Tropical Life Sciences Research, 27(1), 93–114, 2016

Sequencing Crop Genomes: A Gateway to Improve Tropical Agriculture

[1]Gincy Paily Thottathil[*], [1,2]Kandakumar Jayasekaran and [1,3]Ahmad Sofiman Othman

[1]Centre for Chemical Biology, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia
[2]Temasek Life Sciences Laboratory, National University of Singapore, Singapore 117543
[3]School of Biological Sciences, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

Abstrak: Perkembangan agrikultur di tropika ketinggalan di belakang perkembangan di kawasan temperat disebabkan kekurangan teknologi maju dan pelbagai faktor biotik dan abiotik. Untuk menghadapi permintaan yang meningkat terhadap makanan dan lain-lain produk berteraskan tumbuhan, varieti tanaman yang lebih baik perlu dihasilkan. Untuk membiak varieti yang lebih baik ini, pemahaman yang baik tentang genetik tanaman diperlukan. Dengan adanya teknologi penjujukan DNA generasi terkehadapan, banyak genom tanaman telah dijujuk. Kepentingan utama diberikan kepada tanaman makanan termasuk bijirin, tanaman ubi, sayuran, dan buahan. Maklumat jujukan DNA amat penting untuk mengenal pasti gen utama yang mengawal ciri agronomi penting dan untuk menentukan kevariabelan genetik antara kultivar. Namun, penjujukan semula DNA secara besar-besaran serta kajian pengekspresan gen perlu dilakukan untuk meningkatkan dengan ketara pemahaman kita dalam genetik tanaman. Aplikasi pengetahuan dari genom, transkriptom, kajian pengekspresan gen, dan epigenetik akan membolehkan perkembangan varieti yang lebih baik dan mungkin mengarah kepada revolusi hijau kedua. Aplikasi teknologi penjujukan DNA generasi terkehadapan dalam menambah baik tanaman, limitasinya, prospek masa hadapan, dan ciri-ciri penting projek genom tanaman diulas di sini.

Kata kunci: Agrikultur Tropika, Genom Tanaman, Penjujukan Genom

Abstract: Agricultural development in the tropics lags behind development in the temperate latitudes due to the lack of advanced technology, and various biotic and abiotic factors. To cope with the increasing demand for food and other plant-based products, improved crop varieties have to be developed. To breed improved varieties, a better understanding of crop genetics is necessary. With the advent of next-generation DNA sequencing technologies, many important crop genomes have been sequenced. Primary importance has been given to food crops, including cereals, tuber crops, vegetables, and fruits. The DNA sequence information is extremely valuable for identifying key genes controlling important agronomic traits and for identifying genetic variability among the cultivars. However, massive DNA re-sequencing and gene expression studies have to be performed to substantially improve our understanding of crop genetics. Application of the knowledge obtained from the genomes, transcriptomes, expression studies, and epigenetic studies would enable the development of improved varieties and may lead to a second green revolution. The applications of next generation DNA sequencing technologies in crop improvement, its limitations, future prospects, and the features of important crop genome projects are reviewed herein.

Keywords: Tropical Agriculture, Crop Genome, Genome Sequencing

[*]Corresponding author: gincyccb@gmail.com

INTRODUCTION

Tropical countries are generally underdeveloped compared to temperate countries. Poor agricultural productivity is a major reason for the underdevelopment of tropical countries (Gallup & Sachs 2000). The tropics are the centre of origin and domestication for many important crops. However, colonial rule in many developing tropical countries has reduced the number of crops to a few export commodities (Morales 2009), and the improvement of most of the staple food crops has received the least attention. Lack of technological adoption and various abiotic and biotic factors contribute to the decline in agricultural productivity in the tropics. An integrated approach using improved crop varieties and fertilisers and pesticides led to the green revolution in late 1960s, which could protect many of the developing countries against famine. Hybridisation has emerged in the 1960s to 1980s as a powerful breeding tool that gave rise to many high yielding crop varieties (Guimaraes 2009). A greater understanding of genetics, together with technological advancement led to the development of transgenic crops in 1990s (Mannion & Morse 2013). Transgenic technology was widely accepted initially, as it allows the transfer of one or a few desirable genes, in contrast to conventional breeding methods, in which undesired genes may also be transferred. Several transgenic varieties have been commercialised, including, insect resistant cotton, herbicide tolerant soybean, and virus resistant papaya (Mannion & Morse 2013). However, currently transgenic crops are controversial, especially genetically modified (GM) foods, as they may cause food allergies and may transfer antibiotic resistance to bacteria living in the gut (Mannion & Morse 2013). Environmentalists are concerned about the gene flow from transgenic plants to the wild varieties and the ecological imbalance that may be caused by transgenic plants with insecticidal proteins and herbicide resistance genes. Transgenic crops are not allowed in many countries, and transgenic research and field trials are not encouraged. Consequently, a different approach that can meet both the limitations of conventional breeding and the drawbacks of the transgenic approach is necessary for crop improvement. The advancements in sequencing technologies in recent years has revolutionised the field of genetics and opened a new era in crop breeding. The wealth of knowledge obtained from genome, transcriptome, gene expression profiles and epigenetic studies will help improve our understanding of underlying gene regulatory networks to empower a systematic improvement of crop breeding. Here, we review the applications in crop improvement for next generation sequencing technologies, discussing the limitations and future prospects of research. We also review the important crop genomes sequenced thus far.

HOW NEXT GENERATION SEQUENCING HELPS CROP IMPROVEMENT

Identification and exploitation of genetic variation is the basis of plant breeding. Traditional selection based on phenotype is tedious and time consuming. Molecular markers help to associate the phenotype with genotype. Many DNA

based molecular markers have been developed for major crops during the past decades and used for detecting the genetic variation among the cultivars (Varshney *et al*. 2009). Marker assisted selection has been carried out in the progeny, which allows the early selection of desired progeny. DNA markers such as restriction fragment length polymorphism (RFLP), random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), single sequence repeat (SSR), and single nucleotide polymorphisms (SNPs) have been identified and applied to improve breeding of several plants (Salgotra *et al*. 2014). However, most of the conventional markers (RFLP, RAPD, AFLP, SSR) are selectively neutral, as they are located in non-coding and non-regulatory regions. When such markers are used for marker assisted selection, there will be chances of false positives, due to genetic recombination (Salgotra *et al*. 2014). Gene based functional nucleotide polymorphism, if identifiable within the gene of interest, is more powerful and reliable. It is more advantageous than conventional markers, as there is no recombination between the marker and the gene of interest. Therefore, there is no information loss over time. Markers that allow for the identification of allelic variation of a particular trait are more valuable in crop breeding (Salgotra *et al*. 2014).

The recent advances in genome sequencing through next generation sequencing (NGS) technologies provide opportunities to develop millions of novel markers, as well as the identification of agronomically important genes (Edwards & Batley 2010). SNPs now dominate over other molecular marker applications, with the advancement in sequencing technology. Traditionally, PCR amplification is performed for the genomic region of interest from multiple individuals representing the diversity in a population, followed by sequencing. The sequences were then aligned to identify polymorphisms (Edwards & Batley 2010). This approach is expensive and time consuming. Now, large quantities of sequences generated through NGS platforms, together with the development of in silico methods, enable cheaper and more efficient SNP discovery. This approach also allows for the identification of functional indels (insertions or deletions), including partial or complete deletions of genes and different numbers of repeat motifs within SSRs (Salgotra *et al*. 2014). These markers have been used for the development of molecular genetic and physical maps, and for identifying the genes or quantitative trait loci controlling economically important traits (Varshney *et al*. 2009).

