

Genome-wide Development and Physical Mapping of SSR Markers in Sugar Beet (*Beta vulgaris* L.)

Ibrahim CELİK^{1,2*}

Highlights:

- A total of 22500 Sugar beet specific SSR primers were developed
- SSR markers increased the number of available sugar beet specific SSR markers by 55-fold
- A total of 102 SSR markers were found to be putatively associated with carbohydrate synthesis

ABSTRACT:

Sugar beet (*Beta vulgaris* L.) is one of the important sugar crops. Thus, development of molecular tools to understand molecular mechanism of agronomic traits such as root yield and sucrose content by using comprehensive genome analysis is important for sugar beet molecular breeding. A prime to such a comprehensive genome analysis is high throughout marker development. In the present study, mining of sugar beet genome for simple sequence repeats (SSRs) revealed 37704 motifs. Dinucleotide repeats were the most abundant accounting for 69.3 %. A total of 22500 SSR primers were developed and these markers were physically mapped in sugar beet genome. The markers distributed over nine chromosomes and the map had a higher resolution (16.73 kb/SSR). Also 102 SSR markers were found to be putatively associated with carbohydrate synthesis. The present study is the first report of genome-wide development of SSR markers and construction of SSR based physical map in sugar beet. SSR markers increased the number of available sugar beet specific SSR markers by 55-fold. SSR markers developed in the present study are valuable molecular genetic tools for comprehensive genome analysis to facilitate sugar beet molecular breeding.

Keywords:

- Genomics
- *Beta vulgaris* L
- Genomic SSR functional annotation

^{1,2} Ibrahim CELİK (Orcid ID: 0000-0002-6205-0930), ¹ Pamukkale University, Department of Agricultural and Livestock Production, Çal Vocational School of Higher Education, Denizli, Türkiye. ² Pamukkale University, Plant Genetics and Agricultural Biotechnology Application and Research Center (PAU BIYOM), Denizli, Türkiye

*Corresponding Author: Ibrahim CELİK, e-mail: icelik@pau.edu.tr

INTRODUCTION

Sugar beet (*Beta vulgaris* L.) is one of the important sugar crops cultivated worldwide. Approximately 22% of the world sugar need has been met by the crop (Fugate et al., 2014). Also, beet pulp which is a by-product of sugar production process from sugar beet is used as livestock feed (Kelly, 1983). In addition to importance of the crop for human diet and animal feed, sugar beet is an energy crop due to high content of sucrose provide efficient ethanol production (Panella, 2010). Thus, development of sugar beet cultivars with higher root yield and sucrose content is essential for sustainable sugar production (Wang et al., 2018). Comprehensive genome analysis of the crop is essential to reveal the molecular genetic mechanisms of these complex traits for efficient sugar beet breeding.

Cutting-edge comprehensive genome analysis methods such as QTL (Quantitative Trait Locus) mapping require a high number of locus-specific molecular markers distributed throughout the genome. Although general markers systems such as amplified fragment length polymorphism (AFLP), inter simple sequence repeat (ISSR) and random amplified polymorphic DNA (RAPD) were used for molecular characterization of sugar beet germplasm and linkage map construction, these marker systems were not efficient in linkage and QTL mapping (Jiang and Zeng, 1997). Thus, development of high-throughput locus specific molecular markers is necessary for efficient sugar beet genome analysis.

SSR markers are commonly used in plant genomics due to their codominant nature, locus specificity and high polymorphism rates (Li et al., 2002). However, crop specific SSR markers must be developed from expressed sequence tag (EST) and genomic sequences. In sugar beet four SSR markers were developed by Mörchen et al., (1996) using enriched genomic library for repeat types. In another study, 114 SSR marker were developed from enriched genomic library by Rae et al.. (2000). Genomic libraries of *B. vulgaris* ssp *maritima* and *B. vulgaris* ssp *vulgaris* were also used for SSR marker development and a total of 41 and 201 genomic SSR markers were developed by Laurent et al., (2007), respectively. In addition to genomic SSR markers, a total of 53 and 288 genic SSR markers were developed from EST sequences of sugar beet in two studies (Silva et al., 2012; Fugate et al., 2014). Despite these promising genomic studies in sugar beet, the number of available SSR markers (402 genomic and genic SSR markers) were not sufficient for comprehensive sugar beet genome analysis

