Construction of a high-density linkage map and graphical representation of the arrangement of transcriptome-based unigene markers on the chromosomes of onion, *Allium cepa* L.

Satoshi Fujito[1], Turgut Yigit Akyol[2,3], Takuya Mukae[4], Tadayuki Wako[1,5], Ken-ichiro Yamashita[1,6], Hikaru Tsukazaki[1,7], Hideki Hirakawa[8], Keisuke Tanaka[9], Yoko Mine[10], Shusei Sato[2*] and Masayoshi Shigyo[4*]

Abstract

Background: Genomic information for *Allium cepa* L. is limited as it is heterozygous and its genome is very large. To elucidate potential SNP markers obtained by NGS, we used a complete set of *A. fistulosum* L.-*A. cepa* monosomic addition lines (MALs) and doubled haploids (DHs). These were the parental lines of an *A. cepa* mapping population for transcriptome-based SNP genotyping.

Results: We mapped the transcriptome sequence reads from a series of *A. fistulosum-A. cepa* MALs onto the unigene sequence of the doubled haploid shallot *A. cepa* Aggregatum group (DHA) and compared the MAL genotype call for parental bunching onion and shallot transcriptome mapping data. We identified SNP sites with at least four reads on 25,462 unigenes. They were anchored on eight *A. cepa* chromosomes. A single SNP site was identified on 3,278 unigenes and multiple SNPs were identified on 22,184 unigenes. The chromosome marker information was made public via the web database Allium TDB (http://alliumtdb.kazusa.or.jp/). To apply transcriptome based genotyping approach for genetic mapping, we gathered RNA sequence data from 96 lines of a DHA × doubled haploid bulb onion *A. cepa* common onion group (DHC) mapping population. After selecting co-dominant SNP sites, 16,872 SNPs were identified in 5,339 unigenes. Of these, at least two SNPs with identical genotypes were found in 1,435 unigenes. We developed a linkage map using genotype information from these unigenes. All unigene markers mapped onto the eight chromosomes and graphical genotyping was conducted based on the unigene order information. Another 2,963 unigenes were allocated onto the eight chromosomes. To confirm the accuracy of this transcriptome-based genetic linkage map, conventional PCR-based markers were used for linkage analysis. All SNP - and PCR-based markers were mapped onto the expected linkage groups and no inconsistency was found among these chromosomal locations.

* Correspondence: shuseis@ige.tohoku.ac.jp; shigyo@yamaguchi-u.ac.jp
[2]Graduate School of Life Sciences, Tohoku University, 2-1-1 Katahira, 980-8577 Sendai, Miyagi, Japan
[4]Graduate School of Sciences and Technology for Innovation, Yamaguchi University, 1677-1 Yoshida, 753-8515 Yamaguchi, Yamaguchi, Japan
Full list of author information is available at the end of the article

Fujito *et al. BMC Genomics*   (2021) 22:481

Page 2 of 8

Conclusions: Effective transcriptome analysis with unique *Allium* resources successfully associated numerous chromosome markers with unigene information and a high-density *A. cepa* linkage map. The information on these unigene markers is valuable in genome sequencing and useful trait detection in *Allium*.

Keywords: *Allium cepa* L., Bulb onion, Linkage map, Shallot, Transcriptome-based SNP genotyping

Background

The genus *Allium* comprises economically important vegetable crops such as bulb onion (*A. cepa* L.), garlic (*A. sativum* L.), bunching onion (*A. fistulosum* L.), leek (*A. porrum* L.), and numerous wild species (Hanelt, 1990) [1]. Bulb onion is a major vegetable crop worldwide. According to the FAOSTAT database, global bulb onion production was ~ 96 million t and ranked second after tomato in terms of vegetable crop cultivation in 2018 [2]. *Allium cepa* L. consists of the common onion (bulb onion) and the Aggregatum (shallot) groups (Jones and Mann, 1963) [3]. Shallot is also an important vegetable crop and is cultivated mainly in Europe, Southeast Asia, and Africa. Though it differs morphologically and ecologically from bulb onion, both are easily crossed (Astley et al., 1982) [4]. Shallot has a short growing period and is resistant to *Fusarium oxysporum* (Vu et al., 2012) [5]. Hence, analysis of its genome might generate valuable information applicable to bulb onion breeding. The latter is time-consuming and labor-intensive as bulb onion is a biennial and heterogeneous because of severe inbreeding depression. To facilitate bulb onion breeding efforts, then, it is necessary to develop effective methods such as DNA marker-assisted selection.

Various DNA markers have been developed for *Allium* species. Simple sequence repeat (SSR) markers have been used to construct linkage maps as they are co-dominant and useful for PCR-based detection (Baldwin et al., 2012; Bradeen and Havey, 1995; Ipek et al., 2005; Martin et al., 2005; McCallum et al., 2012; van Heusden et al., 2000a, 2000b; Wilkie et al., 1993; Tsukazaki et al., 2006, 2007, 2008, 2011, 2015; Fischer and Bachman, 2000; Song et al., 2004) [6–20]. However, only ~ 500 SSR markers have been developed for bulb onion so far (Fischer and Bachman, 2000; Kuhl et al., 2004; McCallum et al., 2006; Tsukazaki et al., 2008, 2011; Martin et al., 2005) [9, 16, 17, 19, 21, 22]. These are inadequate for precise genetic analysis and DNA marker-assisted selection in bulb onions.