Advancements in NGS enabled the development of high-density genetic maps. Genetic mapping places the markers in linkage groups based on their co-segregation. The genetic map predicts the linear arrangement of markers in a chromosome based on the recombination frequency between genetic loci in a population derived from crosses of genetically diverse parents (Edwards & Batley 2010). The enormous sequence data obtained through NGS technologies have enabled the improvement of genetic maps by increasing marker density. Thousands of markers may be located in different linkage groups. It helps to localise corresponding scaffolds on the map, thus enabling the possibility of complete genome mapping (Perez-de-Castro *et al.* 2012). It also helps to replace traditional quantitative trait locus (QTL) mapping with association mapping,

because QTL provide a wide genome range within which the gene is located, whereas association maps mark traits with high resolution.

The sequence data obtained from genomes and transcriptomes, together with their expression profiles that are associated with different physiological conditions, will help to identify the genes determining different traits. These data enable the unravelling of the regulatory mechanisms behind different traits, and help to elucidate the complete pathway. These data also enable the identification of allelic variations in candidate genes controlling important agronomic traits, which is crucial for successful breeding programmes. Identification of the key genes underlying a trait enables the transfer of the trait to another cultivar or species by genetic modification; alternatively, these traits may be incorporated into a cultivar by marker-assisted selection (Edwards & Batley 2010). Furthermore, the analysis of copy number variations among and between species will contribute to the understanding of the mechanism of heterosis (Bolger *et al*. 2014). In addition to the sequence variation, epigenetic changes are also responsible for heritable traits (Bevan & Uauy 2013). Advancement in sequencing technologies allows for the survey of genome-wide epigenetic variation at high resolution through techniques such as bisulfite sequencing (Bi-seq), methylated DNA immunoprecipitation sequencing (MeDIP-seq), and methylation-sensitive restriction enzyme sequencing (MRE-seq) (Bell & Spector 2011).

Low agricultural productivity in the tropics can be explained by problematic soil due to humidity, rain fall variability, limited irrigation potential, pest and disease loads, and net photosynthetic potential differences (Gallup & Sachs 2000). The lack of freezing temperatures in the tropics favours an increased number of agricultural pests. Although the tropics are warmer and sunnier, it is generally cloudy, thus sunlight is blocked, which is disadvantageous for photosynthesis (Gallup & Sachs 2000). Also, night-time temperature is generally high, which causes high respiration and slows the rate of plant growth (Gallup & Sachs 2000). Identification of genes associated with disease resistance and other abiotic stress management would be particularly important for improving tropical agriculture. The knowledge obtained from genomes, transcriptomes, gene expression studies, and epigenetic variation studies would help to develop crop varieties that are capable of overcoming the disadvantages of tropical climates. Finally, one possible impact of genomics on plant breeding could be the development of a systems breeding approach, which integrates gene function information and regulatory networks to predict and estimate the contributions of genetic and epigenetic variations to phenotypes and field performance (Bevan & Uauy 2013).

A GLIMPSE INTO THE SEQUENCED CROP GENOMES

Following the genome sequencing of the model plant *Arabidopsis*, a number of crop species have been sequenced, many being important to tropical countries (Table 1). Most of the genome assemblies are in draft stage and extensive work is ongoing in the direction of closing the gaps and re-sequencing. In addition to

the genome sequence, transcriptomes and expressions profiles are also available for many crops. The large genome size and polyploidy exhibited by many crop species impedes the sequencing and further analysis. A high percentage of repeat elements is also a major hurdle in genome assembly. However, a platform has been established for many important crops and further research could lead to more information for application in crop breeding.

Sequencing Food Crops: An Endeavour to Reduce Hunger

The recent surge in plant genome sequencing is primarily aimed to reduce hunger. Among the sequenced plant genomes, most are food crops that are especially important for tropical countries. These crops include different cereals, pulses, tuber crops, vegetables, fruits, and oil plants. Functional markers have been developed for many of these crops and genes controlling agronomically important traits have been identified. However, re-sequencing and gene expression studies are continuing to be completed for a comprehensive understanding of genetic mechanism behind each trait and to identify allelic variations. In addition to the sequenced crops, many genome projects are underway or at the planning stage.

Three thousands rice genomes to feed billions

Rice (*Oryza sativa*) is the most important crop, as staple food for more than half of the world's population (Yu *et al*. 2002). It is the main food crop in most of the tropical countries. Rice cultivation occupies 11% of the world's total arable land and it is a source of income for more than 100 million people around the world (Guimaraes 2009). *O. sativa* has two major sub species, indica and japonica. Japonica varieties are usually cultivated in temperate regions, while indica varieties are important for the tropics. A third sub species, javanica is also cultivated in the tropics and is also known as tropical japonica. Glaszmann (1987) classified *O. sativa* into six groups; indica, japonica, aromatic, aus, rayada, and ashina, based on isozymes.

In the 1960s significant attention was given to the genetic improvement of rice, which preceded the green revolution. The main breeding goals were to increase yield, grain quality, resistant to blast disease, and drought tolerance (Guimaraes 2009). In the subsequent years, many high-yielding, semi-dwarf varieties (e.g., IR8) were developed by hybridisation. Mutation breeding was also popular for the development of new rice varieties. Biotechnological tools such as anther culture and protoplast fusion were shown to be promising tools in rice breeding (Guimaraes 2009). Several transgenic rice species (e.g., Golden rice) were also produced in 1990s (Khush & Brar 2003). In addition, different types of molecular markers were developed for rice and marker assisted selection has been employed in breeding programmes (Chen *et al*. 2000). A high-density rice genetic map was constructed with 2,275 markers (Harushima *et al*. 1998).

Table 1: Features of major sequenced crop genomes.

Scientific name	Common name	Economic importance	Top producing countries	Haploid chromosome number	Estimated genome size (Mb)	Assembly size (Mb)	Number of gene predictions	Repeat (%)	Reference
Azadirachta indica	Neem	Pesticides, medicine	India, Burma, Indonesia, Pakistan, Philippines	12	364.00	–	20,000	13.03	Krishnan *et al.* (2012)
Beta vulgaris	Sugar beet	Sugar production	Russia, France, USA, Germany, Ukraine	9	714.00 – 758.00	567.00	27,421	63.00	Dohm *et al.* (2014)
Brassica napus	Rapeseed	Oil, animal feed, biodiesel	Canada, China, India, France, Germany	19	1130.00	892.70	1,01,040	34.80	Chalhoub *et al.* (2014)
Brassica oleracea var. capitata	Cabbage	Food (vegetable)	China, India, Russia, Japan, South Korea	9	630.00	535.50	45,758	38.80	Liu *et al.* (2014b)
Brassica rapa	Chinese cabbage	Food (vegetable)	China, India, Russia, South Korea, Japan	10	529.00	283.80	41,174	39.50	The Brassica rapa Genome Sequencing Project Consortium (2011)
Cajanus cajan	Pigeon pea	Food	India, West Africa, Nigeria	11	833.07	605.78	48,680	51.67	Varshney *et al.* (2011)
Camelina sativa	Camelina	Oil, animal feed, biodiesel	Europe, Asia, North America	20	785.00	641.45	89,418	28.00	Kagale *et al.* (2014)
Carica papaya	Papaya	Food (fruit, vegetable)	India, Brazil, Mexico, Negeria, Indonesia	9	372.00	271.00	24,746	52.00	Ming *et al.* (2008)
Cannabis sativa	Marijuana	Drug	Cultivation is illegal in most of the countries	10	~820.00	534.70	30,000	–	Van Bakel *et al.* (2011)
	Hemp	Fibre, oil	China, France, Chile, Russia, Turkey			220.80	–	–	
Capsicum annum	Hot pepper	Spice	China, Turkey, Mexico, Nigeria, Spain	12	3,480.00	3,060.00	34,903	76.40	Kim *et al.* (2014)

