeune, W Dillen, V Mironov, R Peerbolte, W Broekaert

CropDesign NV, Technologiepark 3, 9052 Gent, Belgium. (Website: <http://www.cropdesign.com>).

CropDesign's technology platform closes the application gap between functional genomics and the development of improved or novel agronomic traits. Today, functional genomics has largely concentrated on elucidating the biochemical and cellular function of genes using different approaches, such as transcriptome analysis, proteome analysis, and loss-of-function mutant collections. Though these approaches have been very successful and have led to a much better understanding of plant gene functions, predicting the potential agronomic value of genes remains difficult, particularly with regard to complex traits such as growth rate, organ size, fertility, seed yield and harvest index. Trait development therefore still requires the testing of thousands of genes and gene combinations and the direct analysis of the effects of these genes and gene combinations on the trait of interest. CropDesign's TraitMill™ platform has been specifically designed for this purpose:

- i) Using high-throughput cloning systems, expression levels of genes and gene combinations can be modulated in entire plants or in selected tissues.
- ii) An industrialized plant transformation system generates the tens of thousands of transgenic plants containing the gene and gene combination constructs.
- iii) Trait evaluation has been automated through 'walking plant systems' for plant transport, digital

imaging tools for plant evaluation and proprietary image analysis software for data production and statistical breakdown of results.

The TraitMill™ currently employs rice for high throughput gene testing, rice being both a key crop and a cereal model species. Arabidopsis is also studied for traits that are thought to have common mechanisms in monocots and dicots, such as tolerance against abiotic stress. CropDesign's research interests are focused around genes that modulate the size, shape and growth of specific plant tissues, as well as the plant as a whole. Also the signalling pathways through which environmental cues impinge on plant growth and development are the target of functional genomics research at CropDesign. CropDesign's position in plant growth and development has been built in collaboration with different leading research institutes. Several of these collaborative programmes aim at the discovery and functional characterisation of genes involved in cell cycle control. Cell cycle is the control mechanism that regulates cell division, the fundamental means by which organisms grow and propagate. Cell cycle genes determine when, where and at which rate, cells undergo division and multiplication. CropDesign has observed different phenotypes with modified expression of cell cycle

genes, including enhanced growth, overproliferation of leaf organs, dwarfism and anthocyanin accumula-

tion. A phenotypic description of these transgenic lines will be presented.

## **Generation and application of Arabidopsis protein chips**

B Kersten, T Feilner, S Wehrmeyer, A Possling, H Lehrach, DJ Cahill

Department Lehrach, Max Planck Institute of Molecular Genetics, Ihnestrasse 73, D-14195 Berlin, Germany

The recent completion of the Arabidopsis genome sequencing now provides a basis for the further systematic analysis of gene expression and function. Therefore complex proteomic approaches using high-throughput technologies are useful. To complement 'classical' proteomics, protein array technology is emerging as a tool to profile and functionally characterize proteins. In this study, we applied different strategies for cloning of Arabidopsis genes into E.coli expression vectors for IPTG-inducible expression of His6-tagged proteins: (i) high-throughput sub-cloning of open reading frames using gene-specific primers, (ii) generation of cDNA-expression libraries. The gene expression was performed in small volume (2 ml) in 96 well microtitre plates.

Recombinant proteins were purified in high-throughput and robotically arrayed onto coated glass slides. Using anti-RGS-His antibody followed by a fluorescence-labeled secondary antibody, the proteins could be specifically detected on the chips with a low background. The protein chips were used so far for the characterization of the specificity and cross-reactivity of commercial available monoclonal antibodies and new polyclonal sera. Using a subset of 95 different proteins including several Dof- and Myb- related transcription factors pilot studies are under way to test further applications of Arabidopsis protein chips, such as studies on protein phosphorylation, on protein-protein or protein-DNA interactions.

## **Gene survey of Eimeria spp. of domestic fowl using open reading frame ESTs (ORESTES) and development of an automatic pipeline for sequence analysis**

AM Durham,<sup>1</sup> FGT Matsunaga,<sup>2</sup> H Sakata,<sup>2</sup> PH Ahagon,<sup>2</sup>  
S Fernandez,<sup>2</sup> AMBN Madeira,<sup>2</sup> A Gruber<sup>2</sup>

<sup>1</sup>Instituto de Matemática e Estatística <sup>2</sup>Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo. Av. Prof. Orlando Marques de Paiva 87, São Paulo SP, 05508-000, Brazil. Phone: 55 11 3091-7705; Fax: 55 11 3091-7829; E-mail: argrub@usp.br

