

Review

Challenges to genome sequence dissection in sweetpotato

Sachiko Isobe*, Kenta Shirasawa and Hideki Hirakawa

Kazusa DNA Research Institute, Kazusa-Kamatari 2-6-7, Kisarazu, Chiba 292-0818, Japan

The development of next generation sequencing (NGS) technologies has enabled the determination of whole genome sequences in many non-model plant species. However, genome sequencing in sweetpotato (*Ipomoea batatas* (L.) Lam) is still difficult because of the hexaploid genome structure. Previous studies suggested that a diploid wild relative, *I. trifida* (H.B.K.) Don., is the most possible ancestor of sweetpotato. Therefore, the genetic and genomic features of *I. trifida* have been studied as a potential reference for sweetpotato. Meanwhile, several research groups have begun the challenging task of directly sequencing the sweetpotato genome. In this manuscript, we review the recent results and activities of large-scale genome and transcriptome analysis related to genome sequence dissection in sweetpotato under the sections as follows: *I. trifida* genome and transcript sequencing, genome sequences of *I. nil* (Japanese morning glory), transcript sequences in sweetpotato, chloroplast sequences, transposable elements and transfer DNA. The recent international activities of *de novo* whole genome sequencing in sweetpotato are also described. The large-scale publically available genome and transcript sequence resources and the international genome sequencing streams are expected to promote the genome sequence dissection in sweetpotato.

Key Words: sweetpotato, *Ipomoea*, genome, transcript.

Introduction

The development of next generation sequencing (NGS) technologies has drastically changed plant genomic and genetic studies. High-throughput and low-cost genome sequencing technologies have enabled the determination of whole genome sequences in many non-model plant species. The number of plant species for which the whole genomes have been sequenced is increasing each year. More than 50 plant genomes were published by 2013 (Michael and Jackson 2013), and 50 other genomes were sequenced by 2015 (Michael and VanBuren 2015). Though whole genome sequencing is becoming a popular approach, only a few genome sequences have been published in polyploidy species, such as octoploid strawberry (Hirakawa *et al.* 2014) and hexaploid bread wheat (The International Wheat Genome Sequencing Consortium 2014), tetraploid oilseed rape (Chalhoub *et al.* 2014), tetraploid upland cotton (Li *et al.* 2015a) and tetraploid mustard (Yang *et al.* 2016a). The limited number of whole genome sequences for polyploidy plant species can be attributed to the difficulty of distinguishing genome sequences among homoeologous chromosomes. To overcome this limitation, the genomes of ances-

tor diploid species are often sequenced and used as a reference for polyploidy species (Bertioli *et al.* 2016, D'Hont *et al.* 2012, Ling *et al.* 2013, Shulaev *et al.* 2011, Wang *et al.* 2012), thereby reducing the complexity of the genome structure in polyploidy species.

Sweetpotato (*Ipomoea batatas* (L.) Lam) is a hexaploid species with 90 chromosomes ($2n = 6X = 90$) and a large genome size of 4.8–5.3 pg/2C nucleus (Ozias-Akins and Jarret 1994). Geographically sweetpotato is considered to have originated between the Yucatan Peninsula of Mexico and the mouth of the Orinoco River in Venezuela (Austin 1988, McDonald and Austin 1990). It is the only species cultivated in the genus *Ipomoea* series *Batatas*, and thirteen wild species are thought to be closely related to sweetpotato (Austin 1988). However, no definitive conclusions have been reached as to the evolutionary origin and genome structure of sweetpotato. Nishiyama (1971) hypothesized that sweetpotato was originated from hexaploid *I. trifida* (H.B.K.) Don., which in turn was derived from hybridization of the diploid *I. leucantha* Jacq. and tetraploid *I. littoralis* Blume. Nearly two decades later, Shiotani and Kawase (1989) generated a synthesized hexaploid *I. trifida* by hybridization of the diploid (K221, B₁B₁, $2n = 30$) and tetraploid (K222, B₂B₂B₂B₂, $2n = 60$) *I. trifida* lines, and concluded that the genome structure of sweetpotato was autohexaploid with respect to the B genome. Another hypothesis was that the genome of sweetpotato was derived from the ancient polyploidization of diploid *I. triloba* (A

Communicated by Makoto Tahara

Received November 28, 2016. Accepted January 23, 2017.

