

The First 50 Plant Genomes

Todd P. Michael* and Scott Jackson

FIFTY-FIVE plant genomes have been published to date representing 49 different species (Table 1 includes PubMed IDs for complete reference). What have we learned from the first wave of plant genomes? It has been said that plant genome papers (and genome papers in general) are dry and lack “biology” and that the days of high impact plant genome papers are drawing to a close unless they explore significant biology. However, with each new genome, earlier observations are refined and plant genome papers continue to reveal novel aspects of genome biology. For example, the tomato and banana genome papers refined current thinking on the whole genome duplications (WGD) that shaped dicot and monocot genome evolution (D’Hont et al., 2012; Tomato Genome Consortium, 2012). These observations were enabled not only by high quality genome assemblies but also by a greater number of genomes available for comparisons. In addition, the initial round of plant genomes enabled the first generation of functional genomics that helped to define the roles of hundreds of genes, provided unprecedented access to sequence-based markers for breeding, and provided glimpses into plant evolutionary history. More genomes, representing the diverse array of species in Viridiplantae are still required to gain a full understanding of plant genome structure, evolution, and complexity.

It All Started with a Wild Mustard Plant

Since the publication in 2000 of the model *Arabidopsis thaliana* genome in the journal *Nature*, the number of genomes has steadily increased, peaking in 2012 with 13 publications (Fig. 1A). At this current trajectory there should be hundreds of plant genome publications over the next several years. Genome papers have been quite formulaic with a description of the assembly, gene numbers, repeats, WGDs, over and under-represented gene families, and finally, some aspect of novel biology, usually with a focus on transcription factors. Genomes have been published in 12 different journals with 38 of the 55 (69%) published genomes appearing in *Nature* journals (*Nature*, *Nature Genetics*, *Nature Biotech*, and *Nature Communications*); *Science* is second with six published genomes. As we see from the most recent publication of the *Capsella rubella* genome paper, the genome paper is shifting from a formulaic approach to a focus on how the genome elucidates novel biological aspects, such as the evolution of selfing to an outcrossing mating system (Slotte et al., 2013). The trend toward biology is quite positive and necessitated by demands for publication in high impact journals. However, the plant community is just at the beginning of exploring the diversity of plant genomes, and the rigor of the genome paper model with the associated in-depth exploration of genome features provides an essential foundation for the plant research community.

One of the forces driving the rapid increase in fully sequenced plant genomes is the exponential decrease in cost and speed of genome sequencing fueled by high throughput DNA sequencing (Schatz et al., 2012). More than half of the published genomes have been sequenced entirely or partly using Sanger technology (Table 1), which provides long

Published in The Plant Genome 6
doi: 10.3835/plantgenome2013.03.0001in
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

T.P. Michael, IBIS Bioscience, Abbott Laboratories, Carlsbad, CA;
S. Jackson, The Center for Applied Genetic Technologies, Univ.
of Georgia, Athens, GA. Received 21 July 2013. *Corresponding
author (toddpmichael@gmail.com).

Table 1. Published plant genomes.[†]

