

THE ROLE OF PLANT GENOMICS IN BIOTECHNOLOGY

A.Varma and N. Srivastava*

*B. V. Patel Pharmaceutical Education & Research Development (PERD) centre,
Sarkhej – Gandhinagar Highway, Thaltej, Ahmedabad, Gujarat, India*

Keywords: Genome, synteny, genomics, bioinformatics, sequencing, transgenic, molecular farming

Contents

1. Understanding the plant genome.
 - 1.1 Extra Nuclear Genomes
 - 1.2 Structure, Size and Organization
 - 1.3 Synteny and Collinearity of Plant Genomes
2. The science of ‘Genomics’
 - 2.1 Genome Sequences – the Raw Material
 - 2.2 Structural and Functional Genomics
3. Arabidopsis Genome sequence – The signpost of plant genomics
 - 3.1 Arabidopsis will not feed the world
 - 3.2 The International Rice Genome Sequencing Project
4. Applying genomics to agricultural biotechnology
 - 4.1 Agrobiotechnology Contributions
 - 4.1.1 The Flavr Savr Tomato
 - 4.1.2 The Golden Rice
 - 4.1.3 Bt Crops
 - 4.1.4 Stress Tolerant, Herbicide and Disease Resistant Crop Varieties
 - 4.2 Farming Pharmaceuticals
5. Concluding remarks
- Acknowledgements
- Glossary
- Bibliography
- Biographical Sketch

Summary

There has been a tremendous growth in the plant biotechnology sector owing to the development of technological advances and enormous genomic resources. Plant genomics has evolved as a discipline of its own, comprising sub-disciplines like proteomics, metabolomics, bioinformatics and systems biology [see also – *Plant Genome Mapping and Strategy*; – *Plant Feedstock Genomics*]. The collaborative efforts of researchers worldwide have accomplished the mammoth tasks of sequencing the entire genomes of important plants and deciphering the functional aspects is underway. Right from the understanding of structure and functioning of large genomes of plants to the development of methods to manipulate them for benefit of mankind, plant genomics provides researchers with the essential tools for the purpose. At this juncture we try to give a brief overview of complexity of plant genomes and their manipulation with the aid of genomic resources.

1. Understanding the 'plant genome'

The essence of every characteristic of a species, from morphology to timing of senescence resides in the information stored in a species DNA; its genome. The term 'genome' was coined by Winkler in 1920 and the definition was modified by Rieger and his coworkers in 1991 that stated that 'in eukaryotes the basic (monoploid) chromosome set, consisting of a species specific number of linkage groups and the genes contained therein' is what is its genome. Eukaryotic genomes are much more complex than prokaryotic genomes owing to the complexity of the organization and functioning and further, plant genomes are much more complex than any other eukaryotic genomes.

1.1 Extra Nuclear Genomes

Plant cells incorporate two more genomes in addition to the nuclear genome. While the mitochondrial genome is present both in plant as well as animal cells, chloroplasts are an integral part of plant cellular organization and have a genome of their own. Both of these organelles are thought to have evolved from prokaryotes that were taken up by eukaryotic cells in symbiotic associations very early in their evolution. They have retained several features of their prokaryotic ancestors and one of them is the circular, double stranded DNA molecule as hereditary material. The chloroplast genomes are highly conserved among plant species (between 120 – 160 kbp in size) and encode most of the genes involved in protein synthesis or photosynthesis along with complete set of rRNAs and tRNAs. Plant mitochondrial genomes however are much larger than those of mammals and yeasts (from 200 kb in *Brassica* spp. to over 2500 kb in *Cucumis melo*), and contain genes for protein synthesis and respiratory electron transport. The traits that are conferred to the mitochondrial or chloroplast genes can be easily followed by their uniparental fashion of inheritance as many of these are in commercial use. A well-known example of cytoplasmic inheritance is the 'Cytoplasmic Male Sterility', best studied in male sterile lines of maize. However, in the process of evolution, many important genes from the organelles have been transferred to the nuclear genome which holds the majority of the genetic information.

