

Trilateral Research Association of Sweetpotato (TRAS) *Ipomoea trifida* and *I. batatas* Sequencing and Crop Improvement Efforts

Sachiko Isobe, Ung-Han Yoon, Qinghe Cao,
Sang-Soo Kwak, Masaru Tanaka, Daifu Ma,
and Qingchang Liu

Abstract

East Asia is an important region of sweetpotato production and consumption. To promote exchange among scientists studying sweetpotato in East Asia, the Trilateral Research Association of Sweetpotato (TRAS) was established in 2004 by sweetpotato scientists from China, South Korea, and Japan. The TRAS genome sequencing consortium was formally launched in 2014 and established a haploid-resolved and chromosome-scale de novo assembly of autohexaploid sweetpotato genome sequences. Before constructing

the genome, we created chromosome-scale genome sequences in *Ipomoea trifida* using a highly homozygous accession, ‘Mx23Hm’, with PacBio RSII and Hi-C reads. Haploid-resolved genome assembly was performed for the sweetpotato (*I. batatas*) cultivar ‘Xushu 18’ by hybrid assembly with Illumina paired-end (PE) and mate-pair (MP) reads, 10X genomics reads, and PacBio RSII reads. Then, 90 chromosome-scale pseudomolecules were generated by aligning the scaffolds onto a sweetpotato linkage map. In total, 34,386 and 175,633 genes were identified on the assembled nucleic genomes of

S. Isobe (✉)

Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu 292-0818, Chiba, Japan
e-mail: sisobe@g.ecc.u-tokyo.ac.jp

U.-H. Yoon

Genomics Division, National Institute of Agricultural Sciences, RDA, 370 Nongsaengmyeong-Ro, Deokjin-Gu, Jeonju 54874, Korea

Q. Cao · D. Ma

Xuzhou Institute of Agricultural Sciences in Jiangsu Xuhuai District, Sweetpotato Research Institute, China Agricultural Academy of Sciences, Kunpeng Road 2, Jiawang District, Xuzhou 221131, China

S.-S. Kwak

Plant Systems Engineering Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-Ro, Yuseong-Gu, Daejeon 34141, Korea

M. Tanaka

Kyushu Okinawa Agricultural Research Center, National Agriculture and Food Research Organization, Yokoichicho 6651-2, Miyakonojo 885-0091, Miyazaki, Japan

Q. Liu

Key Laboratory of Sweetpotato Biology and Biotechnology, Ministry of Agriculture and Rural Affairs, College of Agronomy and Biotechnology, China Agricultural University, Beijing 100193, China

S. Isobe

Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

I. trifida and sweetpotato, respectively. The assembled genome sequences have been used for genetic and RNA-Seq analysis for agronomically important traits. The assembled genome sequences are expected to continue to contribute to genetic and genomic analysis and promote sweetpotato breeding.

Keywords

Sweetpotato · *I. trifida* · Genome · Assembly · TRAS

3.1 Introduction

Sweetpotato (*Ipomoea batatas* (L.) Lam) is widely cultivated and consumed worldwide, with a global production of 86.4 million tons in 2022 (FAO STAT). China is the leading producer, contributing 54% to the world's total production. Sweetpotato is also a popular crop in neighboring countries such as Japan and South Korea, and research on the breeding and cultivation of sweetpotato has been actively conducted in the region. To promote exchange among scientists studying sweetpotato in East Asia, the Trilateral Research Association of Sweetpotato (TRAS) was established in 2004 by sweetpotato scientists from China, South Korea, and Japan.

The inaugural symposium took place in Mokpo, South Korea, and subsequent symposiums have been held approximately every two years, rotating among the three countries. Nine international symposiums have been held to date, with the most recent one taking place in September 2022 in Xuzhou, China.

At the 5th International Sweetpotato Symposium held on Jeju Island, South Korea in 2012, agreement was reached among the three countries to undertake the construction of a reference genome for sweetpotato. After subcommittee meetings in Tokyo and Jeju in 2013, the TRAS genome sequencing consortium was formally launched in Beijing, 2014. The consortium consists of six organizations: the Jiangsu Xuzhou Sweetpotato Research Center, CAAS (China), China Agricultural University (China), Rural Development Administration (Korea), Korea Research Institute of Bioscience and Biotechnology (Korea), Institute of Sweetpotato Research, National Agriculture and Food Research Organization (Japan), and Kazusa DNA Research Institute (Japan). The composition and roles of the consortium are shown in Table 3.1.