	Scientific name	Common name	Year	Type	Division or monocot/dicot	Chr (#)	Size	Assembled	Assem	Gene (#)	Repeat	scaffold N50	contig N50	Sequencer types	Journal	PMID
							—Mb—	%			%		kb			
1	<i>Arabidopsis thaliana</i>	arabidopsis	2000	model	dicot	5	125	115	92	25,498	14	NA	NA	Sa	Nature	11130711
2	<i>Oryza sativa</i>	rice	2002	crop	monocot	12	430	362	84	59,855	26	12	7	Sa	Science	11935017
3	<i>Oryza sativa</i>	rice	2002	crop	monocot	12	420	389	93	61,668	NA	NA	NA	Sa	Science	11935018
4	<i>Oryza sativa</i>	rice	2005	crop	monocot	12	389	371	95	37,544	26	NA	NA	Sa	Nature	16100779
5	<i>Populus trichocarpa</i>	black cottonwood	2006	crop	dicot	19	485	410	84	45,555	NA	3100	126	Sa	Science	16973872
6	<i>Vitis vinifera</i>	grape	2007	crop	dicot	19	475	487	103	30,434	41	2065	66	Sa	Nature	17721507
7	<i>Physcomitrella patens</i>	moss	2008	model	bryophyta	27	510	480	94	35,938	16	1320	292	Sa	Science	18079367
8	<i>Vitis vinifera</i>	grape	2007	crop	dicot	19	505	477	95	29,585	27	1330	18	Sa,4	PlosOne	18094749
9	<i>Carica papaya</i>	papaya	2008	crop	dicot	9	372	370	99	28,629	43	1000	11	Sa	Nature	18432245
10	<i>Lotus japonicus</i>	lotus	2008	model	dicot	6	472	315	67	30,799	56	NA	NA	Sa	DNA Research	18511435
11	<i>Sorghum bicolor</i>	sorghum	2009	crop	monocot	10	818	739	90	34,496	62	62,400	195	Sa	Nature	19189423
12	<i>Cucumis sativus</i>	cucumber	2009	crop	dicot	7	367	244	66	26,682	24	1140	20	Sa,I	Nature Genetics	19881527
13	<i>Zea mays</i>	maize	2009	crop	monocot	10	2300	2048	89	32,540	85	76	40	Sa	Science	19965430
14	<i>Glycine max</i>	soybean	2010	crop	dicot	20	1115	973	87	46,430	57	47,800	189	Sa	Nature	20075913
15	<i>Brachypodium distachyon</i>	brachypodium	2010	model	monocot	5	272	272	100	25,532	21	59,300	348	Sa	Nature	20148030
16	<i>Ricinus communis</i>	castor bean	2010	crop	dicot	10	320	326	102	31,237	50	561	21	Sa	Nature Biotechnology	20729833
17	<i>Malus x domestica</i>	apple	2010	crop	dicot	17	742	604	81	57,386	67	1542	13	Sa,4	Nature Genetics	20802477
18	<i>Jatropha curcas</i>	jatropha	2010	crop	dicot	NA	380	286	75	40,929	37	NA	4	Sa,	DNA Research	21149391
19	<i>Theobroma cacao</i>	cocoa	2011	crop	dicot	10	430	327	76	28,798	24	473	20	Sa,4,I	Nature Genetics	21186351
20	<i>Fragaria vesca</i>	strawberry	2011	crop	dicot	7	240	210	87	34,809	23	1361	NA	4,S,I	Nature Genetics	21186353
21	<i>Arabidopsis lyrata</i>	lyrata	2011	model	dicot	8	207	207	100	32,670	30	24,500	227	Sa	Nature Genetics	21478890
22	<i>Selaginella moellendorffii</i>	spikemoss	2011	non-model	lycopod	NA	110	213	193	22,285	38	1700	120	Sa	Science	21551031
23	<i>Phoenix dactylifera</i>	date palm	2011	crop	monocot	18	658	381	58	28,890	40	30	6	I	Nature Biotechnology	21623354
24	<i>Solanum tuberosum</i>	potato	2011	crop	dicot	12	844	727	86	39,031	62	1318	31	Sa,4,I	Nature	21743474
25	<i>Thellungiella parvula</i>	thellungiella	2011	model	dicot	7	140	137	98	30,419	8	5290	NA	4,I	Nature Genetics	21822265
26	<i>Cucumis sativus</i>	cucumber	2011	crop	dicot	7	367	323	88	26,587	NA	319	323	Sa,4	PlosOne	21829493
27	<i>Brassica rapa</i>	chinese cabbage	2011	crop	dicot	10	485	284	59	41,174	40	1971	27	I	Nature Genetics	21873998
28	<i>Cannabis sativa</i>	hemp	2011	crop	dicot	?	820	787	96	30,074	NA	16	2	4,I	Genome Biology	22014239
29	<i>Cajanus cajan</i>	pigeon pea	2011	crop	dicot	11	833	605	72	48,680	52	516	22	Sa,I	Nature Biotechnology	22057054
30	<i>Medicago truncatula</i>	medicago	2011	model	dicot	8	454	262	58	62,388	31	1270	NA	Sa,4,I	Nature	22089132
31	<i>Setaria italica</i>	setaria	2012	model	monocot	9	490	423	86	38,801	46	1007	25	I	Nature Biotechnology	22580950
32	<i>Setaria italica</i>	setaria	2012	model	monocot	9	510	397	80	35,471	40	47,300	126	Sa	Nature Biotechnology	22580951
33	<i>Solanum lycopersicum</i>	tomato	2012	crop	dicot	12	900	760	84	34,727	63	16,467	87	Sa,4,S,I	Nature	22660326
34	<i>Cucumis melo</i>	melon	2012	crop	dicot	12	450	375	83	27,427	NA	4680	18	Sa,4,I	PNAS	22753475
35	<i>Linum usitatissimum</i>	flax	2012	crop	dicot	15	373	318	85	43,484	24	132	20	I	Plant Journal	22757964
36	<i>Musa acuminata malaccensis</i>	banana	2012	crop	monocot	11	523	472	90	36,542	44	1311	43	Sa,4,I	Nature	22801500
37	<i>Gossypium raimondii</i>	cotton D	2012	crop	dicot	13	880	775	88	40,976	60	2284	45	I	Nature Genetics	22922876
38	<i>Azadirachta indica</i>	neem	2012	crop	dicot	NA	364	NA	NA	20,169	13	452	1	4,I	BMC Genomics	22958331
39	<i>Hordeum vulgare</i>	barely	2012	crop	monocot	7	5100	4980	98	30,400	84	NA	NA	NA	Nature	23075845

(cont'd)

Table 1. Continued.

