Biotechnology Advances 29 (2011) 199–209

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Genomic resources in horticultural crops: Status, utility and challenges

Humira Sonah, Rupesh K. Deshmukh, Vinay P. Singh, Deepak K. Gupta, Nagendra K. Singh, Tilak R. Sharma *

National Research Centre on Plant Biotechnology, IARI, Pusa Campus, New Delhi-110012, India

ARTICLE INFO

Article history:
Received 13 February 2010
Received in revised form 4 September 2010
Accepted 26 September 2010
Available online 19 November 2010

Keywords:
EST
Genomic resources
Horticultural crops
Positional cloning
QTL
Sequencing status

ABSTRACT

A significant review of status and availability of genomic resources in horticultural crops can be utilized for the efficient exploitation of the current research in developing improved varieties and also defining future goals. In this review, we describe the current genomic resources available in major horticultural crops and utility of the genomic and genic sequence information for isolating and characterizing novel useful genes and designing new DNA markers. We have found that these genomic resources have been utilized for both basic and applied research; however the progress is relatively slow. Recent advances in automation and high throughput techniques used in decoding plant genomes play an important role to speed up the genomic research. With the establishment of genome and transcriptome sequencing projects for several horticultural crops, huge wealth of sequence information have been generated. These sequence information have been used extensively for analyzing and understanding genome structures and complexities, comparative and functional genomics and to mine useful genes and molecular markers. However, certain limitations present a number of challenges for the generation and utilization of genomic resources in many important crops.

Contents

1. Introduction

Horticulture is one of the important sectors of agriculture which consists of fruits, flowers, vegetables, spices, plantation crops, medicinal and aromatic plants. The importance of horticultural crops is widely acknowledged in many aspects of innovation, production, quality maintenance, for uplifting economic condition of farmers, entrepreneurs and in providing nutritional security to the people. With the growing population, demand for horticultural products is gradually increasing. Latest genomic technologies can be effectively used in horticultural crop improvement programmes. Developing genomic resources like whole genome sequences, expressed sequence tags (ESTs), genomic survey sequences (GSS) and high throughput genome sequences (HTGs) are required to maintain the growth of these crops and associated value added

---

* Corresponding author. Tel.: +91 112584 1787; fax: +91 112584 3984.
E-mail address: trsharma@nrcpb.org (T.R. Sharma).

0734-9750/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.biotechadv.2010.11.002

opportunities. Genomic resources for model horticultural crops are increasing with great pace, however many of them are still not being exploited. Many genomic resources like whole genome sequence, large number of ESTs, large-insert genomic libraries, plenty of molecular markers and high-density genetic maps are available. These resources have been used for sequencing and annotation, mapping and cloning of genes or quantitative trait loci (QTL), and marker assisted selection (MAS) in important horticultural crops. However, as compared to cereals and other crops, the progress of genomic studies in horticultural crops is relatively slow. Availability of next generation sequencing (NGS) technologies like FLX-454, Illumina, SOLiD and Helicose have brought hopes to generate genomic resources for many more horticultural crops in few years time. Therefore, the horticulture breeders should equip themselves to make use of this extensive genome information in their varietal development programmes. The objectives of this review were to take stock of availability of genomic resources in horticultural crops, compile this information at one place, make sequence information useful for breeders and to identify potential future challenges which one can face while making proper use of genomic resources.

2. Genome sequencing projects in horticultural crops

Many projects were started for the genome sequencing of horticultural crops and few of them are and many more will be available in public domain in near future (Table 1). The most noticeable projects are Tomato genome sequencing project (www.sgn.cornell.edu/about/tomato), Potato genome sequencing consortium, (www.potatogenome.net), Papaya genome sequencing project (www.asgpb.mhpcc.hawaii.edu/papaya/), Grape genome sequencing project (www.vitaceae.org), Floral genome sequencing project (www.fgp.bio.psu.edu/) and many more. Along with conventional methods, these projects are well equipped with advanced sequencing tools, to ensure maximum coverage with high quality sequence and cost efficient methodology. The conventional DNA sequencing method referred as di-deoxynucleotide sequencing, or more commonly, Sanger's method of DNA sequencing provides large enough read length with quality sequence but it is time consuming and labour intensive. With the availability of next generation sequencing (NGS) technologies for DNA sequencing like FLX454 (Roche), Solexa (Illumina) and SOLiD (Applied Biosystems), there has been tremendous increase in the sequence database of several organisms. For instance, a single run of the 454 FLX machine on an average yields around 400,000 reads of 400 bp each, at a reasonable price in a few days time. Due to the large number of reads generated by the 454 DNA sequencing technology, it is possible to reveal the expression of many rare transcripts that would not be covered with Sanger's technology (Cheung et al., 2006). The generated sequence data are being analyzed for characterization of genes and validation of their functions through comparative and functional genomics approaches. For the application of genomics in crop improvement several high-throughput methods, genomic platforms and strategies are currently available to the plant community. Genomic research has great potential to revolutionize the molecular biology research in horticultural crops in many ways.

