Research Paper

QTL analysis and GWAS of agronomic traits in sweetpotato (*Ipomoea batatas* L.) using genome wide SNPs

Emdadul Haque¹⁾, Hiroaki Tabuchi¹⁾, Yuki Monden²⁾, Keisuke Suematsu¹⁾, Kenta Shirasawa³⁾, Sachiko Isobe³⁾ and Masaru Tanaka^{*1)}

¹⁾ Kyushu Okinawa Agricultural Research Center, NARO, 6651-2 Yokoichi-cho, Miyakonojo, Miyazaki 885-0091, Japan

²⁾ Graduate School of Environmental and Life Science, Okayama University, 1-1-1 Tsushimanaka, Kita-ku, Okayama, Okayama 700-8530, Japan

³⁾ Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

While sweetpotato (*Ipomoea batatas* L.) improvement has generally been done by field-based selection, molecular genetic studies on traits of interest, i.e., molecular markers are needed for enhancing the breeding program of this world's 7th most important crop, as such markers facilitate marker-assisted selection. Here, we performed a combined approach of QTLs analyses and GWAS of storage root β -carotene content (BC), dry-matter (DM) and starch content (SC) using the genetic linkage maps constructed with 5,952 and 5,640 SNPs obtained from F₁ progenies between cultivars 'J-Red' and 'Choshu'. BC was negatively correlated with DM ($r = -0.45$) and SC ($r = -0.51$), while DM was positively correlated with SC ($r = 0.94$). In both parental maps, a total of five, two and five QTL regions on linkage groups 7 and 8 were associated with BC, DM and SC, respectively. In GWAS of BC, one strong signal ($P = 1.04 \times 10^{-9}$) was observed on linkage group 8, which co-located with one of the above QTL regions. The SNPs markers found here, particularly for β -carotene, would be useful base resources for future marker-assisted selection program with this trait.

Key Words: sweetpotato, β -carotene, root dry-matter, starch, ddRAD SNPs, QTL, GWAS.

Introduction

Sweetpotato (*Ipomoea batatas* (L.) Lam) is the world's seventh most important crop (FAO 2009) and used as food, feed, nutrition and cash crop. Due to the wild and resilient nature, the crop has become the most affordable sources of carbohydrates, vitamin, and fiber particularly in the developing countries (Low *et al.* 2007, Woolfe 1992). In addition, in Japan, a variety of sweetpotato products, such as steamed and dried sweetpotato, liquor (Shochu), confectionary, extracted starch, pigments etc. have been produced from sweetpotato (Katayama *et al.* 2017). These new uses of sweetpotato have stimulated breeders to develop better adapted, disease- and insect-resistant, as well as high β -carotene, high anthocyanin, high storage root dry-matter and high-yielding varieties to address nutritional and industrial needs in Japan as well as in the world. However, due to the high level of heterozygosity and hexaploid nature, breeding of this important polyploid crop is far more challenging compared with many of self-

pollinating diploid crops. Especially, large numbers of genotypes must be screened in the early stage of the breeding, because of diverse combination of parental chromosomes in the male and female gametes. For example, approximately 50,000 seedlings are yearly subjected to the first cycle of conventional breeding program in Kyushu Okinawa Agricultural Research Station, one of the sweetpotato breeding site of Japan. From these 50,000 seedlings, averagely one cultivar is released after 10 years through extensive field evaluation and consecutive line selection (Kumagai 2001). The application of molecular markers which are diagnostic of alleles of genes linked to the favorable traits is expected to dramatically reduce the labors and periods required to the screening of clones in the early stage of the conventional breeding program, thus enabling shortening of the breeding period as well as finding promising clones from wider genetic variations. However, to date, DNA markers that enables efficient selection in sweetpotato breeding has not been reported. This is largely because of lack of molecular genetic study on agricultural traits of sweetpotato.

Molecular genomics and genetic study have facilitated several molecular breeding strategies including high-density genetic map enabling quantitative trait locus (QTL) identification or association mapping (e.g. GWAS) to

Communicated by Kenji Katayama

Received June 18, 2019. Accepted December 8, 2019.

First Published Online in J-STAGE on May 19, 2020.

*Corresponding author (e-mail: mtanaka@affrc.go.jp)

