Genetic analyses of anthocyanin content using polyploid GWAS followed by QTL detection in the sweetpotato (Ipomoea batatas L.) storage root

Emdadul Haque1, Eiji Yamamoto2, Kenta Shirasawa2, Hiroaki Tabuchi1, Ung-Han Yoon3, Sachiko Isobe2 and Masaru Tanaka1

1Kyushu Okinawa Agricultural Research Center, NARO, 6651-2 Yokoichi-cho, Miyakonojo, Miyazaki 885-0091, Japan
2Kazusa DNA Research Institute, 2-6-7 Kazusakamatari, Kisarazu, Chiba 292-0818, Japan
3National Institute of Agricultural Sciences, 370 Nongsaengmyeong-ro, Jeonju 54874, Korea

Corresponding author: M. Tanaka, E-mail: mtanaka@affrc.go.jp, Phone: +81-986-24-4274, Fax: +81-986-24-4283

Received on March 26, 2020; Accepted on July 30, 2020

Abstract: Genetic studies on the purple-fleshed sweetpotato (Ipomoea batatas L.), which is rich in anthocyanin (AN) in the storage root, were performed by polyploid GWAS based on the allele dosage probability using 59,675 SNPs obtained from 94 F1 progenies between the cultivars ‘Konaishin’ (which has a high yield but no AN) and ‘Atemurasaki’ (which has a high AN content but low to moderate yield). The distribution of relative AN content was highly biased, with 60% of clones showing a low to undetectable level (A530 < 0.5). Fifty-nine SNPs from six signals on homologous groups (HGs) 3, 5 (one major and one smaller signal), 7, 13, and 15 were strongly associated with the relative AN content. Twelve SNPs from the major signal and one from the smaller signal of HG 5 were further detected by QTL analysis. In a database search of the AN biosynthesis gene, transcription factors IbMYB1 and IbWD40 and AN structural genes IbF3H and IbDFR were located on HG 5, suggesting that an SNP marker or markers from HG 5 might be tightly linked to candidate gene(s) homologous to one of these transcription factors and AN structural genes as a major factor in determining AN accumulation in the storage roots. These results would enhance our understanding of the underlying genetic basis of AN accumulation in the storage roots of sweetpotatoes, and the SNP markers found here, especially 13 SNPs from HG 5, would be a potential platform for future marker-assisted selection for breeding high-AN sweetpotato varieties.

Keywords: anthocyanin, ddRAD SNPs, polyploid GWAS, QTLs, root drymatter, sweetpotato

Introduction

The sweetpotato (Ipomoea batatas (L.) Lam), the world’s seventh most important food crop (FAO 2018), is increasingly used in the food industry. Anthocyanin (AN), a color pigment which belongs to a class of flavonoid synthesized via the phenylpropanoid pathway, is a major functional component of sweetpotato storage roots. Sweetpotato storage roots with purple flesh contain a majority of ANs, whereas those with other flesh colors such as white, yellow, or orange contain little or no AN (Tanaka et al. 2017). In Japan, purple-fleshed sweetpotatoes rich in AN have been used for various purposes, for example as principal sources of natural food colorant due to their superior color characteristics, for table use, in ‘shochu’ (sweetpotato liquor) production, and for their high value added to a wide range of food items because of their potential ability to counteract some lifestyle-related diseases (Tanaka et al. 2017). In Japan, improvements of the AN-rich purple-fleshed sweetpotato have been a key breeding interest, and several cultivars have already been developed (Sakai et al. 2010, Katayama et al. 2017, Tanaka et al. 2017). Because of the heterozygosity and hexasomic hexaploid nature of the sweetpotato, leading to diverse combinations of parental chromosomes in the male and female
In the current study, we have taken advantage of new bioinformatic tools that were developed for polyploid genetics in the form of ‘NGS-based SNP analysis’ to generate a completely integrated SNP dataset in 15 pseudo-homologous groups (HGs) for both parents instead of an independent HG for each parent (TRAS sweetpotato genome sequencing consortium 2019). We have also made use of ‘polyploid GWAS’ using the allele dosage probability of each SNP calculated from the read count of NGS data, thus enabling the detection of the allele-dose-dependent relation between SNPs and target traits, which could not be detected using the previous GWAS method involving only homo and hetero genotypes of each SNP (Yamamoto et al. 2020). Thus, the initial goal of this study is to utilize polyploid GWAS to identify a wider range of SNP loci linked to storage root AN with the aim of enhancing our understanding of the underlying genetic basis of AN accumulation in the storage roots of sweetpotatoes. Next, we provide a useful resource base for the development of molecular markers from these candidate SNP loci, with the prime goal of a marker-assisted selection program. We quantified the relative AN content in a segregating population consisting of 94 F₁ progenies between the cultivars ‘Konaishin’ and ‘Akemurasaki’. We report on the polyploid GWAS method using 59,675 SNPs and compare the GWAS result with QTL analysis.