Coccidiosis of the domestic fowl is an important enteric disease caused by seven different protozoan species of the genus Eimeria. The most studied species, E. tenella, presents a genome of circa 60 million base pairs distributed in 14 chromosomes. Genomics studies are being carried out, including an EST project (Washington University - USA) and a 5-fold genome coverage (Sanger Institute, UK). Our laboratory initiated an alternative EST sequencing project based on ORESTES (open reading frame ESTs) methodology, which involves the construction of multiple minilibraries by RT-PCR using arbitrary primers under a very low stringency. This method generates reads biased towards the central region of the genes, with a better normalization than that normally obtained by using conventional cDNA libraries. The data gener-

ated in this study will be complementary to the 5' and 3' end EST sequences already available, and will also contribute to map exon-intron boundaries using genomic data. In order to process the sequencing reads, a pipeline for automatic analysis was constructed using Perl language. This pipeline can be used in any platform and is totally modular. Each module acts as a shell for different programs like Cross\_Match, Phred, Phrap, Cap3, Blast, etc., thus allowing a multi-step configuration using a single setup file. Specific modules for overall read quality evaluation, end trimming and chimera finding were also developed. We expect that this pipeline will be very useful to anyone involved on DNA sequencing projects. Further implementations for partially automatic annotation and database integration are underway.

## **Molecular markers in sugarcane breeding: new tools for an old art**

MT Cornides Hernandez

Centro Nacional de Investigaciones Científicas (CNIC), Habana, Cuba.  
Contact: Cornide Hernández MT (cornide@biocnic.cneuro.edu.cu)

The Poaceae family covers an ample spectrum of adaptation to present ecological conditions. Among the

tropical grasses, sugarcane is an important crop with satisfactory prospects to diversify its products. Mod-

ern plant breeding is aimed at obtaining new cultivars with specific and general adaptation mechanisms, to overcome new diseases and to increase the productivity in areas affected by abiotic stresses. An important limiting factor is the reduced genetic diversity available. Wide hybridization is still the main way to introduce new variability, but it is a long lasting process and the resulting hybrids frequently present chromosome irregularities in number and transmission. Thus, the real amount of genetic variation present in these progenies is masked by those phenomena and the phenotypic selection is hazardly conducted. Increased efficiency of introgression and cultivar selection in sugarcane can be expected through the use of molecular markers as auxiliary tools in the genetic diversity assessment and selection (MAS). Results obtained by

applying molecular tools to identify and widen the genetic basis of the sugarcane breeding program in Cuba are presented. Success and constraints of gene mapping and QTL evaluation strategies currently used are also discussed. On the other hand, nuclear repetitive DNA sequences have been successfully employed in genetic and taxonomic studies in the Poaceae family. The comparative study of these patterns could result in a better understanding of the origins (genetic vs. epigenetic) of some relatively frequent somaclonal variants that have already proven their agronomic performance stability. The study of sugarcane and rice salinity and drought tolerant somaclone collections is discussed. The interrelation with other labs aimed to develop differential expression products for mapping purposes (ESTs) is also presented.

## **Predicting the structures of membrane proteins**

J Schonbrun

University of Washington, Seattle, USA. Contact: Schonbrun J (jschon@u.washington.edu)

We computationally construct models of integral membrane proteins by piecing together fragments of previously solved protein structures. The fragments are chosen to have locally similar sequence profiles to the protein of interest. A large number

of structures are made using a Monte Carlos based assembly process, and good candidates are selected using a scoring system that is based on the statistical properties of known membrane protein structures.

## **Identification of novel genes in response to ethylene and submergence stress by transcriptional profiling**

A De Paepe, F Blanco, AM de Almeida, L Holigue,<sup>1</sup> D Van Der Straeten

University of Gent, Gent, Belgium; <sup>1</sup> Universidad Catolica de Chile, Santiago, Chile;  
Contact: Van Der Straeten D (dostr@gengenp.rug.ac.be)

Ethylene has multiple effects in development and in response to numerous forms of stress. Powerful genetic screens led to the isolation of Arabidopsis mutants in ethylene signaling. The pathway has been resolved in much detail, although some presumed intermediates remain to be identified. A number of ethylene response genes have been characterized in tomato, tobacco, and Arabidopsis. In addition, the ethylene signal has been shown to interact with several endogenous as well as exogenous factors controlling downstream gene expression. Genome-wide approaches are particularly well suited to study the overlap in responses to different signals in order to reveal the intricate network at the basis of the phe-

notype under given conditions. We have used Arabidopsis as a model to study the ethylene response at the transcriptome level, using cDNA - AFLP. The purpose is to identify novel early ethylene response genes, and to highlight possible differences in the effect of low and high concentrations of ethylene. To establish the role of novel genes in ethylene response we will conduct a functional analysis using Arabidopsis insertion mutants. Application of the CATMA array will extend the analysis to the entire genome in the near future. In parallel to the study in Arabidopsis, we have initiated a transcriptome analysis in rice, upon submergence stress. Preliminary data will be discussed.

