**Title:**

**Chromosomal organization and sequence diversity of genes encoding lachrymatory factor synthase in *Allium cepa* L.**

**Authors:**

Noriya Masamura*[,§], John McCallum**, Ludmila Khrustaleva[§§], Fernand Kenel**, Meegham Pither-Joyce**, Jinji Shono[§] Go Suzuki[†], Yasuhiko Mukai[†], Naoki Yamauchi*[,‡], Masayoshi Shigyo*[,‡]

* The United Graduate School of Agricultural Science, Tottori University, Tottori 680-8553, Japan

[§] Somatech Center, House Foods Corporation, 1-4 Takanodai, Yotsukaido, Chiba 284-0033, Japan

** The New Zealand Institute for Plant and Food Research Limited, Private Bag 4704, Christchurch, New Zealand

[§§] Timiryazev Agricultural Academy, Russian Academy of Sciences, Moscow, 127550, Russia

[†] Division of Natural Science, Osaka Kyoiku University, Kashiwara, 582-8582, Japan

[‡] Faculty of Agriculture, Yamaguchi University, Yamaguchi 753-8515, Japan

**Running title:**

Chromosomal organization and sequence diversity of genes encoding lachrymatory factor synthase in *Allium cepa* L.

**Key words:**

*Allium cepa,* alien monosomic addition lines, genetic mapping, BAC-FISH, lachrymatory factor synthase

**Corresponding author:**

Masayoshi Shigyo

Faculty of Agriculture, Yamaguchi University,

1677-1 Yoshida, Yamaguchi 753-8515, Japan

TEL&FAX: +81-839335842

e-mail: shigyo@yamaguchi-u.ac.jp

## ABSTRACT

Lachrymatory factor synthase (LFS) catalyzes the formation of lachrymatory factor, one of the most distinctive traits of bulb onion (*Allium cepa* L.). Therefore, we used LFS as a model for a functional gene in a huge genome, and examined the chromosomal organization of LFS in *A. cepa* by multiple approaches. The first level analysis completed the chromosomal assignment of LFS gene to the chromosome 5 of *A. cepa* via the use of a complete set of *A. fistulosum*-shallot (*A. cepa* L. Aggregatum group) monosomic addition lines. Subsequent use of an $F_2$ mapping population from the interspecific cross *A. cepa* x *A. roylei* confirmed the assignment of an LFS locus to this chromosome. Sequence comparison of two BAC clones bearing LFS genes, LFS amplicons from diverse germplasm and expressed sequences from a doubled haploid line revealed variation consistent with duplicated LFS genes. Furthermore the BAC-FISH study using the two BAC clones as a probe showed that LFS genes are localized in the proximal region of the long arm of the chromosome. These results suggested that LFS in *A. cepa* is transcribed from at least two loci and that they are localized on chromosome 5.


## INTRODUCTION

The onion nuclear genomes is notable for its great size, 17 pg or 15 Gbp per 1C (Arumuganathan and Earle 1991; Bennett and Leitch 1995; Ricroch *et al.* 2005), one of the largest among cultivated plants. There are limited genomic resources available in onion caused by the huge genome size and genetic studies are further complicated because it is a biennial, out-crossing, and highly heterozygous species. However, an EST resource and PCR based map (Kuhl *et al.* 2005; Martin *et al.* 2005) has been recently developed. Combination with these resources and chromosome

addition lines (Shigyo *et al.* 1996) has revealed chromosomal location and genetic map position of genes responsible for important properties such as carbohydrate accumulation (McCallum *et al.* 2006; Masuzaki *et al.* 2006; Yaguchi *et al.* 2008) and flavonoid biosynthesis (Masuzaki *et al.* 2006). The physical distribution of AFLP markers along Allium chromosomes has been studied via the integration of recombination and physical maps in a trihybrid population, *A. cepa* x (*A. roylei* x *A. fistulosum*) (Khrustaleva *et al.* 2005). Direct physical mapping of genes on onion chromosomes is limited due to the genome abundance with repetitive elements (Stack and Comings 1979; Pearce *et al.* 1996). Fluorescence *in situ* hybridization (FISH) was successfully applied for the detection of specific loci using large genomic clones as probes mostly in plant species with small gene-rich genomes, such as *Arabidopsis thaliana* (Koornneeff *et al.* 2003) or rice (Jiang *et al.* 1995). However, in some cases using repetitive DNA that blocks probe hybridization of repetitive sequences allows for the detection of genes inserted in a Bacterial Artificial Chromosome (BAC) (Lamb *et al.* 2007; Szinay *et al.* 2008).