**Materials and Methods**

### Plant materials and growing condition

*I. batatas* cv. Konaishin (KNS) and cv. Akemurasaki (AKM) and 94 F₁ progenies (hereafter referred to as KAF₁) were used for this study. KNS is a novel cultivar for starch production released in 2018 with no AN content and extraordinarily high yield. On the other hand, AKM is a cultivar with significantly high AN content (Sakai et al. 2010), low to moderate yield, and moderate starch content.

Seedling multiplication, pot cultivation and harvesting of roots of KNS, AKM and their KAF₁ were carried out under greenhouse conditions in Kyushu Okinawa Agricultural Research Center, NARO (KARC/NARO), Miyakonojo, Miyazaki prefecture, Japan, in 2018 as described previously (Haque et al. 2020). In brief, for seedling multiplication, stem cuttings of the lateral branches were nurtured in a temperature-controlled greenhouse. For developing storage roots, pot cultivation was carried out in another greenhouse using 6L plastic pots (Kaneya Co. LTD., Aichi, Japan, height 20.5 cm × inner...
diameter 20.5 cm) from July to November in 2018. The experiment was divided into three-unit plots, each containing 96 pots. Randomization was done within the plots. Watering was done every day using a shower for the first seven days and followed by 500 mL of water per pot twice a week. Rotation of the pots was conducted weekly to minimize the differences in the micro-environment. Air temperature and soil temperature (7 cm below the soil surface) were recorded on a monthly basis from three spots in the greenhouse with a thermo recorder (T&D Co. LTD., Matsumoto, Japan). There was no temperature difference among the three locations in the greenhouse. The average monthly minimum/maximum daily temperatures (data from the center spot in the greenhouse) of the air and soil were 23.7/45.5 and 25.5/42.1, 20.7/45.8 and 22.9/41.0, 21.6/34.7 and 23.3/31.3, 13.2/29.2 and 15.2/28.9, and 9.1/25.9 and 11.3/25.0°C for July, August, September, October, and November, respectively.

Measurement of relative AN content

Two to four medium-sized storage roots were selected for the measurement of relative AN content. Then the thickest parts of the selected storage roots (approximately 40 g/plant) with skin (the relative AN content of the skin was very low) were cut into fine strips with a food processor (slicer coupled thinner) (Fig. 1 shows the storage root flesh color in handmade slices). First, AN was extracted from 2.0 g of the sliced sample in 20 mL of 0.5% sulfuric acid contained in a 50 mL Falcon tube. Samples were kept at 4°C in a dark chamber. For the measurement of relative AN, 0.5 mL of sample was diluted (8 times) into MCILVAINE buffer (0.1 M citric acid, 0.2 M Na2HPO4, pH 3.0). After gentle mixing with vortexing, the relative AN content was analyzed as absorbance at 530 nm (A530) using a Pharma spectrophotometer (UV-1700, Shimadzu Co., Kyoto, Japan).

ddRAD-Seq analysis to generate SNP loci in 15 pseudochromosomes

ddRAD-Seq analysis was performed according to the method of Shirasawa et al. (2016). In sequence analysis, a total of 471,617,154 bases were obtained. The average number of bases per individual was 4,912,679 (1.0%). The mapping rate to the I. trifida pseudomolecule (TRAS sweetpotato genome sequencing consortium 2019) was 84.8% on average. Based on the sequence alignment data, 104,881 candidate SNP loci were identified after filtering. These SNPs were further filtered with max-missing (the percentage of individuals that were called out of all individuals). A total of 59,675 SNPs derived using a max-missing value of 0.8 (20%) were selected for further analyses.

The HiSeq reads of ddRAD-Seq libraries are

![Fig. 1. Images of sliced storage roots in the Konaishin (KNS) × Akemurasaki (AKM) progenies (KAF_1). Photographs are from one replication as a representative of three replications.](image-url)
available in the DDBJ Sequence Read Archive under accession number DRA009153.

Polyploid GWAS

A total of 59,675 SNPs underwent allele dosage estimation using the method described previously (Yamamoto et al. 2020) with the following criteria: ploidy = 6, minimum read depth (dp) = 10, maximum dp = 1000, maximum missing = 0.5, maximum frequency = 0.95, round up = 1.00, cut off = 0.05, and read error probability = 0.001. The resulting 59,649 SNPs were then subjected to GWAS according to the method described (Yamamoto et al. 2020). The Manhattan and quantile-quantile (QQ) plots were created with ‘manhattan’ and ‘qq’, respectively, in qqman of R (Turner 2014). The P-values of the SNPs from GWAS underwent sequential Bonferroni correction, where an adjusted $P$-value of 0.05 was used to set the significant threshold level.