First Published Online in J-STAGE on February 24, 2017.

*Corresponding author (e-mail: sisobe@kazusa.or.jp)

genome), diploid *I. trifida* (B genome) and tetraploid *I. tabascana* (B genome; Reddy *et al.* 2007). Though several species have been nominated as candidate ancestor species, a number of cytology and molecular studies have suggested that *I. trifida* is the closest wild species to sweetpotato (Huang and Sun 2000, Jarret and Austin 1994, Komaki *et al.* 1998, Roullier *et al.* 2013, Srisuwan *et al.* 2006). Accordingly, the genetic and genomic features of *I. trifida* have been studied as a potential reference for sweetpotato. Meanwhile, several research groups have begun the challenging task of directly sequencing the sweetpotato genome. In this manuscript, we review the recent results and activities of large-scale genome and transcriptome analysis related to genome sequence dissection in sweetpotato.

I. trifida genome and transcript sequencing

An AFLP linkage map of *I. trifida* was constructed using an F₁ mapping population derived from a cross between the diploid lines 0431-1 and Mx23-4 (Nakayama *et al.* 2010). The 0431-1 map consists of 618 loci on 17 linkage groups while the Mx23-4 map comprises 163 loci on 15 linkage groups. The two maps were considered a first step in the construction of a diploid reference linkage map of sweetpotato. Interestingly, the paternal parent, Mx23-4, showed self-compatibility despite *I. trifida* was considered a self-incompatible species in most respects, and a single descendant selfed line of Mx23-4 was developed at the Kyushu Okinawa Agricultural Research Center, National Agriculture and Food Research Organization (NARO/KARC). Mx23Hm, an S₁₁ line derived from Mx23-4, was subjected to genome sequencing along with 0431-1, the maternal line of the AFLP linkage map (Table 1, Hirakawa *et al.* 2015). The high homozygosity in Mx23Hm was confirmed by polymorphic analysis with 14 simple sequence repeat (SSR) markers and analysis of Kmer frequency. Paired-end (PE) and mate-pair (MP) libraries were constructed from the genome DNAs of the two lines, and genome sequencing was carried out on an Illumina HiSeq platform. The total length of the assembled sequences of Mx23Hm, designated as ITR_r1.0, was 513 Mb and consisted of 77,400 scaffold sequences with an N50 length of 42,586 bp, while the sequence number and length of the 0431-1 assembled sequences (ITRk_r1.0) were 181,194 and 712 Mb, respectively. The larger number and total length of the assembled 0431-1 genome were attributed to the high heterozygosity of this line. The two assembled sequences were classified as “core candidates” (common to the two lines) or “line-specific,” and 240 Mb (Mx23Hm) and 353 Mb (0431-1) were classified into core candidate sequences. The numbers of putative genes identified from the assembled genomes were 62,407 (62.4 Mb) and 109,449 (87.2 Mb) in Mx23Hm and 0431-1, respectively. The total of 1,464,173 single-nucleotide polymorphisms and 16,682 copy number variations (CNVs) were also identified by comparing the two assembled genomes. Although the quality of the assembled genomes was

not high, the results make an important contribution to the progress of genomic and genetic studies of both *I. trifida* and sweetpotato. All the generated data are available on the Sweetpotato GARDEN (<http://sweetpotato-garden.kazusa.or.jp/>) database.

In addition, the transcriptome sequences of *I. trifida* were recently published by Cao *et al.* (2016). A total of 66,329,578 PE reads derived from root, leaf, stem and flower tissues were sequenced by using the Illumina platform, and *de novo* assembly generated 90,684 transcripts. The transcripts were annotated by similarity searches against the NCBI NR database (<http://www.ncbi.nlm.nih.gov/>), the gene ontology (GO) terms (<http://www.geneontology.org/>), Kyoto encyclopedia of genes and genomes (KEGG) pathways (<http://www.genome.jp/kegg/>), known transcription factors and protein kinases. The obtained sequences were used for demonstration of the SSR marker design and cloning of a potential drought-tolerance gene, ItWRKY1.