The advent of next-generation sequencing (NGS) has realized the accumulation of large amounts of sequence information and the identification of numerous single-nucleotide-polymorphisms (SNPs) to develop markers in plants with large genomes (Takahagi et al., 2016) [23]. NGS has been used to generate SNP markers in bulb onions via transcriptomic and selected genomic regions (Duangjit et al., 2013; Jo et al., 2017; Choi et al., 2020)

[24–26]. Numerous SNPs were identified by these approaches. However, only 597, 202, and 319 SNP markers were anchored on each genetic map, respectively, because of parental line heterozygosity in the mapping population and a lack of reliable reference genome sequences. For the effective use of NGS technology for plants without reference genome sequences, plant materials with sufficient homozygosity must be applied to the parental lines of the mapping population.

Doubled haploid (DH) techniques use chromosome doubling of haploid plants to generate materials that are homozygous enough for genetic analysis (Alan et al.,2003; Ajisaka et al., 2001; Jia et al., 2005) [27–29]. For *Allium*, we developed shallot and bulb onion DH lines and their F$_1$ hybrids for use in genetic analysis (Abdelrahman et al., 2015; Wako, 2016) [30, 31]. We also developed several bunching onion (*Allium fistulosum* L.)-shallot monosomic addition lines (MALs) (Shigyo et al., 1996) [32]. These have been used to assign genetic linkage maps to *A. cepa* chromosomes by seeking shallot-type alleles among the eight MALs (van Heusden et al., 2000b; Martin et al., 2005) [9, 12]. The combination of these plant resources could enhance potential SNP genotyping by NGS.

Here, we performed a transcriptome analysis on MALs to generate information about chromosome-specific unigene markers. We conducted transcriptomics on the F$_2$ population derived from a cross between shallot and bulb onion DH lines. We also constructed a high-density genetic linkage map by elucidating the potential SNP sites generated by NGS.

Results and discussion

Unigene chromosome anchoring by SNP genotyping via MAL RNA sequencing

In our previous study, we carried out transcriptome analysis of a complete set of MALs using RNA samples isolated from the leaf, root, and bulb (Abdelrahman et al., 2017) [33]. In the evaluation process of expression level, we obtained transcriptome reads of each MAL, and parental bunching onion and doubled haploid shallot (DHA) were mapped onto unigene data set constructed using DHA bulb transcriptome as a reference. During this analysis, we realized that SNPs between bunching onion and DHA could be identified in the mapping data, and these SNPs could be applied for anchoring the unigene sequences onto shallot chromosomes. Theoretically, in genotype calling, MALs having the shallot

chromosome on which the target gene is allocated become heterozygous, whereas the remaining MALs become homozygous of bunching onion allele. Therefore, we performed transcriptome based genotyping using previously accumulated data (Abdelrahman et al., 2017) [33], as advanced mapping data applications. SNP sites with alternative homozygous calls in bunching onion and reference homozygous calls in DHA were selected by comparing the genotype call of the transcriptome mapping data between the MAL parental lines (bunching onion and shallot). Among 56,161 DHA unigenes, sites with ≥ 4 reads coverage in all eight MALs were identified on 25,462 unigenes (Table 1). Of these, one SNP was identified in 3,278 unigenes, which could be anchored onto a chromosome as a heterozygous genotype were called one MAL. On the contrary, multiple SNP sites were identified in 22,184 unigenes. Of these, 21,996 could be allocated to single physical chromosomes. Extrachromosomal MALs with heterozygous genotypes are consistent with chromosomal unigene locations with multiple SNPs. For the remaining 188 unigenes, ≥ 2 multiple SNPs were ambiguous. There were heterozygous genotypes in eight MAL types and/or parental homozygous genotype(s). The corresponding gene may have been downregulated and the shallot gene had partial homology. These unigenes were assigned to the chromosome based on other marker(s) with a single heterozygous genotype in MALs with the "R" indication and mapped by representative SNPs. A total of 25,462 unigenes were anchored on eight chromosomes. There were 4,513 unigenes on chromosome 2 and only 2,169 unigenes on chromosome 8.

DHA unigene information has been made public through the web database 'Allium Transcriptome Database' (TDB) at http://alliumtdb.kazusa.or.jp. We have added the list of chromosome markers in this data base (Table S1). These anchoring markers are useful in genome sequencing projects.

SNP detection in *Allium cepa* doubled haploids

To expand the transcriptome based genotyping approach for the construction of genetic linkage map and genetic markers, we accumulated transcriptome data of the F_2 plants derived from a cross between the *A. cepa* DH lines (DHA for shallot × DHC for bulb onion). RNA sequence of leaf samples was collected from 96 F_2 plants

(population A) of the mapping population and from DHC. Using the Illumina sequencing platform, 10–25 million 100 base paired end sequence reads were accumulated for each plant, and the obtained reads were mapped onto DHA bulb unigene data set used in the MAL transcriptome analysis. As the parental lines were doubled haploid, genotyping of the F_2 plants from the mapping population should be classified as reference (DHA) homozygous, alternative (DHC) homozygous, and heterozygous. The intraspecific SNPs identified by mapping DHC reads with ≥ 2 reads coverage on all 96 lines were selected for genotyping. Selecting co-dominant SNP sites with heterozygous genotypes among the 96 lines identified 16,872 SNP sites in 5,339 unigenes. One SNP site was identified on 2,109 unigenes. These genotypes were used for map calculation with an "O" indication meaning that one SNP site was supported. Of the 3,230 unigenes with multiple SNP sites, ≥ 2 SNP sites with identical genotype patterns on the 96 lines were identified on 1,435 unigenes. These patterns were selected as the solid genotype (S) of the corresponding unigenes. For the remaining 1,795 unigenes, inconsistencies between the homozygous and heterozygous calls were identified among the 96 lines. The representative genotype (R) was created by selecting the most abundant genotype in each of the 96 lines.