identify such molecular markers. While the technology is efficient in diploid and self-pollinated species, the hexaploidy genome and allogamous nature ($2n=6x=90$) of the cultivated sweetpotato has become the main obstacle for DNA marker development in this important crop. To date, relatively few AFLP (Cervantes-Flores *et al.* 2011, Nakayama *et al.* 2012), AFLP and SSR (Yu *et al.* 2014, Zhao *et al.* 2013) and SSR (Yada *et al.* 2017, Zhang *et al.* 2016) DNA marker-based genetic studies have been conducted in sweetpotato. Among these studies, Cervantes-Flores *et al.* (2011) first identified QTL regions for agronomic traits including β -carotene content, dry-matter content and starch content in a hexaploid sweetpotato mapping population derived from a cross between Tanzania, a white-fleshed, high dry-matter African landrace, and Beauregard, an orange-fleshed, low dry-matter sweetpotato cultivar. They constructed two parental maps consisted of 726 and 947 single-dose AFLP markers ordered into 90 and 86 linkage groups (LGs) for Beauregard and Tanzania, respectively (Cervantes-Flores *et al.* 2008). For dry-matter (Zhao *et al.* 2013) and starch (Yu *et al.* 2014), two parental maps were constructed consisted of 2077 markers (1936 AFLP and 141 SSR) and 1954 markers (1824 AFLP and 130 SSR) ordered into 90 LGs for Xushu 18 and 90 LGs for Xu 781. These markers, however, are still not enough to evenly cover the 90 chromosomes of sweetpotato. Furthermore, due to the unavailability of genome sequence, these results cannot be applied to other F_1 populations.

The numerous genetic markers generated by next-generation sequencing (NGS) and their high-density genetic map covering whole genome is becoming a popular tool to develop the agronomically important trait-linked DNA markers. Until recently, this was not reported due to the absence of reference genome sequence for cultivated hexaploid sweetpotato. However, after publishing a draft genome sequence of diploid *I. trifida* ($2n=2x=30$), the close ancestor of cultivated sweetpotato (Hirakawa *et al.* 2015), followed by diploid *I. triloba* (<http://sweetpotato.plantbiology.msu.edu>), a new era started in the field of sweetpotato genetic study. Very recently, a firsthand high-density SNP genetic map based on NGS technology was reported in hexasomic hexaploid sweetpotato (Shirasawa *et al.* 2017).

In case of traditional QTL mapping, particularly for high-density genetic map covering whole genome, the QTL regions can be quite large, incorporating too many genes to investigate as potential candidate genes. On the other hand, genome wide association studies (GWAS) can narrow down the candidate regions, however, can have high rates of false positive errors (Zhu *et al.* 2008). Therefore, validation of the results of GWAS (Korte and Farlow 2013) as well as QTLs analyses is necessary. The combination of GWAS and QTLs analyses can compensate for the limitations of each approach, enabling the identification of loci controlling agronomically important quantitative traits. Such combined approaches have been successfully used to

identify candidate genes controlling agronomical traits in several plants (Crowell *et al.* 2016, Han *et al.* 2018, Tian *et al.* 2011). The approach was, for the first time, used in sweetpotato for the identification of responsive SNPs of nematode resistance (Sasai *et al.* 2019).

The current study was destined to identify genomic regions controlling several agronomic traits for sweetpotato aiming at helping marker-assisted selection (MAS) program. We used the genetic linkage map constructed with 5,952 and 5,640 SNPs obtained from 52 F_1 progenies between cultivars ‘J-Red’ and ‘Choshu’ (Sasai *et al.* 2019). These SNPs includes simplex and double simplex dosages of markers. We performed a combine approach of QTLs analyses and GWAS of storage root relative β -carotene content (BC), dry-matter content (DM) and starch content (SC).

Materials and Methods

Plant materials

I. batatas cv. ‘J-Red’ (J) and cv. ‘Choshu’ (C) and their F_1 (JCF₁) population consisting of 52 progenies were used for this study. J is an orange-fleshed cultivar with extremely high carotene content, high yield, moderate DM, and strong resistance to root-knot nematode (Yamakawa *et al.* 1998). On the other hand, C is a white-fleshed local variety of Japan with no or undetectable carotene content, low yield, moderate DM and weak to root-knot nematode (Sasai *et al.* 2019, Tabuchi *et al.* 2017).

These plants were multiplied using cuttings of the lateral branches in a temperature-controlled greenhouse located in Kyushu Okinawa Agricultural Research Center, NARO (KARC/NARO).

Growing condition and BC, DM and SC phenotyping

Cultivation of J, C and 52 F_1 in pots was carried out only one year in a greenhouse (The roof of the greenhouse was made up with glasses, the fences were made up with nets) using a 6 L plastic pot (Kaneya Co. LTD., Japan, height 20.5 cm $\times$ inner diameter 20.5 cm) at KARC/NARO from May to September in 2018 (**Supplemental Fig. 1A**). Experiment was divided into three-unit plots each containing 54 pots. Randomization was done within plots. The bottom of each pot was filled with a 2-cm layer of bora soil (pumice soil). The pots were then filled with 2.64 kg andosol. The lower two nodes of a stem cutting with seven non-folded leaves was inserted into soils. One and half grams of an 8N-12P-20K fertilizer was applied per pot at 10 days and 8 weeks after transplantation. Watering was done every day using shower for the first seven days and followed by 500 mL water per pot twice in a week. Air and soil (7 cm below soil surface) temperature were recorded monthly basis with a thermo recorder (T&D Co. LTD., Japan) (**Supplemental Table 1**). Rotation of pots was conducted weekly to minimize the differences in micro-environment. In addition, a phenotypic comparison was conducted with field-grown data of 52 JCF₁.