Genome sequences of *I. nil* (Japanese morning glory)

I. nil is another possible reference species of sweetpotato, it has been used as a model plant in genetics because of its large number of mutant lines. The draft genome sequences covering 98% of the 750 Mb genome were recently published by Hoshino *et al.* (2016) with a scaffold N50 of 2.88 Mb. Of the assembled genomes, 91.4% were anchored to 15 pseudo-chromosomes. These are the first pseudo-chromosomes constructed for the genus *Ipomoea*. Gene prediction based on the transcript sequences of the leaves, flowers, embryos, stems, roots and seed coats generated a total of 42,783 gene models. Phylogenetic analysis with a 1,353 single copy gene estimated that the divergence of *I. nil* from the other Solanales members (Solanaceae) occurred approximately 75.25 Myr ago. Though the genetic distance of *I. nil* to sweetpotato is longer than that of *I. trifida*, the high quality draft genome sequences would be a useful reference source for genetic and genomic sweetpotato analysis.

Transcript sequences in sweetpotato

The large-scale cDNA sequences in sweetpotato were first reported by You *et al.* (2003) for the identification of genes related to the initiation of storage root development. A total of 2,859 cDNA clones derived from early stage storage roots were assembled into 483 clusters and 442 singletons. By the year 2010, approximately 22,000 expressed sequence tag (EST) sequences had been registered (Schafleitner *et al.* 2010).

First transcript sequencing using the NGS platform (454 pyrosequencing) was reported in 2010 (Schafleitner *et al.*) for designing gene-based microsatellite markers. The 524,209 transcript sequence reads derived from cDNA collections of stems and leaves from drought-stressed sweetpotato were

Table 1. Summary of Large-scale genome and transcriptome sequence resources in sweetpotato and *I. trifida*

Species	Type	Cultivar or Line	DNA/RNA Extracted organ	Assembled Sequences				Sequence platform	Assembler	Reference
				Number	Total length (bp)	Average length (bp)	N50 (bp)			
<i>I. trifida</i>	Genome	Mx23Hm	Young leaves	77,400	512,990,885	6,628	42,586	Illumina HiSeq2000	SOAPdenovo2 r223 GapCloser 1.10 SSPACE2.0	Hirakawa <i>et al.</i> (2015)
	Genome	0431-1	Young leaves	181,194	712,155,587	3,930	36,283	Illumina HiSeq2000	SOAPdenovo2 r223 GapCloser 1.10 SSPACE2.0	Hirakawa <i>et al.</i> (2015)
	Transcript	DLP4597	root, leaf, stem and flower	90,684	–	–	1,784	Illumina HiSeq2000	Trinity	Cao <i>et al.</i> (2016)
Sweetpotato (<i>I. batatas</i> (L.) Lam)	ESTs	Jinhongmi	young storage roots	2,859	–	–	–	Shimadzu RISA-384	–	You <i>et al.</i> (2003)
	Transcript	Tanzania	Shoots	66418 (31,685 contigs + 34,733 singletons)	25,048,392 (contigs)	790 (contigs)	–	Roche 454 FLX	NGen	Schafleitner <i>et al.</i> (2010)
	Transcript	Jingshu 6	swelling tuberous roots	473,238 (contigs)	65,542,302 (contigs)	138 (contigs)	118 (contigs)	TITANIUM	SOAPdenovo	Xie <i>et al.</i> (2012)
	Transcript	Xushu 18	young leaves, mature leaves, stems, fibrous roots, initial tuberous roots, expanding tuberous roots, harvest tuberous roots	121,084 (Scaffolds)	36,640,480 (Scaffolds)	303 (Scaffolds)	393 (Scaffolds)	Illumina HiSeq2000	Velvet v1.0.12 SOAPdenovo v1.04 CAP3	Tao <i>et al.</i> (2012)
	Transcript	Xushu 18	whole opened flowers and previously published transcripts in leaves, stems and roots	58,277 (Transcripts)	27,970,067 (Unigenes)	476 (Unigenes)	552 (Unigenes)	Illumina GA II		
Weiduoli HVB-3	Transcript	Xushu 18	storage roots	128,052	41,13M	321	509			
	Transcript	Xushu 18	Young leaves	70,412	–	628	895	Illumina HiSeq2000	SOAPdenovo v1.3 Oases v0.1.20 CAP3	Tao <i>et al.</i> (2013)
	Transcript	Xushu 18	Young leaves	1,557,001 (contigs)	91,371,759 (contigs)	58 (contigs)	58 (contigs)	Illumina HiSeq2000	Trinity, Inch-worm, Chrysalis, Butterfly	Li <i>et al.</i> (2015b)
Chloroplast	Transcript	Xushu 18	Young leaves	35,909 (Unigenes)	19,150,802 (Unigenes)	533 (Unigenes)	669 (Unigenes)	Illumina HiSeq2000 and GAI	Edena v2.1.1 SOAPdenovo2 r240	Yan <i>et al.</i> (2015)
	Transcript	Xushu 18	Young leaves	1	161,303	–	–		Velvet v1.0.12	
BAC-end sequences	Transcript	Xu 781	Young leaves	11,542	7,595,261	658	–	ABI PRISM 3730 DNA Analyzer	–	Si <i>et al.</i> (2016)
	Transcript	Xu 781	Young leaves	11,542	7,595,261	658	–			