Genetic linkage map construction and physical chromosome assignment

We used the solid co-dominant genotype information obtained from 1,435 unigenes in population A to plot a genetic linkage map with JoinMap v. 4.0 (Kyazma BV, Wageningen, The Netherlands). By applying the LOD 5 cutoff, all tested markers were assigned to eight linkage groups. Based on the unigenes with anchored chromosome information, all of these could be anchored to each of the eight bulb onion chromosomes. No inconsistency was detected between each linkage group and assigned chromosome. Hence, this linkage map was reliable.

A graphical genotype list was constructed according to the unigene order information. A total of 610 genotype blocks were assigned based on the patterns of the tested 96 lines (Table S2). The remaining unigenes with "O"- and "R"-coded genotypes were allocated to the most probable genotype block and permitted genotype inconsistencies for ≤ 10 lines. A total of 1,537 "O"-marked and

Table 1 DHA unigenes with SNP(s) detected on MALs

		Chr. 1	Chr. 2	Chr. 3	Chr. 4	Chr. 5	Chr. 6	Chr. 7	Chr. 8	Total
Multiple SNPs	allocated to single chromosome	3,404	3,900	3,148	2,538	2,588	2,546	2,032	1,840	21,996
	including representative SNP	30	35	33	21	16	16	18	19	188
One SNP		486	578	455	397	357	376	319	310	3,278
Total		3920	4,513	3,636	2,956	2,961	2,938	2,369	2,169	25,462

Fujito *et al. BMC Genomics* (2021) 22:481

Page 4 of 8

1,426 "R"-marked unigenes were allocated onto the genotype blocks (Table 2, Supplemental file). As the inconsistency of genotyping in R-marked tends to be a homozygous call in heterozygous allele, lower expression level and/or allele bias in expression could be potential sources of errors in transcriptome based genotyping.