Scientific name	Common name	Year	Type	Division or monocot/dicot	Chr (#)	Size	Assembled	Assem	Gene (#)	Repeat	scaffold N50	contig N50	Sequencer types	Journal	PMID
						—Mb—	%			%		kb			
40 <i>Pyrus bretschneideri</i>	pear	2013	crop	dicot	17	527	512	97	42,812	53	541	36	I	Genome Research	23149293
41 <i>Citrullus lanatus</i>	watermelon	2012	crop	dicot	11	425	354	83	23,440	45	2380	26	I	Nature Genetics	23179023
42 <i>Triticum aestivum</i>	wheat	2012	crop	monocot	21	17,000	3800	22	94,000	80	NA	1	4	Nature	23192148
43 <i>Gossypium raimondii</i>	cotton D	2012	crop	dicot	13	880	738	84	37,505	61	18,800	136	Sa,4,I	Nature	23257886
44 <i>Prunus mume</i>	chinese plum	2012	crop	dicot	8	280	237	85	31,390	45	578	32	I	Nature Communications	23271652
45 <i>Cicer arietinum</i>	chickpea	2013	crop	dicot	8	738	532	72	28,269	49	39,990	24	Sa,I	Nature Biotechnology	23354103
46 <i>Hevea brasiliensis</i>	rubber tree	2013	crop	dicot	18	2150	1119	52	68,955	72	3	NA	4,S,I	BMC Genomics	23375136
47 <i>Phyllostachys heterocyda</i>	moso bamboo	2013	non-model	monocot	24	2075	2051	99	31,987	59	329	12	I	Nature Genetics	23435089
48 <i>Oryza brachyantha</i>	rice relative	2013	non-model	monocot	12	300	263	88	32,038	29	1013	20	I	Nature Communications	23481403
49 <i>Prunus persica</i>	peach	2013	crop	dicot	8	265	227	86	27,852	37	27,400	214	Sa	Nature Genetics	23525075
50 <i>Aegilops tauschii</i>	wheat DD	2013	crop	monocot	7	4360	4244	97	43,150	66	58	5	4,I	Nature	23535592
51 <i>Triticum urartu</i>	wheat AA	2013	crop	monocot	7	4940	4660	94	34,879	67	64	3	I	Nature	23535596
52 <i>Nelumbo nucifera</i>	ancient lotus	2013	non-model	dicot	8	929	804	87	26,685	57	3400	39	I	Genome Biology	23663246
53 <i>Utricularia gibba</i>	bladderwort	2013	non-model	dicot	16	77	82	106	28,500	3	95	26	4,I	Nature	23665961
54 <i>Picea abies</i>	norway spruce	2013	crop	gymnosperm	12	19,600	12,000	61	28,354	NA	NA	NA		Nature	23698360
55 <i>Capsella rubella</i>	capsella	2013	non-model	dicot	8	219	135	62	26,521	NA	15,100	134	Sa	Nature Genetics	23749190

¹Abbreviations: Sa, Sanger; 4, Roche/454; S, SOLiD; I, Illumina; NA, not reported in primary publication; kb, kilobases; Mb, megabases; Chr, chromosome; PMID, PubMed ID

high quality ~1000 base pair (bp) reads. Sanger sequencing requires a cloning step and is time consuming with an expensive price tag, although the final result is usually high quality depending on the genome. When 454 came onto the scene in the early 2000s the cost of sequencing dropped an order of magnitude (US\$200K vs. US\$2 M) encouraging the emergence of consortia and funding for the sequencing of new genomes. Grape was the first genome published in 2007 using a combination of 454 and Sanger, and now there are at least 18 genomes that have used varying amounts of 454 sequence. Illumina and SOLiD sequencing changed the paradigm yet again providing very short reads (35–150 bp) at yet another order of magnitude lower cost than 454. Only two genome projects have used SOLiD for genome sequencing (strawberry and tomato); however, Illumina has played an exclusive role in 12 genomes, and was used in combination with other technologies in another 17 genomes. Third generation sequencing technologies such as Pacific Bioscience (PacBio) promise long (>5 kb) single molecule reads that would greatly improve assembly of repeat rich plant genomes. PacBio long reads show great promise in resolving regions that the other sequencing technologies have problems with (skewed GC, homopolymers), but throughput and accuracy are two issues that still require attention. However, new sequencing technologies are only part of the future of plant genomes since tried and true methods, such

as BACs (bacterial artificial chromosomes), are finding a place in hybrid sequencing approaches such as in the highly heterozygous pear genome (Wu et al., 2013).