assembled with 22,094 published ESTs, and generated 31,685 sets of overlapping DNA segments and 34,733 unassembled sequences. A total of 24,657 putatively unique genes were annotated by BLASTX search with the UniRef100 database (<http://www.uniprot.org>), and 1,661 gene-based microsatellite sequences were identified in the unique genes. The transcript and genome sequences were also used for SNP identification (Meng *et al.* 2015, Xu *et al.* 2015) and its application to tetra-primer Amplification Refractory Mutation System (ARMS)-PCR to identify SNP alleles on agarose gels.

Large-scale *de novo* transcript sequencing was also performed to identify putative genes and for differential gene expression analysis. *De novo* transcriptome sequencing was performed by using RNA extracted from tuberous roots of Jingshu 6, a purple sweetpotato variety (Xie *et al.* 2012). Of the 58,800 obtained unigenes, 40,280 were identified as protein-coding genes. Based on GO and KEGG analysis, at least 3,553 genes were considered to be involved in the biosynthesis pathways of starch, alkaloids, anthocyanin pigments, and vitamins. In addition, a total of 851 SSR were identified in the unigenes.

The first digital gene expression (DGE) analysis in sweetpotato was reported by Tao *et al.* (2012) for seven tissue samples of young leaves, mature leaves, stems, fibrous roots, initial tuberous roots, expanding tuberous roots and harvest tuberous roots by using the Illumina GAI platform. A total of 128,052 transcripts (≥ 100 bp) were subjected to annotation by Blast2GO (<https://www.blast2go.com>), BLASTX (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), GO and KEGG analysis. The research group was also performed for the transcript sequences obtained from the seven tissues in order to clarify the tissue-specific gene expression. Furthermore, Tao *et al.* (2013) performed RNA-Seq by using an Illumina HiSeq 2000 platform for whole opened flowers of sweetpotato, cv. Xushu 18, to identify putative floral-specific and flowering regulatory-related genes. A total of 2,595 and 2,928 putative floral-specific and vegetative-specific transcript sequences, respectively, were obtained and transcripts similar to the key genes in the flowering regulation network of *Arabidopsis thaliana* were identified.

In a later work, Li *et al.* (2015a) performed transcriptome sequencing of an orange-fleshed sweetpotato cultivar, Weiduoli, and its mutant, HVB-3, for application to differentially expressed gene analysis (DEG) related to carotenoid biosynthesis. A total of 58,277 transcripts and 35,909 unigenes were assembled from the Illumina RNA-Seq reads of storage roots. Between the two lines, 874 DEGs were obtained and 22 DEGs and 31 transcription factors were considered to be involved in carotenoid biosynthesis.

Chloroplast sequences

The complete nucleotide sequence of the chloroplast (cp) genome of sweetpotato was reported by Yan *et al.* (2015). A circular molecule of 161,303 bp in length was constructed

as a quadripartite structure with large and small single-copy regions. A total of 145 putative genes were identified, including 94 protein-encoding genes. By comparing the chloroplast sequences of 33 species, including *I. nil* and *I. trifida*, gene-flow events and gene-gain-and-loss events were identified at the intra- and inter-species levels. RNA-editing events and differential expressions of the chloroplast functional genes were also identified by DEG analysis.

