

# Genome-wide SNP discovery and genetic linkage map construction in sunflower (*Helianthus annuus* L.) using a genotyping by sequencing (GBS) approach

Ibrahim Celik · Sabahattin Bodur · Anne Frary · Sami Doganlar

Received: 19 October 2015 / Accepted: 2 September 2016 / Published online: 9 September 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** Recently developed plant genomics approaches (LD mapping and genome-wide selection) require many molecular markers distributed throughout the plant genome. As a result, the availability of an increasing number of markers is essential for maintaining highly efficient and accurate plant breeding programs. In this study, we identified SNP loci in sunflower using a genotyping by sequencing (GBS) approach in an intraspecific  $F_2$  mapping population. A total of 271,445,770 reads were generated by the Genome Analyzer II next-generation sequencing platform and 29.2 % of the reads were aligned to unique locations in the genome. A total of 46,278 SNP loci were identified and 7646 SNP loci were validated in an  $F_2$  population. In addition, a SNP-based linkage map was constructed. This is the first report of SNP discovery in sunflower by GBS. The SNP markers and SNP-based linkage map will be valuable molecular genetics tools for sunflower breeding.

**Keywords** Genetic linkage map · Genotyping · Illumina Genome Analyzer II platform · Next-generation sequencing · Single-nucleotide polymorphism

## Introduction

Many approaches in plant genetics and breeding such as association mapping (AM), QTL linkage mapping and genome-wide selection (GS) require high-throughput marker systems with genome-wide coverage (Mammadov et al. 2012). Single-nucleotide polymorphism (SNP) markers fulfill the criteria for such approaches due to their frequency, reproducibility and genome-wide distribution (Mammadov et al. 2012; Huang and Han 2013). The discovery of a large number of SNPs in sunflower (*Helianthus annuus* L.) is necessary because sunflower is an agronomically important oilseed crop and the application of molecular markers will improve the efficiency and accuracy of selection in sunflower breeding. Two studies described the development of SNP loci by sequencing the transcriptomes of eight sunflower lines using the Roche 454 GS FLX platform (10,640 SNPs; Bachlava et al. 2012) and by sequencing of six sunflower inbred lines using restriction site-associated DNA sequencing (RAD-Seq) and Illumina paired-end sequencing technology (16,467 SNPs; Pegadaraju et al. 2013). In addition, Livaja et al. (2016) recently discovered 20,502 high-quality SNP loci by amplicon sequencing.

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s11032-016-0558-8) contains supplementary material, which is available to authorized users.

---

I. Celik · A. Frary · S. Doganlar (✉)  
Department of Molecular Biology and Genetics, Izmir  
Institute of Technology, Urla, Izmir, Turkey  
e-mail: samidoganlar@iyte.edu.tr

S. Bodur  
Agromar, Karacabey, Bursa, Turkey

Genotyping by sequencing (GBS) is a high-throughput and practical SNP discovery approach which involves sequencing of a plant population using multiplexing (bar coding) and next-generation sequencing (NGS) technologies (Kumar et al. 2012). GBS can be used for SNP discovery and genotyping of plants by mapping the sequence reads of individual plants on a reference genome (Elshire et al. 2011; Huang and Han 2013). To date, the GBS approach has been popular and applied to many plant species for SNP discovery and genotyping. These species include barley (Elshire et al. 2011), *Miscanthus sinensis* (Ma et al. 2012), *Rubus idaeus* (Ward et al. 2013) and maize (Romy et al. 2013). Also GBS was applied in an F<sub>2</sub> population of oil palm to identify 21,471 SNPs and to construct a linkage map containing 1085 SNPs (Pootakham et al. 2015).

Although approximately 27,000 SNP markers were developed by sequencing of a few sunflower accessions (Bachlava et al. 2012; Pegadaraju et al. 2013), no high-throughput SNP discovery study has been performed in sunflower using GBS. Thus, the first aim of this study was to discover SNPs in the sunflower genome using such an approach while the second goal was to construct a linkage map from these markers.