The most distinctive attribute of onion is the tearing property conferred by lachrymatory factor (LF; propanethial *S*-oxide). LF is formed from 1-propenylsulphenic acid by lachrymatory factor synthase (LFS) (Imai *et al.* 2002). 1-propenylsulphenic acid is a putative reaction product derived from 1-propenyl cysteine sulfoxides (trans-PRENCSO) by alliinase. Suppressing LFS gene turns off the conversion of 1-propenylsulfenic acid to LF, which in turn increases the yield of thiosulfinates thought to be responsible for onion flavor, and health beneficial properties of onion. Eady *et al.* (2008) previously demonstrated large shifts in organosulfur secondary compound profiles in onions in which LFS activity was suppressed by RNAi. These studies suggest that LFS is an important target for molecular breeding in onion. LFS cDNAs have been cloned from other five lachrymatory

Allium species (*A. ampeloprasum*, *A. cepa* Aggregatum group, *A. chinese*, *A. fistulosum* and *A. porrum*) and all recombinant proteins from them showed LFS enzymatic activity. In addition, strong homologies were observed in these LFS cDNA and no homologous sequences were yielded in GenBank search (Imai *et al.* 2005; Masamura *et al*. 2012). These results suggested that the LFS gene is strongly conserved among lachrymatory Allium species and is only distantly related to proteins in other higher plant taxa. Because LF is such a bioactive and distinctive compound, it is likely that strong selective forces have acted on LFS genes during evolution and domestication of Allium.

In this study, as a model of functional gene in huge genome, we determined the genome organization of LFS genes by sequence analysis, genetic mapping and physical methods in order to understand evolution of LFS in Allium and contribute to targeting molecular breeding and mutagenesis approaches for manipulating onion quality.

## MATERIALS AND METHODS

### Genetic analyses of LFS gene by using monosomic addition line and mapping

**population:** The plant materials were a complete set of *A. fistulosum*-shallot monosomic addition lines [$2n = 2x + 1 = 17$, FF+1A (plant number 130), FF+2A (141), FF+3A (5), FF+4A (10), FF+5A (26), FF+6A (308), FF+7A (324), FF+8A (240) ] and parental control plants, Japanese bunching onion (*A. fistulosum* cv. Kujyo-hoso, $2n = 2x = 16$, FF) and shallot (*A. cepa* Aggregatum group 'Chiang Mai', $2n = 2x = 16$, AA) (Shigyo *et al.* 1996). They were grown in an experimental field at Yamaguchi University (34°N, 131°E).

Genomic DNA were extracted from the frozen base of leaf sheath tissues by DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). PCR were performed an 25 µl reaction mixture containing 2.5 µl

of template, 1:10 or 1:100 diluted cDNA or genomic DNA (20 ng/µl), 0.125 µl of *Taq* polymerase [Ampli*Taq* GOLD (5 U/µl), Applied Biosystems, Foster City, CA, USA], 2.5 µl 10x PCR buffer, 1.5 µl of $MgCl_2$ (25 mM), 0.5 µl of forward primer (25 µM) and 0.5 µl of reverse primer (25 µM) of primer set (cepaLFS) and 2.5 µl of dNTP mixture (2 mM). Nucleotide sequences of the primers were shown in Table 1. PCR was carried out in GeneAmp 2400 or GeneAmp 9600 (Applied Biosystems, Foster City, CA, USA) with the following amplification program, an initial heating to activate the *Taq* polymerase for 94° for 10 min, followed by 35 cycles of 94° for 1 min, 65° or 68° for 1 min and 72° for 1 min, and then a final elongation at 72° for 10 min. The PCR products were detected by electrophoresis in 2% agarose gels.

DNA templates and genetic map data from the interspecific Allium cross *A. cepa* x *A. roylei* were used as described by van Heusden *et al.* (2000a, b). Primer sets used in this study are shown in Table 1. Design, PCR and analysis methods for SSCP and SSR markers were described previously (McCallum *et al.* 2006; 2007; 2008). Linkage analysis was performed using Joinmap Ver. 4 (van Ooijen 2006).