Several genetic linkage maps were constructed by mainly RFLP (Restriction Fragment Length Polymorphism), RAPD (Random-Amplified Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism) and SRAP (Sequence Related Amplified Polymorphism) markers (Pillen et al., 1992; Pillen et al., 1993; Barzen et al., 1992; Barzen et al., 1995; Schneider et al., 2002; McGrath et al., 2007; Laurent et al., 2007; Wang et al., 2014;). However, these maps had low resolution. The map had highest resolution was constructed by Wang et al., (2018) by using Specific-Locus Amplified Fragment sequencing (SLAF-seq) technology. Although the map contained 3287 single nucleotide polymorphism (SNP) markers, these markers cannot be used for genome analysis unless converted to Cleaved Amplified Polymorphic Sequences (CAPS) or kompetitive alle specific PCR (KASP) assays. To overcome these limitations in sugar beet genomics, development of genome-wide SSR markers and construction of SSR-based physical map in sugar beet was performed. These current SSR markers can be used in comprehensive sugar beet genome analysis.

MATERIALS AND METHODS

Genome-wide SSR Mining

Genome of sugar beet was downloaded from NCBI genome database (RefBeet-1.2.2, GCF_000511025.2). Unplaced scaffolds and chloroplast genome were excluded and all chromosome sequences were used for SSR mining. five repeat types (from dinucleotide to hexanucleotide repeats) had more than six repeats were identified using GMATA high-throughput SSR marker development pipeline (Wang and Wang, 2016).

Development of SSR Markers

PCR primers targeting SSRs in the genome were designed using Primer3 software provided in GMATA pipeline software. PCR product size and optimum annealing temperature of primers were 120–400 bp and 60°C, respectively.

Functional Annotation of SSR Markers

The sequence file containing identified SSRs were converted in FASTA format and functional annotation and gene ontology analysis was performed to determine molecular function of SSRs using Go Feat software (Araujo et al., 2018).

Physical Map Construction

Primer positions of each chromosome (Mb) were recorded and used in physical map construction. Also, in silico PCR for all SSR markers was performed by using e-PCR software provided GMATA pipeline with default parameters (Schuler, 1997). The map was drawn using MapChart software (Voorrips, 2002). Maps of each chromosome were analyzed and reported separately due to containing a large number of markers.

Supplementary Data

All Supplementary Data files (Online Resource 1, 2, 3, 4 and 5 contain Figure S1, S2, S3, S4, S5 and Table S1; Sugar beet specific genomic SSR markers; Physical map of SSR markers; In silico PCR results of SSR markers; Functional annotation of SSR markers, respectively) data can be downloaded from figshare database (https://figshare.com/articles/dataset/Supplementary_Data/21303021)

RESULTS AND DISCUSSION

Genome-wide SSR Mining

The genome of sugar beet comprises 567 Mb was screened for SSRs. A total of 37704 SSRs were identified over nine chromosomes. SSR length ranged from 12 to 2680 nt with a mean value of 28.5 ± 0.22 nt. The most abundant SSR types were dinucleotide SSRs, which comprised more than half of the total SSRs (69.3 %), followed by trinucleotide repeats (24.6%). The rest of repeat types were rare and accounted for less than 4% of all SSRs: 3.8%, 1.2% and 1.2% for pentanucleotide, hexanucleotide and tetranucleotide repeats, respectively (Table 1). Frequency analysis of the repeat motifs demonstrated that TA/TA AT/AT TC/GA and AG/CT were the most abundant SSR motifs which accounted for 21.6%, 17.2%, 13.2% and 10% of all SSRs, respectively (Table 2).

Flanking sequences of identified SSR repeats were used to design PCR primers. As results, a total of non-redundant 22500 primers (59.6%) targeting 24959 (66.2%) SSR loci were developed. Primer pairs could not have designed due to lack of desired primer properties. Detailed primer data are given in Online Resource 2.