## Materials and methods

### Plant materials

An intraspecific F<sub>2</sub> population consisting of 300 individuals was generated by self-pollinating F<sub>1</sub> hybrids from a cross of *H. annuus* cv. RHA 436 × *H. annuus* cv. H08 M1. Plants were grown in the field in Bursa, Turkey, in 3 m × 0.3 m rows with row spacing of 0.7 m. Soil was fertilized with 15:15:15 fertilizer (25 kg ha<sup>-1</sup>) and nitrogen (N), phosphorus (P) and potassium (K) (3.75 kg ha<sup>-1</sup> for each element). Total genomic DNA was isolated from leaf tissue of the two parents (RHA 436 and H08 M1) and 300 F<sub>2</sub> individuals using a CTAB method (Doyle and Doyle 1990). Genomic DNA of individuals was quantified using Qubit<sup>TM</sup> quantitation assay (Life Technologies) and agarose gel electrophoresis. Genomic DNA of the two parents (RHA 436 and H08 M1) and 93 F<sub>2</sub> individuals selected from the population for the best DNA quality and quantity were used for GBS.

### Genotyping by sequencing (GBS)

Genomic DNA of the 95 sunflower genotypes (93 F<sub>2</sub> individuals plus 2 parents) was subjected to GBS by the University of Wisconsin Biotechnology Center (<https://www.biotech.wisc.edu/>, Madison, WI USA). Genomic DNA digestion, common and bar code adaptor ligation, sample pooling and amplification for sequencing library construction were performed according to protocols described by Elshire et al. (2011). Single-end sequencing of the 95-plex library was performed with the Genome Analyzer II next-generation sequencing platform (Illumina Inc. San Diego, CA).

### GBS data analysis and SNP identification

The FASTQ file format generated from raw sequence reads by the CASAVA 1.8.2 software package (Illumina Inc.) was used for further processing in the GBS discovery pipeline of TASSEL software (Glaubitz et al. 2014). The FASTQ and sample key file (containing the plate layout and bar codes for each genotype) were used as input files in the pipeline. Sequence reads containing the expected bar codes followed by an *Ape*KI cut site remnant (CWGC) were trimmed to 64 bases using the FastqToTagCountPlugin of the pipeline. Reads containing unidentified bases (N) were excluded from analysis. The bar-coded sequence reads were sorted and collapsed into unique sequence tags with counts using the plug-in with default parameters. The only exception was that minimum number of times a tag must be present to be output was altered to 3. After filtration using a minimum count threshold of 3, tag count files were merged into a master file using the MergeMultipleTagCountPlugin. The master tags were converted to FASTQ format by the TagCountToFastqPlugin using default parameters except the minimum count of reads for a tag was 3. The tags were then aligned to the sunflower draft genome (Celera\_14libs\_sspace2\_ext-final.scaffolds.split.fasta.bz2) ([http://sunflowergenome.org/early\\_access/repository/main/genome/](http://sunflowergenome.org/early_access/repository/main/genome/)) using the Bowtie 2 plug-in with default parameters (Langmead and Salzberg 2012). Overall, these steps of the pipeline were used to filter and correct most of the sequencing errors (Glaubitz et al. 2014). The output of the alignment was converted to a “Tags On Physical Map” (TOPM) file. This file contained information

about the physical positions of the master tags which had the best unique alignment with the genome as generated by the SAMConverterPlugin. In addition, a “Tags by Taxa” (TBT) file which contained sorted and demultiplexed reads according to their bar code adapters was generated by the FastqToTBTPlugin. Finally, the TOPM and TBT files were used for SNP calling using the TagsToSNPByAlignmentPlugin. Parameters used for SNP calling and filtration are given in Table S1. SNPs were recorded in a HapMap file for each linkage group. MergeDuplicateSNPsPlugin was used to merge the duplicate SNPs in the HapMap file. A physical map of the identified SNPs was drawn using Mapchart software (Voorrips 2002).

### Construction of linkage map

The SNP markers developed in this study were further filtered for linkage map construction. SNP markers with more than 10 % missing data and which deviated significantly from the expected 1:2:1 Mendelian segregation ratio (Chi-square goodness-of-fit test,  $p < 0.01$ ) were excluded from linkage analysis. The linkage map was constructed using JoinMap version 3.0 software (Ooijen and Voorrips 2001) with  $F_2$  population-type parameters. Linkage groups were determined using a minimum LOD score of 3.0 and maximum recombination ratio of 0.5. Default parameters of the maximum likelihood algorithm were used for ordering of markers in linkage groups. Map distances were calculated using the Kosambi mapping function. Correlations between the linkage and physical maps were calculated for each linkage group using the positions of SNPs to compare the maps using PASW software (Norusis 2010).