**Screening of BAC clones containing LFS gene and its molecular chracterization:** The partial BAC library of onion (Suzuki *et al.* 2001) was used for the PCR screening with onion LFS specific primers amplifying a 459 bp fragment comprising most of the LFS ORF. Nucleotide sequence of the primer set (LFSorf) used were shown in Table 1. PCR was performed with Ampli-*Taq* GOLD (Applied Biosystems, Foster City, CA, USA) for 40 cycles of denaturation for 1 min at 94° annealing for 1 min at 58° and extension for 1 min at 72° followed by a final extension for 7 min at 72°.

BAC DNA purified by QIAGEN plasmid Midi kit (Qiagen, Hilden, Germany) was digested with *Eco*RI or *Hin*dIII and separated on 0.7% agarose gel. After electrophoresis, DNA was transferred to HybondN$^{+}$ membrane (GE Healthcare UK Ltd., Buckinghamshire, UK). The digoxigenin (DIG)-labeled LFS specific DNA probe (452 bps) was prepared by PCR reaction using PCR DIG Labeling Mix (Roche Diagnostics, Manheim, Germany). Hybridization was carried out in 5x SSC, 50% formamide at 38.5° for overnight. The membrane was washed in 0.5x SSC, 0.1% SDS at 65° for 30 min. After blotting and washing, we detected the positive signal with fluorescence according to the instruction of the DIG fluorescent detection Kit (Roche Diagnostics, Manheim, Germany) with ECF (GE Healthcare UK Ltd., Buckinghamshire, UK) as the substrate.

Purified BAC clone 2E8/10 was nebulised and shotgun cloned into vector pSmartHC and 752 clones were sequenced from both ends (mean sequence length *ca.* 700 bp) by GATC Biotech Ltd Germany (Konstanz, Germany http://www.gatc-biotech.com). Trace files were assembled using the Staden package ver. 1.6 (Staden 1996) and Sequencher® (Gene Codes Corporation, Ann Arbor, MI, USA). The shotgun sequencing of BAC clone 4F10/155 was performed at Hokkaido System Science Co. Ltd (Sapporo, Japan) with the GS FLX Titanium system (454 Life Sciences, Roche Branford, CT, USA). The GS FLX Titanium generated ca.8.5 million bases (about 30 thousand reads) giving 85-fold coverage of the insert DNA of the clone. The reads were assembled into contigs with GS De Novo Assembler software (454 Life Sciences, Roche Branford, CT, USA). The contigs (~20 kb) containing LFS from both BAC clones were annotated using BLAST (Altschul *et al.* 1997) searches against Genbank plant protein and nucleotide databases. BLASTN comparison between contigs was visualized using genoPlotR (Guy *et al.* 2010). Tandem repeats were identified

using EMBOSS quicktandem (Rice *et al*. 2000). Sequences were submitted to Genbank (Accessions JN798503 and JN798504).

**BAC FISH analysis:** Mitotic chromosomes were prepared from young root meristems of bulb onion (Allium cepa cv. Khalcedon). Onion Cot-100 DNA was prepared as described by Zwick et al. (1997) with some modification. Total genomic DNA was isolated using CTAB method (Rogers and Bendich 1988) and was sonicated to a fragment size of about 1 kb. Sheared DNA was denatured in 0.3 M NaCl at 95º for 10 min, and then let it re-anneal at 65º for 31 h 40 min. The exact renaturation time was calculated taking in account the initial DNA concentration and percent of G-C fraction in onion genome. The remaining ssDNA was digested with S1 endonuclease (Fermentas, Burlington, ON, Canada, http://www.fermentas.com) with final concentration 1 U /µg for 90 min at 37º. The reaction was stopped and DNA was extracted by adding 300 µl chloroform/iso-amylalcohol (24:1). Purified BAC DNA was labeled with digoxigenin (DIG)-11-dUTP by Nick-translation standard protocol (Roche Diagnostics, Manheim, Germany). Slide pre-treatment was performed according to common FISH procedure. Hybridization was carried out with the mixture consisted of 50% (v/v) deionized formamide, 10% (w/v) dextran sulphate, 2x SSC, 0.25% sodium dodecyl sulphate, 12.5 ng/µl probe DNA and 0.3 µg/µl Cot-100. Stable sites of the probe hybridization were detected with anti-DIG FITC antibody. Chromosomes were counterstained with DAPI in Vectashield antifade (Vector Laboratories, Burlingame, CA, USA). Slides were examined under a Zeiss Axio Imager microscope (Carl Zeiss MicroImaging, Jena, Germany). Selected images were captured using a digital camera Axio Cam MRm. Image processing and thresholding were performed using AxioVision. Final image optimization was performed using PhotoshopTM (Adobe

Inc., San Jose, CA, USA). Detailed experimental conditions are presented in Supporting information, File S1.