Pots were harvested at 4 months after transplantation. After recording the number and weight of the storage roots, 2–4 medium sized storage roots were selected for the measurement of BC, DM and SC. Then the thickest part of the selected storage roots (approximately 40 g/plant) were cut into fine strips with a food processor. For the measurement of relative BC, 2.0 g of the sample was homogenized in 20 mL acetone using polytron (PT-MR 3100, Littau, Switzerland). After keeping overnight at 4°C in a dark chamber, the extract was centrifuged at 3,000 rpm for 5 min. BC was analyzed as absorbance at 455 nm (A_{455}) using a Pharma spectrophotometer (UV-1700, Shimadzu, Japan).

For DM measurement, 10 g of the strips were oven dried at 80°C for 2–3 days. For SC, 10 g sample was first blended with 50 mL water. The blend mix was filtered through two different sieved sieves (Tokyo Screen Co. LTD., Japan), upper sieve was 250 µm and lower one was 63 µm. The blended residue was rinsed with additional 200 mL water so that no starch remains in the sieves. After keeping overnight at room temperature, the supernatant was carefully discarded without breaking the starch pellet. The starch pellet was weighed after oven-dried at 80°C for 2–3 days.

Supplemental Fig. 1B, 1C show images of storage root fresh weight and DM measurement.

QTLs mapping

QTL controlling the contents of DM, SC and BC were independently detected for J and C. Firstly, the quantitative score of BC, DM and SC were subjected to QTL analysis using Kruskal-Wallis (KW) methods of MapQTL6.0 (Van Ooijen 2009). Map data from Sasai *et al.* (2019) was used in this analysis with J-parental map consisting of 5,952 and C-parental map consisting of 5,640 ddRAD-SNPs. The initial bi-parental maps on 90 LGs were constructed with 113 JCF₁ population (Sasai *et al.* 2019) using the OneMap program. During maintaining process over a period, the original progeny size was reduced due to damages or diseases infection, and we decided to use 52 individuals out of 113 JCF₁ for the genetic analyses of BC, DM and SC. As reported by Sasai *et al.* (2019), initially there were 6,341 markers in J parental map (SNP: 5,952, retrotransposon-based: 228, SSR: 161), yielding a linkage map with a total length of 13,247.0 cM and marker density of a single LG was 2.09 cM. There were 6,008 markers in C parental map (SNP: 5,640, retrotransposon-based: 192, SSR: 176), yielding a linkage map with a total length of 12,241.8 cM and averaged marker density of a LG was 2.04 cM. Other details are available in Sasai *et al.* (2019). QTLs were detected at the criteria of $P < 0.01$. Next, mixed model interval mapping was performed to detect the locations and effects of the putative QTLs for BC, DM and SC. Threshold values were calculated automatically by mixed model default system with 1,000 permutation test. Typical threshold value was LOD 3.0. The QTL graphs were prepared

using MapChart 2.32 (Voorrips 2002) using LOD scores of interval mapping. Genotype score of each QTLs was calculated from the mean of BC, DM and SC of progenies with homozygous and heterozygous SNPs genotype, respectively. The associated SNP markers for each significantly detected region were listed in **Supplemental Table 2**, and markers nearest to the peak, most significantly associated with the trait and co-associated with GWAS signals were selected as candidate SNPs which were listed in **Table 1**.

GWAS

GWAS was conducted with 13,574 filtered SNPs using a general linear model in Trait Analysis by aSSociation, Evolution, and Linkage (TASSEL5.2.49; Bradbury *et al.* 2007). The Manhattan plots of the GWAS signals were constructed using ‘qqman’ R package (Turner, unpublished). The approaches which were applied to combine the GWAS data and QTL data are as follows. P -values of GWAS signals for each significantly detected QTL regions were listed (**Supplemental Table 2**). Next, two candidate SNP markers nearest to the QTL peak were consulted with their P -values of GWAS (**Supplemental Table 2**). In most cases, the SNP markers with highest P -values of GWAS was given priority to combine QTLs and were listed in **Table 1**.

Results

BC

The distribution of BC in the storage root of the JCF₁ was highly biased (**Fig. 1A**), with 52% clones showing low levels ($A_{455} < 0.1$). The average A_{455} was 1.36 and 0.02 for J and C, respectively (**Supplemental Fig. 2A**). Transgressive segregation was observed in the JCF₁ progeny, with some progeny exhibited higher level and others lower level of BC than their parental clone. BC showed negative correlation with DM ($r = -0.45$) and SC ($r = -0.51$) (**Table 2**).