Transposal elements and transfer DNA

Transposable elements (TEs) affect genetic diversity through the replications and movements in a genome. An active element of the Ty1-copia retrotransposon family was identified in the sweetpotato genome by Tahara *et al.* (2004). Monden *et al.* (2014) further developed screening methods for long terminal repeat (LTR) retrotransposons that show high insertion polymorphisms in strawberry. By using this approach, Monden *et al.* (2015) identified a large number of *Rtsp-1* retrotransposon insertion sites, and constructed a linkage map by using retrotransposon insertion polymorphisms. Meanwhile, Yan *et al.* (2014) performed a large-scale TE identification by *de novo* assembly of four published sweetpotato transcriptome databases. A total of 1,405 TEs were identified, including 883 retrotransposons and 552 DNA transposons. Illumina DGE profiling of seven tissues of Xushu 18 revealed that 107 TEs were expressed in all seven tissues, while 417 TEs were expressed in one or more tissues.

The report of Kyndt *et al.* (2015) created a sensation in the realm of sweetpotato research. While assembling small interfering RNAs, Kyndt *et al.* discovered that the sweetpotato genome contained *Agrobacterium* transfer DNAs (T-DNAs) with expressed genes and that sweetpotato was a naturally transgenic crop. They performed simple and quantitative PCR, Southern blotting, genome walking and bacterial artificial chromosome (BAC) library screening and sequencing, and revealed that two T-DNA regions were expressed in different tissues of sweetpotato. One of the T-DNA regions was present in 291 tested sweetpotato accessions, but not in *I. tabascani*, *I. trifida* or *I. triloba*. Therefore it was considered that *Agrobacterium* infection occurred in evolutionary times and the T-DNA provided traits that were selected for during domestication.

Challenges to genome sequence dissection in sweetpotato

BAC-end sequences (BESs) were just recently published by Si *et al.* (2016). Both ends of 8,310 BAC clones randomly selected from the 240,384 clones were sequenced by the Sanger method (ABI PRISM 3730 DNA Analyzer). The total length obtained was 7,595,261 bp, average length of 658 bp. Based on the analysis of BESs, the sweetpotato genome was estimated as consisting of 10.0% of coding regions, 18.3% of sweetpotato-unique repetitive DNA and

12.17% of known repetitive DNA, including 7.37% LTR retrotransposons, 1.15% Non-LTR retrotransposons and 1.42% Class II DNA transposons etc.

The genome sequencing of sweetpotato was recently reported in a non-peer review journal (Yang *et al.* 2016b). The authors of that report proposed construction of haplotype-resolved genome sequences by performing SNP phasing. A total of ~824 Mb assembly was generated from at least 40-fold monoploid genome coverage obtained by Illumina HiSeq 2500 and Roche GS FLX+ platforms. Though the total length of the assembled sequences was less than 30% of the total length of the haplotype genome (approximately 3 Gb), the possibility of such assembly is worth considering.

To our knowledge, further two international research groups have tried to develop genomic resources for sweetpotato (Fei 2016, Yoon *et al.* 2015). One is the Trilateral Research Association of Sweetpotato (TRAS) genome sequencing consortium. The consortium was launched in 2012 and consists of six organizations, the Jiangsu Xuzhou Sweetpotato Research Center (China), China Agricultural University (China), Rural Development Administration (Korea), Korea Research Institute of Bioscience and Biotechnology (Korea), National Agriculture and Food Research Organization (Japan) and Kazusa DNA Research Institute (Japan). The consortium members agreed on the genome sequencing of sweetpotato cultivar Xushu 18 and performed *de novo* whole genome sequencing, transcript analysis, and linkage map construction by using Illumina and PacBio sequences.