(continued on next page)

Table 1: (*continued*)

Scientific name	Common name	Economic importance	Top producing countries	Haploid chromosome number	Estimated genome size (Mb)	Assembly size (Mb)	Number of gene predictions	Repeat (%)	Reference
Cicer arietinum	Chickpea	Food	India, Australia, Pakistan, Turkey, Burma	8	~738.00	532.29	28,269	49.41	Varshney *et al.* (2013)
Citrullus lanatus	Water melon	Food (fruit)	Nigeria, Cameroon, Sudan, Congo, Central African Republic	11	~425.00	353.50	23,440	45.20	Guo *et al.* (2013)
Citrus clementina	Clementine mandarin	Food (fruit)	Brazil, USA, India, China, Mexico	9	367.00	301.40	24,533	45.00	Wu *et al.* (2014)
Citrus sinensis	Sweet orange	Food (fruit)	Brazil, USA, India, China, Mexico	9	367.00	320.50	29,445	20.50	Xu *et al.* (2013)
Coffea canephora	Robusta coffee	Food	Vietnam, Brazil, India, Indonesia	11	710.00	568.60	25,574	50.00	Denoeud *et al.* (2014)
Cucumis melo	Melon	Food (fruit)	China, Turkey, Iran, Brazil, Egypt	12	450.00	375.00	27,427	19.70	González *et al.* (2010)
Cucumis sativus	Cucumber	Food (vegetable)	China, Iran, Turkey, Russian Federation, USA	7	367.00	243.50	26,682	24.00	Huang *et al.* (2009a)
Elaeis guineensis	Oil palm	Edible oil	Indonesia, Malaysia, Thailand, Nigeria, Colombia	16	1,800.00	1,535.00	34,802	57.00	Singh *et al.* (2013b)
Eragrostis tef	Tef	Food	Eritrea, Ethiopia	20	772.00	672.00	–	14.00	Cannarozzi *et al.* (2014)
Eucalyptus grandis	Eucalyptus	Wood, biofuel, medicine	China, India, Brazil, South Africa, Kenya	11	640.00	605.00	36,796	50.00	Myburg *et al.* (2011)
Fragaria vesca	Strawberry	Food (fruit)	USA, Spain, Japan, S.Korea, Mexico	7	240.00	209.8	34,809	16.00	Shulaev *et al.* (2011)
Glycine max	Soybean	Food	USA, Brazil, Argentina, China, India	20	1,115.00	950.00	46,430	57.00	Schmutz *et al.* (2010)

(continued on next page)

Table 1: (*continued)*

Scientific name	Common name	Economic importance	Top producing countries	Haploid chromosome number	Estimated genome size (Mb)	Assembly size (Mb)	Number of gene predictions	Repeat (%)	Reference
Musa acuminata	Banana	Food (fruit)	India, China, Philippines, Brazil, Ecuador	11	523.00	472.20	36,542	43.72	D'Hont *et al.* (2012)
Nicotiana tabacum	Tobacco	Smoking	China, India, Brazil, USA, Indonesia	12	4,500.00	3,700.00	90,000	72.00–78.00	Sierro *et al.* (2014)
Oryza sativa- spp indica	Rice	Food	China, India, Indonesia, Vietnam, Thailand	12	430.00	466.00	46,022–55,615	42.20	Yu *et al.* (2002)
Oryza sativa-spp japonica					420.00	389. 80	37,544	35.00	Goff *et al.* (2002)
Phaseolus vulgaris	Common bean	Food	India, Brazil, Burma, China, USA	11	587.00	473.00	27,197	45.37	Schmutz *et al.* (2014)
Phoenix dactylifera	Date palm	Food (fruit)	Egypt, Iran, Saudi Arabia, Pakistan, Iraq	18	671.00	605.40	41,660	21.99	Al-Dous *et al.* (2011); Al-Mssallem *et al.* (2013)
Phyllostachys heterocycla	Moso bamboo	Building material, furniture, paper	India, Brazil, China, Indonesia, Laos	24	2,075.00	2,050.00	31,987	59.00	Peng *et al.* (2013)
Populus trichocarpa	Poplar	Wood, paper	USA, Canada	19	485.00	410.00	45,555	44.00	Tuskan *et al.* (2006)
Prunus mume	Chinese plum/Mei	Food (fruit)	China, Serbia, USA, Romania, Chile	8	280.00	237.00	31,390	45.00	Zhang *et al.* (2012b)
Prunus persica	Peach	Food (fruit)	China, Italy, USA, Spain, Greece	8	265.00	226.60	27,852	29.60	The International Peach Genome Initiative (2013)
Pyrus bretschneideri	Pear	Food (fruit)	China, Italy, USA, Argentina, Spain	17	527.00	512.00	42,812	53.10	Wu *et al.* (2013)
Ricinus communis	Castor bean	Oilseed	India, China, Brazil, Ethiopia, Paraguay	10	320.00	350.00	31,237	50.33	Chan *et al.* (2010)
Setaria italica	Foxtail millet	Food, fodder, biofuel	India, China	9	490.00	423.00	38,801	46.00	Zhang *et al.* (2012a); Bennetzen *et al.* (2012)

(continued on next page)