QTLs analyses on both parental maps reveals the presence of five regions having significant effects on the variation of BC (**Table 1**), of which three were on the J parental map (**Fig. 2A**), while two were on the C parental map (**Fig. 3A**). In the J parental map, two out of three loci had positive effect on BC and were associated with markers Itr_sc001596_10966 on the linkage group LG_Ib07_4 ($K^* = 6.3$, $P < 0.05$) and Itr_sc000066_68839 on LG_Ib08_4 ($K^* = 24.2$, $P < 0.0001$) (**Table 1**, **Fig. 2A**). The third locus and its associated marker Itr_sc000227_94993 on LG_Ib08_6 ($K^* = 8.0$, $P < 0.005$) exhibited a negative effect on BC. In the C parental map, two loci were observed to have significant and negative effects on the variation of BC (**Table 1**, **Fig. 3A**). The associated markers were Itr_sc000227_94989 on LG_Ib08_3 ($K^* = 8.0$, $P < 0.005$) and Itr_sc000066_1754 on LG_Ib08_3 ($K^* = 4.02$, $P < 0.05$).

GWAS identified 10 highly significant SNPs for BC ($P = 5.96 \times 10^{-7}$ – 1.04×10^{-9}) (**Supplemental Table 2**). Remarkably, all these SNPs were detected on the same QTL region located on LG 8 in the J parental map. SNP

Table 1. QTLs controlling BC, DM and SC in JCF1 progeny

Traits and maps	Locus			Kruskal-Wallis analysis		Interval-mapping		Genotype score ^f				<i>P</i> -value of GWAS
	LG ^a	Marker name	Position ^b	K* ^c	Sig. ^d	LOD	PVE ^e	Homo	Hetero	MD	D ^g	
BC_J	Ib07_4	Itr_sc001596_10966	67.75	6.32	**	4.69	68.0	0.19	0.59	0.40	+	1.06 × 10 ⁻³
	Ib08_4	Itr_sc000066_68839	164.21	24.22	*****	7.74	49.6	0.09	0.79	0.70	+	1.04 × 10 ^{-9h}
	Ib08_6	Itr_sc000227_94993	65.68	8.03	****	3.24	29.3	0.79	0.20	0.59	–	3.09 × 10 ⁻⁴
BC_C	Ib08_3	Itr_sc000227_94989	565.29	8.03	****	3.83	28.7	0.79	0.20	0.59	–	3.09 × 10 ⁻⁴
		Itr_sc000066_1754	620.05	4.02	**	3.21	24.8	0.66	0.17	0.49	–	6.04 × 10 ⁻⁵
DM_J	Ib07_4	Itr_sc000546_59175	77.20	11.46	*****	3.04	25.2	26.93	24.00	2.93	–	3.45 × 10 ⁻⁴
	Ib08_4	Itr_sc000057_157946	57.15	10.64	****	3.31	50.9	27.46	24.82	2.64	–	0.0045
SC_J	Ib07_4	Itr_sc000546_59175	77.20	11.72	*****	2.93	25.5	16.25	13.77	2.48	–	4.74 × 10 ⁻⁴
	Ib08_4	Itr_sc000057_157946	57.15	14.95	*****	3.31	50.9	16.97	14.34	2.63	–	7.67 × 10 ⁻⁴
SC_C	Ib07_2	Itr_sc000393_73439	90.03	8.71	****	3.07	23.8	14.65	16.66	2.01	+	0.0074
		Itr_sc004815_16417	118.76	8.46	****	3.20	24.7	14.33	16.37	2.04	+	0.0061
	Ib07_4	Itr_sc000216_53020	0.00	7.80	***	3.30	25.3	16.29	14.19	2.10	–	0.0015

^a LG indicates linkage group.

^b shows position of marker in cM from top.

^c K* shows Kruskal-Wallis test results.

^d **, ***, ****, *****, ***** and **** indicate significant differences between genotypes by KW analysis at *P* < 0.05, 0.01, 0.005, 0.001, 0.0005 and 0.0001 levels, respectively.

^e PVE indicates the percentage in phenotype of variance explained in each QTL.

^f Arithmetic mean of BC, RDM and SC content of progenies according to the homozygous (Homo) and heterozygous (Hetero) genotypes of the SNPs markers, respectively, and the mean differences (MD) between the genotypes (homo, hetero).

^g Direction (D) indicates SNP genotypes with increasing or decreasing effect on BC, DM and SC levels. “+” means SNP genotypes (hetero) resembles that of J or C with increasing effect on traits. “–” means SNP genotypes (hetero) resembles that of J or C with decreasing effect on traits.

^h indicates SNPs with significant *P*-values after Bonferroni correction.

Fig. 1. Histogram of β -carotene (BC), root dry-matter (DM) and starch content (SC) in the J-Red (J) × Choshu (C) progenies (JCF₁). (A) Histogram of BC (absorbance at 455 nm; A₄₅₅) in 52 JCF₁ and their parents. (B) Histogram of DM (%). (C) Histogram of SC (%). Arrows indicate positions of parental cultivars.