To confirm transcriptome-based genetic linkage map accuracy, we applied conventional PCR-based markers to the same F_2 population (A). The PCR-based SSR and InDel markers were previously reported (Fischer and Bachmann, 2000; Kuhl et al., 2004; Martin et al., 2005; McCallum et al., 2012; Tsukazaki et al., 2006, 2007, 2008, 2011; Wako, 2016) [9, 10, 14–17, 19, 21, 31] and used in the present study. Thirty-three markers were polymorphic between DHA and DHC. Fourteen InDel polymorphisms were detected for the sequence comparisons between DHA and DHC in Allium TDB. We designed primer sets that included these polymorphism sites and amplified them by PCR. We used 47 PCR-based markers in a linkage analysis on another F_2 population (B). All linkage groups were assigned to eight physical chromosomes in MALs confirmed by amplification. These marker locations matched those in previous reports (Tsukazaki et al., 2008, 2011, 2015; Masuzaki et al., 2006a, 2006b; Wako, 2016) [16–18, 31, 34, 35]. We selected 14 reliable PCR-based markers covering all eight chromosomes, applied them to population A, and integrated them onto the transcriptome-based genetic linkage map. The reconstructed map consisted of eight linkage groups with all SNP solid and PCR-based markers covering 936.6 cM. The average marker interval was 0.65 cM. All PCR-based markers were integrated onto positions corresponding to those on population B. The latter was based on a PCR marker-based linkage map. No contradiction in marker location was caused by using common markers between these maps and another linkage map previously constructed with a gynogenic population (C) derived from the same F_1 hybrid between DHA and DHC with the exception of some on Chr. 2 and Chr. 5 (Fig. 1) (Wako, 2016) [31]. We also compared the genetic maps against a published transcriptome-based SNP marker analysis (Duangjit et al., 2013) [24]. Comparison of the positions of 137 SNP markers on sequences overlapping in both analyses revealed that the anchored chromosomes and relative positions were consistent for all SNP markers (Table S3). Therefore, our transcriptome-based genetic linkage map is reliable.
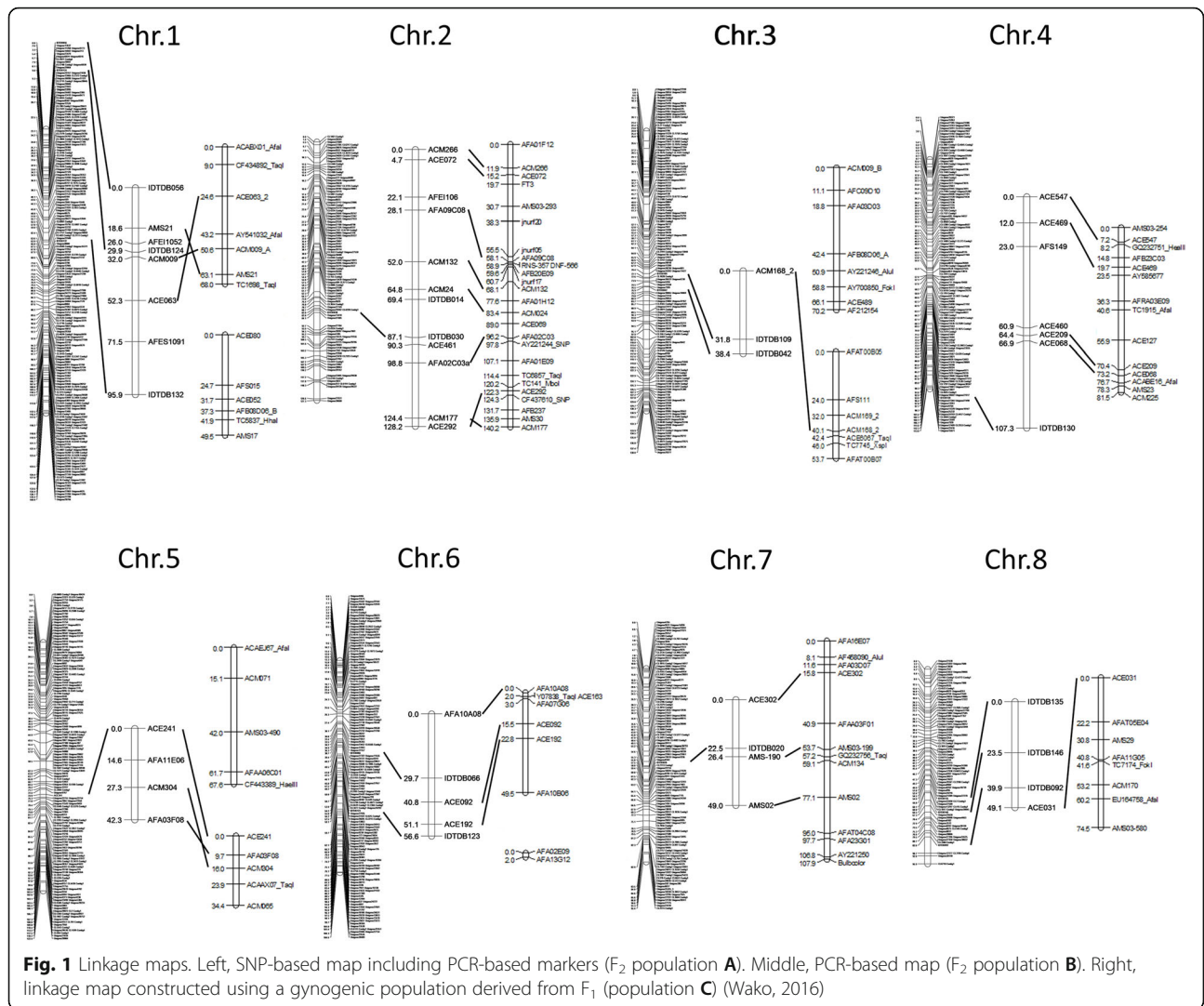
MALs have been used extensively to assign DNA markers to physical chromosomes (van Heusden et al., 2000b; Martin et al., 2005; Tsukazaki et al., 2008) [9, 12, 16]. Here, we identified chromosome-specific SNPs by comparing transcriptome data with MALs. For the first time, we used F_2 populations from *Allium* DH parental lines. The parental line has each homozygous allele. SNPs between the parental lines DHA and DHC are easily detected. Transcriptome data from the DH lines efficiently found SNPs (Baldwin et al., 2012) [6] and we obtained abundant and reliable SNP information here. We constructed a reliable genetic map based on S-marked SNP markers. No inconsistency was found between the physical chromosome assignments and S-labeled markers in the linkage group. The genetic map comprised 1,435 SNP markers, one bulb onion SSR marker, and 13 InDel markers and covered 936.6 cM. To our knowledge, this map has the highest number of markers to date. Integrated linkage maps include markers associated with phenotypic characteristics for the nuclear male fertility restoration loci of cytoplasmic male sterility (Chr. 2) and bulb color (Chr. 7) (Wako, 2016) [31]. Shallot is a genetic breeding resource for bulb onion as it produces certain distinctive chemical compounds such as saponins conferring pathogen resistance (Shigyo et al., 1997; Abdelrahman et al., 2017) [33, 36]. By combining these DH lines with linkage map information, progress is anticipated in *Allium* molecular breeding by marker-assisted selection for several agronomic bulb onion traits.

Conclusions

In the present study, we constructed a high-density linkage map in *Allium cepa* using numerous SNP markers obtained from the transcriptome information of the *Allium* DH lines and the MALs. As DH techniques depress inbreeding, they are useful for making homozygous pure lines that resemble inbred lines (Abdelrahman et al., 2015) [30]. Though bulb onion and shallot have different characteristics, they both belong to *A. cepa* and are easy to cross (Astley et al., 1982) [4]. The MALs have all *A. fistulosum* chromosomes and one *A. cepa* chromosome (Shigyo et al., 1996) [32]. We performed a transcriptome analysis to identify unigenes and assign them to physical

Table 2 Genetic markers developed for each chromosome

	Chr. 1	Chr. 2	Chr. 3	Chr. 4	Chr. 5	Chr. 6	Chr. 7	Chr. 8	Total
Solid (S)	270	126	200	170	199	198	170	102	1,435
Representative(R)	175	416	195	163	148	123	110	96	1,426
One (O)	239	278	229	192	135	186	148	130	1,537
Total	684	820	624	525	482	507	428	328	4,398