Table 1. Frequency of SSR types

Motif length	Number of SSRs	Frequency (%)
Dinucleotide	26114	69.3
Trinucleotide	9292	24.6
Pentanucleotide	1417	3.8
Hexanucleotide	447	1.2
Tetranucleotide	434	1.2

Table 2. Frequency of SSR motif types

SSR motif	Number of SSRs	% of SSR motif
TA/TA	8153	21.6
AT/AT	6501	17.2
TC/GA	4977	13.2
AG/CT	3756	10.0
AAT/ATT	2143	5.7
TTA/TAA	1785	4.7
ATA/TAT	1755	4.7
CA/TG	1414	3.8
AC/GT	1233	3.3
CAA/TTG	644	1.7
GTT/AAC	438	1.2
TTCAG/CTGAA	411	1.1
ACA/TGT	402	1.1
CAC/GTG	241	0.6
ATG/CAT	227	0.6
ATC/GAT	209	0.6
GGT/ACC	206	0.5
TCAGA/TCTGA	203	0.5
TGG/CCA	192	0.5
TGA/TCA	174	0.5
TTC/GAA	144	0.4
TCT/AGA	108	0.3
CTT/AAG	102	0.3

The sugar beet genome was screened for simple sequence repeats. Dinucleotide repeat types were the most abundant. Previous studies performed in *Cucumis melo*, *Daucus carota*, *Sorghum bicolor*, *Oryza sativa*, *Arabidopsis thaliana* and *Medicago truncatula* also reported that dinucleotide is the most abundant after mononucleotide repeats (Sonah et al., 2011; Zhu et al., 2016; Uncu and Uncu, 2020). Mononucleotide repeats were not identified in the present study due to insufficient polymorphism rate. In contrast to genomic SSRs, trinucleotide repeats in the most frequent in ESTs of sugar beet (Fugate et al., 2014). The reason of that is negative selection pressure on dinucleotide repeats in coding region due to resulting frameshift mutations (Cloutier et al., 2009). Frequency of SSR motif types of sugar beet was similar to previous studies performed in *Cucumis melo*, *Daucus carota*, *Sorghum bicolor*, *Arabidopsis thaliana*, *Medicago truncatula* and *Populus trichocarpa* reported that AT/AT and TA/TA were the most prevalent motif types. However, AG/CT repeat was the most prevalent motif type in *Brachypodium distachyon* and *Oryza sativa*. The reason of species-specific frequency of some repeats types was not clear due to undefined evolution mechanisms of SSRs in plant genomes (Sonah et al., 2011). In the present study, SSR markers were developed targeting more than half of SSR identified in the genome. These new SSR primers increased number of available sugar beet specific SSR markers by 55-fold.

Physical Map Construction

Positions of SSR markers designed in the present study were used to construct a genome-wide SSR marker based physical map (Online Resource 1, Figure S1, S1-5). As results, a total of 22500 SSR markers were mapped over nine chromosomes. Number of SSR markers ranged from 1636 to

3755 and size of the chromosomes ranged from 26.54 to 60.93 Mb. Largest chromosome (60.93 Mb) had the highest number of markers (3755). The smallest chromosome (26.54 Mb) had lowest number of markers (1,636) (Table 3). There was strong correlation between size of chromosomes and number of markers ($R^2=0.997$). Total size of the SSR-based map covered 78.1% of sequenced chromosomes (481.95 Mb) and 52.7% of the estimated total genome (714 Mb) of sugar beet.

Table 3. Numbers and frequencies of SSR markers in sugar beet genome.