Sweetpotato 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). When de novo assembly is performed in polyploid species, it is common to advance the analysis by referencing

Table 3.1 Organizational overview and role of the TRAS genome sequencing consortium

Organization	Country	Role
The Jiangsu Xuzhou Sweetpotato Research Center, CAAS	China	Development of a S1 ‘Xushu 18’ S1 mapping population, Illumina genome sequence collection, DenovoMAGIC assembly
College of Agronomy and Biotechnology, China Agricultural University	China	‘Xushu 18’ preparation, Illumina genome sequence collection, DenovoMAGIC assembly
National Institute of Agricultural Sciences, RDA	Korea	PacBio and Illumina sequence collection, de novo assembly of PacBio and Illumina reads, gene prediction and annotation, organelle genome assembly
Biological Resource Center, KRIBB	Korea	Illumina genome sequence collection, DenovoMAGIC assembly
Kyushu Okinawa Agricultural Research Center, NARO	Japan	Development of a S1 ‘Xushu 18’ S1 mapping population, <i>I. trifida</i> material preparation
Kazusa DNA Research Institute	Japan	Illumina genome sequence collection, <i>I. trifida</i> genome assembly, ‘Xushu 18’ S1 preparation of mapping population, DenovoMAGIC assembly, comparative analysis

the genome of closely related diploid species (Kyriakidou et al. 2018). Sweetpotato is the only species in the genus *Ipomoea* that is cultivated as a crop; among the genus's wild species, thirteen are thought to be closely related to sweetpotato (Austin 1988). Although no definitive conclusions have been reached as to the evolutionary origin and genome structure of sweetpotato, *I. trifida* (H.B.K.) Don. has been considered a likely diploid progenitor of sweetpotato (Nishiyama 1971).

In 2012, when genome sequence analysis was first proposed as an appropriate project for TRAS, the genome sequences of diploid species of *Ipomoea* had not yet been published. Therefore, the consortium decided to conduct genome analysis, not only for the hexaploid sweetpotato but also for related diploid species. For genome assembly and transcriptome analysis in *I. batatas*, we used the Chinese variety 'Xushu 18', which is a leading variety in China, bred at Xuzhou Institute of Agricultural Sciences in Jiangsu Xuhuai District and released in 1977.

3.2 Genome Assembly of *I. trifida* 'Mx23Hm'

Whole-genome sequencing and assembly was first performed for two *I. trifida* lines, a selfed line, 'Mx23Hm', and a heterozygous line, '0431-1' (Hirakawa et al. 2015). The whole-genome de novo assembly was conducted using Illumina paired-end (PE) and mate-pair (MP)

libraries. The assembled genome was initially employed for genetic analysis, such as SNP detection, serving as the first reference genome for *I. trifida*. However, due to the assembly being based solely on short reads, the scaffolds exhibited fragmentation, and connectivity at the chromosomal scale was lacking.

In order to obtain chromosome-scale scaffold sequences, the RDA group obtained a total length of 64.26 Gb PacBio subreads from 'Mx23Hm' and conducted whole genome de novo assembly. De novo assembly was conducted with subreads using the SMRTMAKE assembly pipeline (Chin et al. 2013), and a total of 2881 contigs were generated with a total length of 495.7 Mb. The 2881 contigs were polished with Illumina reads, and chromosome-scale scaffolding was then performed by HiRise (Putnam et al. 2016) with 471 M Hi-C reads. The 15 chromosome-scale scaffolds and the chr0 sequences were designated as Itr_r2.2 (Table 3.2). The total length of Itr_r2.2 was 502.2 Mb, including total lengths of 460.77 Mb for 15 pseudomolecules and 41.47 Mb for the chr0 scaffold. Itr_r2.2 covered 97.4% of the 'Mx23Hm' genome, when the genome size was considered to be 515.8 Mb (Hirakawa et al. 2015), while the cover ratio of the 15 chromosome-scale scaffolds was 89.3%. The ratio of complete BUSCOs was 98.5%, including 93.4% of single-copy genes and 5.1% of duplicated genes (Simão et al. 2015). The ratios of fragmented and missing BUSCOs were 0.8% and 0.7%, respectively. A total of 34,386