Fujito *et al. BMC Genomics* (2021) 22:481

Page 5 of 8

Fig. 1 Linkage maps. Left, SNP-based map including PCR-based markers (F$_2$ population **A**). Middle, PCR-based map (F$_2$ population **B**). Right, linkage map constructed using a gynogenic population derived from F$_1$ (population **C**) (Wako, 2016)

chromosomes. To this end, we compared shallot DH and MAL transcriptome data. We then used the F$_2$ mapping population between bulb onion DH and shallot DH to detect SNP sites. A total of 16,872 SNP sites were identified on 5,339 unigenes. Of these, 1,435 were selected as the solid genotype of the corresponding unigenes. By constructing a linkage map with SNP solid markers, all markers were mapped and the locations between the physical chromosomes and linkage groups were consistent. The number of SNPs located on the linkage map was much higher than those previously reported. Thus, the linkage map resolution was high. Furthermore, linkage maps integrated with PCR-based markers are now available. Shallots produce chemical compounds conferring resistance to certain bulb onion diseases (Abdelrahman et al., 2017) [33]. Hence, connecting phenotype and genotype information is a holistic approach towards *Allium* gene expression analysis for plant breeding and an effective, low-cost method of developing novel disease-resistant *Allium* varieties.

Methods

Plant materials and genetic cross

F$_2$ plants generated by a cross between shallot and bulb onion doubled haploid (DH) lines (DHA and DHC, respectively) were used to construct a linkage map. DHA was derived from 'Chiang Mai' shallot in Thailand while DHC was derived from long-day 'Sapporo-ki' onion cultivar in Japan (Abdelrahman et al., 2015) [30]. For the transcriptome and conventional PCR-based marker analyses, populations A (96 individuals) and B (186 individuals) were created. They were raised in the greenhouses of Yamaguchi University (34°11" N; 131°28" E) and the Institute of Vegetable and Floriculture Science of NARO (34°61" N; 136°25" E) in Japan.

For the *A. cepa* unigene chromosome assignments via interspecific SNP detection, a complete set of MALs of *Allium fistulosum* L. with eight single shallot extrachromosomes (Shigyo et al., 1996) [32] was used (Abdelrahman et al., 2017) [33].

Fujito *et al. BMC Genomics*   (2021) 22:481

Page 6 of 8

Transcriptome sequencing

The strategy for the accumulation of RNA sequencing data of MALs and *A. fistulosum* and construction of DHA bulb unigene data set has been described previously (Abdelrahman et al., 2017) [33].

For RNA sequencing of 96 F$_2$ plants and DHC, total RNA was isolated from leaf samples using the RNeasy plant mini kit (QIAGEN Sciences, Germantown, MD, USA). RNA quality was checked with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Samples with RNA integrity number (RIN) > 8.0 were selected for further use. The cDNA library was generated with a TruSeq™ RNA sample preparation kit (Illumina, San Diego, CA, USA) in accordance with the manufacturer's instructions. Sequencing was performed on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA).

SNP detection and selection

RNA sequence reads were filtered with PRINSEQ v. 0.20.4 (Schmieder and Edwards, 2011) [37] and fastx_clipper in FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The filtered single-end reads were mapped onto DHA unigene sequences in end-to-end mode with Bowtie v. 22.1.0 (Langmead et al., 2009) [38]. The sequence alignment/map (SAM) format files were converted into BAM format with SAMtools v. 0.1.19 (Li et al., 2009) [39]. The BAM files were subjected to SNP calling with the mpileup option of SAMtools31 v. 0.1.19 and the mpileup2snp option of VarScan v. 2.3 to obtain a variant call format (VCF) file containing the SNP information.

PCR-based marker analysis

Total DNA was prepared from individual plants according to the method of Song et al. (2004) [20]. To detect polymorphisms between DHA and DHC, SSR markers derived from the *A. cepa* genome (Fischer and Bachmann, 2000) [19], the *A. fistulosum* genome (Ohara et al., 2005; Song et al., 2004; Martin et al., 2005; Tsukazaki et al., 2006, 2007, 2008, 2011; Wako, 2016) [9, 14–17, 20, 31, 40], and *A. cepa* expressed sequence tags (ESTs) (Jakse et al., 2005; Kuhl et al., 2004; Martin et al., 2005; Tsukazaki et al., 2008; 2011) [9, 16, 17, 21, 41] and other CAPS and SCAR markers (Kuhl et al., 2004; Masuzaki et al., 2006a, 2006b; McCallum et al., 2006; Yaguchi et al., 2008) [21, 22, 34, 35, 42] were screened. Moreover, polymorphisms were detected between DHA and DHC nucleotide sequences obtained from transcriptome information in Allium Transcriptome DataBase (AlliumTDB) (Abdelrahman et al., 2017) [33] by sequence comparison with BLASTn. InDels > 6 bp were selected and primer sets including the polymorphism site were designed with Primer 3 (http://bioinfo.ut.ee/primer3/) (IDTDBxxx). The markers were applied to the F$_2$ population and parental lines. PCR was performed in a 10-μL reaction mixture containing 10 ng template DNA, 0.2 μM of each primer, 0.2 mM of each dNTP, and 5 μL GoTaq Master Mix (Promega Corp., Madison, WI, USA). Amplification was performed for 35 cycles after initial denaturation at 94 ℃ for 4 min. Each cycle consisted of 15 s at 94 ℃, 30 s at 55 ℃, 1 min at 72 ℃, and 4 min at 72 ℃. The PCR products were evaluated by 2 % (w/v) agarose gel electrophoresis. Certain forward primers were fluorescently labeled with 6-FAM, NED, PET, or VIC dyes (Applied Biosystems, Foster City, CA, USA) before PCR. The PCR products were loaded onto a capillary DNA sequencer (ABI3730; Applied Biosystems, Foster City, CA, USA) and analyzed with GeneMapper v. 3.0 (Applied Biosystems, Foster City, CA, USA).

