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. 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.
is the closest wild species to sweetpotato, 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 *S*1 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
| 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. trifida | Genome    | Mx23Hm          | Young leaves            | 77,400              | 512,990,885      | 6,628               | 42,586                 | Illumina HiSeq2000     |
|         | Genome     | 0431-1          | Young leaves            | 181,194             | 712,155,587      | 3,930               | 36,283                 | Illumina HiSeq2000     |
|         | Transcript | DLP4597         | root, leaf, stem and flower | 90,684              | –                | –                   | 1,784                  | Illumina HiSeq2000     |
| Sweetpotato (I. batatas (L.) Lam) ESTs | Jinhongmi   | young storage roots | 2,859                  | –                  | –                 | –                   |                        |
| Transcript | Tanzania  | Shoots           |                         | 6641.8 (31,685 contigs + 34,733 singletons) | 25,048,392 (contigs) | 790 (contigs)       | –                      | Shimadzu RISA-384      |
| Transcript | Jingshu 6 | swelling tuberous roots | 473,238 (contigs)     | 65,542,302 (contigs) | 138 (contigs)    | 118 (contigs)       | –                      | Roche 454 FLX TITANUM  |
| Transcript | Xushu 18 | young leaves, mature leaves, stems, fibrous roots, initial tuberous roots, expanding tuberous roots, harvest tuberous roots whole opened flowers and previously published transcripts in leaves, stems and roots | 128,052 | 41.13M | 321 | 509 | | |
| Transcript | Xushu 18 | whole opened flowers and previously published transcripts in leaves, stems and roots | 70,412 | – | 628 | 895 | | |
| Transcript | Weiduoli HVB-3 | storage roots | 1,557,001 (contigs) | 58,277 (Transcripts) | 161,303 | – | – | Trinity, Inchworm, Chrysalis, Butterfly |
| Chloroplast | Xushu 18 | Young leaves | 1,571,759 (contigs) | 34,741,399 (Transcripts) | 19,150,802 (Unigenes) | – | – | Edena v2.1.1 |
| BAC-end sequences | Xu 781 | Young leaves | 11,542 | 7,595,261 | 658 | – | – | ABI PRISM 3730 DNA Analyzer |

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

Transposable 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 Rsp-I 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. tabascana, 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 of polyploidy, 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.

Knykt, T., D. Quispe, H. Zhai, R. Jarret, M. Ghislain, Q. Liu, G. Ghesyen 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 tetramer primer ARMS-PCR in sweet-potato. 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 allopolyloid genomes: no longer formidable. Genome Biol. 16: 27.

Monden, Y., N. Fujii, K. Yamaguchi, K. Ikeo, Y. Nakazawa, 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. Mag. 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 (2015) Identification of genes possibly related to storage root induction in sweetpotato. FEBS Lett. 536: 101–105.

Srivastava, S., D. Shachak 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.

Shahzad, S., A. Kazi and J. Wang (2013) Dis-entangling the origins of cultivated sweet potato (Ipomoea batatas L.) and its wild relatives through the cytogenetic approaches. Plant Sci. 171: 424–433.

Si, Z., B. Du, J. Hua, 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. Sihasak and S. Sokumni (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.

Yang, J., M.K. You, M.C. Suh, B.C. Jeong, J.S. Shin and 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.

Zhang, Z. Yun-gang and D.F. Ma (2015) A genome-wide BAC library of sweet potato [Ipomoea batatas]. PLoS ONE 10: e0124083.

Zhang, X. et al. (2011) The genome of woodland strawberry (Fragaria vesca). Nat. Genet. 43: 109–116.

Zhang, Y., D. Liu, X. Wang, C. Ji, F. Cheng, B. Liu, Z. Hu, S. Chen, D. Pental and M. Zhang (2016a) The genome sequence of allopolyloid Brassica juncea and analysis of differential homoeolog gene expression influencing selection. Nat. Genet. 48: 1225–1232.

Zhang, 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.

Zhang, Z. Yun-gang and D.F. Ma (2015) A rapid and efficient method for detecting sweetpotato (Ipomoea batatas L.) SNP markers. Mol. Plant Breed. 13: 891–897.