## Results

### GBS

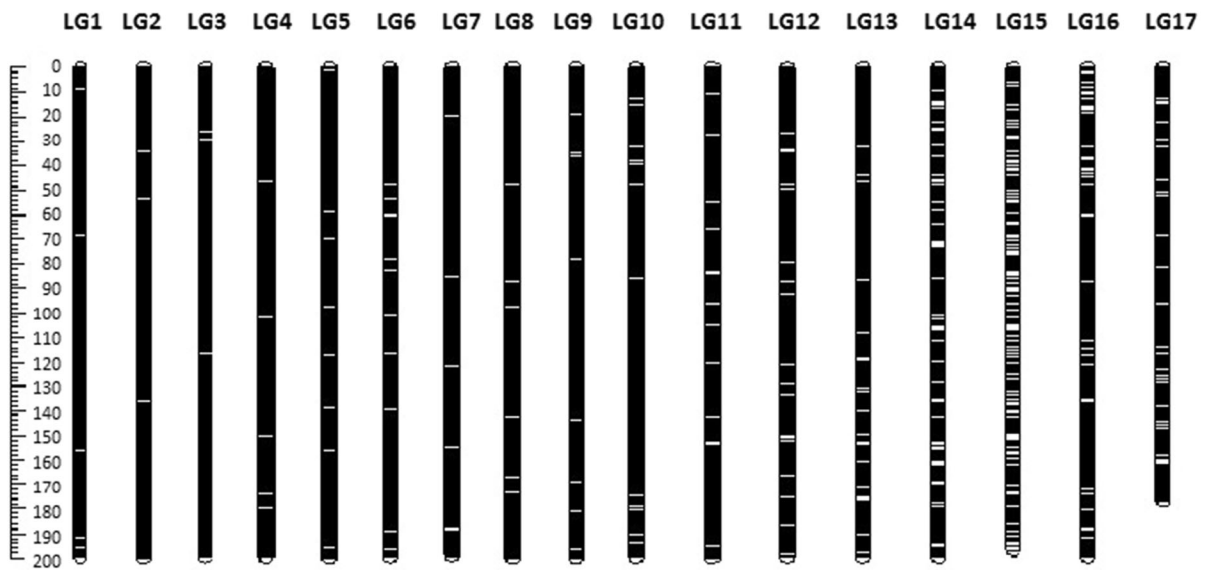
A total of 93 individuals and the two parents of the  $F_2$  population were subjected to GBS and analysis. A total of 271,445,770 sequence reads were generated from the 95-plex library Genome Analyzer II next-generation sequencing platform. From these reads, 1,208,784 tags were generated. While 29.2 % (353,304) of the sequence tags were uniquely aligned

to the sunflower genome, 14.2 % (172,196) of the tags were aligned to multiple positions. A majority of the sequence tags (56.5 %, 683,284) could not be aligned to the genome. Only those that were uniquely aligned to the genome were used for SNP calling by the TASSEL software pipeline. Tags averaged 64 bases and covered 22.6 Mb. Sequences of the merged GBS tags can be accessed at <https://figshare.com/s/7e40b56de730e94f9687>.

### GBS data analysis and SNP identification

A total of 46,278 SNP loci were discovered in the sunflower genome based on the locations of the unique alignment of merged tags (353,304) in the genome which contained 17 linkage groups (LG1–LG17). Thus, the SNP loci identified in this study were distributed on all 17 linkage groups. Physical mapping of the identified SNPs in the sunflower genome showed that the SNPs were well distributed with high coverage (Fig. 1). After filtration of the SNPs according to TASSEL software parameters (Table S1), a total of 9535 (20.6 %) SNPs were retained. Average number of filtered SNPs per tag was 0.03. While the frequency (resolution) of unfiltered SNPs was 72.2 kb/SNP (1 SNP every 72.2 kb), the frequency of filtered SNPs was 350.6 kb/SNP (Table 1). For both unfiltered and filtered SNPs, LG2 had the most SNPs: 3849 and 879 unfiltered and filtered SNPs, respectively. As a result, LG2 had the highest resolution among chromosomes. LG15 had the fewest SNPs with 834 and 91 unfiltered and filtered SNPs, respectively) (Table 1). Most of the SNPs (58.7 %) identified in this study were transition mutations (A/G or C/T) (Table 2). The most observed substitution type was A/G (29.7 %). The least common substitution type was C/G transversion (6.9 %). The observed transition/transversion ratio was 1.42.

The filtered SNPs were further analyzed to identify only SNPs that were polymorphic for the parents of the  $F_2$  population. In this way, a total of 7646 (16.5 % of the total unfiltered SNPs) polymorphic SNPs were identified. Frequency (resolution) of polymorphic SNPs between parents was 437 kb/SNP (1 SNP every 437 kb). Physical mapping of these SNPs showed that they were well distributed in the sunflower genome with some gaps on LGs 14, 15, 16 and 17 (Fig. 2).