Fig. 2. QTLs on linkage maps and associated SNPs for the J parental map. (A) Significant QTLs for BC, DM and SC, which were listed in **Supplemental Table 2**, are shown in boxes. The selected *SNP markers* were listed in **Table 2**, are shown in triangle. QTLs with positive effect are indicated by “+” sign, and those with negative effects were indicated by “-” sign. Position of each QTLs were shown in parenthesis. (B) Manhattan plots from GWAS. Redline is significance threshold of 5% Bonferroni correction. QTLs related to BC, DM and SC are indicated. $-\log_{10}(P)$ for each QTLs are shown in parenthesis.

Table 2. Correlation among traits in the 52 JCF₁ progenies of the J × C

Trait ^a	BC	DM	SC
BC	1		
DM	-0.45	1	
SC	-0.51	0.94	1

^a All the root traits were from physiological maturity under greenhouse condition.

marker Itr_sc000066_68839 was highly associated (**Fig. 2B**) with the variation in BC ($P=1.04 \times 10^{-9}$) and explained 49.6% of the BC variance in this JCF₁ progeny (**Table 1**). The remaining two QTLs in the J parental map (**Fig. 2B**) and two in the C parental map (**Fig. 3B**) were also showed association with weak GWAS signals ($P=1.06 \times 10^{-3}$ – 6.04×10^{-5}) (**Figs. 2B, 3B**).

DM

The DM of the JCF₁ progeny was distributed normally (**Fig. 1B**). DM ranged from 17 to 32% with a progeny mean of 26% (**Supplemental Fig. 2B**). The average DM was

26.7 and 23.3% for J and C, respectively. Transgressive segregation was observed in the JCF₁ progeny. Strong positive correlation was found between DM and SC ($r=0.94$) (**Table 2**).

In the J parental map, two regions having significant and negative effects on the variation of DM were revealed by QTL analysis (**Table 1, Fig. 2A**). As shown in **Table 1**, the associated markers for these QTLs were Itr_sc000546_59175 on LG_Ib07_4 ($K^*=11.5$, $P<0.0001$) and Itr_sc000057_157946 on LG_Ib08_4 ($K^*=10.6$, $P<0.005$). No strong GWAS signals for these DM QTLs were identified, however, one of DM QTL showed association with weak GWAS signal (Itr_sc000546_59175, $P=3.45 \times 10^{-4}$) (**Fig. 2B**). No QTL region was observed for the variations of DM in the C parental map (**Fig. 3**).

SC

The SC of the JCF₁ progeny was distributed normally (**Fig. 1C**). SC ranged from 8 to 21 % with a population mean of 15% (**Supplemental Fig. 2C**). The average SC was 16.0 and 13.1% for J and C, respectively. Like DM, transgressive segregation was also observed for SC in the

Fig. 3. QTLs on linkage maps and associated SNPs for the C parental map. (A) Significant QTLs for BC, DM and SC, which were listed in Supplemental Table 2, are shown in boxes. The selected SNP markers were listed in Table 2, are shown in triangle. QTLs with positive effect are indicated by “+” sign, and those with negative effects were indicated by “-” sign. Position of each QTLs were shown in parenthesis. (B) Manhattan plots from GWAS. Redline is significance threshold of 5% Bonferroni correction. QTLs related to BC, DM and SC are indicated. $-\log_{10}(P)$ for each QTLs are shown in parenthesis.

JCF₁ progeny.

QTLs analyses on both parental maps reveals the presence of five regions having significant effects on the variation of SC (Table 1). In the J parental map (Fig. 2A), two regions, both were common in DM, were detected to have negative effects on the variation of SC. The associated two markers showed quite good statistical features except for Itr_sc000546_59175 where the LOD value was slightly lower (LOD=2.93) (Table 1). In the C parental map (Fig. 3B), two out of three regions had positive effects on the variation of SC and the associated markers were Itr_sc000393_73439 on LG_Ib07_2 ($K^*=8.7$, $P<0.005$) and Itr_sc004815_16417 on LG_Ib07_2 ($K^*=8.5$, $P<0.005$) (Table 1). The third region showed negative effect on the variation of SC and the associated marker was Itr_sc000216_53020 on LG_Ib07_4 ($K^*=7.8$, $P<0.01$). In the J parental map, both of SC QTL regions showed association with weak GWAS signals (Itr_sc000546_59175, $P=4.74 \times 10^{-4}$ and Itr_sc000057_157946, $P=7.67 \times 10^{-4}$) (Fig. 2B), while in the C parental map, none of the SC QTL region showed association with GWAS signals (Fig. 3B).