Table 3.2 Statistics on the assembled *I. trifida* 'Mx23Hm' (Itr_r2.2)

	Chr01–15 + Chr0	Chr01–15	Gene
Number of sequences	16	15	34,386
Total length of sequences (bp)	502,237,654	460,770,816	36,202,112
N50 length (bp)	31,779,616	31,779,616	1455
GC%	37.2	36.5	46.4
N%	1.3	0.18	0.00
BUSCO v5.2.2 embryophyta_odb10, Number of BUSCOs = 1614			
Complete	98.5	98.5	82.6
Complete single	93.4	93.5	78.5
Complete double	5.1	5	4.1
Fragmented	0.8	0.8	5.1
Missing	0.7	0.7	12.3

gene sequences were predicted on the Itr_r2.2 genome based on ab initio and evidence-based gene models.

3.3 Genome Assembly of *I. batatas* ‘Xushu18’

When the TRAS genome sequencing consortium started the whole-genome de novo assembly of *I. batatas* ‘Xushu 18’ in 2012, long-read sequencing was expensive, and its utilization in genome assembly was not realistic. Consequently, our approach involved the use of Illumina short reads for sequencing, and the PE and MP sequences shown in Table 3.3 were obtained.

The genome size of ‘Xushu 18’ was estimated as 2.6 Gb on the basis of the distribution of distinct k-mers ($K=17$) identified by jellyfish (Marçais and Kingsford 2011) with a total length of 215.7 PE read. The results of genome size estimation have varied across studies. For example, Ozias-Akins and Jarret (1994) reported that the 2C content of the sweetpotato nucleus was 4.8–5.3 pg/2C, while Srisuwan et al. (2019) reported it as 3.1–3.3 pg/2C. Given that the haploid genome size of the diploid *I. trifida* haploid

is around 500 Mb, it is reasonable to assume that the genome size of sweetpotato is around 3 Gb/2C. Therefore, it was considered that the use of jellyfish (2.6 Gb) led to an underestimation due to the influences of homologous sequences across homoeologous chromosomes.

De novo whole-genome assembly was performed with Illumina short reads using three assembly tools. However, the N50 length ranged from 347 to 1598 bp, indicating significant fragmentation (Table 3.4).

Two approaches were then used for haploid-resolved genome sequence assembly: that is, DenovoMAGIC (NRGene, Israel) for Illumina and 10X Genomics reads and Falcon-unzip (PacBio) for PacBio reads (Yoon et al. 2022). The total length of primary contigs and haplotigs was 1.8 Gb (N50=325.5 Kb) and 336 Mb (N50=44.9 Kb), respectively (Table 3.5), while total and N50 lengths assembled by DenovoMAGIC were 2.4 Gb and 2150 Kb, respectively. The shorter total lengths in PacBio and DenovoMAGIC assembly are considered to be due to the integration of sequences across homoeologous chromosomes. Consequently, hybrid assembly with the Illumina scaffolds and PacBio reads were then performed by NRGene, and a total of 110,708 sequences were generated

Table 3.3 Sequenced illumina short reads of ‘Xushu 18’

Library	Platform	Insert size (base)	Total length (Gb)	Coverage (X) ^a
Paired-end	MiSeq, HiSeq	500	615.6	205
Mate-Pair	MiSeq, HiSeq	2000	23.3	8
Mate-Pair	MiSeq	5000	25.6	9
Mate-Pair	HiSeq	8000	141.9	47
Mate-Pair	MiSeq	10,000	22	7
Mate-Pair	HiSeq	15,000	135.3	45
Mate-Pair	HiSeq	20,000	145	48

^aThe coverage of the total read length obtained concerning a genome size of 3 Gb

Table 3.4 Results of de novo assembly of ‘Xushu 18’ with illumina short reads

Assembly tool	Newbler 3.0	SOAP de novo 2	Platanus
Number of sequences	1,443,167	8,491,291	3,112,396
Total length (bp)	985,580,805	2,662,222,474	1,213,108,214
Max length (bp)	4312	88,150	1,009,350
N50 length (bp)	1598	347	911
GC%	35.2	36.1	37.0