Chromosome	Number of markers	Map distance (Mb)	Resolution (kb/Marker)
Chromosome 1	2089	34.91	16.71
Chromosome 2	2413	40.38	16.73
Chromosome 3	1636	26.54	16.22
Chromosome 4	1940	33.00	17.01
Chromosome 5	3110	52.44	16.86
Chromosome 6	3755	60.93	16.23
Chromosome 7	2572	44.10	17.15
Chromosome 8	2295	38.77	16.89
Chromosome 9	2690	45.27	16.83
Total	22500	376.34	16.73

Total resolution of the map was 16.73 kb/SSR (1 SSR marker every 16.73 kb). Resolution of chromosomes ranged from 16.22 to 17.15 kb/SSR. Chromosome 3, 6 and 1 had slightly higher resolution (16.22, 16.23 and 16.71 kb/SSR) than total map resolution (16.73 kb/SSR). Chromosome 2 had resolution as much as total map resolution. Rest of the chromosomes (2, 4, 5, 7, 8 and 9) has slightly lower resolution than total map resolution (Table 3 and Online Resource 3). Although SSR markers equally distributed in the genome, there were gaps between markers. Size of the gaps ranged from 0.012 kb to 210.94 kb with a mean value of 16.72 kb. Minimum gap size was 0.012 kb for all chromosomes except for chromosome 3 and 8 (0.013 kb).

Table 4. Average and range of gaps in SSR based map in sugar beet.

Chromosome	Minimum gap size (kb)	Maximum gap size (kb)	Average gap size (kb)	Number of gaps 100 kb <
Chromosome 1	0.012	196.79	16.70	15
Chromosome 2	0.012	149.00	16.74	15
Chromosome 3	0.013	197.17	16.23	13
Chromosome 4	0.012	160.77	16.99	10
Chromosome 5	0.012	161.09	16.86	21
Chromosome 6	0.012	151.97	16.23	19
Chromosome 7	0.012	210.94	17.15	22
Chromosome 8	0.013	156.25	16.88	11
Chromosome 9	0.012	195.76	16.83	24
Total				150

Maximum gap size ranged from 149 kb (chromosome 2) to 210.94 kb (chromosome 7). Chromosome 7 with the 17.15 kb gap had the largest average gap size. There were 150 gaps larger than 100 kb in the genome. Number of these gaps were ranged from 10 (chromosome 4) to 24 (chromosome 9) (Table 4 and Online Resource 3).

This map constructed in the present study had a higher number of markers than linkage maps constructed by Laurent et al., (2007) containing 284 SSR markers, McGrath et al. (2007) containing 331 AFLP and SSR markers and Wang et al. (2014) containing 561 SRAP and 114 SSR markers. The present map also has more markers than the most comprehensive linkage map containing 3283 SLAF-seq based SNP markers constructed by Wang et al., (2018). In addition, the present map had higher map resolution than previous linkage maps due to large number of markers. However, the map lower resolution than SSR based physical map of carrot (*Daucus carota* L.) (5.38 kb/SSR) and cucumber (*Cucumis sativus* L.) (0.55 kb/SSR) (Cavagnaro et al., 2010; Uncu and Uncu, 2020). Even though SSR

markers well distributed in the genome, there are some gaps as expected. Other marker systems needed to be developed targeting these gaps such as SNP markers that have higher frequencies than SSR markers.

Linkage maps are valuable genomic resources to reveal genome structure and organization, single gene and QTL mapping. However, accuracy of linkage maps depends on several parameters such as population types and size of mapping populations. In addition, meta-QTL analysis is difficult due to insufficient number of common markers between linkage maps. Thus, functionality of linkage mapping decreased in plant genome analysis since genome of many agronomic plants were available. Plant genome sequences provide an efficient consensus map framework for mapping all the markers linked to genes and QTLs for meta-analysis of various mapping studies to reveal complex molecular mechanisms. In future study, all developed sugar beet specific markers can be mapped in the genome. Also, physical maps are very useful in QTL mapping studies performed in unbalanced population such as IBL (Inbred Backcross Lines) in which impossible to perform linkage analysis (Celik et al., 2017).

In silico PCR for primers were performed to determine the number of targets. A total of 20202 (89.3% of all primers) SSR markers targeted single locus while 1,185 (5.3%) primers targeted two loci in the genome. Rest of the markers (1,213) targeted more than two loci in the genome. In silico PCR results of all markers are given in Online Resource 4. In silico PCR analysis showed that most of the primers are single-locus markers. The number of these single locus markers was still sufficient for comprehensive genome analysis and resolution of the map with single locus markers did not change much (18.6 kb/SSR, 11% decreased). Although multilocus SSR markers targeting duplicated loci in the genome can be used in plant genome analysis, analysis of PCR fragments is more complicated.