Discussion

In Japan, the development of high value-added products from sweetpotato has flourished (Katayama *et al.* 2017), which urges improved breeding tools to facilitate sweetpotato improvement with desired traits including BC, DM and SC. NGS based high-density genetic map covering whole genome enabled us to identify firsthand multiple SNP DNA markers linked with these traits and thereby inheritance on the LGs 7 and 8 as revealed by QTLs analyses in combination with GWAS. Earlier marker trait associated on these complex traits were done with AFLP and/or SSR markers (Cervantes-Flores *et al.* 2011, Yu *et al.* 2014, Zhao *et al.* 2013) through relatively low-density linkage map-based QTLs or association studies (Yada *et al.* 2017, Zhang *et al.* 2016). For example, in the case of Cervantes-Flores *et al.* (2008), the maps based on AFLP markers were consisted of 726 markers and 947 markers ordered into 90 LGs for Beauregard (covered 5276 cM with an average marker distance of 4.8 cM) and 86 LGs for Tanzania

(covered 5,792 cM with an average marker distance of 4.5 cM). For dry-matter (Zhao *et al.* 2013) and starch (Yu *et al.* 2014), two parental maps were based on AFLP and SSR markers ordered into 90 LGs for Xushu 18 (covered 8,184.5 cM with an average marker distance of 3.9 cM) and 90 LGs for Xu 781 (covered 8,151.7 cM with an average marker distance of 4.2 cM). Our results are latest toward MAS in the early cycles of breeding to save breeding resources and increase chance to get new breeding line.

BC

Three QTL regions exhibited significant effect for BC in the J parental map, two of them with positive effect and one with negative effect (Table 1). In the C parental map, two QTL regions, both having negative effect were detected. The existence of several QTL regions agreed with the fact that there are several key enzymes involved in the biosynthetic pathway of BC (Cloutault *et al.* 2012).

Out of the two QTLs with positive effects in the J parental map, the region found in LG_Ib08_4 showed strong association with GWAS signal (associated SNP Itr_sc000066_68839) (Fig. 2B), suggesting that this region might contains major gene that affect BC. The positive effect loci found in the orange-fleshed J is interesting and suggesting that their associated markers are correlated with the higher levels of BC. In contrast, negative effect QTLs were identified in the C parental map (on the LG_Ib08_3) which have only low amount of BC (Supplemental Fig. 1C). Such kinds of QTLs have also been reported in sweetpotato and other plants (Beyer *et al.* 2002, Cervantes-Flores *et al.* 2011, Ducreux *et al.* 2005, Ye *et al.* 2000). These QTL regions should suppress the orange-fleshed phenotype in C, resulting in white flesh (Supplemental Fig. 1C). The reason why the third QTL region found in the J parental map (LG_Ib08_6) exhibited negative effect (Table 1) is unknown at present. Interestingly, the associated SNPs (Itr_sc000227_94993, Itr_sc000227_94981, Itr_sc000227_94989 and Itr_sc000227_94982) with this region (LG_Ib08_6) was also commonly observed in the LG_Ib08_3 in C parental map (Supplemental Table 2).

Finally, BC was suggested to be negatively correlated with DM and SC (Table 2). This observation suggesting that genotypes containing higher levels of BC are more likely to have lower levels of SC and therefore lower DM (Cervantes-Flores *et al.* 2011, Yada *et al.* 2017). Interestingly, one putative SNP (LG_Ib07_4, Itr_sc001596_10966) having positive effect found for BC (LG_Ib07_4) is also mapped for SC (LG_Ib07_4) in the J parental map with negative effect (Supplemental Table 2). This can be explained by negative correlation between BC and SC.

DM and SC

The highly positive correlation between DM and SC ($r=0.94$) found in this study agreed with the previous report (Cervantes-Flores *et al.* 2011, Yada *et al.* 2017). This can be confirmed by our findings that the two putative

QTLs found in LGs 7 and 8 for DM were also mapped to the same region for SC in the J parental map (Table 1, Fig. 2).

The multiple QTL regions found in this study for DM and SC (Table 1, Figs. 2, 3) agreed with previous report (Cervantes-Flores *et al.* 2011) and suggest that multiple key enzymes involved in biosynthetic pathway of SC (Buléon *et al.* 1998, Kreuze *et al.* 2009) as well as DM.

All the QTL regions found for DM and SC in the J parental map showed negative effect, while all the QTL regions, except for one region on LG_Ib07_4, found for SC in the C parental map showed positive effect (Table 1). These data are not coinciding with the higher and lower levels of DM and SC observed in J and C, respectively (Fig. 1B, 1C, Supplemental Fig. 2B, 2C). One of the explanations is that the actual difference between J and C was not so high for these traits, nonetheless relatively large variance in these traits in the F_1 progenies.

Data validation and future directions

Good transgressive segregation is important for efficient breeding of sweetpotato. Thus, our both parents might be good cross combination. Because SC only accounts for 60% of the DM (Brabet *et al.* 1998, Woolfe 1992), QTLs more than SC should have appeared for the variation of DM, which might be involved in the synthesis of other parts of DM such as fibers, soluble sugars, proteins, etc. (Woolfe 1992). Contrastingly, in our study, more QTLs were observed for SC than DM. One reason might be the number of JCF₁ progeny used in this study is relatively less (52 clones) to reach the hypothesis. It is also possible that total amount of DM is mostly controlled by the genes related to distribution of carbohydrates to storage roots, while SC in DM is further controlled by the genes specifically involved in starch biosynthesis, resulting in higher number of QTLs for SC. The smaller annual fluctuation of DM compared to SC observed in previous studies (Cervantes-Flores *et al.* 2011, Sakai 1964) support this idea.