The other research group is The Genomic Tools for Sweetpotato (GT4SP) Improvement Project for Sub-Saharan Africa (SSA). This project was founded by the Bill & Melinda Gates Foundation with a funding amount of more than 12M USD (<http://www.sweetpotatoknowledge.org/project/genomic-tools-for-sweetpotato-improvement-gt4sp/>). However, it is not the aim of this project to sequence the sweetpotato genome. Rather, the project seeks to conduct high quality *I. trifida* genome sequencing and to develop a genome sequence-based marker platform for sweetpotato improvement. Seven organizations are involved in the project: the International Potato Center (CIP), two universities in the USA, an institute in the USA, an institute in Australia and a research organization and institute in the SSA.

Polyploidy is commonly observed in plant species and can be advantageous (Comai 2005). Despite the importance the genome assembly of polyploid species has been left behind because of its complexity. Meanwhile, novel approaches for genome assembly applicable in polyploidy species were recently proposed (Aguilar and Istrail 2013, Ming and Man Wai 2015). The key point of the approaches is distinction of haploid genomes by using local haplotypes. Together with the development of bioinformatics approaches, the large-scale publically available genome and transcript sequence resources and the international genome sequencing streams are expected to promote the genome sequence dissection in sweetpotato.

Literature Cited

- Aguilar, D. and S. Istrail (2013) Haplotype assembly in polyploid genomes and identical by descent shared tracts. *Bioinformatics* 29: i352–360.
- Austin, D.F. (1988) The taxonomy, evolution and genetic diversity of sweetpotato and related wild species. *In: Exploration, Maintenance and Utilization of sweet potato Genetic Resources International Potato Center (CIP), Lima*, pp. 27–60.
- Bertioli, D.J., S.B. Cannon, L. Froenicke, G. Huang, A.D. Farmer, E.K. Cannon, X. Liu, D. Gao, J. Clevenger and P. Izua-Akins (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48: 438–446.
- Cao, Q., A. Li, J. Chen, Y. Sun, J. Tang, A. Zhang, Z. Zhou, D. Zhao, D. Ma and S. Gao (2016) Transcriptome sequencing of the sweet potato progenitor (*Ipomoea trifida* (H.B.K.) G. Don.) and discovery of drought tolerance genes. *Trop. Plant Biol.* 9: 63–72.
- Chalhoub, B., F. Denoeud, S. Liu, I.A. Parkin, H. Tang, X. Wang, J. Chiquet, H. Belcram, C. Tong, B. Samans *et al.* (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345: 950–953.
- Comai, L. (2005) The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6: 836–846.
- D'Hont, A., F. Denoeud, J.M. Aury, F.C. Baurens, F. Carreel, O. Garsmeur, B. Noel, A. Bocs, G. Droc, P. Wincker *et al.* (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488: 213–217.
- Fei, Z. (2016) Genome sequencing and annotation of sweetpotato wild progenitors. *Proceeding of PAG XXIV W897*.
- Hirakawa, H., K. Shirasawa, S. Kosugi, K. Tashiro, S. Nakayama, M. Yamada, M. Kohara, A. Watanabe, Y. Kishida, T. Fujishiro *et al.* (2014) Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Res.* 21: 169–181.
- Hirakawa, H., Y. Okada, H. Tabuchi, K. Shirasawa, A. Watanabe, H. Tsuruoka, C. Minami, S. Nakayama, S. Sakamoto, M. Kohara *et al.* (2015) Survey of genome sequences in a wild sweet potato, *Ipomoea trifida* (H. B. K.) G. Don. *DNA Res.* 22: 171–179.
- Hoshino, A., V. Jayakumar, E. Nitasaka, A. Toyoda, H. Noguchi, T. Itoh, T. Shin-I, Y. Minakuchi, Y. Koda, A.J. Nagano *et al.* (2016) Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. *Nat. Commun.* 7: 13295.
- Huang, J.C. and M. Sun (2000) Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series Batatas (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theor. Appl. Genet.* 100: 1050–1060.
- Jarret, R.L. and D.F. Austin (1994) Genetic diversity and systematic relationship in sweetpotato (*Ipomoea batatas* (L.) Lam.) and related species as revealed by RAPD analysis. *Genet. Resour. Crop Evol.* 41: 165–173.
- Komaki, K., H.N. Regmi, K. Katayama and S. Tamiya (1998) Morphological and RAPD pattern variations in sweetpotato and its closely related species. *Breed. Sci.* 48: 281–286.
- Kyndt, T., D. Quispe, H. Zhai, R. Jarret, M. Ghislain, Q. Liu, G. Gheysen and J.F. Kreuze (2015) The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: An example of a naturally transgenic food crop. *Proc. Natl. Acad. Sci. USA* 112: 5844–5849.
- Li, F., G. Fan, C. Lu, G. Xiao, C. Zou, R.J. Kohel, Z. Ma, H. Shang,