Most of the plants chosen to be sequenced to date fit specific criteria such as size of research community, model organisms or economically important, small genome size, ploidy (diploid), availability of inbred lines (low heterozygosity), access to genetic and physical maps, expressed sequence tags (EST)/transcriptome and other genomic tools. Seventy-three percent (40) of the plant genome publications have been on crop species and some of these crop species double as model systems while several were sequenced purely for research such as *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Brachypodium distachyon*, *Physcomitrella patens* (moss), and *Selaginella moellendorffii* (spikemoss). Most (94%) genomes sequenced to date are Angiosperms, of which 36 are dicots and 16 are monocots, while only one gymnosperm (spruce), one bryophyte (moss), and one lycophyta (club-moss) have been sequenced (Table 1). Much of the early decisions about which genomes to sequence were driven by the Department of Energy Joint Genome Institute (JGI), which resulted in the publication and public availability (phytozome) of eleven of the highest quality plant genomes. The Beijing Genome Institute (BGI) has contributed consistently over the years starting with the rice genome, then ten additional genomes primarily based on Illumina technology, and now they have announced

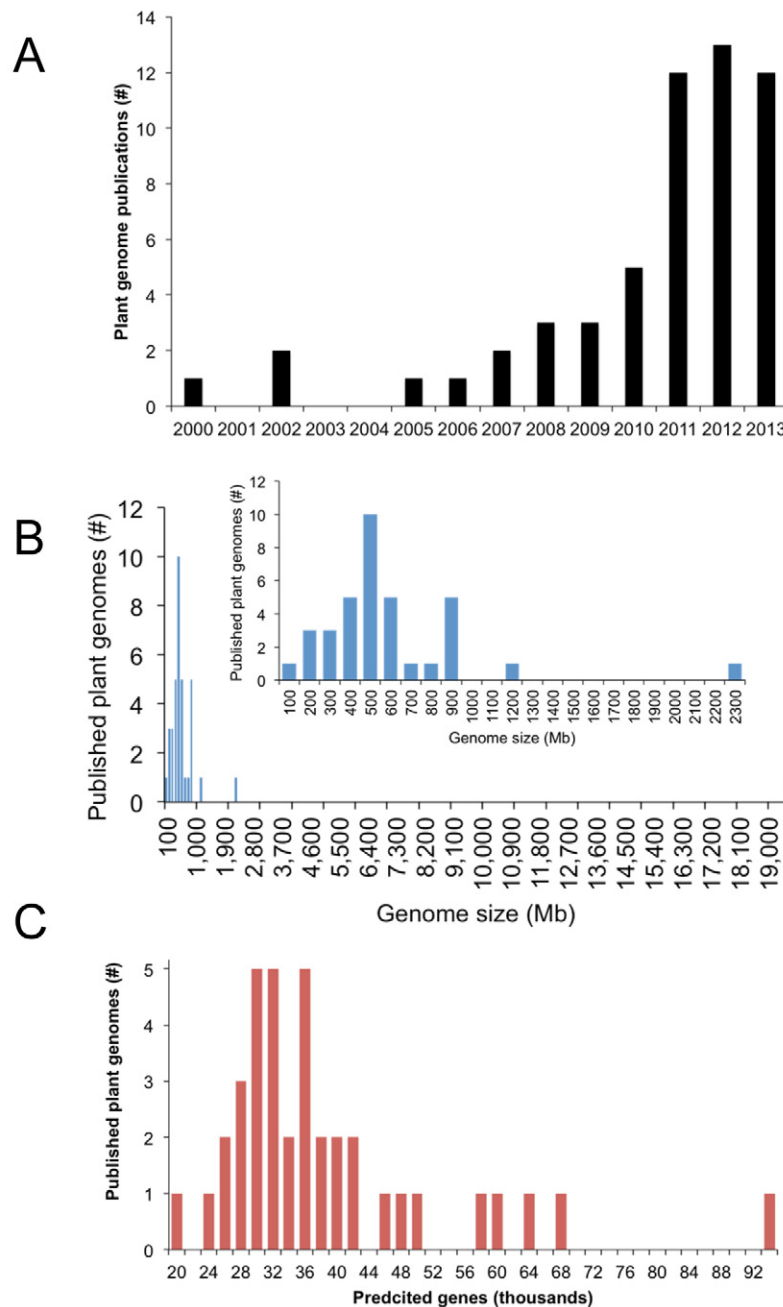


Figure 1. Published plant genome statistics. (A) Number of plant genomes sequenced since *Arabidopsis thaliana* in 2000 by year. (B) Published plant genome size distribution with insert focused on median genome size between 77 and 2300 Mb. (C) Predicted gene number across published plant genomes.

a large-scale plant genome sequencing project. However, a “1000 plant genome project” analogous to that in other communities has yet to emerge.