**Re-sequencing of LFS Amplicons:** DNA was isolated as described previously (McCallum *et al.* 2006) from *Allium roylei* and the following onion genotypes selected on the basis of previous genetic diversity (McCallum *et al*. 2008) and pungency phenotypes: 'W202A', 'Texas Grano 438', 'BYG15-23', 'Alisa Craig 43', 'Colossal Grano', 'Early Longkeeper P12', 'W429A', 'Houston Grano', 'Tearless $F_1$', 'Faridpuri'. PCR reactions were performed in a 15 µl volume and contained 0.5 µM of primers LFS5L and R (Table 1) , 200 µM of dNTP's, 1.5 mM of $Mg^{2+}$, 10–30 ng of template DNA, 0.375 U of ThermoPrime*Taq* polymerase (Thermo Scientific, Massachusetts, USA) and 1x the manufacturers ReddyMix ™ PCR buffer. Reactions were performed in a Gene Amp® PCR System 9700 thermocycler (Applied Biosystems, Foster City, CA, USA). Cycling conditions were 2 min at 95° followed by a touchdown which included four cycles of 30 sec at 95°, 30 sec of 62° to 59° where this annealing temperature decreased by 1° each cycle and 30 sec at 72°. This touchdown was followed by 36 cycles of 30 sec at 95°, 30 sec at 58° and 30 sec at 72° and a final extension of 10 min at 72°. PCR products were visualized by electrophoresis on a 1% LE agarose plus 1% NuSieve (FMC Bioproducts, Rockland, ME, USA) agarose gel stained with ethidium bromide.

PCR products were then ligated into the pCR® 4 TOPO® Vector using a TOPO® TA Cloning® Kit (Invitrogen, Carlsbad, CA, USA) and transformed into DH5α by heat shock. Twelve colonies for each of the 12 onion lines were selected after growth on LB containing Amplicilin 100 µg/ml. These were streaked on Amplicilin 100 µg/ml and tested by PCR (see above). Eight colonies containing LFs amplicons from each of the twelve lines were used to inoculate a 1 ml LB (Ampicillin 100 µg/ml) overnight culture. Plasmids were then isolated using a Perfectprep® Plasmid

Isolation Kit (Eppendorf, Hamburg, Germany). LFS plasmid inserts were then Sanger-sequenced with M13 forward primer using Big Dye® Terminators v3.1 (Applied Biosystems, Foster City, CA, USA). Sequencing reactions were run on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Traces were trimmed and aligned using Geneious software (Drummond *et al.* 2011) and submitted into Genbank (Accession numbers HQ738844–HQ738919). Alignments were cropped to the region corresponding to bases 368–632 of Genbank Acc AB089203 and maximum likelihood trees were constructed using PHYML (Guindon and Gascual 2003) with 100 bootstraps. Nucleotide diversity calculations were performed using DnaSP (Librado and Rozas 2009).

**Identification of LFS Transcripts by 454 Sequencing:** Raw 454 flowgram files were used from Genbank BioProject accession PRJNA60277 for shoot transcriptome sequencing of doubled haploid onion line 'CUDH2150' (Cornell University; Genbank BioSample 138247) and 'Nasik Red' (USDA-ARS PI 271311; Genbank BioSample 138248). Reads were aligned to the LFS reference sequence (AB089203) using Roche gsMapper software using minimum overlap identity of 98% in the assembly step and minimum overlap length of 300 for 'DH2150' (Titanium reads) and 150 for 'Nasik Red' (GS-FLX reads).