- X. Ma, J. Wu *et al.* (2015a) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol.* 33: 524–530.
- Li, R., H. Zhai, C. Kang, D. Liu, S. He and Q. Liu (2015b) *De Novo* transcriptome sequencing of the orange-fleshed sweet potato and analysis of differentially expressed genes related to carotenoid biosynthesis. *Int. J. Genomics* 2015: 843802.
- Ling, H.Q., S. Zhao, D. Liu, J. Wang, H. Sun, C. Zhang, H. Fan, D. Li, L. Dong, J. Tao *et al.* (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496: 87–90.
- McDonald, J.A. and D.F. Austin (1990) Changes and additions in *Ipomoea* section *Batatas* (Convolvulaceae). *Brittonia* 42: 116–120.
- Meng, K., X.U. Jia-lei, L. Qiang, L. Ya-ju, W. Xin, T. Wei, Y. Hui, Z. Yun-gang and D.F. Ma (2015) Development of SNP markers using RNA-Seq technology and tetraprimer ARMS-PCR in sweetpotato. *J. Integr. Agric.* 15: 60345–60347.
- Michael, T.P. and S. Jackson (2013) The first 50 plant genomes. *Plant Genome* 6: 1–7.
- Michael, T.P. and R. VanBuren (2015) Progress, challenges and the future of crop genomes. *Curr. Opin. Plant Biol.* 24: 71–81.
- Ming, R. and C. Man Wai (2015) Assembling allopolyploid genomes: no longer formidable. *Genome Biol.* 16: 27.
- Monden, Y., N. Fujii, K. Yamaguchi, K. Ikeo, Y. Nakazawa, T. Waki, K. Hirashima, Y. Uchimura and M. Tahara (2014) Efficient screening of long terminal repeat retrotransposons that show high insertion polymorphism via high-throughput sequencing of the primer binding site. *Genome* 57: 245–252.
- Monden, Y., T. Hara, Y. Okada, O. Jahana, A. Kobayashi, H. Tabuchi, S. Onaga and M. Tahara (2015) Construction of a linkage map based on retrotransposon insertion polymorphisms in sweetpotato via high-throughput sequencing. *Breed. Sci.* 65: 145–153.
- Nakayama, H., M. Tanaka and Y. Takahata (2010) An AFLP-based genetic linkage map of *Ipomoea trifida* (H.B.K.) G. Don., a diploid relative of sweetpotato, *I. batatas* (L.) Lam. *Trop. Afr. Develop.* 54: 9–16.
- Nishiyama, I. (1971) Evolution and domestication of the sweet potato. *Bot. Maj.* Tokyo 84: 377–387.
- Ozias-Akins, P. and R.L. Jarret (1994) Nuclear DNA content and ploidy levels in the genus *Ipomoea*. *J. Am. Soc. Hortic. Sci.* 119: 110–115.
- Reddy, U.K., G.T. Bates, J. Ryan-Bohac and P. Nimmakayala (2007) Sweetpotato. In: Kole, C. (ed.) *Genome mapping and molecular breeding in plants*, Springer-Verlag, Berlin, pp. 237–247.
- Roullier, C., A. Duputié, P. Wennekes, L. Benoit, V.M. Fernández Bringas, G. Rossel, D. Tay, D. McKey and V. Lebot (2013) Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLoS ONE* 8: e62707.
- Schaffleitner, R., L.R. Tincopa, O. Palomino, G. Rossel, R.F. Robles, R. Alagon, C. Rivera, C. Quispe L. Rojas, J.A. Pacheco *et al.* (2010) A sweetpotato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. *BMC Genomics* 11: 604.
- Shiotani, I. and T. Kawase (1989) Genomic structure of the sweet potato and hexaploids in *Ipomoea trifida* (H.B.K.) DON. *Japan. J. Breed.* 39: 57–66.