**Fig. 1** Physical map of 46,278 unfiltered SNPs in sunflower genome. *Black-colored* locations were highly saturated with SNPs. Detailed information about physical map and SNPs is available in supplemental Excel file

**Table 1** Numbers and frequencies of unfiltered, filtered and polymorphic SNP loci in sunflower linkage groups

	Size of LG (Mb)	Number of SNPs			Frequency (kb) of SNPs		
		Unfilt.	Filt. <sup>a</sup>	Poly. <sup>b</sup>	Unfilt.	Filt. <sup>a</sup>	Poly. <sup>b</sup>
LG 1	199.6	3735	832	681	53.4	239.9	293.1
LG 2	198.1	3849	879	685	51.5	225.4	289.3
LG 3	199.1	3427	803	653	58.1	247.9	304.9
LG 4	199.6	3639	757	611	54.9	263.7	326.7
LG 5	199.2	3332	694	536	59.8	287.1	371.7
LG 6	197.7	3224	668	535	61.3	295.9	369.5
LG 7	198.1	3329	796	654	59.5	248.9	303.0
LG 8	198.5	3240	618	494	61.3	321.2	401.9
LG 9	200	3244	703	566	61.7	284.5	353.4
LG 10	199.9	2865	623	519	69.8	320.9	385.2
LG 11	195.9	2830	677	542	69.2	289.3	361.4
LG 12	197.7	2379	547	441	83.1	361.5	448.4
LG 13	195.6	2020	354	272	96.9	552.7	719.3
LG 14	199.9	1470	185	136	136	1080.5	1469.8
LG 15	195.8	834	91	66	234.7	2151.3	2966.2
LG 16	196.1	1609	211	185	121.9	929.5	1060.1
LG 17	171.8	1252	97	70	137.2	1770.8	2453.8
Total	3342.7	46278	9535	7646	72.2	350.6	437.2

<sup>a</sup> Number and frequency of SNPs after the filtration using TASSEL software parameters

<sup>b</sup> Number and frequency of filtrated SNPs based on missing data (<20 %) which were also polymorphic for parents

Linkage map construction

After elimination of SNPs with too much missing data (more than 10 %) or segregation distortion, a total of

1548 SNPs were retained for linkage map construction. A total of 6098 SNPs were excluded from linkage map construction due to high levels of missing data and segregation distortion. Most of the SNPs (73.8 %

**Table 2** Types of substitutions represented by the identified SNP loci

Transition/transversion	Number of SNPs	Frequency (%)
Transition		
C/T	13,439	29.0
A/G	13,723	29.7
Total	27,162	58.7
Transversion		
C/G	3192	6.9
A/C	5035	10.9
G/T	5043	10.9
A/T	5846	12.6
Total	19,116	41.3

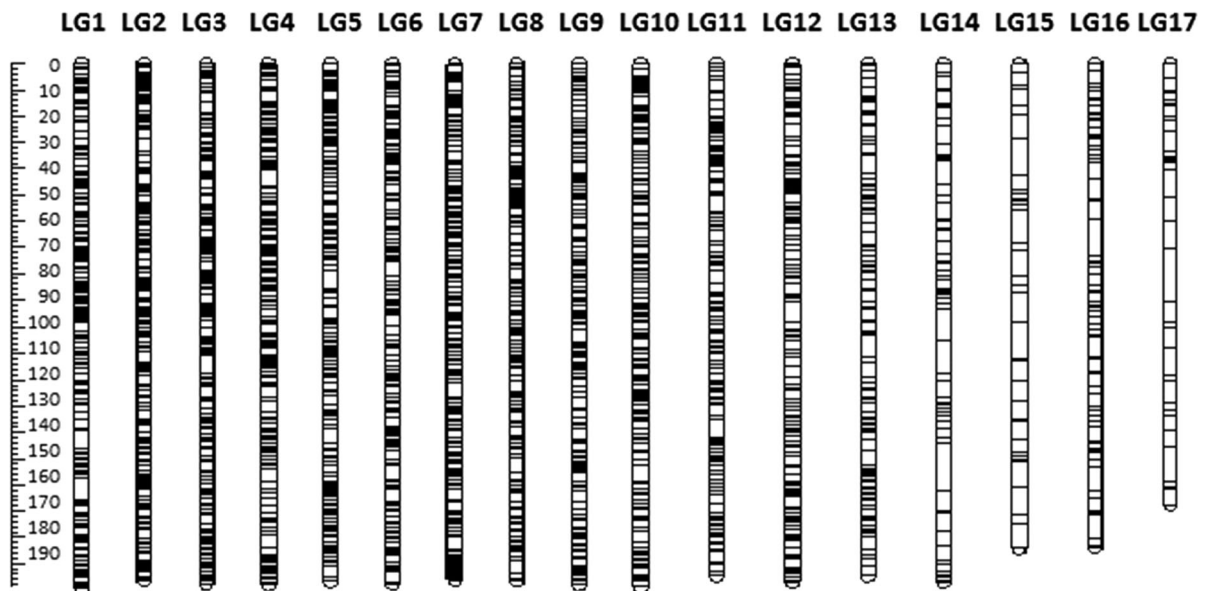
of the total SNPs polymorphic between the parental lines) were excluded due to high levels of missing data (more than 10 %), the rest (26.2 %) were excluded due to segregation distortion. A genetic linkage map consisting of 817 SNPs distributed over 17 linkage groups was constructed (Fig. 3). The total distance of the map was 2472 cM. The size of the linkage groups varied from 17 cM (LG15 with 7 SNPs) to 236 cM (LG11 with 236 SNPs) (Table 3). Average distance between SNPs in the map was  $3.0 \pm 0.2$  cM. The genetic linkage map was compared to the physical