Table 1: (*continued)*

Scientific name	Common name	Economic importance	Top producing countries	Haploid chromosome number	Estimated genome size (Mb)	Assembly size (Mb)	Number of gene predictions	Repeat (%)	Reference
Solanum lycopersicum	Tomato	Food (vegetable)	China, India, USA, Turkey, Egypt	12	900.00	760.00	34,727	63.28	The Tomato Genome Consortium (2012)
Solanum melongena	Eggplant	Food (vegetable)	China, India, Iran, Egypt, Turkey	12	1126.00	833.10	85,446	70.40	Hirakawa *et al.* (2014)
Solanum tuberosum	Potato	Food	China, India, Russia,Ukrain, USA	12	844.00	727.00	39,031	62.20	The Potato Genome Sequencing Consortium (2011)
Sorghum bicolor	Sorghum	Food, beverage	USA, China, Brazil, Mexico, Indonesia	10	~730.00	698.00	27,640	62.00	Paterson *et al.* (2009)
Theobroma cacao	Cocoa	Food	Cote d Ivoire, Indonesia, Ghana, Nigeria, Cameroon	10	430.00	326.90	28,798	25.70	Argout *et al.* (2011)
Triticum aestivum	Bread wheat	Food	USA, France, Canada, Australia, Argentina	21	17,000.00	3,800.33	94,000–96,000	80.00	Brenchley *et al.* (2012)
Vaccinium macrocarpon	Cranberry	Food (fruit)	USA, Argentina, Chile, Netherlands	12	470.00	420.00	36,364	5.60	Polashock *et al.* (2014)
Vigna radiata	Mungbean	Food	India, China, Myanmar	11	579.00	431.00	22,427	43.00	Kang *et al.* (2014)
Vitis vinifera	Grape	Food (fruit), beverage	China, Italy, USA, Spain, France	19	475.00	487.00	30,434	41.40	The French–Italian Public Consortium for Grapevine Genome Characterization (2007)
Zea mays	Maize	Food	USA, China, Brazil, Argentina, Mexico	10	2,300.00	2,048.00	32,540	85.00	Schnable *et al.* (2009)
Ziziphus jujuba	Jujube	Dry fruit, medicine	China	12	444.00	437.65	32,808	49.49	Liu *et al.* (2014a)

The development of NGS technology enabled fast-forward genetic studies in rice (Huang *et al*. 2013b). The International Rice Genome Sequencing Project (IRGSP) started in 1997, and included representation from 11 countries (International Rice Genome Sequencing Project 2008). The 12 chromosomes of *O. sativa* were distributed among the groups from 11 different countries (China, Japan, India, United States of America, United Kingdom, Taiwan, Korea, Thailand, France, Brazil, and Canada) (Eckardt 2000). Some private firms also contributed to the rice genome sequencing. In 2000, Monsanto completed a draft of the rice genome and made it available to IRGSP (Eckardt 2000). IRGSP aimed to obtain a high quality, map-based sequence of the rice genome using cultivar Nipponbare of *O. sativa* ssp. japonica. IRGSP adopted the clone-by-clone shotgun sequencing strategy so that each sequenced clone can be associated with a specific position on the genetic map (http://rgp.dna.affrc.go.jp/IRGSP/index.html). In addition, two independent groups published the draft genome of both indica (Yu *et al*. 2002) and japonica (Goff *et al*. 2002) sub-species using whole genome shotgun strategy. The genome assembly of the indica sub-species was 466 Mb in size with an estimated 46,022 to 55,615 genes (Yu *et al*. 2002). The genome size of *O. sativa* ssp. japonica was estimated to be 420 Mb and the assembly covered 93% of the genome with 32,000–50,000 gene predictions. Only 49.4% of predicted rice genes had homologs in *Arabidopsis thaliana*, whereas 80.6% of predicted *A. thaliana* genes were represented in rice genome (Yu *et al*. 2002). IRGSP released a high-quality map-based draft sequence in December 2002. They completed the rice genome sequencing in December 2004 and a high quality map-based sequence of the entire genome was published (International Rice Genome Sequencing Project 2005) and is available in public databases. The genome size was found to be 389 Mb, comprising 37,544 protein coding genes. The transposon content was estimated to be 35%, and 80,127 polymorphic sites were identified that distinguishes japonica and indica. Sequence and physical maps for individual chromosomes were also published, including chromosome 1 (Sasaki *et al*. 2002), chromosome 4 (Feng *et al*. 2002), chromosome 10 (The Rice Chromosome 10 Sequencing Consortium 2003), chromosome 3 (The Rice Chromosome 3 Sequencing Consortium 2005), chromosome 11 and 12 (The Rice Chromosomes 11 and 12 Sequencing Consortia 2005) and chromosome 5 (Cheng *et al*. 2005).

The various rice genome projects released an enormous amount of invaluable information and laid a strong foundation for rice genomics. These data were used to elucidate a major QTL for rice grain production, *Gn1a*, which was later identified as a cytokinin oxidase/dehydrogenase, an enzyme that degrades cytokinin (Ashikari *et al*. 2005). Later, the transcription factor controlling the expression of *Gn1a* was identified to be a zinc finger transcription factor, DST (draught and salt tolerance) (Li *et al*. 2013), which has been reported to regulate drought and salt tolerance in rice (Huang *et al*. 2009b). A genome-wide association study in a population of 950 world-wide rice varieties, including both indica and japonica subspecies, identified 32 loci associated with flowering time and 10 loci were associated with grain-related traits (Huang *et al*. 2012). However, more QTLs have to be mapped and the genetic variability between the cultivars and novel alleles from diverse germplasm has to be identified to improve

breeding programmes. The International Rice Research Institute (IRRI), the Chinese Academy of Agricultural Sciences (CAAS) and the Beijing Genome Institute (BGI) have undertaken a re-sequencing of 3,024 rice varieties to uncover the allelic variation. Alignment to the reference, japonica Nipponbare genome identified variants at over tens of millions loci. Variant calling with other reference genomes is underway (McNally 2014). The re-sequencing of 3,000 rice genomes would be the second milestone in rice genomics. Systematic mining of these data would help to link phenotypic variation to functional variation. Future crop breeding programmes should consider the effects of climate change and loss of arable land. As this project comprised rice varieties from different geographical locations, including many indigenous varieties, it can address questions on the genetic variations linked to climate and geographical factors. The results would lead to the generation of some of the most valuable data for rice breeding, eventually leading to the development of superior varieties with improved yield, high nutritional quality and improved tolerance towards diseases, pests, different soil conditions, and stresses such as draught and flood, to feed billions, especially the populations of developing tropical countries.

More than Food: Other Economically Important Crop Genomes

In addition to food crops, a few other economically important crops were also sequenced (Table 1). Some of these crops are highly valuable, governing the economy of tropical countries. Systematic mining and utilisation of these data would help to develop varieties with higher yield and tolerance to biotic as well as abiotic stresses, and would boost up the economy of many tropical countries.

Rubber and oil palm genomes: Promises to Malaysian economy

Natural rubber (NR) is a unique biopolymer used in the manufacture of over 50,000 products world-wide (Nair 2010). *Hevea brasiliensis* (rubber tree) is the major source of NR. The rubber tree originated from the amazon basin and has been domesticated in other tropical countries. Today, rubber cultivation is mainly performed in the Asian countries, which account for 93% of the world's supply. Malaysia has 4th position in NR production, after Thailand, Indonesia, and Vietnam. NR production in Malaysia was in its peak during the early 20th century; however, rubber plantation area has been gradually decreasing over the past 10 years. The rubber cultivation area reduced to 1.02 million ha in 2011 (Economic Transformation Programme [ETP] 2012). Decreased yield and susceptibility to diseases are the major challenges for rubber cultivation. Several high yielding clones were developed by the Malaysian rubber board and by rubber research centres in other countries. However, global demand for NR is increasing and to cope with this demand, genetically improved clones with more productivity have to be developed. In addition to NR, rubber wood is used as a source of timber with export value.