1.2 Structure, Size and Organization

The structure and organization of plant genomes is similar to animal and yeast genomes at the elementary level. The division of genomic DNA into independent chromosomes is a fundamental feature of genomic architecture. The size and number of chromosomes varies amongst different plant species leading to variation in genome size. The tremendous diversity found in the genome size is one of the most startling features of plant genomes. In angiosperms alone the nuclear DNA content varies approximately 1000 folds (1C = approx. 0.1 to 127.4 pg with 1 pg = 980 Mb), where 1C is the amount of DNA in the unreplicated gametic nucleus of an organism. The C-value is referred to as a key character in biology and is highly affected by polyploidy found in plant species, e.g. *Triticum aestivum* is a hexaploid and its C-value (1C = 17.33 pg) is approximately three times larger than those of related diploid taxa in Triticeae e.g. *Hordeum vulgare* and *Aegilops* spp. Conversely, when comparing closely related species like maize (*Zea mays*) and sorghum (*Sorghum bicolor*), it is observed that though both have same number of chromosomes, the maize genome is >3 times larger than sorghum.

Nevertheless, there is a lack of direct relationship between genome size and organism complexity, what is known as the C-value paradox.

In plant kingdom, *A. thaliana* has the smallest genome size (7×10^7 bp), while *Fritillaria assyriaca* of lily family has a genome of 1×10^{11} bp. Hybridization studies based on renaturation kinetics clearly indicated that the ‘extra DNA’ that the big plant genomes inherit, is made up of non-coding and repetitive DNA sequences. Plant genomes can contain from 10% to 80% of repetitive DNA, which has accumulated over the millennia. It is the differences that exist between the sizes of these repetitive sequences that form the basis of large variations in genome sizes of closely related plant species. These repeats are found in plant genomes in either clustered or short-period interspersed arrangement. The clustered repeat elements mostly comprise ribosomal genes, centromeric and telomeric sequences whereas the interspersed elements are variable number of tandem repeats (VNTRs) or minisatellite DNAs. Also there are dispersed repeats that comprise the mobile elements or transposons. Transposons and retrotransposons are found scattered throughout the genome and show various degrees of sequence divergence, e.g. more than a dozen dispersed repeats occupy the region surrounding the *Adh1* gene in maize. Different levels of transposable element amplification along with polyploidy have been the major factors responsible for variation in plant genome size. Although most of the non-coding part of the genome comprises of repetitive sequences, not all the repetitive DNA is non-coding. The ribosomal genes are an excellent example of clustered genes, which are repeated thousands of times in the Nucleolar Organizer Regions (NORs) and represent one of the largest families of repetitive DNA. Repetitive DNA (once considered as of no use) contributes to the character, function and structure of chromosomes and plays an important role in genome organization.

The coding regions of a typical plant genome are the unique and low-copy sequences that make together the majority of the actual gene number of a plant. Plant genes are structurally similar to animal or yeast genes (having exons and intronic sequences) with a few minute differences like in regulatory signals and splicing mechanisms. Plants appear to have largely evolved independently when it comes to transcriptional control. Mechanisms like alternative splicing significantly contribute to protein diversity in different groups of plants. Events of reverse transcription and frequent insertions/deletions of introns have contributed a lot in the evolution of plant gene structure as well.

1.3 Synteny and Collinearity of Plant Genomes

Gene mapping projects have revealed that large segments of chromosomes are often conserved among related species (e.g. maize and sorghum). This collinearity of loci is called **synteny**. Comparative studies of wheat, maize, and rice and other grass species with common DNA probes has revealed remarkable conservation of gene content and gene order over substantial evolutionary distances of radiation of *Poaceae*. There have been comparative studies within monocot and dicot groups, but very few studies have been performed on synteny between the two monophyletic groups which appear to have diverged from a common ancestor between 120 and 200 mya. From the results, it appears that though it is not possible to infer large-scale gene orders in rice on the basis

of *Arabidopsis* genome, microcollinearity between individual rice genes and their homologues in *Arabidopsis* validate the strategy of annotating individual genes. Therefore it becomes apparent that the genome sequencing projects are based on the premise that knowledge of the whole sequence of *Arabidopsis* will aid in the isolation from the crop plants of agronomically important genes that bear homology to *Arabidopsis* genes.