map of sunflower, and the correlation between SNP positions was calculated. A moderate level of correlation was determined for all linkage groups with  $r = 0.70$  except for linkage groups LG4 and LG10 ( $r = 0.39$  and  $0.20$ , respectively).

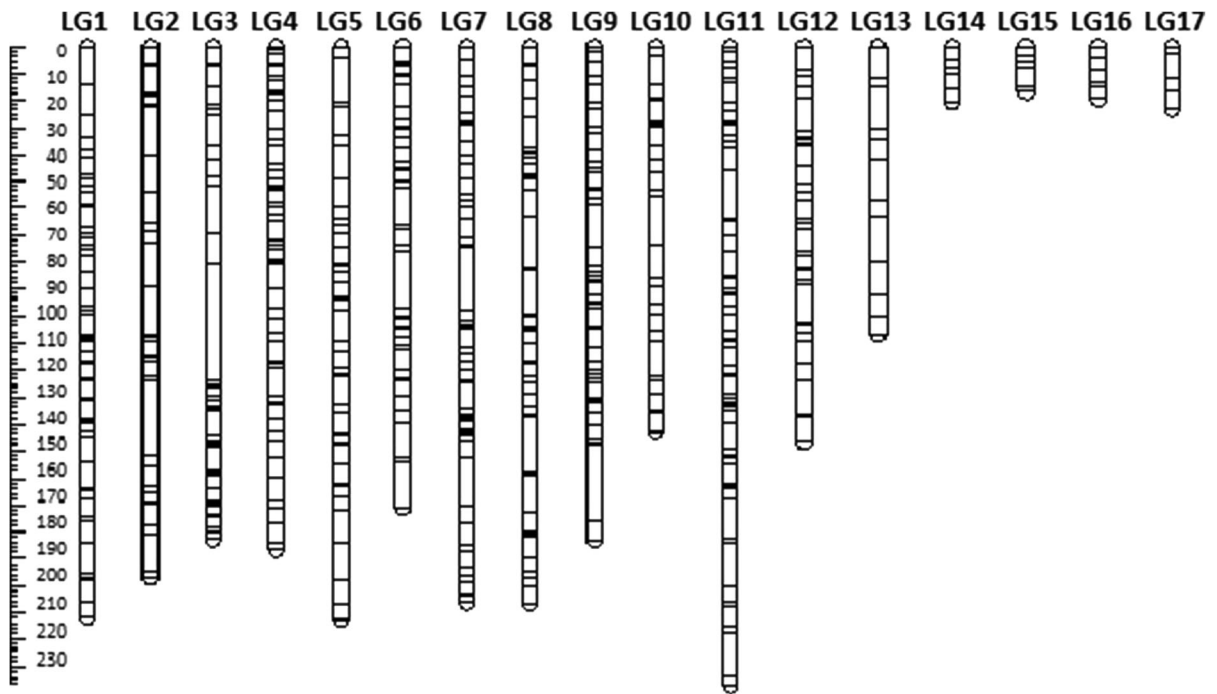
## Discussion

### GBS and SNP identification

The genotyping by sequencing approach has been applied to many species for SNP discovery. Although the numbers of sequence reads and percentages of uniquely aligned reads were variable in previous work, comparison indicates that the sunflower GBS performed in this study was efficient in terms of read number. For example, total reads of *Rubus idaeus* generated by GBS (135,776,036) (Ward et al. 2013) were lower than those we obtained for sunflower. In addition, the percentage of reads that passed filtration and uniquely aligned to the genome in our study (29.2 %) was higher than for GBS analysis of *Miscanthus sinensis* (23 %) for which many more reads were obtained (415,694,046) (Ma et al. 2012). Differences in read numbers among studies can be due to the number of individuals used in GBS and/or



**Fig. 2** Physical map of 7646 SNPs found to be polymorphic for parental lines. Detailed information about physical map and SNPs is available in supplemental Excel file



**Fig. 3** SNP-based genetic linkage map. Detailed information about genetic linkage map is available in supplemental Excel file

**Table 3** Summary of SNP loci in linkage map

LGs	Size of LGs (cM)	Number of SNPs	Frequency of SNPs (per cM)
LG 1	212	86	2.47
LG 2	197	48	4.10
LG 3	183	48	3.81
LG 4	187	82	2.28
LG 5	212	57	3.72
LG 6	171	57	3.00
LG 7	206	78	2.64
LG 8	207	57	3.63
LG 9	183	70	2.61
LG 10	143	45	3.18
LG 11	236	80	2.95
LG 12	147	54	2.72
LG 13	106	21	5.05
LG 14	21	8	2.63
LG 15	17	7	2.43
LG 16	20	13	1.54
LG 17	24	6	4.00
Total	2472	817	3.03

technical performance of the sequencing chemistry. GBS of *Miscanthus sinensis* used 192 individuals (Ma et al. 2012) while our study and that of *Rubus idaeus*

(Ward et al. 2013) used 95 individuals. In another study, GBS analysis of an oil palm  $F_2$  population consisting of 108 individuals generated a much higher