Towards molecular breeding, several molecular markers have been developed for rubber tree and a saturated genetic linkage map was published based on RFLP, AFLP, microsatellite, and isozyme markers (Lespinasse *et al*. 2000). The same group published another linkage map for the *H. brasiliensis* cultivar MDF 180, which is resistant to South American leaf blight, and the QTL

for resistance was mapped (Le Guen *et al*. 2011). Expressed sequence tags (EST) were generated from rubber latex, which provided more insights into rubber biosynthesis (Ko *et al*. 2003; Chow *et al*. 2007). With the advent of next generation sequencing technologies, several transcriptome sequencing projects have been completed and have been made available in the public domain (Triwitayakorn *et al*. 2011; Xia *et al*. 2011; Chow *et al*. 2012; Gébelin *et al*. 2012; Li *et al*. 2012; Lertpanyasampatha *et al*. 2012; Pootakham *et al*. 2012; Tang *et al*. 2013; Salgado *et al*. 2014). To obtain more insight into the noncoding regions and their regulatory roles, the draft genome of *H. brasiliensis* was published recently (Rahman *et al*. 2013). The assembly comprises 1.1 Gb of scaffolds of the estimated 2.1 Gb of genome. Approximately 78% of the genome was estimated to be repetitive DNA. A total of 68,955 gene models were predicted, of which 12.7% are unique to *H. brasiliensis*. Most of the genes associated with rubber biosynthesis, rubber wood formation, disease resistance and allergenicity have been identified. The genomic information together with transcriptomes provides a good foundation for the genetic studies and crop improvement.

Rubber yield depends mainly on three factors—the number of laticifer rings, the rate of sucrose loading into the laticifers and the rate of isopentenyl diphosphate (IPP) polymerisation on the rubber particle. Systematic mining of genomic and transcriptomic information together with further expression studies would help to identify the key genes associated with the above aspects, which could be utilised in breeding clones with higher yield. A major impairment to rubber cultivation is its susceptibility to various diseases. Genomic and transcriptomic studies have identified the disease resistant genes and further studies would reveal more insights into the rubber tree's genetic interaction with specific pathogens, leading to the development of disease resistant clones. Moreover, rubber cultivation is geographically restricted to a few regions. Increasing the area of rubber cultivation is another important approach to increase rubber production to cope with the global demand. Systematic mining of genomic and transcriptomic data would lead to the identification of genes imparting resistance to various geographical ailments and would lead to the development of clones suitable for various agro-climatic regions.

Oil palm (*Elaeis guineensis*) is the principal source of palm oil. Palm oil is a food ingredient and is also used to produce biodiesel and other industrially important products. In addition, palm biomass is used to generate renewable energy, fuels, and biodegradable products. Oil palm is a native plant to west and central Africa, and domesticated in South East Asia in the 19th century (Gerritsma & Wessel 1997). Malaysia is the second largest producer of palm oil, after Indonesia. Indonesia and Malaysia produce approximately 85% of the world's palm oil. The palm oil industry is one of the key economic drivers of these countries. In Malaysia, the oil palm planted area reached 5.23 million hectares by 2013 (Malaysian Palm Oil Board [MPOB] report, May 2013). Malaysia's palm oil sector is targeted to boost the country's total gross national income (GNI) by RM 125 billion to RM 178 billion by 2020 (ETP 2012).

Oil palm breeding has been revolutionised by the discovery of a single gene inheritance for shell thickness. The gene shows co-dominant monogenic inheritance, and has been exploited in breeding programmes (Sambanthamurthi

et al. 2009). With the advancement of genomics technology, the generation of ESTs, genetic mapping and application of DNA chip technology have been employed (Sambanthamurthi *et al*. 2009). A linkage map was constructed comprising 17 linkage groups with 117 RFLP loci, 384 AFLP markers and 23 SSR markers (Singh 2005). Several QTLs and the fruit colour genes (*vir*) have been successfully tagged in the linkage map. The markers associated with shell thickness have been identified; however, the closest marker linked to the shell thickness loci was mapped approximately 5 cM away from the shell thickness loci, far away to allow for an error free selection of the trait in the nursery (Sambanthamurthi *et al*. 2009). The ESTs also provided a platform for large-scale functional analysis of the genes using microarrays.

With the recent surge in next generation sequencing, the 1.8 Gb *E. guineensis* genome was sequenced with a combination of Roche/454 and Sanger bacterial artificial chromosome (BAC) end sequencing (Singh *et al*. 2013b). In addition, transcriptome data from 30 tissues and a draft sequence of the South American oil palm, *Elaeis oleifera* were reported. A total of 34,802 genes were predicted, including oil biosynthesis genes, homologues of WRINKLED1 (WRI1), and other transcriptional regulators, which are highly expressed in the kernel (Singh *et al*. 2013b). In the subsequent studies, the gene responsible for the shell thickness (*SHELL*) was identified and mapped (Singh *et al*. 2013a), delivering the opportunity for further exploitation in breeding programmes. Recently, an SNP based high density linkage map was constructed using genotyping by sequencing approach, and 3 QTL affecting trunk height and a single QTL associated with fruit bunch weight were identified (Pootakham *et al*. 2015). The sequence information provides the opportunity to mine other key genes responsible for higher productivity and resistance to biotic and abiotic stress.

A major criticism against oil palm cultivation is that oil palms are grown in rainforest regions and a large area of precious virgin forests is felled for oil palm plantation. This criticism will be more severe in the future, as the global demand for palm oil is increasing. Extending oil palm cultivation to less suitable areas is the only way to overcome this problem. However, this would severely affect the productivity and thereby the economy of Malaysia, the country currently with the highest cultivation of oil palms. Utilising the vast resource of genome sequence information, it is possible to identify the genes providing resistance against the adverse soil and environmental conditions in these areas, which would help to breed suitable varieties for these regions. The genome sequence could be a rich resource for oil palm breeders and could be an important step towards the sustainable production of palm oil to meet the global demand, and for the sustainability of Malaysian economy.

LIMITATIONS AND FUTURE DIRECTIONS

The advancement in sequencing technology has revolutionised the field of genetics, enabling the mass sequencing of genomes and transcriptomes. Taking advantage of the new technologies, many crop genomes have been sequenced. However, this research is still in its embryonic stage. Many crop genome assemblies are still in the draft stage. A high percentage of repeats in many plant genomes makes it difficult to assemble the short reads from the NGS platforms. Failure to capture the information embedded in the repetitive fraction of the genome is a major drawback, as it may have key roles in regulatory aspects (Feuillet *et al*. 2011). Heterozygosity and polyploidy also add to the difficulties. The redundancy created by 2 or more sets of genes can affect the accuracy of genome assembly (Feuillet *et al*. 2011). Scientists are trying to close the gaps in the assembly using a non-gridded BAC library approach. Launching third-generation sequencing platforms such as Pacific Biosciences would be promising to obtain longer reads for the assembly of whole chromosomes. Purification of individual chromosomes and using them for shotgun sequencing or construction of BAC libraries is also a powerful method to obtain the complete genome assembly (Bolger *et al*. 2014). Another shortcut to improve the assembly is the approximate ordering and positioning of genes, uses the synteny information from related species (Bolger *et al*. 2014). Extensive re-sequencing is needed for the detection of SNPs. The cost of sequencing is the major hurdle here. However, the cost has been reduced considerably in recent years and is expected to be cheaper in the near future. Sequence capture and targeted sequencing is advantageous in this respect as it is cost-effective and helps to find the variants in the selected genomic region. More reliable and user-friendly software have to be developed for more precise data analysis.