Allium cepa linkage map construction by two methods and map comparison

Linkage analysis was performed with JoinMap v. 4.0 (van Ooijen, 2006) [43]. The Kosambi function was used to obtain the cM values (Kosambi, 1944) [44]. The DNA markers used to construct the *A. cepa* linkage map are listed in Table S4. Eight different *A. fistulosum-A. cepa* MALs were used to determine the corresponding physical chromosomes for the linkage groups. The linkage maps were compared according to the SNP and PCR-based markers by using the 14 anchor markers on both maps.

Abbreviations

DH: Doubled haploid; NGS: Next-generation sequencing; PCR: Polymerase chain reaction; MAL: Monosomic addition lines; SSR: Simple sequence repeat; InDel: Insertion-deletion; EST: Expressed sequence tag; SNP: Single-nucleotide polymorphism; CAPS: Cleaved amplified polymorphic sequence; SCAR: Sequence characterized amplified region

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07803-y.

Additional file 1.
Additional file 2.
Additional file 3.
Additional file 4.
Additional file 5.

Authors' contributions

All authors have read and approved the submitted version of the manuscript, and have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. Authors' contribution are, (1) Conceptualization: TW, SS, MS; (2) Methodology: HH, YM, SS, MS.; (3) Investigation: SF, TYA, TM, KY, HT, KT; (4) Writing-Original Draft; SF, TA, SS, MS; (5) Writing-Review and Editing: SF, TYA, TM, TW, KY, HT, HH, YM, SS, MS.

Availability of data and materials

RNA-Seq data for the MALs and F_2 mapping population (DHA × DHC) are available in the DDBJ sequence read archive under Accessions No. DRA005096 and DRA009194, respectively. The DHA unigene information has been made public through the web database 'Allium Transcriptome DataBase' (AlliumTDB) located at http://alliumtdb.kazusa.or.jp.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Institute of Vegetable and Floriculture Science, National Agriculture and Food Research Organization (NARO), 360 Kusawa, Ano, 514-2392, Tsu, Mie, Japan. [2]Graduate School of Life Sciences, Tohoku University, 2-1-1 Katahira, 980-8577 Sendai, Miyagi, Japan. [3]Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10, DK-8000 Aarhus C, Denmark. [4]Graduate School of Sciences and Technology for Innovation, Yamaguchi University, 1677-1 Yoshida, 753-8515 Yamaguchi, Yamaguchi, Japan. [5]Ministry of Agriculture, Forestry and Fisheries, 1-2-1 Kasumigaseki, Chiyoda- ku, 100-8950 Tokyo, Japan. [6]Western Region Agricultural Research Center, NARO, 1-3-1 Senyu-cho, 765-8508 Zentsuji-shi, Kagawa, Japan. [7]Tohoku Agricultural Research Center, NARO, 4 Akahira, Shimo-kuriyagawa, 020-0198 Morioka, Iwate, Japan. [8]Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu 292-0818, Chiba, Japan. [9]The NODAI Genome Research Center, Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, 156-8502 Tokyo, Japan. [10]Department of Agriculture, Tokyo University of Agriculture, 1737 Funako, 243-0034 Atsugi-shi, Kanagawa, Japan.