Plant Genomes Both Large and Small

Plant genome sizes span several orders of magnitude from the carnivorous corkscrew plant (*Genlisea aurea*) at 63 megabases (Mb) to the rare Japanese *Paris japonica* at 148,000 Mb (Bennett and Leitch, 2011). The smallest published genome is the carnivorous bladderwort (*Utricularia gibba*) at 82 Mb, while the largest, the Norway Spruce (*Picea abies*), stands by itself at 19,600 Mb, compared to the second largest of

maize at 2300 Mb and the overall median of 480 Mb (Table 1, Fig. 1B). Access to high quality reference genomes confirmed that long terminal repeats (LTRs) retrotransposons are a primary driver of the dramatic size range in plants (El Baidouri and Panaud 2013). For the large barley genome (5100 Mb), where retrotransposons are abundant and more recently active, a powerful genomics resource was generated through an alternative “gene-ome” approach by anchoring a high quality genespace assembly on a deep physical map merged with high-density genetic maps (International Barley Genome Sequencing Consortium, 2012). In contrast, large gymnosperm genomes have highly diverged ancient repeats,

which could make assembling these genomes tractable with current sequencing and assembly technologies (Kovach et al., 2010). The smallest reported conifer genome is the same size as maize and the median genome size is 9700 Mb, which is why a large push to sequence gymnosperms may have to wait for the next wave of sequencing technologies with increased read length and decreased price. As the community moves forward to choose the next round of genomes to sequence, the Kew Genome Size database will continue to provide a rich resource of non-model/non-crop species to investigate (Bennett and Leitch, 2011).

One measure of genome assembly quality is the contiguity or the length of contigs and scaffolds at which 50% of the assembly can be found; this is commonly referred to as N50. Sorghum, *Brachypodium distachyon*, soybean, and foxtail millet have the top four scaffold contiguities with 62.4, 59.3, 47.8, and 47.3 Mb respectively and all four were sequenced using Sanger as part of the JGI pipeline (Table 1). However, the genome with the ninth largest scaffold N50 is the tomato genome at 16 Mb, which was predominantly assembled using 454. Each scaffold is comprised of thousands of contigs and contig length generally drives the completeness and quality of the gene predictions. Not surprising, the 11 JGI assemblies based on Sanger have the top contig N50 ranging from 347 to 119 kilobases (kb), while the median contig N50 for all assemblies is 25.6 kb. Illumina based assemblies, primarily from BGI, have a similar median length (25.9 kb), which reflects their comprehensive strategy that makes use of different sized sequencing libraries. Another measure of a genome assembly is the amount of the genome captured in the assembly. Of the published genomes, the median genome assembly captured 85% of the predicted genome size, which is usually estimated by flow cytometry or more recently by k-mer depth analysis. The remaining fraction of the genome not assembled generally represents the highly repetitive portion of the genome such as high copy number ribosomal repeats, centromeres, telomeres, and transposable elements. Therefore an average plant genome assembly captures 85% of the genome space in thousands of contigs with an N50 of 20 kb and tens of scaffolds with an N50 of 1 Mb.

Annotation of any genome, but particularly plant genomes, is difficult especially as the definition of what constitutes a gene continues to evolve. Many parts of the genome are ‘expressed’ in that RNAs are formed, but do not correspond to traditional genes in that they are not translated to a protein. However, most annotated plant genomes have between 20,000 and 94,000 genes with a median predicted gene count of 32,605 (Table 1, Fig. 1C). Differences between genomes most likely lies in the tools used for annotation and how relaxed the annotators were in calling genes as well as lineage-specific genes and gene family expansions. Genomes produced by next generation sequencing typically have smaller contig and scaffold sizes that complicate annotation as genes may not exist on single contigs but may be broken across contigs, thus inflating the number of annotated genes (e.g., pigeon pea, Varshney

et al., 2012). Further complicating annotation is that there are many expressed non-coding RNAs that are functionally important (Eddy, 2001), but not considered genes in a traditional sense. Small RNA precursors are often not included in a genome annotation, but are important for plant development and silencing of TEs (Arikiti et al., 2013). Small RNAs and other non-coding RNAs are often annotated and curated separately from genome annotations in small, boutique databases. Long-term, however, one goal should be to combine these various sources of information into a single database/annotation making it easier for the biologist to pull together relevant information needed for forming hypotheses.

Plant genomes are packed, and often obese, with transposable elements (TEs) (Bennetzen 2000), which contain protein-coding sequences that are often annotated as genes. In rice, for instance, it was estimated that only 40,000 of the more than 55,000 annotated genes are really genes and that the other 10,000 to 15,000 are TEs—usually low copy TEs as high copy elements are relatively easy to find (Bennetzen et al., 2004). TEs include various families that move via copy-and-paste (class I) and cut-and-paste (class II) mechanisms. Copy-and-paste TEs can dramatically increase the size of a genome such as occurred in a relative of rice with a genome nearly two-fold larger than rice (Piegu et al., 2006). Transposon biology is an intriguing area of research and relies on relatively complete genomes so that TEs are captured in sequence contigs and can be accurately annotated. Schemes for classification of TEs have been agreed on (Wicker et al., 2007), but annotation of non-LTR TEs is complicated by the lack of structural clues that allow routine ab initio prediction (El Baidouri and Panaud, 2013). Another complication is that in genomes produced by short read DNA sequencing technology, TEs are often missed in the assembly due to their repetitive nature. Genomes sequenced to date range from 3 to 85% repetitive sequence (Table 1; median 43%), with TEs, specifically cut-and-paste TEs (LTRs), comprising the majority of that sequence. Capturing and annotating these genomic components is important as it is becoming increasingly clear that TEs can be domesticated to function in gene regulation and as structural components of the genome.