- Shulaev, V., D.J. Sargent, R.N. Crowhurst, T.C. Mockler, O. Folkerts, A.L. Delcher, P. Jaiswal, K. Mockaitis, A. Liston, S.P. Mane *et al.* (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43: 109–116.
- Si, Z., B. Du, J. Huo, S. He, Q. Liu and H. Zhai (2016) A genome-wide BAC-end sequence survey provides first insights into sweetpotato (*Ipomoea batatas* (L.) Lam.) genome composition. *BMC Genomics* 17: 945.
- Srisuwan, S., D. Sihachakr and S. Siljak-Yakovlev (2006) The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. *Plant Sci.* 171: 424–433.
- Tahara, M., T. Aoki, S. Suzuka, H. Yamashita, M. Tanaka, S. Matsunaga and S. Kokumai (2004) Isolation of an active element from a high-copy-number family of retrotransposons in the sweetpotato genome. *Mol. Genet. Genomics* 272: 116–127.
- Tao, X., Y.H. Gu, H.Y. Wang, W. Zheng, X. Li, C.W. Zhao and Y.Z. Zhang (2012) Digital gene expression analysis based on integrated *de novo* transcriptome assembly of sweet potato [*Ipomoea batatas* (L.) Lam.]. *PLoS ONE* 7: e36234.
- Tao, X., Y.H. Gu, Y.S. Jiang, Y.Z. Zhang and H.Y. Wang (2013) Transcriptome analysis to identify putative floral-specific genes and flowering regulatory-related genes of sweet potato. *Biosci. Biotechnol. Biochem.* 77: 2169–2174.
- The International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345: 1251788.
- Wang, K., Z. Wang, F. Li, W. Ye, J. Wang, G. Song, Z. Yue, L. Cong, H. Shang, S. Zhu *et al.* (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* 44: 1098–1103.
- Xie, F., C.E. Burklew, Y. Yang, M. Liu, P. Xiao, B. Zhang and D. Qiu (2012) *De novo* sequencing and a comprehensive analysis of purple sweet potato (*Ipomoea batatas* L.) transcriptome. *Planta* 236: 101–113.
- Xu, J.L., Q. Li, M. Hou, Y.J. Liu, X. Wang, W. Tang, H. Yan, Y.G. Zhang and D.F. Ma (2015) A rapid and efficient method for detecting sweetpotato (*Ipomoea batatas* L.) SNP markers. *Mol. Plant Breed.* 13: 891–897.
- Yan, L., Y.H. Gu, X. Tao, X.J. Lai, Y.Z. Zhang, X.M. Tan and H. Wang (2014) Scanning of transposable elements and analyzing expression of transposase genes of sweet potato [*Ipomoea batatas*]. *PLoS ONE* 9: e90895.
- Yan, L., X. Lai, X. Li, C. Wei, X. Tan and Y. Zhang (2015) Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata*]. *PLoS ONE* 10: e0124083.
- Yang, J., D. Liu, X. Wang, C. Ji, F. Cheng, B. Liu, Z. Hu, S. Chen, D. Pental and M. Zhang (2016a) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* 48: 1225–1232.
- Yang, J., M.H. Moeinzadeh, H. Kuhl, J. Helmuth, P. Xiao, G. Liu, J. Zheng, Z. Sun, W. Fan, G. Deng *et al.* (2016b) The haplotype-resolved genome sequence of hexaploid *Ipomoea batatas* reveals its evolutionary history. *bioRxiv*. doi: 10.1101/064428.
- Yoon, U.H., J.C. Jeong, S.S. Kwak, J.W. Yang, T.H. Kim, H.U. Lee, S.S. Nam and J.H. Hahn (2015) Current status of sweetpotato genomics research. *J. Plant Biotechnol.* 42: 161–167.
- You, M.K., C.G. Hur, Y.S. Ahn, M.C. Suh, B.C. Jeong, J.S. Shin and J.M. Bae (2003) Identification of genes possibly related to storage root induction in sweetpotato. *FEBS Lett.* 536: 101–105.