References

1. Hanelt P. Taxonomy, evolution, and history. In: Rabinowitch HD, Brewster LL, editors. Onions and allied crops I. Boca Raton: CRC Press; 1990. pp. 1–26.
2. Rome FAOSTATFAO. 2018. http://faostat3.fao.org/browse/Q/QC/E. Accessed 5 October 2020.
3. Jones HA, Mann LK. Onions and their allies. London: Leonard Hill Ltd; 1963.
4. Astley D, Innes NL, Van der Meer QP. Genetic resources of *Allium* species: A global report. Rome: IBPGR Secretariat; 1982.
5. Vu HQ, El-Sayed MA, Ito S, Yamauchi N, Shigyo M. Discovery of a new source of resistance to *Fusarium oxysporum*, cause of Fusarium wilt in *Allium fistulosum*, located on chromosome 2 of *Allium cepa* Aggregatum group. Genome. 2012;55:797–807. https://doi.org/10.1139/g2012-065.
6. Baldwin S, Revanna R, Thomson S, Pither-Joyce M, Wright K, Crowhurst R, Fiers M, Chen L, Macknight R, McCallum JA. A toolkit for bulk PCR-based marker design from next-generation sequence data: application for development of a framework linkage map in bulb onion (*Allium cepa* L.). BMC Genome. 2012;13:637. https://doi.org/10.1186/1471-2164-13-637.
7. Bradeen JM, Havey MJ. Restriction fragment length polymorphisms reveal considerable nuclear divergence within a well-supported maternal clade in *Allium* Section Cepa (Alliaceae). Amer J Bot. 1995;82:1455–62. https://doi.org/10.1002/j.1537-2197.1995.tb12683.x.
8. Ipek M, Ipek A, Almquist SG, Simon PW. Demonstration of linkage and development of the first low-density genetic map of garlic, based on AFLP markers. Theor Appl Genet. 2005;110: 228 – 36. https://doi.org/10.1007/s00122-004-1815-5.
9. Martin WJ, McCallum J, Shigyo M, Jakše J, Kuhl JC, Yamane N, Pither-Joyce M, Gokce AF, Sink KC, Town CD, Havey MJ. Genetic mapping of expressed sequences in onion and in silico comparisons with rice show scant colinearity. Mol Genet Genom. 2005;274:197–204. https://doi.org/10.1007/s00438-005-0007-6.
10. McCallum J, Baldwin S, Shigyo M, Deng Y, van Heusden S, Pither-Joyce M, Kenel F. AlliumMap-A comparative genomics resource for cultivated *Allium* vegetables. BMC Genom. 2012;13:168. https://doi.org/10.1186/1471-2164-13-168.
11. van Heusden AW, van Ooijen JW, Vrielink-van Ginkel R, Verbeek WHJ, Wietsma WA, Kik C. A genetic map of an interspecific cross in *Allium* based on amplified fragment length polymorphisms (AFLP™) markers. Theor Appl Genet. 2000a;100:118–26. https://doi.org/10.1007/s001220050017.
12. van Heusden AW, Shigyo M, Tashiro Y, Vrielink-van Glinkel R, Kik C. AFLP linkage group assignment to the chromosomes of *Allium cepa* L. via monosomic addition lines. Theor Appl Genet. 2000b;100:480–6. https://doi.org/10.1007/s001220050062.
13. Wilkie SE, Isaac PG, Slater RJ. Random amplified polymorphic DNA (RAPD) markers for genetic analysis in *Allium*. Theor Appl Genet. 1993;86:497–504. https://doi.org/10.1007/BF00838566.
14. Tsukazaki H, Fukuoka H, Song YS, Yamashita K, Wako T, Kojima A. Considerable Heterogeneity in Commercial F_1 varieties of Bunching Onion (*Allium fistulosum*) and Proposal of Breeding Scheme for Conferring Variety Traceability Using SSR Markers. Breed Sci. 2006;56:321–6. https://doi.org/10.1270/jsbbs.56.321.
15. Tsukazaki H, Nunome T, Fukuoka H, Kanamori H, Kono I, Yamashita K, Wako T, Kojima A. Isolation of 1,796 SSR clones from SSR-enriched DNA libraries of bunching onion (*Allium fistulosum*). Euphytica. 2007;157:83–94. https://doi.org/10.1007/s10681-007-9398-2.
16. Tsukazaki H, Yamashita K, Yaguchi S, Masuzaki S, Fukuoka H, Yonemaru J, Kanamori H, Kono I, Hang T, Shigyo M, Kojima A, Wako T. Construction of SSR-based chromosome map in bunching onion (*Allium fistulosum*). Theor Appl Genet. 2008;117:1213. https://doi.org/10.1007/s00122-008-0849-5.
17. Tsukazaki H, Yamashita K, Yaguchi S, Yamashita K, Hagihara T, Shigyo M, Kojima A, Wako T. Direct determination of the chromosomal location of bunching onion and bulb onion markers using bunching onion–shallot monosomic additions and allotriploid-bunching onion single alien deletions. Theor Appl Genet. 2011;122:501–10. https://doi.org/10.1007/s00122-010-1464-9.
18. Tsukazaki H, Yaguchi S, Sato S, Hirakawa H, Katayose Y, Kanamori H, Kurita K, Itoh T, Kumagai M, Mizuno S, Hamada M, Fukuoka H, Yamashita K, McCallum JA, Shigyo M, Wako T. Development of transcriptome shotgun assembly-derived markers in bunching onion (*Allium fistulosum*). Mol Breed. 2015;35:55. https://doi.org/10.1007/s11032-015-0265-x.
19. Fischer D, Bachmann K. Onion microsatellites for germplasm analysis and their use in assessing intra- and interspecific relatedness within the subgenus Rhizirideum. Theor Appl Genet. 2000;101:153–64. https://doi.org/10.1007/s001220051464.
20. Song YS, Suwabe K, Wako T, Ohara T, Nunome T, Kojima A. Development of microsatellite markers in Japanese bunching onion (*Allium fistulosum* L.). Breed Sci. 2004;54:361–5. https://doi.org/10.1270/jsbbs.54.361.
21. Kuhl JC, Cheung F, Yuan Q, Martin W, Zewdie Y, MaCallum J, Catanach A, Rutherford P, Sink KC, Jenderk M, Prince JP, Town CD, Havey MJ. A unique set of 11,008 onion expressed sequence tags reveals expressed sequence and genomic differences between the monocot orders Asparagales and Poales. Plant Cell. 2004;16:114–25. https://doi.org/10.1105/tpc.017202.
22. McCallum J, Clarke A, Pither-Joyce M, Shaw M, Butler R, Brash D, Scheffer J, Sims I, van Heusden S, Shigyo M, Harvey MJ. Genetic mapping of a major gene affecting onion bulb fructan content. Theor Appl Genet. 2006;112:958–67. https://doi.org/10.1007/s00122-005-0199-5.
23. Takahagi K, Uehara-Yamaguchi Y, Yoshida T, Sakurai T, Shinozaki K, Mochida K, Saisho D. Analysis of single nucleotide polymorphisms based on RNA sequencing data of diverse bio-geographical accessions in barley. Sci Rep. 2016;6:33199. https://doi.org/10.1038/srep33199.
24. Duangjit J, Bohanec B, Chan AP, Town CD, Havey MJ. Transcriptome sequencing to produce SNP-based genetic maps of onion. Theor Appl Genet. 2013;126:2093–101. https://doi.org/10.1007/s00122-013-2121-x.