Another challenge is that the functions of many genes identified by genome sequencing remain unknown and the genetic control of the majority of agronomic traits has yet to be determined. Global research in *A. thaliana* has revolutionised the understanding of basic mechanisms in plant development, adaptation and tolerance to abiotic and biotic stresses. As the basic pathways are common to all plants, *Arabidopsis* genes can be used as candidate genes for identifying orthologs in crops. However, such translational biology is complex and inefficient for disease resistance. This is because, there are two resistance mechanisms; pathogen associated molecular pattern-triggered immunity (PTI) and effector-triggered immunity (ETI), of which ETI is specific to each species (Feuillet *et al*. 2011). Moreover, several crop plants are polyploids with more complex regulation between homoeologous genes, which may obscure the orthologous relationship between models and polyploid crop genomes (Feuillet *et al*. 2011). Gene expression profiling of different physiological responses by microarray or RNA-seq can provide clues to the functionality of genes. However, complete characterisation is needed before attempting gene transfer. The negative pleiotropic side effects also have to be considered (Salgotra *et al*. 2014). A complete and precise knowledge of the sequence, expression and functions of the genes has to be obtained before translating them into application through breeding. This decade should focus on acquiring knowledge and the

application of the knowledge acquired would be expected in the coming decades in the form of improved varieties of crops with better yield and resistant to biotic and abiotic stress.

CONCLUSION

Advancement in sequencing technologies has had a great impact on crop genetics, enabling the sequencing of genomes and transcriptomes of several crops. Although, reference genomes have been obtained for many important crops, massive re-sequencing and gene expression studies are essential to identify the key genes responsible for a desired trait and to find its allele variability. Utilisation of this knowledge in crop breeding would empower the development of better crop varieties and may lead to a second green revolution. This would reduce the hunger of billions and revolutionise the economies of developing tropical countries.

ACKNOWLEDGEMENT

We acknowledge with thanks the financial support from APEX funding (1002-PCCB-910206) (Malaysia Ministry of Higher Education) and eScience Fund (305-PCCB/613228), Ministry of Science, Technology and Innovation, Malaysia, provided to the Centre for Chemical Biology, Universiti Sains Malaysia. We thank Universiti Sains Malaysia for providing postdoctoral fellowships to the authors.

REFERENCES

Al-Dous E K, George B, Al-Mahmoud M E, Al-Jaber M Y, Wang H, Salameh Y M, Al-Azwani E K *et al*. (2011). De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology* 29(6): 521–527.

Al-Mssallem I S, Hu S, Zhang X, Lin Q, Liu W, Tan J, Yu X *et al*. (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nature Communications* 4: 2274 doi:10.1038/ncomms3274.

Argout X, Salse J, Aury J M, Guiltinan M J, Droc G, Gouzy J, Allegre M *et al*. (2011). The genome of *Theobroma cacao*. *Nature Genetics* 43(2): 101–108.

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles E R, Qian Q, Kitano H and Matsuoka M. (2005). Cytokinin oxidase regulates rice grain production. *Science* 309(5735): 741–745.

Bell J T and Spector T D. (2011). A twin approach to unraveling epigenetics. *Trends in Genetics* 27(3): 116–125.

Bennetzen J L, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli A C, Estep M *et al*. (2012). Reference genome sequence of the model plant Setaria. *Nature Biotechnology* 30(6): 555–561.

Bevan M W and Uauy C. (2013). Genomics reveals new landscapes for crop improvement. *Genome Biology* 14(6): 206–216.

Bolger M E, Weisshaar B, Scholz U, Stein N, Usadel B and Mayer K F. (2014). Plant genome sequencing—Applications for crop improvement. *Current Opinion in Biotechnology* 26: 31–37.

Brenchley R, Spannag M, Pfeifer M, Barker G L A, D'Amore R, Allen A M, McKenzie N *et al.* (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491(7426): 705–710.

Cannarozzi G, Plaza-Wüthrich S, Esfeld K, Larti S, Wilson Y S, Girma D, De Castro E *et al*. (2014). Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). *BMC Genomics* 15(1): 581–600.

Chalhoub B, Denoeud F, Liu S, Parkin I A, Tang H, Wang X, Chiquet J *et al*. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345(6199): 950–953.

Chan A P, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A *et al.* (2010). Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology* 28(9): 951–956.

Chen S, Lin X H, Xu C G and Zhang Q F. (2000). Improvement of bacterial blight resistance of 'Minghui 63', an elite line of hybrid rice, by molecular marker-assisted selection. *Crop Science* 40(1): 239–244.

Cheng C-H, Chung M-C, Liu S-M, Chen S-K, Kao F-Y, Lin S-J, Hsiao S-H *et al.* (2005). A fine physical map of the rice chromosome 5. *Molecular Genetics and Genomics* 274(4): 337–345.

Chow K-S, Mat-Isa M N, Bahari A, Ghazali A K, Alias H, Mohd-Zainuddin Z, Hoh C C and Wan K L. (2012). Metabolic routes affecting rubber biosynthesis in *Hevea brasiliensis* latex. *Journal of Experimental Botany* 63(5):1863–1871.

Chow K-S, Wan K L, Isa M N M, Bahari A, Tan S H, Harikrishna K and Yeang H Y. (2007). Insights into rubber biosynthesis from transcriptome analysis of *Hevea brasiliensis* latex. *Journal of Experimental Botany* 58(10): 2429–2440.

Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot, R, Pietrella M, Zheng C *et al*. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345(6201): 1181–1184.

D'Hont A, Denoeud F, Aury J-M, Baurens F-C, Carreel F, Garsmeur O, Noel B *et al*. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488 (7410): 213–217.

Dohm J C, Minoche A E, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, Rupp O *et al*. (2014). The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505(7484): 546–549.

Eckardt N A. (2000). Sequencing the rice genome. *Plant Cell* 12(11): 2011–2018.

Economic Transformation Programme (ETP). (2012). *Economic Transformation Programme annual report 2012.* http://etp.pemandu.gov.my/annualreport2014/ (accessed on 30 May 2015). Putrajaya: Performance Management and Delivery Unit, Prime Minister's Department, Malaysia.

Edwards D and Batley J. (2010). Plant genome sequencing: Applications for crop improvement. *Plant Biotechnology Journal* 8(1): 2–9.

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y *et al*. (2002). Sequence and analysis of rice chromosome 4. *Nature* 420(6913): 316–320.

Feuillet C, Leach J E, Rogers J, Schnable P S and Eversole K. (2011). Crop genome sequencing: Lessons and rationales. *Trends in Plant Science* 16(2): 77–88.

Gallup J L and Sachs J D. (2000). Agriculture, climate, and technology: Why are the tropics falling behind? *American Journal of Agricultural Economics* 82(3): 731–737.

Gébelin V, Argout X, Engchuan W, Pitollat B, Duan C, Montoro P and Leclercq J. (2012). Identification of novel microRNAs in *Hevea brasiliensis* and computational prediction of their targets*. BMC Plant Biology* 12(1): 18.

Gerritsma W and Wessel M. (1997). Oil palm: Domestication achieved? *Netherlands Journal of Agricultural Science* 45(4): 463–475.

Glaszmann J C. (1987). Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics* 74(1): 21–30.