Table 3.5 Status of whole genome assembly in *I. batatas* ‘Xushu 18’

Assemble	DenovoMAGIC+PacBio	DenovoMAGIC	Falcon unzip+Hybrid
Read types	Illumina + 10X + Pacbio	Illumina + 10X	PacBio + Illumina
Number of sequences	110,708	163,164	6,026
Total length (bp)	2,907,440,085	2,374,736,154	1,839,235,287
N50 length (bp)	3,144,671	2,150,559	510,412

with 2.91 Gb length. The total length was close to the estimated genome size of sweetpotato, and the result suggested that hybrid assembly using Illumina DenovoMAGIC scaffolds and PacBio reads is effective for haploid-resolved assembly in autopolyploidy species.

To create chromosome-scale scaffolds, an S_1 linkage map was constructed using the variants identified on the *I. trifida* genome. The dd-RAD-Seq sequences of 437 S_1 individuals were mapped onto 520 scaffolds comprising the ‘Mx23Hm’ Hi-C scaffolds. A total of 534 scaffolds were aligned on the linkage map as 90 chromosome-level scaffolds. With 109,896 unplaced scaffolds, the 90 chromosome-level scaffolds were designated as IBA_r1.0. The total length of IBA_r1.0 was 2907.4 Mb, consisting of 2168.4 Mb at the chromosome level and 738.9 Mb unplaced scaffolds (Table 3.6). The ratio of complete BUSCOs assembly on IBA_r1.0 was 99.5%, including 1.7% of single-copy genes and 97.8% of duplicated genes. A total of 175,633 gene sequences were predicted for the Itr_r2.2 genome based on ab initio and evidence-based gene models.

The genome sequences of the 90 chromosome-level scaffolds were then compared with *I. trifida* genome sequences (Itr_r2.2). There was clear macro-synteny between *I. batatas* and the diploid species (Fig. 3.1).

3.4 Application of Assembled Genome Sequences for Crop Improvement and Future Prospects

The Itr_r2.2 and IBA_r1.0 genome sequences are available on Plant GARDEN (Itr_r2.2: <https://plantgarden.jp/ja/list/t35884/genome/t35884.G002>, IBA_r1.0: <https://plantgarden.jp/ja/list/t4120/genome/t4120.G001>) and have already been used for genomic and genetic analysis. For example, Suematsu et al. (2022) reported identification of a major QTL for root thickness in *I. trifida* using a QTL-Seq approach. A BC₁F₁ population derived from crosses between ‘Mx23Hm’ and ‘0431-1’ was used for the analysis, and a major QTL for root thickness (*qRTI*) was identified on chr06 of the

Table 3.6 Statistics on the assembled *I. batatas* ‘Xushu 18’ (IBA_r1.0) genome sequences and genes

	All	Chromosome-scale	Unplaced	Gene
Number of sequences	109,986	90	109,896	175,633
Total length of sequences (bp)	2,907,375,442	2,168,449,239	738,926,203	181,148,040
N50 length (bp)	23,279,505	26,187,252	17,387	1,380.00
GC%	36.1	35.7	37.2	46.2
N%	0.07	0.14	0.34	0
BUSCO v5.2.2 embryophyta_odb10, Number of BUSCOs = 1614				
Complete	99.5	99.4	56.1	97.2
Complete single	1.7	2.5	35.5	9.5
Complete double	97.8	96.9	20.6	87.7
Fragmented	0.2	0.2	6.8	1.7
Missing	0.3	0.4	37.1	1.1