## **Strategies for the identification and isolation of novel genes involved in plant resistance to biotic and abiotic factors**

O Borrás - Hidalgo

Center for Genetic Engineering and Biotechnology, Havana, Cuba.  
Contact: Borrás - Hidalgo O (orlando.borras@cigb.edu.cu)

Modern-day plants are products of eons of evolution from primal living organisms in response to abiotic and biotic environmental changes. Plants, in

nature, are generally resistant to most pathogens. The ability of a pathogen to produce a disease in a host plant is usually the exception, not the rule. The

interactions between plants and pathogens are specific, complex and dynamic. On the other hand, in the natural environment, plants often grow under unfavorable conditions, such as drought, salinity, chilling, freezing, high temperature, flooding, or strong light. These conditions are known collectively as abiotic stresses, and any of them can delay growth and development, reduce productivity and, in extreme cases, cause the plant to die. Our group carried out several strategies for the discovery, characterization and functional analysis of plant genes involved in triggering, signaling and responding to biotic and abiotic factors. We use suppression subtractive

hybridization (SSH) and cDNA - AFLP to generate cDNA libraries highly enriched for sequences up-regulated specifically by resistance gene activation. Moreover, the clones are sequenced and compared to sequences in international databases using BLAST searches. The expression of genes is studied using Northern analyses as well as temporally measuring it under several conditions to clearly identify sequences up-regulated specifically in the resistance response. On the other hand, candidate genes are cloned into a virus vector for virus induced gene silencing (VIGS). Furthermore, these are assayed for a compromised resistance response.

## **Emerging technologies: challenges and opportunities in studies of host-pathogen interactions in forest tree species**

AE Ekramoddoullah

Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, Victoria, Canada.  
Contact: Ekramoddoullah AE (aekramoddoul@pfc.forestry.ca)

Root pathogens and rust diseases can cause extensive damage to Canadian forest tree species. Loss of growth, although difficult to visualize, is substantial over the life of a tree. Understanding host - pathogen interactions is important in managing yield loss and can aid in the identification of disease-resistant trees. However, studying the interactions involving forest pathogens offers both challenges and opportunities. Some of these issues are described from our perspective through working on the white pine - blister rust pathosystem and

pathogens that infect the roots. Several resistance mechanisms to the white pine blister rust fungus, *Cronartium ribicola*, have been identified in pine. At the molecular level, several defence responsive proteins and their genes have been characterized. Some of these are identified to be potential candidates for markers associated with resistance or susceptibility. Current research activities and future directions and application of technologies to isolate and characterize resistance genes in white pine are discussed.

## **Genetic dissection of transcriptional regulation in budding yeast**

R Brem

Fred Hutchinson Cancer Research Center, Seattle, USA. Contact: Brem R (rbrem@fhcrc.org)

Our work applies classical linkage analysis techniques to a modern quantitative trait: mRNA expression levels as measured by microarray. We used this paradigm to study a cross between a laboratory strain and a wild strain of *Saccharomyces cerevisiae*. Over 1500 genes were differentially expressed between these parent strains. Among haploid F1's, expression levels of 570

genes were linked to one or more different loci, with most expression levels showing complex inheritance patterns. The *loci* detected by linkage fell largely into two categories: cis-acting modulators of single genes and trans-acting modulators of many genes. We found eight such trans-acting loci, each affecting the expression of a group of seven to 94 genes of related function.