Functional Annotation of SSR Markers

Functional annotation of flanking sequences contained SSR markers was performed based on three gene ontology terms. Molecular functions of 7465 (33.52%) SSRs were predicted. ATP, nucleic acid, DNA and metal ion binding were the most common molecular functions with 706, 535, 407 and 303 SSRs, respectively. A total of 33 SSR markers were found to be putatively associated with carbohydrate synthesis (29 carbohydrate binding, three carbohydrate derivative binding and one carbohydrate transmembrane transporter activity). A total of 3532 (15.86%) SSR markers were assigned to biological processes. The most common biological processes were DNA integration, regulation of transcription and transmembrane transport with 173, 136 and 91 hits, respectively. A total of 87 SSR markers found to be putatively associated with carbohydrate synthesis in six biological process categories. Also, cellular component of 11277 (50.63%) SSR markers were determined. As result, integral component of membrane and nucleus were the most common with 8950 and 446 hits, respectively. Gene ontology data of SSR markers are given in Online Resource 5.

The present study performed functional annotation of sequences contained SSR markers. Molecular function and biological processes of most of the SSR markers (66.48% and 84.14%, respectively) were not identified. The SSR markers had no significant hits to GO terms might have targeted non-coding sequences. Although functional classification of sugar beet transcriptome was reported by Fugate et al., (2014), the present study performed functional annotation of genome-wide SSR markers for the first time. Functional annotation also revealed a total of 102 SSR markers putatively associated with carbohydrate synthesis for the first time (Online Resource 1, Table S1). These markers can be used in mapping studies for revealing the molecular genetic mechanism of carbohydrate synthesis which is the most important trait in sugar beet.

CONCLUSION

In the present study, sugar beet genome was screened for SSR and a total of 22500 SSR markers were developed distributed over nine chromosomes. The study increased the number of available SSR markers by 55-fold and generated a high-resolution map. This is the first study of genome-wide mining of SSRs and high-resolution SSR based physical map. The map is a valuable genomic resource for molecular breeding of sugar beet and the markers can be used in comprehensive genome analysis such as QTL mapping.

Conflict of Interest

The author declared that there is no conflict of interest.

REFERENCES

- Araujo, F. A., Barh, D., Silva, A., Guimaraes, L., & Ramos, R. T. J. (2018). GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. *Scientific reports*, 8(1), 1794.
- Barzen, E., Mechelke, W., Ritter, E., Seitzer, J. F., & Salamini, F. (1992). RFLP markers for sugar beet breeding: chromosomal linkage maps and location of major genes for rhizomania resistance, monogerm and hypocotyl colour. *The Plant Journal*, 2(4), 601-611.
- Barzen, E., Mechelke, W., Ritter, E., Schulte-Kappert, E., & Salamini, F. (1995). An extended map of the sugar beet genome containing RFLP and RAPD loci. *Theoretical and Applied Genetics*, 90, 189-193.
- Cavagnaro, P. F., Senalik, D. A., Yang, L., Simon, P. W., Harkins, T. T., Kodira, C. D., ... & Weng, Y. (2010). Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC genomics*, 11, 1-18.
- Celik, I., Gurbuz, N., Uncu, A. T., Frary, A., & Doganlar, S. (2017). Genome-wide SNP discovery and QTL mapping for fruit quality traits in inbred backcross lines (IBLs) of *Solanum pimpinellifolium* using genotyping by sequencing. *BMC genomics*, 18(1), 1-10.
- Cloutier, S., Niu, Z., Datla, R., & Duguid, S. (2009). Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). *Theoretical and Applied Genetics*, 119, 53-63.
- Fugate, K. K., Fajardo, D., Schlautman, B., Ferrareze, J. P., Bolton, M. D., Campbell, L. G., Zalapa, J. (2014). Generation and Characterization of a Sugarbeet Transcriptome and Transcript-Based SSR Markers. *The Plant Genome*, 7(2), plantgenome2013-11.
- Jiang, C., Zeng, Z. B. (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica*, 101, 47-58.
- Kelly P, 1983. Sugar beet pulp—a review. *Animal Feed Science and Technology*, 8(1): 1-18.
- Laurent, V., Devaux, P., Thiel, T., Viard, F., Mielordt, S., Touzet, P., Quillet, M. C. (2007). Comparative effectiveness of sugar beet microsatellite markers isolated from genomic libraries and GenBank ESTs to map the sugar beet genome. *Theoretical and applied genetics*, 115, 793-805.
- Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular ecology*, 11(12), 2453-2465.
- McGrath, J. M., Trebbi, D., Fenwick, A., Panella, L., Schulz, B., Laurent, V., ... & Murray, S. C. (2007). An open-source first-generation molecular genetic map from a sugarbeet × table beet cross and its extension to physical mapping. *Crop Science*, 47, S-27.