**Amplification of LFS haplotypes in the FF+5A:** PCR were performed an 30 µl reaction mixture containing 3.0 µl of genomic DNA (20 ng/µl) purified from two different plants [FF+5A(26) and FF+5A(71)] of FF+5A or shallot parent, 0.15 µl of *Taq* polymerase (Ex*Taq*(5 U/µl), TakaraBio, Ohtsu, Japan), 3.0 µl 10x Ex buffer, 0.15 µl of forward primer (25 µM) and 0.15 µl of reverse primer (25 µM) of primer set (LFShaplo) and 2.4 µl of dNTP mixture (2.5 mM). Nucleotide sequences of the primers were shown in Table 1. PCR was carried out in GeneAmp 2400

1 or GeneAmp 9600 (Applied Biosystems, Foster City, CA, USA) with the following amplification
2 program, an initial heating to activate the *Taq* polymerase for 94° for 3 min, followed by 35 cycles of
3 94° for 1 min, 64.9° for 1 min and 72° for 1 min. The 5 µl of PCR reactions were subjected to
4 restriction endonuclease digestion with *Nsi*I and *Sph*I enzymes and the fragments were analyzed by
5 electrophoresis in 2.0% agarose gels.

Table 1 Previously unpublished primer sets used in this study.

| Primer Set | Marker type | Genbank Accession Number | Forward Primer<br>Reverse Primer |
|---|---|---|---|
| cepaLFS | — | AB089203 | ACAAAGCCAGAGCAAGCATGGACA<br>CTGCAAACCTCTTCGATTTTCTGACCTATC |
| LFS5 | Heteroduplex | AB089203 | GCACTAGAACTTGCAAAAAGCA<br>TGAGATAGGTCAGAAAATCGAAGA |
| ACP052 | SSCP | CF445004,<br>CF445805,<br>CF445805 | TTCCCTCCTCACTCCCTACA<br>CGACCACAAACACAAGCAAC |
| ACM295 | SSR | CF445600 | AGATCCGTCCCATGAAACT<br>GATCCGCTTCTGAAATCTCG |
| ACM021 | SSR | CF448154 | AAAACCCTCAACATCTCACTCC<br>TCTCTTCTTCCTCGTCCTGC |
| ACM076 | SSR | CF449018 | ATTAGAAACATCCATCGCCG<br>CGCGATCATCATTTTCCATA |
| ACP003 | SSCP | BE205590 | AAGCTCTTAAAGCTGCTGATGG<br>ATGCACGATAGCACAAGACATC |
| LFSorf | — | AB089203 | ACAAAGCCAGAGCAAGCATGGACA<br>CTGCAAACCTCTTCGATTTTCTGACCTATC |
| LFShaplo | — | AB089203 | ATAGTGGAGGGTCCTGAGCA<br>ACACAACACTCAGTCTTACTTATT |

1

2

## RESULTS

**Chromosomal mapping of LFS gene in *A. cepa*:** From the genomic DNA of alien monosomic addition lines, the expected size amplicon was observed only in the alien monosomic addition line FF+5A (Figure 1). Therefore, we assigned LFS of *A. cepa* to chromosome 5A. Genetic

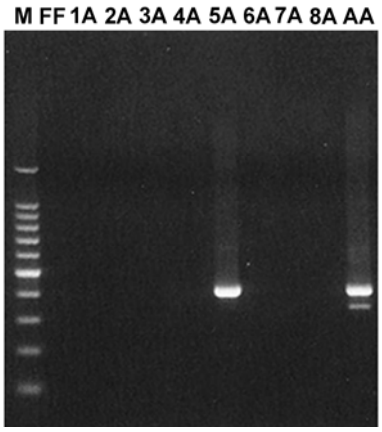


**Figure 1**

mapping of polymorphisms detected by heteroduplex analysis of LFS amplicons in the *A. cepa* x *A. roylei* interspecific cross revealed co-segregation with chromosome 5 markers (Figure 2). The LFS marker was linked to several other markers developed to onion ESTs, most notably the SNP marker ACP052. This was designed to partial ESTs showing homology to N terminal regions of group I