2. The science of ‘Genomics’

The term genomics is derived from the term “genome”. First used in 1986, ‘genomics’ portrayed the enterprise that aimed to map and sequence the entire human-genome. Genomics is the field of genome studies and includes intensive efforts to determine the sequence of DNA and RNA using high-throughput sequencing strategies; generating fine-scale genetic maps; microchip arrays; collecting genome variations within a population and ascertaining the transcriptional control of genes; and employing digital technology and computationally intensive analysis to understand the structure, function and evolution of diverse organisms. Genomics provides the essential tools to speed the work of the traditional molecular geneticist and is now a scientific discipline in its own right. Application of genomic technologies in all areas of biology has lowered the barriers that once separated the plant, animal, microbial research communities. Genomes of both prokaryotes and eukaryotes have been subjected to genomic analysis for the benefit of science and man. Here we are focusing on the genomics approach to plants and its applications.

2.1 Genome Sequences – the Raw Material

Genome sequencing is right at the heart of genome sciences and provides the platform for all advance studies. Automated sequencing coupled with capillary electrophoresis facilitates analysis of 96-384 samples in 2-3 hrs which allow extensive parallel analysis of nucleotide sequences. Two basic strategies are applied; the shotgun approach that exhibits ‘clone-by-clone’ methodology and gives long accurate sequences or the ‘partial or draft-type’ whole genome sequencing (WGS) approach which is rapid, simple and compatible with high-throughput equipments. The choice essentially depends on the objective of the study. For smaller genomes with low percentage of repetitive DNA, Bacterial Artificial Chromosome (BAC) clone-by-clone sequencing is the predominant approach. For the larger genomes characteristic of most flowering plants (where repeats are often intermixed with genes) gene enrichment (WGS) approaches have been suggested as a more efficient sequencing strategy. Improvements in liquid handling techniques and computational data analysis have thrusted the technology far enough that the whole genome sequencing projects were completed ahead of schedule and under budget.

2.2 Structural and Functional Genomics

The excitement over genomics is not founded on the production of sequence information but on its utility and applications. Genome sequencing teamed with generation of high-throughput data-sets, bioinformatics tools, functional and comparative studies can provide a road map for the next generation of agricultural

research. The discipline genomics has two integral elements, ‘structural genomics’ that deals notably with markers (landmark sequences), genome maps, cloned genome fragments, sequencing and gene discovery; and ‘functional genomics’ that aims to explain gene functions. Functional genomics is increasingly seen as a way towards understanding genes in isolation and within the complex networks in which genes interact. It is aided by the knowledge of bioinformatics that includes all the computational tools and methods needed to manage genomic data and extract meaningful information from it. A sequence similarity analysis using bioinformatics tools permits the assignment of probable gene function and the identification of genes similar between species. However, elucidating their exact function still requires experimental approaches, especially if the organism shows a high degree of complexity, or like many plants, an extremely large genome (Wheat 16000 Mbs vs. Arabidopsis 180 Mbs). A cost effective way to survey the expressed portion of the genome is EST sequencing.

An expressed sequence tag or EST is a short sub-sequence of a transcribed spliced nucleotide sequence which may be used to identify gene transcripts and is instrumental in gene discovery and gene sequence determination. A comparison of EST databases from different plants, tissues and conditions reveals the diversity in coding sequences between plants. At the same time, however, it provides a global perspective of the similarities in genes for specific processes, such as ripening. The rapidly growing EST database has become an invaluable tool for gene discovery, gene mapping and genome annotation. There are numerous public databases which have generated such a gene index based on ESTs, including NCBI’s Unigene, TIGR Gene Indices, PlantGDB etc. The genes annotated through the cDNA libraries, ORFs and EST databases can then be subjected to a number of strategies have been proposed to probe gene function. And as mentioned previously a particular protein playing a key role in one species has counterparts in other species encoded by similar genes. This highlights the importance of having complete genome sequences for model organisms.