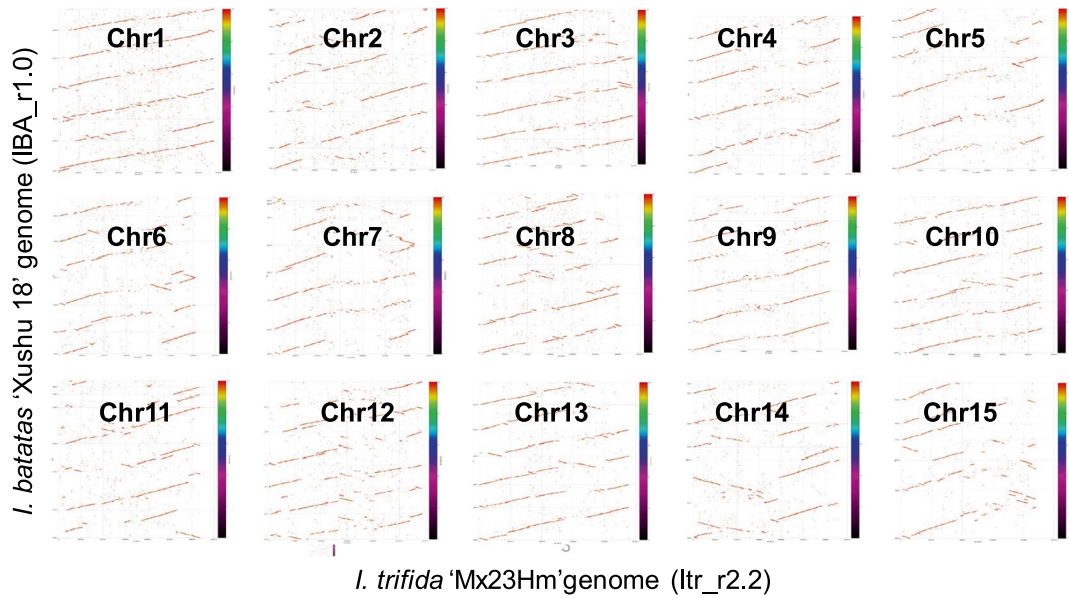


Fig. 3.1 Comparison between *I. batatas* 'Xushu 18' genome (IBA_r1.0) and *I. trifida* 'Mx23Hm' genome (Itr_r2.2) sequences

Itr_r2.2 genome. Haque et al. (2023) reported genetic analysis of starch contents (SC) using 204 F_1 progenies derived from a bi-parental cross between *I. batatas* cultivars, 'Konaishin' and 'Akemurasaki'. Base variants were identified on the Itr_r2.2 genome, and significant QTL for SC were identified on Chr15. One of the candidate genes located on the QTL regions, *IbGBSSI*, was considered to be involved in starch accumulation in sweetpotato root, by the results of qRT-PCR analysis.

For the expression analysis of starch, anthocyanin, and carotenoid genes in *I. batatas* tissues, RNA-Seq analysis was performed on RNAs extracted from the leaves at 42 days after transplantation (DAT), stems at 42 DAT, and roots at 90 DAT (Yoon et al. 2022). The fragments per kilobase of transcript per million mapped reads (FPKM) values were calculated on the genes predicted on the *I. batatas* genome, IBA_r1.0. Significantly high expressions were observed in roots for starch pathway genes. Conversely, in the leaves, the robust expression of genes associated with anthocyanin genes was observed.

Sweetpotatoes are utilized for a diverse range of purposes, including food and processed

products such as starch, distilled spirits and natural colorants. Given the various applications, breeding goals for sweetpotato are diverse, necessitating genetic analyses across a multitude of traits. According to the comprehensive review by Yan et al. (2022), previous genetic analyses have predominantly focused on yield, root development, quality, and biotic resistance. Until recently, genetic analyses were predominantly conducted using the genome sequences of diploid species like *I. trifida*. However, the recent completion of the hexaploid genome sequence now paves the way for more advanced analyses. While the sweetpotato genome structure has been suggested to be either complete auto-hexaploid or auto-allo-hexaploid, elucidating the extent of genome sequence variation among homoeologous chromosomes and the conservation of gene sequences on these chromosomes is a task for the future. This advancement is anticipated to enhance our understanding of how genes governing target traits are regulated across homologous chromosomes, enabling more precise breeding strategies.

In the era of climate change, when food production faces escalating challenges, sweetpotato,

with its relatively stable yields even in marginal lands, is expected to attract greater attention as a source of nutrition. The sweetpotato genomes, including those created by the TRAS consortium, are poised to serve as a crucial information resource for accelerating global sweetpotato breeding efforts. As we anticipate difficulties with food production amid changing climates, leveraging the genomic information of sweetpotato will become crucial for developing resilient crops and ensuring global food security.