There were several merits of our pot cultivation system such as our greenhouse was a glasshouse and the fence of it were made up with nets, we controlled equal amount of soils, water and fertilizers per pot and did rotation per week, resulting in reproducibility among replications for the data taken in this study. However, when compared to field condition, no correlation was found for yield (storage root weight per plant) components (Data not shown). Regarding the yield data, although all the JCF₁ progenies produced storage root, the optimum turnover to yield might have not happen due to the limitation of soil amount and space inside our pot. Thus, our pot system was considered as a pre-breeding system for β -carotene and other qualitative traits but seems to be limited for yield traits which might be another reason to deal the above hypothesis.

In conclusion, this works is a first-hand information on SNP-based DNA markers which would enhance our understanding of the mechanisms underlying the inheritance of

complex traits in sweetpotato. These results are predicted to provide useful resource base for improving the breeding of these traits, particularly β -carotene content, with MAS, and promoting the future discovery of key genes controlling it. However, further analysis would be necessary to confirm the QTLs with greater number of F_1 progenies. Moreover, for yield trait, future study would be interesting to conduct under field condition.

Author Contribution Statement

H.T. developed the JCF₁ population; E.H. and M.T. conceived and designed the experiment; E.H. performed the experiment; E.H., K. Suematsu and M.T. taken the phenotypic data; K. Shirasawa, Y.M. and S.I. performed ddRAD-seq analyses and genetic map construction; E.H., M.T., Y.M. and H.T. conducted QTL analysis and GWAS; E.H. wrote the manuscript. All authors contributed to subsequent revisions.

Acknowledgments

We would like to thank Mrs. A. Tokiwa and the members of the field management staffs of Miyakonojo research station, KARC/NARO for their assistance. This work could not have been accomplished without F_1 progeny derived from J-Red and Choshu by Dr. Hiroki Nakayama (NARO), who passed away in 2012.