QTLs mapping

SNP sites with dp > 10 and an altered allele frequency (AAF) ≥ 0.7500 and ≤ 0.2500 were selected as simplex and double-simplex candidates. These SNPs were subjected to a chi-square test for the expected ratio of 1:1 or 3:1. A total of 28,492 (47.7%) simplex and double-simplex loci fit into the above two segregation patterns. A genetic map for pseudochromosomes from both parents was constructed with Lep-MAP3 (v.3.0) (logarithm of odds (LOD) > 3.685) (Rastas 2017). The workflow of Lep-MAP3 consists of the modules ParentCall2, Filtering2, SeparateChromosomes2, JoinSingles2All, and OrderMarkers2. In the Filtering2 command, markers were filtered based on, e.g., high segregation distortion (data tolerance = 1) and the amount of missing data.

The parameters used for the SeparateChromosomes2 command were: lodLimit = 3.685 distortionLod = 1 informativeMask = 123 sizeLimit = 10 theta = 0.05.

QTLs analyses were done according to Haque et al. (2020) using MapQTL6.0 software. A LOD of 2.5 was selected as a threshold to determine the QTL.

Searching the location of structural genes (SGs) and their transcription factors (TFs) of the AN biosynthesis pathway within the reference genome

To identify the homology regions of SGs associated with the AN biosynthesis pathway, the IbCHS cDNA (AB023791), IbCHI cDNA (EU402467), IbF3’H cDNA (HM460342), IbF3H cDNA (HQ441168), IbDFR cDNA (EF108570), IbANS cDNA (GU598212, FJ478179), and IbUF3GT cDNA (KF056328) were used to query the reference genome database of I. trifida (ITR_r2.2. scaffold) using a local blast search (Genetyx ver.15, GENETYX Co.). For TFs associated with the AN biosynthesis pathway, the IbMYB1 cDNA (AB258984), IbWD40 cDNA (JQ955738), and IbbHLH3 cDNA (KU589265) were used as query sequences. The highly homologous regions for each SG or TF from the reference genome were then confirmed using a BLAST search against the GenBank nucleotide database.

Results

Distribution of AN content in the storage roots

The distribution of relative AN content was highly biased, with 60% (n = 56) of clones showing a low to undetectable level (A530 < 0.5) (Fig. 2). The average A530 values were 12.01 ± 0.70 and 0.10 ± 0.01 for AKM and KNS, respectively. The segregation

![Fig. 2. Frequency distribution of the relative anthocyanin content (AN; A530) in the AKM × KNS progenies (n = 94) and both parents. Arrows show the values (mean ± SD) of AKM (12.01 ± 0.70) and KNS (0.10 ± 0.01).](image)
### Table 1. Summary of the 59 SNPs above the qqman significance thresholds