number of reads (524,508,111) with 88 % of the total reads aligned to unique locations in the oil palm genome (Pootakham et al. 2015). This higher alignment percentage stresses the importance of the genome because oil palm has a complete genome sequence (Singh et al. 2013) while the current assembly of the sunflower genome represents only 85 % of the complete genome. Although the draft genome of sunflower is not complete, the assembled genome is highly accurate and contiguous because most (more than 76 %) of the contigs larger than 5 kb were assigned to the physical and/or genetic map (Kane et al. 2011).

The present study implemented the GBS approach for SNP discovery in sunflower. Although a reasonable number of SNPs were discovered, the study showed that the GBS approach performed in a parental genetic background in sunflower was less efficient than other SNP discovery methods. These alternative approaches discovered many more validated SNPs in sunflower using NGS sequences from a few individuals (10,640 SNPs; Bachlava et al. 2012; 16,467 SNPs; Pegadaraju et al. 2013; 20,502 SNPs; Livaja et al. 2016). The number of SNPs discovered by GBS was relatively low because of the limited number of sequence tags uniquely aligned to the sunflower genome (29.2 %). This was due to the high number of repetitive sequences in the genome which made tag alignments problematic (Treangen and Salzberg 2012). However, when compared with other GBS studies, the number of identified SNPs after filtering in this study (9535) was similar to the numbers reported in *Rubus idaeus* (9143) and soybean (10,120) (Ward et al. 2013; Sonah et al. 2013).

The average frequency of sunflower SNPs identified in this study was 72 kb/SNP (1 SNP every 72 kb); this is higher than soybean (100 kb/SNP) (Sonah et al. 2013) and much lower than the frequency identified in oil palm (1 SNP in 0.665 kb) (Pootakham et al. 2015). Again, the higher frequency of SNPs in oil palm could reflect the higher number of reads and higher coverage of the oil palm alignment (88 %). This result could also indicate that the oil palm genome is more diverse than sunflower due to higher gene flow and less inbreeding in trees than annual plants (Austerlitz et al. 2000).

The number of transition SNPs was higher than transversions as expected because it is well known that random mutations lead to higher transition rates

during evolution (Vogel and Kopun 1977). This higher transition rate was also observed in oil palm. In oil palm and *Miscanthus sinensis*, the most prevalent substitution types were A/G and C/T (Ma et al. 2012; Pootakham et al. 2015). These substitutions were also the most prevalent substitution types identified in this study.