Goff, S A, Ricke D, Lan T, Presting G, Wang R, Dunn M, Glazebrook J *et al*. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296(5565): 92–100.

González V M, Benjak A, Hénaff E M, Mir G, Casacuberta J M, Garcia-Mas J and Puigdomènech P. (2010). Sequencing of 6.7 Mb of the melon genome using a BAC pooling strategy. *BMC Plant Biology* 10(1): 246.

Guimaraes E P. (2009). Rice breeding. In M J Carena (ed.). *Cereals, the banks and the Italian economy*. USA: Springer Science+Business Media LLC, 99–126.

Guo S, Zhang J, Sun H, Salse J, Lucas W J, Zhang H, Zheng Y *et al*. (2013).The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics* 45(1): 51–58.

Harushima, Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T *et al.* (1998). A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* 148(1): 479–494.

Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamaguchi H *et al*. (2014). Draft genome sequence of eggplant (*Solanum melongena* L.): The representative solanum species indigenous to the old world. *DNA Research* 21(6): 649–660.

Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, Zhang L *et al.* (2013a). Draft genome of the kiwifruit *Actinidia chinensis*. *Nature Communications* 4: 2640. doi:10.1038/ncomms3640.

Huang X, Lu T and Han B. (2013b). Resequencing rice genomes: An emerging new era of rice genomics. *Trends in Genetics* 29(4): 225–232.

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W *et al*. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44(1): 32–39.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas W J *et al.* (2009a). The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics* 41(12): 1275–1281.

Huang X Y, Chao D Y, Gao J P, Zhu M Z, Shi M and Lin H X. (2009b). A previously unknown zinc finger protein, DST, regulates drought and salt tolerance in rice via stomatal aperture control. *Genes and Development* 23(15): 1805–1817.

International Rice Genome Sequencing Project. (2008). *International Rice Genome Sequencing Project*. http://rgp.dna.affrc.go.jp/IRGSP/index.html (accessed on 30 May 2015).

_____. (2005).The map-based sequence of the rice genome. *Nature* 436(7052): 793–800.

Kagale S, Koh C, Nixon J, Bollina V, Clarke W E, Tuteja R, Spillane C *et al*. (2014). The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nature Communications* 5: 3706. doi:10.1038/ncomms4706.

Kang Y J, Kim S K, Kim M Y, Lestari P, Kim K H, Ha B K, Jun T H *et al*. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nature Communications* 5: 5443. doi:10.1038/ncomms6443.

Khush G S and Brar D S. (2003). Biotechnology for rice breeding: Progress and impact in sustainable rice production for food security. In D V Tran and R Duffy. *Proceedings of the 20th session of the International Rice Commission*. Bangkok, Thailand, 23–26 July 2002. Rome: Plant Production and Protection Division, Food and Agriculture Organization (FAO), 41–58.

Kim S, Park M, Yeom S I, Kim Y M, Lee J M, Lee H A, Seo A *et al*. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nature Genetics* 46(3): 270–278.

Ko J-H, Chow K-S and Han K-H. (2003). Transcriptome analysis reveals novel features of the molecular events occurring in the laticifers of *Hevea brasiliensis* (para rubber tree). *Plant Molecular Biology* 53(4): 479–492.

Krishnan N M, Pattnaik S, Jain P, Gaur P, Choudhary R, Vaidyanathan S, Deepak S *et al*. (2012). A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genomics* 13(1): 464.

Le Guen V, Garcia D, Doaré F, Mattos C R, Condina V, Couturier C, Chambon A, Weber C, Espéout S and Seguin M. (2011). A rubber tree's durable resistance to *Microcyclus ulei* is conferred by a qualitative gene and a major quantitative resistance factor. *Tree Genetics and Genomes* 7(5): 877–889.

Lertpanyasampatha M, Gao L, Kongsawadworakul P, Viboonjun U, Chrestin H, Liu R, Chen X and Narangajavana J. (2012). Genome-wide analysis of microRNAs in rubber tree (*Hevea brasiliensis* L.) using high-throughput sequencing. *Planta* 236(2): 437–445.

Lespinasse D, Rodier-Goud M, Grivet L, Leconte A, Legnate H and Seguin M. (2000). A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme marker. *Theoretical and Applied Genetics* 100(1): 127–138.

Li D, Deng Z, Qin B, Liu X and Men Z. (2012). De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics* 13(1): 192.

Li S, Zhao B, Yuan D, Duan M, Qian Q, Tang L, Wang B *et al*. (2013). Rice zinc finger protein DST enhances grain production through controlling Gn1a/OsCKX2 expression. *Proceedings of the National Academy of Sciences* 110(8): 3167–3172.

Liu M J, Zhao J, Cai Q L, Liu G C, Wang J R, Zhao Z H, Liu P *et al*. (2014a). The complex jujube genome provides insights into fruit tree biology. *Nature Communications* 5: 5315. 10.1038/ncomms6315.

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin I A, Zhao M *et al*. (2014b). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications* 5: 3930. doi:10.1038/ncomms4930.

Mannion A M and Morse S. (2013). GM crops 1996–2012: *A review of agronomic, environmental and socio-economic impacts*. UK: Centre for Environmental Strategy, University of Surrey & Department of Geography and Environmental Science, University of Reading.

May C Y. (2013). *Overview of the Malaysian oil palm industry 2013*. Kuala Lumpur: MPOB.

McNally K L. (2014). The 3000 rice genome project. *Plant and Animal Genome XXII Conference*. Town and Country Hotel, San Diego, CA, 10–15 January 2014.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw J H, Senin P *et al.* (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452(7190): 991–996.

Morales F J. (2009). Introduction to tropical agriculture and outlook for tropical crops in a globalised economy. In K Del Claro, P S Oliveira and V Rico-Gray (eds.). *Tropical biology and conservation management,* vol. 3. Oxford, UK: EOLSS Publishers Co Ltd., 1–27.

Myburg A, Grattapaglia D, Tuskan G, Jenkins J, Schmutz J, Mizrachi E, Hefer C *et al*. (2011). The *Eucalyptus grandis* Genome Project: Genome and transcriptome resources for comparative analysis of woody plant biology. *IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery*. Arraial d'Ajuda, Bahia, Brazil, 26 June – 2 July 2011. *BMC Proceedings* 5(suppl. 7): I20. doi:10.1186/1753-6561-5-S7-I20.

Nair P K P. (2010). *The agronomy and economy of important tree crops of the developing world*. Burlington, USA: Elsevier.

Natsume S, Takagi H, Shiraishi A, Murata J, Toyonaga H, Patzak J, Takaqi M *et al*. (2014). The draft genome of hop (*Humulus lupulus*), an essence for brewing. *Plant and Cell Physiology* 56(3):428–441.

Paterson A H, Bowers J E, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G *et al*. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 457(7229): 551–556.

Peng Z, Lu Y, Li L, Zhao Q, Feng Q, Gao Z, Lu H *et al*. (2013). The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nature Genetics* 45(4): 456–461.

Pérez-de-Castro A M, Vilanova S, Cañizares J, Pascual L, Blanca J M, Díez M J, Prohens J and Picó B. (2012). Application of genomic tools in plant breeding. *Current Genomics* 13(3): 179–195.