| Marker        | HGs | -log_{10} (P)  | P-value          | Estimated allele dosage | Lep-MAP mapping result |
|---------------|-----|----------------|------------------|-------------------------|------------------------|
| Itr_chr05_27090704 | 5   | 21.95          | 1.11 × 10^{-22}  | m                       |                        |
| Itr_chr05_26718188 | 5   | 21.55          | 2.84 × 10^{-22}  | m                       |                        |
| Itr_chr05_26718670 | 5   | 19.59          | 2.55 × 10^{-20}  | m                       |                        |
| Itr_chr05_26396741 | 5   | 17.99          | 1.03 × 10^{-18}  | s                       | Mapped                 |
| Itr_chr05_26029224 | 5   | 17.41          | 3.89 × 10^{-18}  | s                       | Mapped                 |
| Itr_chr05_25701591 | 5   | 15.77          | 1.69 × 10^{-16}  | s                       | Mapped                 |
| Itr_chr05_24862401 | 5   | 14.76          | 1.75 × 10^{-15}  | m                       |                        |
| Itr_chr05_273879726 | 5  | 14.65          | 2.22 × 10^{-15}  | m                       |                        |
| Itr_chr05_26275802 | 5   | 13.83          | 1.47 × 10^{-14}  | m                       |                        |
| Itr_chr05_26275799 | 5   | 13.83          | 1.50 × 10^{-14}  | m                       |                        |
| Itr_chr05_26275800 | 5   | 13.66          | 2.17 × 10^{-14}  | m                       |                        |
| Itr_chr05_26275798 | 5   | 13.63          | 2.36 × 10^{-14}  | m                       |                        |
| Itr_chr05_26701979 | 5   | 13.33          | 4.68 × 10^{-14}  | m                       |                        |
| Itr_chr05_26275794 | 5   | 13.20          | 6.29 × 10^{-14}  | m                       |                        |
| Itr_chr05_23536069 | 5   | 13.15          | 7.06 × 10^{-14}  | s                       | Mapped                 |
| Itr_chr05_26701812 | 5   | 13.13          | 7.44 × 10^{-14}  | m                       |                        |
| Itr_chr05_26351002 | 5   | 12.84          | 1.43 × 10^{-13}  | s                       | Not mapped             |
| Itr_chr05_27427060 | 5   | 12.84          | 1.44 × 10^{-13}  | s                       | Not mapped             |
| Itr_chr05_27569615 | 5   | 12.65          | 2.25 × 10^{-13}  | m                       |                        |
| Itr_chr05_22832191 | 5   | 12.59          | 2.57 × 10^{-13}  | s                       | Mapped                 |
| Itr_chr05_23551173 | 5   | 12.55          | 2.85 × 10^{-13}  | s                       | Mapped                 |
| Itr_chr05_29358196 | 5   | 11.90          | 1.26 × 10^{-12}  | s                       | Mapped                 |
| Itr_chr05_24960141 | 5   | 11.89          | 1.30 × 10^{-12}  | m                       |                        |
| Itr_chr05_26864099 | 5   | 11.45          | 3.56 × 10^{-12}  | m                       |                        |
| Itr_chr05_26550117 | 5   | 11.30          | 5.04 × 10^{-12}  | m                       |                        |
| Itr_chr05_29385059 | 5   | 11.00          | 9.90 × 10^{-12}  | m                       |                        |
| Itr_chr05_26550111 | 5   | 10.85          | 1.41 × 10^{-11}  | m                       |                        |
| Itr_chr05_29394223 | 5   | 10.75          | 1.77 × 10^{-11}  | s                       | Mapped                 |
| Itr_chr05_24861975 | 5   | 10.17          | 6.70 × 10^{-11}  | ds                      | Mapped                 |
| Itr_chr05_26461536 | 5   | 10.16          | 6.96 × 10^{-11}  | s                       | Mapped                 |
| Itr_chr13_13144677 | 13  | 9.86           | 1.38 × 10^{-10}  | s                       | Not mapped             |
| Itr_chr13_13144678 | 13  | 9.74           | 1.83 × 10^{-10}  | s                       | Not mapped             |
| Itr_chr05_23460474 | 5   | 9.64           | 2.30 × 10^{-10}  | m                       |                        |
| Itr_chr05_25731262 | 5   | 9.63           | 2.32 × 10^{-10}  | m                       |                        |
| Itr_chr13_13144896 | 13  | 9.32           | 4.79 × 10^{-10}  | s                       | Not mapped             |
| Itr_chr05_29451144 | 5   | 9.26           | 5.49 × 10^{-10}  | m                       |                        |
| Itr_chr05_27696681 | 5   | 9.14           | 7.19 × 10^{-10}  | s                       | Not mapped             |
| Itr_chr05_24605925 | 5   | 9.01           | 9.74 × 10^{-10}  | m                       |                        |
| Itr_chr05_24750065 | 5   | 9.00           | 9.98 × 10^{-10}  | m                       |                        |
| Itr_chr05_24927289 | 5   | 8.82           | 1.52 × 10^{-09}  | s                       | Mapped                 |
| Itr_chr05_24927560 | 5   | 8.73           | 1.85 × 10^{-09}  | s                       | Mapped                 |
| Marker                  | Homologous group | -log_{10}(P) | p-value     | Distortion  | dosage status |
|------------------------|------------------|--------------|-------------|-------------|---------------|
| Itr_chr05_25375649     | 5                | 8.63         | 2.32 × 10^{-9} | m           |               |
| Itr_chr05_25375645     | 5                | 8.58         | 2.63 × 10^{-9} | m           |               |
| Itr_chr05_28447653     | 5                | 8.39         | 4.08 × 10^{-9} | ds          | Not mapped    |
| Itr_chr05_26550141     | 5                | 8.15         | 7.10 × 10^{-9} | ds          | Not mapped    |
| Itr_chr05_25439632     | 5                | 8.09         | 8.07 × 10^{-9} | m           |               |
| Itr_chr05_23484740     | 15               | 6.89         | 1.30 × 10^{-7} | ds          | Not mapped    |
| Itr_chr03_30274266     | 3                | 6.56         | 2.78 × 10^{-7} | s           | Not mapped    |
| Itr_chr05_27388420     | 5                | 6.45         | 3.56 × 10^{-7} | m           |               |
| Itr_chr05_27388642     | 5                | 6.32         | 4.73 × 10^{-7} | m           |               |
| Itr_chr05_23460534     | 5                | 6.30         | 4.97 × 10^{-7} | m           |               |
| Itr_chr05_27835691     | 5                | 6.28         | 5.22 × 10^{-7} | ds          | Not mapped    |
| Itr_chr05_27388444     | 5                | 6.23         | 5.89 × 10^{-7} | m           |               |
| Itr_chr05_28124449     | 5                | 6.12         | 7.59 × 10^{-7} | ds          | Not mapped    |
| Itr_chr05_28508098     | 5                | 6.07         | 8.49 × 10^{-7} | ds          | Not mapped    |
| Itr_chr05_27388441     | 5                | 6.05         | 8.83 × 10^{-7} | m           |               |
| Itr_chr07_17548345     | 7                | 6.03         | 9.29 × 10^{-7} | s           | Not mapped    |
| Itr_chr07_7761458      | 5                | 6.01         | 9.73 × 10^{-7} | ds          | Mapped        |