Making Genomes “Functional”

One of the key take homes from the first 49 sequenced plant species is that we still have a lot to learn about the organization of genomes, function of genes, and how to characterize the non-coding space. Each new genome uncovers novel genes specific to a species, and a vast amount of non-coding space that requires methods for ab initio and functional annotation. One specific challenge is how we will leverage a growing number of high throughput technologies, otherwise referred to as “omics” approaches, to functionally annotate features of the plant genome. In this special issue of *The Plant Genome* we highlight several omics studies that have used high throughput approaches such as gen-omics (SNP detection), epigen-omics (methylation) metagen-omics (plant-fungal interactions), and ion-omics (element profiling) to refine our functional understanding of several

key crop genomes (Eichten et al., 2013; Roorkiwal et al., 2013; Ruzicka et al., 2013; Ziegler et al., 2013). As we have seen through the model organism and human ENCODE projects, the layering of omics data exponentially increases the value of a reference genome (Celniker et al., 2009; ENCODE Project Consortium 2012).

While reference genomes provide a starting point, or platform for discovery in a specific species, it only captures a brief moment in the history of that species' diversity and lacks the information content that would enable activities such as molecular breeding and phylogenetic analyses. Roorkiwal et al. (2013, this issue) describe the development of an Illumina BeadXpress SNP genotyping platform for two important crops in the developing world, pigeon pea and chickpea (Roorkiwal et al., 2013). Both pigeon pea and chickpea have lagged behind other crops in their genetic improvement due to a lack of genome and breeding resources that would enable such applications as marker assisted selection (MAS) and phylogenetic screens to identify genetic novelty in wild species. The development of an Illumina BeadXpress SNP genotyping platform provides the opportunity to assess larger populations of plants with an adequate density of markers, which is ideal for breeding applications such as MAS and scans of diversity for disease and abiotic traits.

A prominent feature of plant genomes is their epigenetic landscape. The epigenome encompasses DNA methylation, histone modifications and other modifications not directly encoded in the genome. In general, DNA methylation is thought to mark permanent changes in the genome that must exist over the developmental lifetime of the plant, such as silencing transposable elements in embryonic tissue to protect the fidelity of the genome from transposition. Eichten et al. (2013, this issue) address the question of whether DNA methylation also specifies tissue types in maize. Using genome-wide array and sequencing technologies to assess DNA methylation and gene expression in two maize inbreds, B73 and Mo17, across four tissue types (leaf, immature tassel, embryo and endosperm), the authors find that there are more differentially methylated regions (DMRs) between maize inbreds than in the tissues they sampled (Eichten et al., 2013). The DMRs that were identified between tissue types did not correlate with subsequent expression changes suggesting the DMRs were not in fact functional in specifying tissue type. Despite other plants such as tomato that display tissue and developmentally regulated DMRs (Zhong et al., 2013), this may not be a general phenomenon in other species such as maize, which highlights the need to functionally define genomic elements in specific species.

Genetic screens are still the primary tool for functionally defining features of genomes. Mutant screens have been central in elucidating pathways, uncovering novel functionality of known genes, and allowing the discovery of novel non-coding features such as epigenetic regulation and small RNAs. Ziegler et al. (2013, this issue) describe a powerful high throughput mutant screen for elemental differences between field grown soy plants, which could be applied to any plant species with modestly sized seeds like soy (Ziegler et al., 2013). High throughput elemental

profiling, or ionomics, is an emerging omics platform that provides a glimpse of a plant–soil environment and how that plant is accessing that environment. Ionomics screens have been powerful at detecting genetic factors controlling ion uptake but also have started to shed light on root architecture and morphology. Therefore, this high throughput screen, which is agnostic to plant species, has the potential to functionally characterize a plant organ, the root, which has traditionally been difficult to define genetically and molecularly in a field environment.

An almost uncharacterized area of plant biology is the complement of organisms that live mutually with plant communities, or the metagenome. In many plants, the acquisition of inorganic minerals is facilitated by an active network of mycorrhizal associations between soil fungal species and plant roots. However, assessing how these fungal and plant species interact has been hampered by the fact that many fungal species cannot be cultured. The advent of high throughput sequencing has enabled an unprecedented opportunity to identify the genomic changes induced through these communal relationships. Ruzicka et al. (2013, this issue) use high throughput sequencing to characterize the transcriptomes of both the tomato genome and its arbuscular mycorrhizal fungal symbiont in the field (Ruzicka et al., 2013). Instead of culturing the symbiont, a metagenomic sequencing strategy was employed where RNA from a wild-type tomato plant and a mutant for reduced mycorrhizal colonization were sequenced and bioinformatically separated. This metagenomic analysis revealed a suite of genes for transport and cell wall remodeling required for the symbiotic relationship. Metagenomic sequencing will open up the opportunity to explore additional symbiotic relationships and further functionally characterize aspects of the genome that are not innate to the genome sequence.