Fujito *et al. BMC Genomics*   (2021) 22:481

Page 8 of 8

25. Jo J, Purushotham PM, Han K, Lee H-R, Nah G, Kang B-C. Development of a Genetic map for onion (*Allium cepa* L.) using reference-free genotyping-by-sequencing and SNP assays. Front Plant Sci. 2017;8:1606. https://doi.org/10.3389/fpls.2017.01606.

26. Choi Y, Kim S, Lee J. Construction of an Onion (*Allium cepa* L.) genetic linkage map using genotyping-by-sequencing analysis with a reference gene set and identification of QTLs controlling anthocyanin synthesis and content. Plants. 2020;9(5):616. https://doi.org/10.3390/plants9050616.

27. Alan AR, Mutschler MA, Brant A, Cobb E, Earle ED. Production of gynogenic plants from hybrids of *Allium cepa* L. and *A. roylei* Stearn. Plant Sci. 2003;165:1201–11. https://doi.org/10.1016/S0168-9452(03)00327-3.

28. Ajisaka H, Kuginuki Y, Yui S, Enomoto S, Hirai M. Identification and mapping of a quantitative trait locus controlling extreme late bolting in Chinese cabbage (*Brassica rapa* L. ssp. pekinensis syn. campestris L.) using bulked segregant analysis. Euphytica. 2001;118:75–81. https://doi.org/10.1023/A:1004023532005.

29. Jia G, Chen P, Qin G, Bai G, Wang X, Wang S, Zhou B, Zhang S, Liu D. QTLs for *Fusarium* head blight response in a wheat DH population of Wangshuibai/Alondra's'. Euphytica 2005;146:183 – 91. https://doi.org/10.1007/s10681-005-9001-7.

30. Abdelrahman M, Sawada Y, Nakabayashi R, Sato S, Hirakawa H, El-Sayed M, Hirai MY, Saito K, Yamauchi N, Shigyo M. Integrating transcriptome and target metabolome variability in doubled haploids of *Allium cepa* for abiotic stress protection. Mol Breed. 2015;35:195. https://doi.org/10.1007/s11032-015-0378-2.

31. Wako T. Genetic studies of disease resistance and bolting time based on genomic analysis in Japanese bunching onion (*Allium fistulosum* L.). 2016. Ph.D. Thesis. Tottori University, Koyamachō-Minami, Japan.

32. Shigyo M, Tashiro Y, Isshiki S, Miyazaki S. Establishment of a series of alien monosomic addition lines of Japanese bunching onion (*Allium fistulosum* L.) with extra chromosomes from shallot (*A. cepa* L. Aggregatum group). Genes Genet Syst. 1996;71:363–71. https://doi.org/10.1266/ggs.71.363.

33. Abdelrahman M, El-Sayed M, Sato S, Hirakawa H, Ito S, Tanaka K, Mine Y, Sugiyama N, Suzuki M, Yamauchi N, Shigyo M. RNA-sequencing-based transcriptome and biochemical analyses of steroidal saponin pathway in a complete set of *Allium fistulosum-A. cepa* monosomic addition lines. PLoS ONE. 2017;12(8):e0181784. https://doi.org/10.1371/journal.pone.0181784.

34. Masuzaki S, Shigyo M, Yamauchi N. Complete assignment of structural genes involved flavonoid biosynthesis influencing bulb color to individual chromosomes of the shallot (*Allium cepa* L.). Genes Genet Syst. 2006a;81:255–63. https://doi.org/10.1266/ggs.81.255.

35. Masuzaki S, Shigyo M, Yamauchi N. Direct comparison between genomic constitution and flavonoid contents in *Allium* multiple alien addition lines reveals chromosomal locations of genes related to biosynthesis from dihydrokaempferol to quercetin glucosides in scaly leaf of shallot (*A. cepa* L). Theor Appl Genet. 2006b;112:607–17. https://doi.org/10.1007/s00122-005-0157-2.

36. Shigyo M, Tashiro Y, Iino M, Terahara N, Ishimaru K, Isshiki S. Chromosomal locations of genes related to flavonoid and anthocyanin production in leaf sheath of shallot (*Allium cepa* L. Aggregatum group). Genes Genet Syst. 1997;72:149–52. https://doi.org/10.1266/ggs.72.149.

37. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4. https://doi.org/10.1093/bioinformatics/btr026.

38. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.

39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352.

40. Ohara T, Song Y, Tsukazaki H, Wako T, Nunome T, Kojima A. Genetic mapping of AFLP markers in Japanese bunching onion (*Allium fistulosum*). Euphytica. 2005;144:255–63. https://doi.org/10.1007/s10681-005-6768-5.

41. Jakse J, Martin W, McCallum J, Havey MJ. Single nucleotide polymorphisms, indels, and simple sequence repeats for onion cultivar identification. J Am Soc Hort Sci. 2005;130:912–7. https://doi.org/10.21273/JASHS.130.6.912.

42. Yaguchi S, McCallum J, Shaw M, Pither-Joyce M, Onodera S, Shiomi N, Yamauchi N, Shigyo M. Biochemical and Genetic Analysis of Carbohydrate Accumulation in *Allium cepa* L. Plant Cell Physiol. 2008;49(5):730–9. https://doi.org/10.1093/pcp/pcn048.

43. van Ooijen JW. JoinMap 4.0: Software for the calculation of genetic linkage map in experimental population. Wageningen: Kyazma BV; 2006.

44. Kosambi D. The estimation of map distance from recombination values. Ann Eugen. 1944;12:172–5.

Publisher's Note