**a** 'Itr' indicates, *Ipomoea trifida*, 'chr' indicates the chromosome and 'numeric' indicates the position (bp) of each marker on its respective chromosome.

**b** HGs indicates homologous groups.

**c** 's' and 'ds' indicate markers fit to simplex or double simplex segregation in a $\chi^2$-test, respectively, while 'm' indicates markers showing significant distortion from simplex or double-simplex segregation. The dosage status of these simplex and double-simplex markers was further confirmed by NGS read ratio.

**d** 'Mapped' indicates the 's' and 'ds' SNP markers passed by Lep-MAP3, while, 'Not mapped' indicates the 's' and 'ds' SNP markers which could not be passed by Filtering2, SeparateChromosomes2, and JoinSingles2All commands.

**Fig. 3.** Manhattan plots from GWAS of the relative AN content. The horizontal redline is the significance threshold of 5% Bonferroni correction. QTLs related to the relative AN content are indicated. The $-\log_{10}(P)$ for each QTL is shown in parentheses. 5i and 5ii indicate the major and smaller signals on HG 5, respectively.
ratio of white-fleshed clones (n = 56) and the remaining purple-fleshed clones (n = 38) fit to an expected segregation ratio of 1:1 ($\chi^2 = 3.074, P = 0.08$) or 3:1 ($\chi^2 = 1.89, P = 0.169$) (Fig. 1), which suggests that AN pigmentation of storage roots (including the skin and flesh) in the KAF population was determined by a major gene and inherited in a simplex or double simplex manner. However, A530 in purple-fleshed clones showed a continuous distribution (Fig. 2), suggesting that variation in the relative AN content of storage roots in purple-fleshed clones was regulated by multiple genes.

**GWAS**

GWAS identified six significant signals for the variation of relative AN content in KAF1, with two distinct signals on HG 5 (the major one was indicated as 5i and the smaller one as 5ii), and one signal each on HGs 3, 7, 13, and 15 (Fig. 3). These six signals consisted of a total of 59 SNPs (Table 1); one in HG 3 ($P = 2.78 \times 10^{-11}$), one ($P = 9.73 \times 10^{-07}$) and 52 ($P = 8.83 \times 10^{-07} - 1.11 \times 10^{-12}$) in HG 5, one in HG 7 ($P = 9.29 \times 10^{-07}$), three in HG 13 ($P = 4.79 \times 10^{-10} - 1.38 \times 10^{-10}$), and one in HG 15 ($P = 1.30 \times 10^{-07}$).

**Map construction and QTL analysis**

A high-density map with 15,747 (26.4% of the initial 59,675 SNPs and 55.3% of the simplex and double-simplex SNPs) markers (a total of 3,411 bins) was constructed on 93 linkage groups (LGs), covering a total distance of 4,726 cM. The average length of an LG was 50.8 cM (ranging from 0 to 210.5 cM), and the average number of SNPs in a single LG was 169.3 (2 to 2759). The marker density of an LG was 0.30 cM (0.08 to 2.61 cM).

In the QTL analysis of relative AN content, 13 significant SNPs were detected (Table 2). Of these SNPs, 12 were located in the major GWAS signal 5i on HG 5 (LOD = 5.27 – 13.67, PVE = 24.3 – 62.3%) and one in the smaller signal 5ii on HG 5 (LOD = 7.52, PVE = 43.7%). However, 10 from the major GWAS signal 5i and one from the smaller GWAS signal 5ii were mapped onto a single region of LG 1 (27.57 – 58.08 cM) as four LOD peaks, whereas two SNPs from the major GWAS signal 5i were mapped onto another region at 142.10 – 151.94 cM of LG 12 (Table 2 and Fig. 4). With an increasing effect on ratio of white-fleshed clones (n = 56) and the remaining purple-fleshed clones (n = 38) fit to an expected segregation ratio of 1:1 ($\chi^2 = 3.074, P = 0.08$) or 3:1 ($\chi^2 = 1.89, P = 0.169$) (Fig. 1), which suggests that AN pigmentation of storage roots (including the skin and flesh) in the KAF population was determined by a major gene and inherited in a simplex or double simplex manner. However, A530 in purple-fleshed clones showed a continuous distribution (Fig. 2), suggesting that variation in the relative AN content of storage roots in purple-fleshed clones was regulated by multiple genes.