Polashock J, Zelzion E, Fajardo D, Zalapa J, Georgi L, Bhattacharya D and Vorsa N. (2014). The American cranberry: First insights into the whole genome of a species adapted to bog habitat. *BMC Plant Biology* 14(1): 165–181.

Pootakham W, Chanprasert J, Jomchai N, Sangsrakru D, Yoocha T, Tragoonrung S and Tangphatsornruang S. (2012). Development of genomic-derived simple sequence repeat markers in *Hevea brasiliensis*. *Plant Breeding* 131(4): 555–562.

Pootakham W, Jomchai N, Ruang-areerate P, Shearman J R, Sonthirod C, Sangsrakru D, Tragoonrung S *et al.* (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics* 105(5): 288–295.

Prochnik S, Marri P R, Desany B, Rabinowicz P D, Kodira C, Mohiuddin M, Rodriguez F *et al.* (2012). The cassava genome: Current progress, future directions. *Tropical Plant Biology* 5(1): 88–94.

Rahman A Y A, Usharraj A O, Misra B B, Thottathil G P, Jayasekaran K, Feng Y, Hou S *et al*. (2013). Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics* 14(1): 75.

Salgado L R, Koop D M, Pinheiro D G, Rivallan R, Guen V L, Nicolás M F and de Almeida L G *et al*. (2014). De novo transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers. *BMC Genomics* 15(1): 236.

Salgotra R K, Gupta B B and Stewart Jr C N. (2014). From genomics to functional markers in the era of next-generation sequencing. *Biotechnology Letters* 36(3): 417–426.

Sambanthamurthi R, Singh R, Kadir A P G, Abdullah M O and Kushairi A (2009). Opportunities for the oil palm via breeding and biotechnology. In S M Jain and P M Priyadarshan (eds.). *Breeding plantation tree crops: Tropical species*. New York: Springer Science+Business Media LLC, 377–421.

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J *et al*. (2002). The genome sequence and structure of rice chromosome 1. *Nature* 420(6913): 312–316.

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K *et al*. (2011). Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Research* 18(1): 65–76.

Schafleitner R, Tincopa L R, Palomino O, Rossel G, Robles R F, Alagon R, Rivera C *et al*. (2010). A sweet potato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. *BMC Genomics* 11(1): 604.

Schmutz J, Cannon S B, Schlueter J, Ma J, Mitros T, Nelson W, Hyten D L *et al.* (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463 (7278): 178–183.

Schmutz J, McClean P E, Mamidi S, Wu G A, Cannon S B, Grimwood J, Jenkins J *et al*. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics* 46(7): 707–713. doi:10.1038/ng.3008.

Schnable P S, Ware D, Fulton R S, Stein J C, Wei F, Pasternak S, Liang C *et al.* (2009).The B73 maize genome: Complexity, diversity and dynamics. *Science* 326(5956): 1112–1115.

Shulaev V, Sargent D J, Crowhurst R N, Mockler T C, Folkerts O, Delcher A L, Jaiswa P *et al*. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* 43(2): 109–116.

Sierro N, Battey J N, Ouadi S, Bakaher N, Bovet L, Willig A, Geopfert S, Peitsch M C and Ivanov N V. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nature Communications* 5: 3833. doi:10.1038/ncomms4833.

Singh R. (2005). *Molecular markers and their applications in the construction of genetic linkage maps and analysis of monogenic and quantitative traits in oil palm*. PhD diss., Universiti Putra Malaysia.

Singh R, Low E-T L, Ooi L C-L, Ong-Abdullah M, Ting N-C, Nagappan J and Nookiah R. (2013a). The oil palm *SHELL* gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* 500(7462): 340–344.

Singh R, Ong-Abdullah M, Low E-T L, Mana M A A, Rosli R, Nookiah R, Ooi L C-L *et al*. (2013b). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500(7462): 335–339.

Tang C, Xiao X, Li H, Fan Y, Yang J, Qi J and Li H. (2013). Comparative analysis of latex transcriptome reveals putative molecular mechanisms underlying super productivity of *Hevea brasiliensis*. *PloS One* 8(9): e75307.

The *Brassica rapa* Genome Sequencing Project Consortium. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* 43(10): 1035–1039.

The French–Italian Public Consortium for Grapevine Genome Characterization. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161): 463–467.

The International Barley Genome Sequencing Consortium. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491(7426): 711–716.

The International Peach Genome Initiative. (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics* 45(5): 487–494.

The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355): 189–195.

The Rice Chromosome 3 Sequencing Consortium. (2005). Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. *Genome Research* 15(9): 1284–1291.

The Rice Chromosome 10 Sequencing Consortium. (2003). In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* 300(5625): 1566-1569.

The Rice Chromosomes 11 and 12 Sequencing Consortia. (2005). The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biology* 3: 20. doi:10.1186/1741-7007-3-20.

The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400): 635–641.

Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, Sraphet S, Yoocha T, Sangsrakru D, Chanprasert J *et al*. (2011). Transcriptome sequencing *of Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. *DNA Research* 18(6): 471–482.

Tuskan G A, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N *et al*. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793): 1596–1604.

van Bakel H, Stout J M, Cote A G, Tallon C M, Sharpe A G, Hughes T R and Page J E. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology* 12(10): 102.

Varshney R K, Chen W, Li Y, Bharti A K, Saxena R K, Schlueter J A, Donoghue M T A *et al.* (2011). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotechnology* 30(1): 83–90.

Varshney R K, Nayak S N, May G D and Jackson S A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27(9): 522–530.

Varshney R K, Song C, Saxena R K, Azam S, Yu S, Sharpe A G, Cannon S *et al*. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology* 31(3): 240–248.

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P *et al.* (2010). The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* 42(10): 833–839.

Wai C M, Van Buren R, Wang M, Tang H, Schatz M, Guyot R, Paull R E, Yu Q and Ming R. (2014). Development of pineapple (*Ananas comosus* L.) genetic map with restriction site associated DNA (RAD) markers. *Plant and Animal Genome XXII Conference*. San Diego, CA, 10–15 January 2014.

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L *et al.* (2012b). The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *The Plant Journal* 72(3): 461–73.

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z *et al*. (2012 a). The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* 44(10): 1098–1103.

Wu G A, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X *et al*. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology* 32(7): 656–662.

Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan M A *et al*. (2013). The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Research* 23(2): 396–408.

Xia Z, Xu H, Zhai J, Li D, Luo H, He C and Huang X. (2011). RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. *Plant Molecular Biology* 77(3): 299–308.

Xu Q, Chen L L, Ruan X, Chen D, Zhu A, Chen C, Bertrand D *et al.* (2013). The draft genome of sweet orange (*Citrus sinensis*). *Nature Genetics* 45(1): 59–66.

Yang H, Tao Y, Zheng Z, Zhang Q, Zhou G, Sweetingham M W, Howieson J G *et al*. (2013). Draft genome sequence, and a sequence-defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *PloS one* 8(5): e64799.

Yu J, Hu S, Wang J, Wong K S, Li S, Liu B, Deng Y *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296(5565): 79–92.

Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M *et al*. (2012a). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature Biotechnology* 30(6): 549–554.

Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y *et al.* (2012b). The genome of *Prunus mume*. *Nature Communications* 3: 1318.