- Mörchen, M., Cuguen, J., Michaelis, G., Hänni, C., & Saumitou-Laprade, P. (1996). Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris* L. *Theoretical and Applied Genetics*, 92, 326-333.
- Panella, L. (2010). Sugar beet as an energy crop. *Sugar Tech*, 12, 288-293.
- Pillen, K., Steinrücken, G., Wricke, G., Herrmann, R. G., Jung, C. (1992). A linkage map of sugar beet (*Beta vulgaris* L.). *Theoretical and Applied Genetics*, 84, 129-135.
- Pillen, K., Steinrücken, G., Herrmann, R. G., Jung, C. (1993). An extended linkage map of sugar beet (*Beta vulgaris* L.) including nine putative lethal genes and the restorer gene X. *Plant breeding*, 111(4), 265-272.
- Rae, S. J., Aldam, C., Dominguez, I., Hoebrechts, M., Barnes, S. R., & Edwards, K. J. (2000). Development and incorporation of microsatellite markers into the linkage map of sugar beet (*Beta vulgaris* spp.). *Theoretical and Applied Genetics*, 100, 1240-1248.
- Schneider, K., Schäfer-Pregl, R., Borchardt, D., & Salamini, F. (2002). Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. *Theoretical and Applied Genetics*, 104, 1107-1113.
- Schuler, G. D. (1997). Sequence mapping by electronic PCR. *Genome research*, 7(5), 541-550.
- Silva, D. C., de Souza, M. C. P., Filho, L. S. C. D., dos Santos, J. M., de Souza Barbosa, G. V., & Almeida, C. (2012). New polymorphic EST-SSR markers in sugarcane. *Sugar Tech*, 14, 357-363.
- Sonah, H., Deshmukh, R. K., Sharma, A., Singh, V. P., Gupta, D. K., Gacche, R. N., ... & Sharma, T. R. (2011). Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *Plos one*, 6(6), e21298.
- Wang, M., Li, B., & Wang, H. (2014). Construction of molecular genetic linkage map of sugarbeet. *Acta Agronomica Sinica*, 40(2), 222-230.
- Wang, M., Xu, Y., Wu, Z., Wang, H., Zhang, H. (2018). High-density genetic map construction in sugar beet (*Beta vulgaris* L.) by high-throughput technology. *Sugar tech*, 20, 212-219.
- Wang, X., & Wang, L. (2016). GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in plant science*, 7, 1350.
- Uncu, A. O., & Uncu, A. T. (2020). High-throughput simple sequence repeat (SSR) mining saturates the carrot (*Daucus carota* L.) genome with chromosome-anchored markers. *Biotechnology & Biotechnological Equipment*, 34(1), 1-9.
- Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of heredity*, 93(1), 77-78.
- Zhu, H., Guo, L., Song, P., Luan, F., Hu, J., Sun, X., & Yang, L. (2016). Development of genome-wide SSR markers in melon with their cross-species transferability analysis and utilization in genetic diversity study. *Molecular Breeding*, 36, 1-14.