-
-
-

**TO ACCESS ALL THE 12 PAGES OF THIS CHAPTER,
Visit: <http://www.eolss.net/Eolss-sampleAllChapter.aspx>**

Bibliography

- Bennetzen J.L. (2007). Patterns in grass genome evolution. *Current Opinion in Plant Biology* 10, 176-181. [The article describes the major mechanisms responsible for plant genome structural evolution]
- Bevan M. and Walsh S. (2005). The *Arabidopsis* genome; a foundation for plant research. *Genome Research* 15, 1632-1642. [An excellent review that summarizes progress in *Arabidopsis* research and its implications for crop science]
- Janardhar B.S. (2007). Promising achievements and new challenges in agricultural biotechnology. *Current Science* 85 (10) 1052-1054. [This report on the ‘Asia Pacific Conference on Plant Tissue Culture

and Agribiotechnology 2007' emphasizes on the current scenario, challenges and opportunities in the field of agribiotechnology]

Lal P., Ramachandran V.G., Goyal R. and Sharma R. (2007). Edible vaccines: current status and future. *Indian Journal of Medical Microbiology* 25 (2), 93-102. [This review attempts to discuss the promises and challenges of transgenic plants as edible vaccines]

Paine J.A., Shipton C.A., Chaggar S., Howells R.M., Kennedy M.J., Vernon G., Wright S.Y., Hinchliffe E., Adams J.L., Siverstone A.L. & Drake R. (2005). Improving the nutritional value of golden rice through increased pro-vitamin A content. *Nature Biotechnology* 23 (4), 482-487. [This work describes the quality improvement obtained in the genetically modified golden rice]

Tatusova T., Smith-White B. and Ostell J. (2007). A collection of plant-specific genomic data and resources at NCBI. *Methods in Molecular Biology* 406, 61-88. [The article gives the overview of extensive genomic resources that NCBI provides for plant researchers]

Thro A.M., Parrot W., Udall J.A., Baevs W.D. (2004). Genomics and plant breeding: the experience of the initiative for future agricultural and food systems. *Crop Science* 44, 1893-1919. [This is an overall summary and main points of a symposium held at the annual meeting of the Crop Science Society of America and throws light on the value of genomics to plant breeding]

Vaughan D.A., Ealazs E. and Heslop-Harrison J.S. (2007). From crop domestication to super-domestication. *Annals of Botany* 100, 893-901. [The article describes how the availability of well-characterized germplasm resources in the global network of genetic resources centres (genebanks) can improve crop domestication]

Walley J.W., Crughlan S., Hudson M.E., Covington M.F., Kaspi R., Banu G., Harmer S.L. and Dehesh K. (2007). Mechanical stress induces biotic and abiotic stress responses via a novel *cis*-element. *PloS Genetics* 3 (10), 1800-1812. [The paper describes identification of a novel stress response element with the help of bioinformatic tools]

Yazaki J., Kojima K., Suzuki K., Kishimoto N. and Kikkuchi S. (2004). The Rice PIPELINE: a unification tool for plant functional genomics. *Nucleic Acids Research* 32, D383-D387. [This is a scientific resource that pools publicly available genomic data from various databases]

Biographical Sketches

Astha Varma has done her graduate and post graduate studies in biological sciences. At present she is pursuing her doctoral program in the area of plant molecular biology under the guidance of Dr. Neeta Srivastava (corresponding author). Her research area involves quality improvement of medicinal plants through the development of molecular markers that also includes phytochemical evaluation of marker compounds. She has two publications to her credit and a couple in the review process for publication.

Neeta Srivastava* has post graduate and doctorate degrees in plant sciences. Her research areas are plant biotechnology, molecular biology and phytochemistry and pharmacognosy. Her research interests also include phytopharmacology. Presently she works as a plant scientist in BV Patel PERD Centre, Ahmedabad. She has 44 publications, which include research articles, invited reviews, chapters in book and monographs. She has guided many students for post graduate dissertation. She has three students working with her for doctoral program.

* Corresponding Author