**Table 2. SNPs that were confirmed by QTLs analyses for AN in KAF1 progeny**

| Marker | Polyplod GWAS | QTL analysis | Genotype score |
|--------|---------------|--------------|---------------|
|        | HGs | -log10 ($P$) | ANOVA | LOD | LGs | Position | PVE (%) | KNS | AKM |
| Itr_chr05_24861975 | 5 (si) | 10.17 | 6.70 \times 10^{-11} | *** | 11.7 | LG 1 | 27.57 | 44.2 | 0.48 (G) | 4.69 (K) |
| Itr_chr05_25701591 | 5 (si) | 15.77 | 1.69 \times 10^{-09} | *** | 7.68 | LG 1 | 27.57 | 32.7 | 0.27 (T) | 5.13 (Y) |
| Itr_chr05_26461536 | 5 (si) | 10.16 | 6.96 \times 10^{-11} | *** | 5.27 | LG 1 | 28.11 | 24.3 | 0.78 (C) | 5.44 (Y) |
| Itr_chr05_26396741 | 5 (si) | 17.99 | 1.03 \times 10^{-09} | *** | 9.48 | LG 1 | 40.43 | 37.2 | 0.65 (A) | 5.10 (R) |
| Itr_chr05_24927560 | 5 (si) | 8.73 | 1.85 \times 10^{-09} | *** | 11.77 | LG 1 | 41.51 | 44.7 | 1.24 (T) | 5.13 (W) |
| Itr_chr05_24927289 | 5 (si) | 8.82 | 1.52 \times 10^{-09} | *** | 13.47 | LG 1 | 41.51 | 49.4 | 1.24 (A) | 5.13 (R) |
| Itr_chr05_26029224 | 5 (si) | 17.41 | 3.89 \times 10^{-09} | *** | 12.25 | LG 1 | 44.71 | 46.7 | 1.10 (C) | 5.19 (S) |
| Itr_chr05_23536069 | 5 (si) | 13.15 | 7.06 \times 10^{-09} | *** | 10.57 | LG 1 | 50.06 | 40.5 | 0.93 (C) | 5.39 (Y) |
| Itr_chr05_23551173 | 5 (si) | 12.55 | 2.85 \times 10^{-09} | *** | 9.58 | LG 1 | 54.34 | 39.7 | 0.80 (C) | 4.87 (Y) |
| Itr_chr05_27716458 | 5 (si) | 6.01 | 9.73 \times 10^{-09} | *** | 7.52 | LG 1 | 54.34 | 43.7 | 1.17 (A) | 4.17 (W) |
| Itr_chr05_22832191 | 5 (si) | 12.59 | 2.57 \times 10^{-09} | *** | 9.07 | LG 1 | 58.08 | 35.9 | 0.83 (G) | 4.72 (K) |
| Itr_chr05_29358196 | 5 (si) | 11.90 | 1.26 \times 10^{-11} | *** | 13.67 | LG 12 | 151.94 | 62.3 | 1.22 (C) | 5.18 (Y) |
| Itr_chr05_29394223 | 5 (si) | 10.75 | 1.77 \times 10^{-11} | *** | 13.67 | LG 12 | 151.94 | 62.3 | 0.97 (G) | 4.93 (R) |

a Marker order of the Table 2 was sorted by positions on the physical and genetic maps. b,c,d,e Indicate the 7, 2, 3, and one SNP loci where QTL analysis and genotype scoring were carried out with 94, 93, 92, and 73 KAF genotype data, respectively. f *** indicates significant differences between genotypes by ANOVA at $P < 0.001$. g Arithmetic mean of relative AN content of progenies according to the KNS and AKM genotypes of the SNPs markers, respectively. Alleles determining types of ANs (KNS or AKM) are indicated within the parentheses.
the relative AN content, the 11 SNP markers on LG 1 exhibited genotype scores of 0.27–1.24 and 4.17–5.39 for the genotypes of KNS and AKM, respectively, while the two SNP markers on LG 12 exhibited the genotype score of 0.97–1.22 and 4.93–5.18 for the genotypes of KNS and AKM, respectively (Table 2). It is noteworthy that out of these 13 SNP loci named in Table 2, there were some missing genotype data for a few SNP loci (Table 2), which may have affected the QTL results.

**Locations of SGs and TFs for SNP markers of interest**

By data search, we found that the genes most homologous to *IbF3H*, *IbDFR*, *IbMYB1*, and *IbWD40* were located on HG 5 of *I. trifida* pseudomolecule, while the *IbF3H* was also located on HG 2 (Table 3). Homologous genes of *IbUF3GT* and *IbCHI* were located on HG 10 and HG 11, respectively. The locations of homologous genes of *IbCHS* were on HGs 11, 12, and 14, while homologous genes of *IbANS* and *IbbHLH3* were on HG 13 and HG 15, respectively.