The SNPs identified in this study were further filtered to retain only those SNPs which were polymorphic for the parents. After parental filtration, most of the SNPs (80 % of the filtered SNPs) were retained. Although some gaps were observed in four linkage groups, most of the sunflower genome was well covered with SNPs. Gaps can be filled using the unfiltered SNPs identified in the relevant regions (Fig. 1). In a future study, these SNPs can be utilized with genotyping platforms such as Kompetitive Allele Specific PCR (KASP) (Semagn et al. 2013) for breeder friendly sunflower genotyping.

#### Linkage map construction

The polymorphic SNPs were used to construct a linkage map containing 817 SNP markers. Although this set represented only 1.7 % of the unfiltered SNPs, exclusion of most of the SNPs due to missing data or segregation distortion was also observed in other GBS linkage mapping studies performed in apple (0.9 % of total identified SNPs retained) and oil palm (5 % retained) (Pootakham et al. 2015; Gardner et al., 2014). The total size of the sunflower linkage map (2472 cM) was larger than previous SNP marker maps reported by Talukder et al. (2014) (1444 cM) and Bowers et al. (2012) (1080 cM). The larger total size of the linkage map and SNPs with segregation distortion might be due to genotyping errors caused by the nature of sunflower genome structure. For example, repetitive sequences can limit the alignment efficiency of the sequence tags generated by GBS. Because the sunflower genome is known to contain a large proportion of repetitive sequences (~78 %), this made tag alignments problematic and caused genotyping errors (Kane et al. 2011; Treangen and Salzberg 2012). Although the linkage map constructed in this study had lower marker density than the other two SNP-based maps (0.13 cM/SNP for Talukder's map and 0.29 cM/SNP for Bowers' map), it had higher marker density than the map (5.5 cM/SNP) constructed by Lai et al. (2005). The maps of Talukder

and Bowers had higher resolution because they are consensus maps and were constructed from multiple populations. In addition, one of the two parents of the four mapping populations (Bowers et al. 2012) was a wild accession of sunflower and a non-oilseed landrace while the rest of the parents were sunflower cultivars. Although our linkage map had some unmapped regions similar to the maps of Talukder et al. (2014) and Lai et al. (2005), the SNP markers were generally well distributed.

The number of linkage groups of the map equals the haploid chromosome number of sunflower ( $n = 17$ ) as expected and moderate correlation was determined between the genetic linkage and physical maps. Conflict between linkage and physical maps was also reported in a GBS study in apple with some SNP positions in the linkage map conflicting with their physical map locations (Gardner et al. 2014). This might be due to misassembly of pseudochromosomes or a small population size (Silva et al. 2007). The  $F_2$  population used for linkage map construction had 93 individuals and this might not be sufficient for accurate ordering of large numbers of SNPs (Silva et al. 2007).

In conclusion, this study is the first to demonstrate the implementation of a GBS approach for SNP identification in sunflower. A total of 46,278 SNP loci were identified and 7646 SNP loci were validated in an  $F_2$  population. Although a reasonable number of SNPs was discovered, the relatively low efficiency of GBS in sunflower was revealed. This is most likely due to the high number of repetitive sequences in the genome which resulted in a larger and lower-resolution genetic linkage map. Nevertheless, the SNP markers and SNP-based linkage map constructed in this work will be valuable molecular genetics tools for sunflower breeding.

**Acknowledgments** This study was supported by Grant 5130037 from the Scientific and Technological Research Council of Turkey (TUBITAK). Genotyping by sequencing was performed by the University of Wisconsin Biotechnology Center, Madison, Wisconsin, USA.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical standards** This study involved no work with animal participants by any of the authors.