Future Plant Genomes

The first ~50 plant genomes have provided a glimpse at the gene number, types and numbers of repeats, and how genomes grow and contract. However, we are just at the beginning of defining the functional aspects of plant genomes. To reach the goal of breeding better plants for future food, clothing, and energy, we will need to expand both the species sequenced, the number of species re-sequenced, and the type of omics data layered on genomes. Currently only one gymnosperm has been sequenced and no CAM (Crassulacean acid metabolism) photosynthetic plants have been sequenced. While we have come a long way over the past 13 yr since the publication of the *Arabidopsis* genome, we still have a long way to go before we will be able to engineer the plant of the future.

Current Access to Plant Genomes

General Plant Genome Resources (accessed 24 July 2013)

<http://www.phytozome.net/>
<http://bioinformatics.psb.ugent.be/plaza/>
<http://mips.helmholtz-muenchen.de/plant/genomes.jsp>

<http://www.plantgdb.org/>
<http://pgn.cornell.edu/>
<http://www.gramene.org/>
<http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html>
http://genomeevolution.org/wiki/index.php/Sequenced_plant_genomes

Plant Specific (accessed 24 July 2013)

<http://www.arabidopsis.org/>
<http://www.maizegdb.org/>
<http://www.maizesequence.org>
<http://solgenomics.net/>

Supplemental Information Available

Table 1 is also available for download with the online version of this article.