**Discussion**

Improved breeding tools to facilitate the breeding of sweetpotatoes rich in AN are a priority in Japan, because of the drastic expansion of AN-added products from purple-fleshed sweetpotatoes (Katayama et al. 2017, Tanaka et al. 2017). Here, we have applied the polyploid GWAS method followed by QTL analysis. This enabled us to identify 59 SNP markers linked to the relative AN content on HGs 3, 5 (5i and 5ii), 7, 13, and 15. Specifically, 12 and one...
SNP markers from the above major signal 5i and smaller signal 5ii on HG 5, respectively, showed strong associations with the QTL regions, which explained 24.3 – 62.3% of the relative AN variances in this KAF_1 progeny (Table 2).

In the GWAS based on the physical map the of _I. trifida_ genome, two peaks (5i and 5ii) were detected on HG5. The polyploid GWAS method (Yamamoto et al. 2020) has a merit that error-tolerance genotyping scores, allele dosage probability, calculated from low coverage data are employed for genetic analysis. On the other hand, in genetic mapping, the SNPs at the 5i and 5ii were mapped together in the single region at 27.57 – 58.08 cM of LG 1 as four LOD peaks. This conflict could be arisen by possible chromosome rearrangements between _I. trifida_ and _I. batatas_. Another possibility might be genotyping errors due to the low-coverage data. Genotyping scores were required in the QTL mapping with MapQTL even if the read coverages were insufficient to distinguish three scores, e.g., A, B, and H. To obtain high accuracy scores in polyploid species, cost-consuming high-coverage data would be required. Obviously, this situation is not desired in practical breeding programs. One of the reasons that fewer SNPs were associated with QTL regions might be that the QTL analysis method used in this study was limited to the mode of inheritance of Mendelian markers that resembled the mode of inheritance for the simplex and double simplex dosages of markers. In our analysis, approximately 52.3% of SNP markers were discarded by the χ²-test, i.e., among the 59,675 SNP markers, 31,183 were not likely simplex and double-

simplex markers because the observed segregation pattern was significantly distorted from the expected segregation ratio for simplex or double-simplex markers. Furthermore, only 55.3% (15,747 out of 28,492 SNPs) of the simplex and double-simplex dosages of markers could be mapped by Lep-MAP3. There might be a handful of reasons of why the remaining 44.7% of markers were not mapped, including the autohexaploid genome nature of the sweetpotato, which could easily lead to complex recombination among homologous chromosomes in the male and female gametes. As categorized in Table 1, out of 59 SNPs, 28 were simplex and double-simplex. However, only 13 markers were mapped by Lep-MAP3, and unfortunately the remaining 15 markers were discarded by Lep-MAP3. It is likely that polyploid GWAS minimized the loss of SNP information during the genetic analysis compared to the current QTL analysis. However, GWAS does have the potential for false-positive error, and validation of the results is necessary (Zhu et al. 2008, Korte and Farlow 2013). The number of markers used in the GWAS highly affects its results (Han et al. 2018), particularly in the case of SNP markers where numerous markers are produced by NGS-based genotyping in a single experiment. Thus, the increase in the number of SNPs could easily lead to a genome-wide higher number of false-positive signals. Associations at unlinked, noncausal markers can also arise because of the inheritance pattern of the given trait (qualitative or quantitative), pleiotropy, epistasis, etc. (Platt et al. 2010). Although polyploid GWAS identified a much higher number of SNPs linked to

| Name       | Homologous groups | Position within _I. trifida_ |
|------------|-------------------|----------------------------|
| _IbMYB1_  | 5                 | 2573593-2574056            |
| _IbWD40_  | 5                 | 29809214-29807853          |
| _IbF3H_   | 5                 | 1716060-1716489            |
| _IbDFR_   | 2                 | 4297086-4297516            |
| _IbUF3GT_ | 10                | 1245564-1244686            |
| _IbCHI_   | 11                | 29751152-29751481          |
| _IbCHS_   | 12                | 6374874-6375977            |
| _IbANS_   | 13                | 25777607-25776844          |
| _IbHLH3_  | 15                | 1605017-1602807            |
IbMYB1 is the storage root (Park et al. 2015). Moreover, Mano sweetpotato increased AN pigment accumulation in purple-fleshed KAF through the IbMYB1 may be involved in AN pigmentation to AN accumulation in the storage roots. Thus, IbMYB1-2b, but not IbMYB1-1, were tightly linked IbMYB1-2b, IbMYB1-2a, and IbMYB1-2b, where IbMYB1-2a and IbMYB1-2b, but not IbMYB1-1, were tightly linked to AN accumulation in the storage roots. Thus, IbMYB1 may be involved in AN pigmentation accumulation in purple-fleshed KAF1, through the regulation of SGs (Fig. 1 and Fig. 2). WD40, which is a member of the TF complex MYB-bHLH-WD40 (MBW), has been shown to functionally enhance complex activation rather than to directly participate in the recognition of the target gene promoter (Baudry et al. 2004). Although in transgenic Arabidopsis thaliana seedlings, IbWD40 from the purple-fleshed sweetpotato was suggested to regulate AN biosynthesis (Dong et al. 2014), there is no evidence so far based on a mutant or overexpressed line in the sweetpotato that IbWD40 regulates AN biosynthesis. In light of the above discussion, we postulated that IbMYB1 might be harbored by the HG 5 as a major factor in determining AN accumulation in the storage roots via the regulation of these co-located or other AN biosynthesis genes. Further study is necessary to clarify the above hypothesis.