The whole genome sequencing of model plant *Arabidopsis thaliana* laid the foundation of plant genomics research (AGI, 2000). Since then rapid progress has been made in decoding complete genome sequence of plants with significant developments like generation of high-quality rice genome sequence (IRGSP, 2005), draft genome sequence of poplar (www.eurekalert.org), sorghum (www.phytozome.net/sorghum), grapevine (Jaillon et al., 2007), transgenic papaya (Ming et al., 2008), and Cucumber (Huang et al., 2009). The draft sequence of several other plant species like cassava, potato, tomato, shepherd's purse and peach are currently available in database. Multinational genome projects on *Brassica* (www.brassicagenome.org) and Solanaceous genomes (www.solgenomics.net) are in progress. Such huge sequence data of several species facilitate comparative genomics studies in plants. However, trained human resources in computational biology and efficient bioinformatics tools are required to make use of this information.

Among the *solanaceous* crops, tomato and potato have been used as scientific model for genomic studies. International Solanaceae Genome Project (ISGP) was formulated in 2003 to sequence tomato as

Table 1
Status of sequencing projects in horticultural crops up to August 2010.

Crop	Size (Mb)	Chr.	Status	Centre
Allium cepa	15,000	8	Initiated	USDA-ARS and University of Wisconsin
Aquilegia coerulea	350	–	In progress	Joint Genome Institute(JGI)
Aquilegia formosa	350	7	In progress	DOE Joint Genome Institute
Beta vulgaris	760	9	Initiated	Hokkaido University, Japan, Max Planck Institute, Michigan State University,
Brassica juncea	–	18	Initiated	Multinational Brassica Genome Project
Brassica oleracea	600	9	In progress	TIGR
Brassica rapa	500	10	In progress	*Brassica* IGF Project, The Multinational *Brassica rapa* Sequencing Project
Capsicum annuum	3000	12	Initiated	Korea Research Institute of Bioscience and Biotechnology, Seoul National University, Cornell University
Carica papaya	370	9	Draft assembly	The Papaya Genome Sequencing Consortium
Carya illinoinensis	–	16	Initiated	University of Georgia, USA
Citrus sinensis	380	9	In progress	U.S. Citrus Genomics Initiative, University of California
Coffea arabica	–	11	Initiated	University of Central Florida
Corylus avellana	–	11	Initiated	Department of Horticulture
Cucumis sativus	370	7	Draft assembly	The Cucumber Genome Initiative
Daucus carota	470	9	Initiated	University of Central Florida, Vegetable Crops Research Unit
Kadua centranthoides	–	–	Initiated	Lita Annenberg Hazen Genome Sequencing Center
Lactuca sativa	–	9	Initiated	Research Institute of Innovative Technology for the Earth, Japan, University of California, Davis (UCD)
Malus × domestica	750	17	In progress	IASMA research center, The Horticulture and Food Research Institute of New Zealand
Musa acuminata	600	11	In progress	Global Musa Genomics Consortium (GMGC)
Prunus armeniaca	–	8	Initiated	Instituto Valencianao de Investigaciones Agrarias, INRA Avignon
Prunus dulcis	–	8	Initiated	Beijing Agro-biotechnology Research Center at CUGI
Prunus persica	290	8	Initiated	Clemson University, CUGI
Rosa multiflora	–	7	Initiated	Rosaceae International Genomics Initiative (RosIGI)
Solanum bulbocastanum	–	12	In progress	University of Central Florida, Molecular Biology and Microbiology, USA, Orlando, TIGR
Solanum demissum	–	12	In progress	TIGR
Solanum lycopersicum	950	12	Complete	International Tomato Genome Sequencing Project; solgenomics.net
Solanum melongena	1100	12	Initiated	Cornell University
Solanum tuberosum	840	12	Draft assembly	Potato Genome Sequencing Consortium, TIGR, The Canadian Potato Genome Project
Vitis vinifera	500	19	Draft assembly	International Grape Genome Program, IASMA Research Center