References

- Arikiti, S., J. Zhai, and B.C. Meyers. 2013. Biogenesis and function of rice small RNAs from non-coding RNA precursors. *Curr. Opin. Plant Biol.* 16:170–179. doi:10.1016/j.pbi.2013.01.006.
- Bennett, M.D., and I.J. Leitch. 2011. Nuclear DNA amounts in angiosperms: Targets, trends and tomorrow. *Ann. Bot. (Lond.)* 107(3):467–590. doi:10.1093/aob/mcq258.
- Bennetzen, J.L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* 42(1):251–269. doi:10.1023/A:1006344508454.
- Bennetzen, J.L., C. Coleman, R. Liu, J. Ma, and W. Ramakrishna. 2004. Consistent over-estimation of gene number in complex plant genomes. *Curr. Opin. Plant Biol.* 7:732–736. doi:10.1016/j.pbi.2004.09.003.
- Celniker, S.E., L.A. Dillon, M.B. Gerstein, K.C. Gunsalus, S. Henikoff, G.H. Karpen, M. Kellis, E.C. Lai, J.D. Lieb, D.M. MacAlpine, G. Micklem, F. Piano, M. Snyder, L. Stein, K.P. White, and R.H. Waterston. 2009. modENCODE Consortium. Unlocking the secrets of the genome. *Nature* 459(7249):927–930. doi:10.1038/459927a.
- D'Hont, A., F. Denoeud, J.M. Aury, F.C. Baurens, F. Carreel, O. Garsmeur, B. Noel, S. Bocs, G. Droc, M. Rouard, C. Da Silva, K. Jabbari, C. Cardi, J. Poulain, M. Souquet, K. Labadie, C. Jourda, J. Lenglé, M. Rodier-Goud, A. Alberti, M. Bernard, M. Correa, S. Ayyampalayam, M.R. McKain, J. Leebens-Mack, D. Burgess, M. Freeling, D. Mbéguié-A-Mbéguié, M. Chabannes, T. Wicker, O. Panaud, J. Barbosa, E. Hribova, P. Heslop-Harrison, R. Habas, R. Rivallan, P. Francois, C. Poirion, A. Kilian, D. Burthia, C. Jenny, F. Bakry, S. Brown, V. Guignon, G. Kema, M. Dita, C. Waalwijk, S. Joseph, A. Dievart, O. Jaillon, J. Leclercq, X. Argout, E. Lyons, A. Almeida, M. Jeridi, J. Dolezel, N. Roux, A.M. Risterucci, J. Weissenbach, M. Ruiz, J.C. Glaszmann, F. Quétier, N. Yahiaoui, and P. Wincker. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*. doi:10.1038/nature11241.
- Eddy, S.R. 2001. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* 2:919–929. doi:10.1038/35103511.
- Eichten, S.R., M.W. Vaughn, P.J. Hermanson, and N.M. Springer. 2013. Variation in DNA methylation patterns is more common among maize inbreds than among tissues. *Plant Gen.* 6, doi:10.3835/plantgenome2012.06.0009.
- El Baidouri, M., and O. Panaud. 2013. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol. Evol.* 5(5):954–965. doi:10.1093/gbe/evt025.
- ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74. doi:10.1038/nature11247.
- International Barley Genome Sequencing Consortium, K.F. Mayer, R. Waugh, J.W. Brown, A. Schulman, P. Langridge, M. Platzer, G.B. Fincher, G.J. Muehlbauer, K. Sato, T.J. Close, R.P. Wise, and N. Stein. 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491(7426):711–716. doi:10.1038/nature11543.
- Kovach, A., J.L. Wegrzyn, G. Parra, C. Holt, G.E. Bruening, C.A. Loopstra, J. Hartigan, M. Yandell, C.H. Langley, I. Korf, and D.B. Neale. 2010. The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. *BMC Genet.* 11:420. doi:10.1186/1471-2164-11-420.
- Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Sanyal, H. Kim, K. Collura, D.S. Brar, S.A. Jackson, R.A. Wing, and O. Panaud. 2006. Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genet. Res.* 16:1262–1269. doi:10.1101/gr.5290206.
- Roorkiwal, M., S.L. Sawargaonkar, A. Chitkineni, M. Thudi, R.K. Saxena, H.D. Upadhyaya, M.I. Vales, O. Riera-Lizarazu, and R.K. Varshney. 2013. Single nucleotide polymorphism genotyping for breeding and genetics applications in chickpea and pigeonpea using the BeadXpress platform. *Plant Gen.* 6, doi:10.3835/plantgenome2013.05.0017.
- Ruzicka, D., S. Chamala, F.H. Barrios-Masias, F. Martin, S. Smith, L.E. Jackson, W.B. Barbazuk, and D.P. Schachtman. 2013. Inside arbuscular mycorrhizal roots—Molecular probes to understand the symbiosis. *Plant Gen.* 6, doi:10.3835/plantgenome2012.06.0007.
- Schatz, M.C., J. Witkowski, and W.R. McCombie. 2012. Current challenges in de novo plant genome sequencing and assembly. *Genome Biol.* 13(4):243.
- Slotte, T., K.M. Hazzouri, J.A. Agren, D. Koenig, F. Maumus, Y.L. Guo, K. Steige, A.E. Platts, J.S. Escobar, L.K. Newman, W. Wang, T. Mandáková, E. Vello, L.M. Smith, S.R. Henz, J. Steffen, S. Takuno, Y. Brandvain, G. Coop, P. Andolfatto, T.T. Hu, M. Blanchette, R.M. Clark, H. Quesneville, M. Nordborg, B.S. Gaut, M.A. Lysak, J. Jenkins, J. Grimwood, J. Chapman, S. Prochnik, S. Shu, D. Rokhsar, J. Schmutz, D. Weigel, and S.I. Wright. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45(7):831–835. doi:10.1038/ng.2669.
- Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400):635–641. doi:10.1038/nature11119.
- Varshney, R.K., W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T.A. Donoghue, S. Azam, G. Fan, A.M. Whaley, A. Farmer, R. Tuetja, R.V. Penmetsa, W. Wu, H. Upadhyaya, S.-P. Yang, T. Shah, K.B. Saxena, E. Ward, T. Michael, W. R. McCombie, B. Yang, J.D.G. Jones, C. Spillane, D.R. Cook, G.D. May, X. Xu, and S.A. Jackson. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotech.* doi:10.1038/nbt.2022.
- Wicker, T., F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, and A.H. Schulman. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8(12):973–982. doi:10.1038/nrg2165.
- Wu, J., Z. Wang, Z. Shi, S. Zhang, R. Ming, S. Zhu, M.A. Khan, S. Tao, S.S. Korban, H. Wang, N.J. Chen, T. Nishio, X. Xu, L. Cong, K. Qi, X. Huang, Y. Wang, X. Zhao, J. Wu, C. Deng, C. Gou, W. Zhou, H. Yin, G. Qin, Y. Sha, Y. Tao, H. Chen, Y. Yang, Y. Song, D. Zhan, J. Wang, L. Li, M. Dai, C. Gu, Y. Wang, D. Shi, X. Wang, H. Zhang, L. Zeng, D. Zheng, C. Wang, M. Chen, G. Wang, L. Xie, V. Sovero, S. Sha, W. Huang, S. Zhang, M. Zhang, J. Sun, L. Xu, Y. Li, X. Liu, Q. Li, J. Shen, J. Wang, R.E. Paull, J.L. Bennetzen, J. Wang, and S. Zhang. 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genet. Res.* 23(2):396–408. doi:10.1101/gr.144311.112.
- Ziegler, G., A. Terauchi, A. Becker, P. Armstrong, K. Hudson, and I. Baxter. 2013. Ionomic screening of field-grown soybean identifies mutants with altered seed elemental composition. *Plant Gen.* 6, doi:10.3835/plantgenome2012.07.0012.
- Zhong, S., Z. Fei, Y.R. Chen, Y. Zheng, M. Huang, J. Vrebalov, R. McQuinn, N. Gapper, B. Liu, J. Xiang, Y. Shao, and J.J. Giovannoni. 2013. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat. Biotechnol.* 31(2):154–159. doi:10.1038/nbt.2462.