**Conclusion**

Here, we located 13 SNP markers on HG 5. One of them was assumed to include the candidate gene homologous to any of the TFs and AN SGs that acts as a major factor in determining AN accumulation in the storage root of sweetpotato. These results enhance our understanding of the underlying genetic basis of AN accumulation in the storage roots of sweetpotatoes, and the SNP markers generated in this study could be a platform for future marker-assisted selection for this trait in a sweetpotato breeding program. However, further analysis is necessary to confirm and develop molecular markers for these SNPs.

**Acknowledgments**

The authors wish to thank M. Nishinaka (Institute of Crop Science, NARO, Ibaraki, Japan), Y. Okada (KARC/NARO, Okinawa, Japan), A. Oyanagi (KARC/NARO, Miyazaki, Japan) and Y. Monden (Graduate School of Environmental and Life Science, Okayama University, Japan) for their suggestions. We also thank A. Tokiwa, A. Tsuneyoshi, M. Kamimura (sweetpotato crossing and seed production team) and the members of the field management staff of Miyakonojo research station, KARC/NARO for their assistance.

**References**

Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L 2004 TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. Plant J. 39: 366-380.

Cervantes-Flores JC, Sosinski B, Pecota KV, Mwanga ROM, Catignani GL, Truong VD, Watkins RH, Ulmer MR, Yencho GC 2011 Identification of quantitative trait loci for dry-matter, starch, and beta-carotene content in sweetpotato. Mol. Breed. 28: 201-216.

Dong W, Niu L, Gu J, Gao F 2014 Isolation of a WD40-repeat gene regulating anthocyanin biosynthesis in storage roots of purple-fleshed sweet potato. Acta Physiol. Plant. 36: 1123-1132.

Food and Agricultural Organization (FAO) 2018 FAO Statistics. http://www.fao.org/faostat/en/#data/QC (latest update: September 14, 2020), (accessed September 29, 2020)

Gemenet DC, Pereira GDS, De Boeck B, Wood JC,
Mollinari M, Olukolu BA, Diaz F, Mosquera V, Ssali RT, David M, Kitavi MN, Burgos G, Felde TZ, Ghislain M, Carey E, Swanckaert J, Cion LJ, Fei Z, Hamilton JP, Yada B, Yencho GC, Zeng Z-B, Mwanga ROM, Khan A, Gruneberg WJ, Buell CR 2020. Quantitative trait loci and differentially expressed genes in sweet potato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Huang Y, Wang J 2016. Development and identification of SSR markers for starch properties and β-carotene content in sweet potato. J. Integr. Agric. 15: 2744-2754.

Khan A, Gruneberg WJ, Buell CR 2020. Quantitative trait loci and differential gene expression analyses reveal the basis for negatively-associated β-carotene and starch content in hexaploid sweetpotato [Ipomoea batatas (L.) Lam.]. Theor. Appl. Genet. 133: 925-936.

Kang B-C 2013. A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweet potato. J. Integr. Agric. 12: 2410-2416.

Khan A, Gruneberg WJ, Buell CR 2020. Quantitative trait loci and differential gene expression analyses reveal the basis for negatively-associated β-carotene and starch content in sweetpotato. J. Integr. Agric. 13: 220-227.

Kato H, Yamauchi T, Hayashi M, Nishimura K, Ito K, Spalding J, Tsuchiya Y, Idehara H, Takahashi H, Shibata T, Kitagawa K, Kojima H, Itoh K, Nagai M, Shinozaki K, Yano K 2010. Genome-wide association study in sweetpotato. DNA Res. 17: 131-138.

Kumagai T, Kumagai T, Kai Y, Ishiguro K, Yamakawa O, Rastas P 2017. Lep-MAP 3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics 33: 3726-3732.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.

Kumagai T 2001. Breeding and varieties of sweetpotato. Proc. Natl. Acad. Sci. 117: 13181-13186.