Acknowledgements The TRAS genome sequencing work was performed by Dr. Ung-Han Yoon, Dr. Tae-Ho Lee, Dr. Tae-Ho Kim, Dr. Jang-Ho Hahn and Dr. Byoung Ohg Ahn of the National Institute of Agricultural Sciences, RDA, Dr. Qinghe Cao, Dr. An Zhang, Dr. Shizhuo Xiao, and Dr. Daifu Ma of the Sweetpotato Research Institute, CAAS, Dr. Hong Zhai, Dr. Xiangfeng Wang and Dr. Qingchang Liu of the College of Agronomy and Biotechnology, China Agricultural University, Dr. Ho Soo Kim, Dr. Sul-U Park, Dr. Sang-Soo Kwak of the Plant Systems Engineering Research Center, KRIBB, Dr. Masaru Tanaka, Dr. Hiroaki Tabuchi, Dr. Yoshihiro Okada, and Dr. Yasuhiro Takahata of the Kyushu Okinawa Agricultural Research Center, NARO, Dr. Jae Cheol Jeong of the Biological Resource Center, KRIBB, Dr. Soichiro Nagano of the Forest Tree Breeding Center, FFPRI, Dr. Younhee Shin of Insilicogen, Inc., Dr. Hyeong-Un Lee of the Bioenergy Crop Research Institute, National Institute of Crop Science, RDA, Dr. Seung Jae Lee of DNA Link, Inc. and College of Life Sciences and Biotechnology, Korea University, Dr. Keunpyo Lee of the Technology Cooperation Bureau, RDA, Dr. Jung-Wook Yang of the National Institute of Crop Science, RDA, and Dr. Kenta Shirasawa, Dr. Hideki Hirakawa, and Dr. Hideki Nagasaki and Dr. Sachiko Isobe of Kazusa DNA Research Institute.

References

Austin, DF (1988) The taxonomy, evolution and genetic diversity of sweet potatoes and related wild species. In: Exploration, maintenance, and utilization of sweet potato genetic resources: report of the first sweet

- potato planning conference 1987. International Potato Center, Lima, pp 27–59
- Chin C-S, Alexander DH, Marks P et al (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569
- FAOSTAT. <https://www.fao.org/faostat/>. Accessed Jan 2024
- Haque E, Shirasawa K, Suematsu K et al (2023) Polyploid GWAS reveals the basis of molecular marker development for complex breeding traits including starch content in the storage roots of sweet potato. *Front Plant Sci* 14:1181909
- Hirakawa H, Okada Y, Tabuchi H et al (2015) Survey of genome sequences in a wild sweet potato, *Ipomoea trifida* (H. B. K.) G. Don. *DNA Res* 22:171–179
- Kyriakidou M, Tai HH, Anglin NL et al (2018) Current strategies of polyploid plant genome sequence assembly. *Front Plant Sci* 9:1660
- Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764–770
- Nishiyama I (1971) Evolution and domestication of the sweet potato. *植物学雑誌* 84:377–387
- Ozias-Akins P, Jarret RL (1994) Nuclear DNA content and ploidy levels in the genus *ipomoea*. *J Am Soc Hortic Sci* 119:110–115
- Putnam NH, O'Connell BL, Stites JC et al (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* 26:342–350
- Simão FA, Waterhouse RM, Ioannidis P et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212
- Srisuwan S, Sihachakr D, Martín J et al (2019) Change in nuclear DNA content and pollen size with polyploidisation in the sweet potato (*Ipomoea batatas*, Convolvulaceae) complex. *Plant Biol* 21:237–247
- Suematsu K, Tanaka M, Isobe S (2022) Identification of a major QTL for root thickness in diploid wild sweetpotato (*Ipomoea trifida*) using QTL-seq. *Plant Prod Sci* 25:120–129
- Yan M, Nie H, Wang Y et al (2022) Exploring and exploiting genetics and genomics for sweetpotato improvement: Status and perspectives. *Plant Commun* 3:100332
- Yoon U-H, Cao Q, Shirasawa K et al (2022) Haploid-resolved and chromosome-scale genome assembly in hexa-autoploid sweetpotato (*Ipomoea batatas* (L.) Lam). *bioRxiv* 2022.12.25.521700

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