Literature Cited

- Beyer, P., S. Al-Babili, X. Ye, P. Lucca, P. Schaub, R. Welsch and I. Potrykus (2002) Golden rice: introducing the beta-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. Symposium: plant breeding: a new tool for fighting micronutrient malnutrition. Am. Soc. Nutr. Sci. 132: 506S–510S.
- Babet, C., D. Reynoso, D. Dufour, C. Mestres, J. Arredondo and G.J. Scott (1998) Starch content and properties of 106 sweetpotato clones from the world germplasm collection held at CIP, Peru. CIP Program Rep. 1997–98: impact on a challenging world. International Potato Center, Lima.
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635.
- Buln, A., P. Colonna, V. Planchot and S. Ball (1998) Starch granules: structure and biosynthesis. Int. J. Biol. Macromol. 23: 85–112.
- Cervantes-Flores, J.C., G.C. Yencho, A. Kriegner, K.V. Pecota, M.A. Faulk, R.O.M. Mwanga and B.R. Sosinski (2008) Development of a genetic linkage map and identification of homologous linkage groups in sweetpotato using multiple-dose AFLP markers. Mol. Breed. 21: 511–532.
- Cervantes-Flores, J.C., B. Sosinski, K.V. Pecota, R.O.M. Mwanga, G.L. Catignani, V.D. Truong, R.H. Watkins, M.R. Ulmer and G.C. Yencho (2011) Identification of quantitative trait loci for dry-matter, starch, and β -carotene content in sweetpotato. Mol. Breed. 28: 201–216.
- Haque, Tabuchi, Monden, Suematsu, Shirasawa, Isobe and Tanaka
- Cloutault, J., D. Peltier, V. Soufflet-Freslon, M. Briard and E. Geoffriau (2012) Differential selection on carotenoid biosynthesis genes as a function of gene position in the metabolic pathway: A study on the carrot and dicots. PLoS ONE 7: e38724.
- Crowell, S., P. Korniliev, A. Falco, A. Ismail, G. Gregorio, J. Mezey and S. McCouch (2016) Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. Nat. Commun. 7: 10527.
- Ducreux, L.J.M., W.L. Morris, P.E. Hedley, T. Shepherd, H.V. Davies, S. Millam and M.A. Taylor (2005) Metabolic engineering of high carotenoid potato tubers containing enhanced levels of beta-carotene and lutein. J. Exp. Bot. 56: 81–89.
- Food and Agricultural Organisation (FAO) (2009) FAO Statistics. In: <http://apps.fao.org>.
- Han, K., H.-Y. Lee, N.-Y. Ro, O.-S. Hur, J.-H. Lee, J.-K. Kwon and B.-C. Kang (2018) QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. Plant Biotechnol. J. 16: 1546–1558.
- Hirakawa, H., Y. Okada, H. Tabuchi, K. Shirasawa, A. Watanabe, H. Tsuruoka, C. Minami, S. Nakayama, S. Sasamoto, 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.
- Katayama, K., A. Kobayashi, T. Sakai, T. Kuranouchi and Y. Kai (2017) Recent progress in sweetpotato breeding and cultivars for diverse applications in Japan. Breed. Sci. 67: 3–14.
- Korte, A. and A. Farlow (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9: 29.
- Kreuze, J.F., J.P.T. Valkonen and M. Ghislain (2009) Genetic engineering. In: Loebenstein, G. and G. Thottappilly (eds.) The sweet-potato. Springer, Berlin, pp. 41–63.
- Kumagai, T. (2001) Breeding and varieties of sweetpotato. Farming Japan 35: 11–17.
- Low, J.W., M. Arimond, N. Osman, B. Cungaara, F. Zano and D. Tschirley (2007) A food-based approach introducing orange-fleshed sweet potatoes increased vitamin A intake and serum retinol concentrations in young children in rural Mozambique. J. Nutr. 137: 1320–1327.
- Nakayama, H., M. Tanaka, Y. Takahata, K. Matsui, H. Iwahori, Z. Sano and M. Yoshinaga (2012) Development of AFLP-derived SCAR markers associated with resistance to two races of southern root-knot nematode in sweetpotato. Euphytica 188: 175–185.
- Sakai, K. (1964) Studies on the enlargement of variations and the improvement of selection methods in sweet potato breeding. Bull. Kyushu Agr. Expt. Sta. 9: 247–397.
- Sasai, R., H. Tabuchi, K. Shirasawa, K. Kishimoto, S. Sato, Y. Okada, A. Kuramoto, A. Kobayashi, S. Isobe, M. Tahara *et al.* (2019) Development of molecular markers associated with resistance to *Meloidogyne incognita* by performing quantitative trait locus analysis and genome-wide association study in sweetpotato. DNA Res. 26: 399–409.
- Shirasawa, K., M. Tanaka, Y. Takahata, D. Ma, Q. Cao, Q. Liu, H. Zhai, S.-S. Kwak, J.-C. Jeong, U.-H. Yoon *et al.* (2017) A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweetpotato (*Ipomoea batatas*). Sci. Rep. 7: 44207.
- Tabuchi, H., T. Kuranouchi, A. Kobayashi, Y. Monden, K. Kishimoto, M. Tahara, Y. Okada and H. Iwahori (2017) Southern root-knot nematode race SP6 is divided into two races. Nematol. Res. 47: 29–33.
- Tian, F., P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T.R. Rocheford, M.D. McMullen, J.B. Holland and E.S. Buckler

- (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* 43: 159–162.
- Van Ooijen, J.W. (2009) MapQTL6, Software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V., Wageningen, Netherlands.
- Voorrips, R.E. (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93: 77–78.
- Woolfe, J.A. (1992) Sweetpotato: an untapped resource. Cambridge University Press, New York, p. 643.
- Yada, B., G. Brown-Guedira, A. Alajo, G.N. Ssemakula, E. Owusu-Mensah, E.E. Carey, R.O.M. Mwanga and G.C. Yencho (2017) Genetic analysis and association of simple sequence repeat markers with storage root yield, dry matter, starch and β -carotene content in sweetpotato. *Breed. Sci.* 67: 140–150.
- Yamakawa, O., M. Yoshinaka, T. Kumagai, M. Hidaka, K. Komaki, H. Kukimura and K. Ishiguro (1998) “J-Red”: A new sweetpotato cultivar. *Bull. Kyushu Agr. Expt. Sta.* 33: 49–72.
- Ye, X., S. Al-Babili, A. Klöti, J. Zhang, P. Lucca, P. Beyer and I. Potrykus (2000) Engineering the provitamin A (β -carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287: 303–305.
- Yu, X., N. Zhao, H. Li, Q. Jie, H. Zhai, S. He, Q. Li and Q. Liu (2014) Identification of QTLs for starch content in sweetpotato (*Ipomoea batatas* (L.) Lam.). *J. Integr. Agric.* 13: 310–315.
- Zhang, K., Z. Wu, D. Tang, C. Lv, K. Luo, Y. Zhao, X. Liu, Y. Huang and J. Wang (2016) Development and identification of SSR markers associated with starch properties and β -carotene content in the storage root of sweet potato (*Ipomoea batatas* L.). *Front. Plant Sci.* 7: 223.
- Zhao, N., X.X. Yu, Q. Jie, H. Li, H. Li, J. Hu, H. Zhai, S.Z. He and Q.C. Liu (2013) A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweetpotato. *Mol. Breed.* 32: 807–820.
- Zhu, C., M. Gore, E.S. Buckler and J. Yu (2008) Status and Prospects of Association Mapping in Plants. *Plant Genome* 1: 5–20.