## **Molecular characterization of the p5r gene family in *Digitalis purpurea***

I Gavidia, HU Seitz,<sup>1</sup> P Pérez-Bermúdez<sup>2</sup>

Instituto de Investigaciones Agrobiológicas de Galicia, CSIC, 1: Santiago de Compostela, Spain; Zentrum für Molekularbiologie der Pflanzen, Tübingen, Germany; 2: Departamento de Biología Vegetal, Universidad de Valencia, Burjassot, Spain. Contact: Gavidia I (igavidia@iiag.cesga.es)

*Digitalis* cardiac glycosides are secondary metabolites of major interest in the pharmaceutical industry. At present, the most important use of these products is in the treatment of congestive heart failure because they increase the force of systolic contractions. Cardenolides are 5 - configured steroid derivatives. According to the postulated biosynthetic pathway,

the precursor of this class of secondary metabolites is 5-pregnane - 3,20 - dione, which is formed from progesterone. The enzyme catalyzing this reaction is progesterone 5 - reductase (P5R), the first stereospecific enzyme in this metabolic route. A genomic library from *Digitalis purpurea* was screened with the P5R cDNA as probe. Four positive clones were further

analyzed. The three new genes were named Dp1, Dp2 and Dp3. Sequencing showed that they are homologous, sharing 70 - 80% with the P5R sequence when the coding regions are compared. These genes showed homology to several *Arabidopsis* genes with unknown

function, sharing about 60% identity when their putative amino acid sequences are compared. Northern Blot analysis was done to study mRNA expression levels in various tissues (roots, stem, leaves and flowers) and at different developmental stages.

## **Aquaporin overexpression in legume root - nodule: Why?**

Drevon J

INRA - Agropolis - UMR Sol et Environnement, Montpellier, France.  
Contact: Drevon J (drevonjj@ensam.inra.fr)

Aquaporins are water - channel proteins. In plants they are very diversified, and found in plasmalemma and tonoplast membranes, and the peribacteroid membrane in legume root-nodules. We have examined whether the expression of an aquaporin gene is tissue - specific within *Phaseolus vulgaris* nodules and how it relates with its N<sub>2</sub> - fixing activity. Common bean (BAT477) seeds were surfaced sterilized and inoculated with *Rhizobium tropici* CIAT899, then grown hydroaerobically in glasshouse. Fully active nodules were collected at 35 days and immediately cut in half, fixed and subsequently embedded in methacrylate resin for making 5mm sections. In situ hybridization was performed with 35S radio - labeled RNA probe of a Mip - 1 aquaporin DNA fragment obtained by RT-PCR from a nodule - cortex RNA extract. There was no radioactive signal with the

sense probe. With the antisense probe there was a much higher signal-density in the cortex parenchyma than in the infected zone. In the later, the signal was found only in the non - infected cells. Within the cortex, the signal was higher in the 2 - 4 cell layers between the vascular traces and the infected zone, i.e. the inner - cortex, than in the next 2 - 3 surrounding cell layers, i.e. the middle - cortex. In addition, the signal was surprisingly high in the first 2 cell - layers surrounding the later. An increase in the inner - cortex signal was associated with an increase in nodule permeability under P deficiency. The relation of these results with the improvement of the efficiency in utilization of phosphorus for the symbiotic nitrogen fixation of common and its subsequent effect on the plant growth and productivity will be discussed.

## **Carotenoid biosynthesis in plants: from gene cloning to biotechnology**

J Hirschberg

The Hebrew University of Jerusalem, Jerusalem, Israel. Contact: Hirschberg J (hirschju@vms.huji.ac.il)

Carotenoids are natural pigments that play essential roles in photosynthesis. In higher plants they also furnish attractive colors to flowers and fruits. Carotenoids are indispensable components in human nutrition as precursors of vitamin A and they provide additional health benefits that are attributed to their antioxidant activity in vivo. Industrial applications of carotenoids include food and feed additives, cosmetics and pharmaceutical products. Plant foods are the major source of carotenoids in the human diet. Hence, manipulating carotenoid content and composition in plants may well improve their nutritional value, provide a new source of valuable materials for industry and generate new colored varieties of flowers and fruits. Detailed molecular description of carotenoid biosynthesis in plants has been deficient because of lack of in - vitro assays for most of the enzymes. Using various genetic approaches we have cloned plant genes for

enzymes in the pathway. In recent years we have used a map - based (positional) cloning in order to identify novel genes that affect carotenoid biosynthesis in tomato (*Lycopersicon esculentum*). We have cloned the genes responsible for the fruit color mutations Delta (Del), Beta (B), old-gold (og) and tangerine (t). It was determined that Del is an allele of the gene *Lcy - e*, which encodes lycopene epsilon - cyclase; B encodes a novel type of chromoplast -specific lycopene beta - cyclase; og is a null mutation of B and t encodes carotenoid cis - trans isomerase, a redox - type enzyme, CRTISO, that is structurally related to the bacterial-type phytoene desaturase. Cloning nearly all of the enzymes required for carotenoid biosynthesis in plants enables now to genetically manipulate carotenoid composition in crop plants. Examples of metabolic engineering of carotenoid biosynthesis in fruits and flowers will be presented.