## References

- Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics* 154:1309–1321
- Bachlava E, Taylor CA, Tang S, Bowers JE, Mandel JR, Burke JM, Knapp SJ (2012) SNP discovery and development of a high-density genotyping array for sunflower. *PLoS ONE* 7:e29814. doi:10.1371/journal.pone.0029814
- Bowers JE, Bachlava E, Brunick RL, Rieseberg LH, Knapp SJ, Burke JM (2012) Development of a 10,000 locus genetic map of the sunflower genome based on multiple crosses. *G3 Genes Genom Genet* 2:721–729. doi:10.1534/g3.112.002659
- Doyle JJ, Doyle JE (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi:10.1371/journal.pone.0019379
- Gardner KM, Brown P, Cooke TF, Cann S, Costa F, Bustamante C, Velasco R, Troglio M, Myles S (2014) Fast and cost-effective genetic mapping in apple using next-generation sequencing. *G3 Genes Genom Genet* 4:1681–1687. doi:10.1534/g3.114.011023
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9:e90346. doi:10.1371/journal.pone.0090346
- Huang X, Han B (2013) Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol* 65:1–21. doi:10.1146/annurev-arplant-050213-035715
- Kane NC, Gill N, King MG, Bowers JE, Berges H, Gouzy J, Rieseberg LH (2011) Progress towards a reference genome for sunflower. *Botany* 89:429–437. doi:10.1139/b11-032
- Kumar S, Banks TW, Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. *Int J Plant Genomics* 2012:1–14. doi:10.1155/2012/831460
- Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. *Theor Appl Genet* 111:1532–1544. doi:10.1007/s00122-005-0082-4
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 4:357–359
- Livaja M, Unterseer S, Erath W, Lehermeier C, Wieseke R, Plieske J, Polley A, Luerßen H, Wieckhorst S, Mascher M, Hahn V, Ouzunova M, Schön CC, Ganai MW (2016) Diversity analysis and genomic prediction of Sclerotinia resistance in sunflower using a new 25 K SNP genotyping array. *Theor Appl Genet* 129(2):317–329. doi:10.1007/s00122-015-2629-3
- Ma X-F, Jensen E, Alexandrov N, Troukhan M, Zhang L, Thomas-Jones S et al (2012) High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*. *PLoS ONE* 7:e33821. doi:10.1371/journal.pone.0033821



- Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. *Int J Plant Genomics* 2012:1–11. doi:[10.1155/2012/728398](https://doi.org/10.1155/2012/728398)
- Norusis MJ (2010) PASW statistics 18 guide to data analysis. Prentice Hall Press, Upper Saddle River
- Ooijen JW, Voorrips RE (2001) JoinMap version 3.0: software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen
- Pegadaraju V, Nipper R, Hulke B, Qi L, Schultz Q (2013) De novo sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach. *BMC Genomics* 14:4556. doi:[10.1186/1471-2164-14-556](https://doi.org/10.1186/1471-2164-14-556)
- Pootakham W, Jomchai N, Ruang-areerate P, Shearman JR, Sonthirod C, Sangsrakru D, Tragoonrung S, Tangphat-sornruang S (2015) Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics* 105:288–295. doi:[10.1016/j.ygeno.2015.02.002](https://doi.org/10.1016/j.ygeno.2015.02.002)
- Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ et al (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol* 14:R55. doi:[10.1186/gb-2013-14-6-r55](https://doi.org/10.1186/gb-2013-14-6-r55)
- Semagn K, Babu R, Hearne S, Olsen M (2013) Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol Breed* 33:1–14. doi:[10.1007/s11032-013-9917-x](https://doi.org/10.1007/s11032-013-9917-x)
- Silva LD, Cruz CD, Moreira MA, Barros EGD (2007) Simulation of population size and genome saturation level for genetic mapping of recombinant inbred lines (RILs). *Genet Mol Biol* 30:1101–1108. doi:[10.1590/S1415-47572007000600013](https://doi.org/10.1590/S1415-47572007000600013)
- Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R et al (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500:335–339. doi:[10.1038/nature12309](https://doi.org/10.1038/nature12309)
- Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B et al (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* 8:e54603. doi:[10.1371/journal.pone.0054603](https://doi.org/10.1371/journal.pone.0054603)
- Talukder ZI, Gong L, Hulke BS, Pegadaraju V, Song Q, Schultz Q, Lilli Q (2014) A high-density SNP map of sunflower derived from RAD-sequencing facilitating fine-mapping of the rust resistance gene *R<sub>12</sub>*. *PLoS ONE* 9:e98628. doi:[10.1371/journal.pone.0098628](https://doi.org/10.1371/journal.pone.0098628)
- Treangen TJ, Salzberg SL (2012) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 13(1):36–46. doi:[10.1038/nrg3117](https://doi.org/10.1038/nrg3117)
- Vogel F, Kopun M (1977) Higher frequencies of transitions among point mutations. *J Mol Evol* 9:159–180
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93:77–78. doi:[10.1093/jhered/93.1.77](https://doi.org/10.1093/jhered/93.1.77)
- Ward JA, Bhangoo J, Fernández FF, Moore P, Swanson JD, Viola R, Velasco R, Bassil N, Weber CA, Sargent DJ (2013) Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genomics* 14:1–14. doi:[10.1186/1471-2164-14-2](https://doi.org/10.1186/1471-2164-14-2)