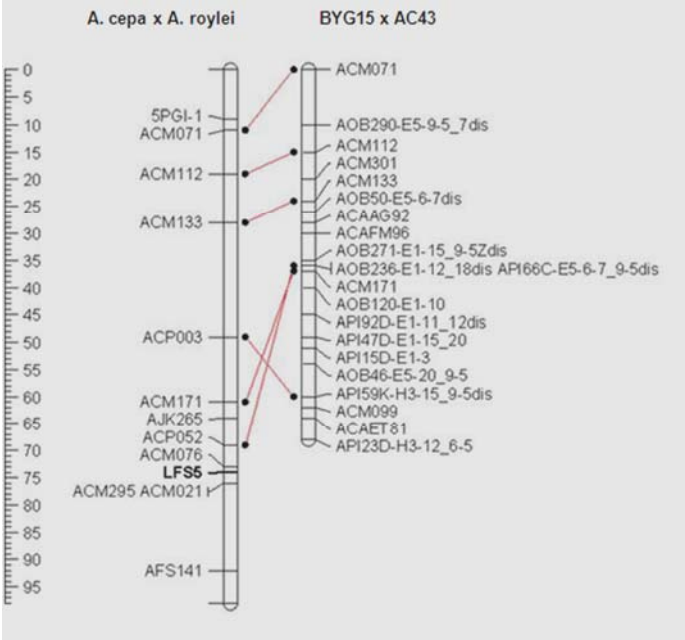


**Figure 2**

sucrose transporters (Braun and Slewinski 2009). The RFLP marker API66C-E5 mapped in the 'BYG15-23 x AC43' intraspecific onion population (King *et al.* 1998) was revealed by the cDNA probe API66 (GenBank Accession BE205593.1) which is homologous to sucrose transporter group I and III proteins. It is not known yet whether ACP052 and API66 markers target the same gene or linked duplications. We previously showed close linkage of API66 markers to a well-supported QTL affecting bulb dry matter (Galmarini *et al.* 2001; Martin *et al.* 2005; Masuzaki *et al.* 2007; McCallum *et al.* 2007), suggesting that the LFS loci are in close linkage with this QTL.

**Selection and characterization of BAC clones containing LFS gene:** PCR screening of the BAC library revealed eight positive clones, from a total of 48,000 clones comprising 0.32 of the onion genome (Suzuki *et al.* 2002). Southern blot analysis of the positive clones was carried out by using a DIG-labeled LFS probe (Figure 3). In *Hin*d III digestion, all eight clones showed signal in similar size but *Eco*RI digestion, showed a larger size in 4F10/155 compared to the other seven clones. This observation suggested that there are at least two LFS loci in onion genome.

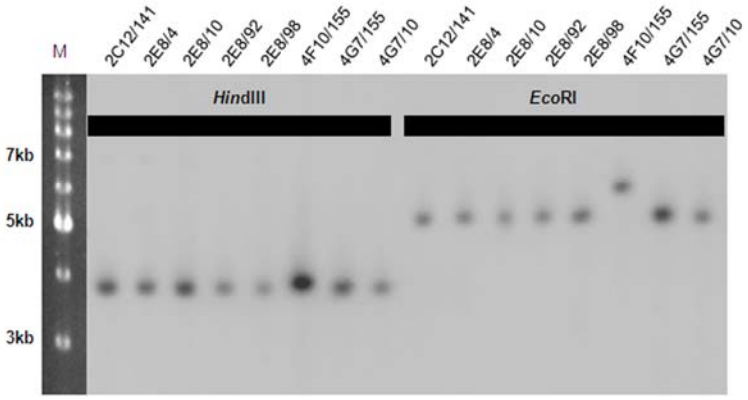


**Figure 3**

**Sequence comparison between two LFS containing BAC clones:** Sanger and 454 shotgun sequencing of clones 2E8/10 and 4F10/155 provided contigs of 19 and 7 which were longer than 500 bp respectively. From Sanger shotgun analysis of 2E8/10, a LFS ORF sequence was detected in a contig containing 8,623 bp and sequence of 20,577 bp (from upper 13 kbp to lower 7 kbp of LFS ORF) was built with average coverage of 5–6 by assembling the contigs. In the 454 contigs (total number of bases were 138,394 bp) of 4F10/155, A LFS ORF sequence was observed in the longest contig containing 71,391 bp.

Nucleotide sequence comparison of the approximately 20 kbp contigs from both BAC exhibited high similarity over 10 kbp flanking the intronless LFS gene. Notable areas of lower similarity within this region were the region immediately upstream of the LFS gene, including its promoter, and a 1.5 kbp insertion in 2E8/10 showing BLASTX similarity to polyproteins (Figure 4). Both contigs shared a 445 bp direct repeat in 1 kbp upstream and a region of homology to other onion BAC sequences in 8–14 kbp upstream of LFS. Regions 3–6 kbp downstream showed BLASTX homology to plant Ty1-copia like elements. This same general pattern of degenerated retroviral elements and transposons was reported in earlier onion BAC sequencing by Jakše *et al.* (2008). Sequence comparison in the region corresponding to LFS transcripts revealed five differences at sites corresponding to the following locations in Genbank Accession AB094593: a

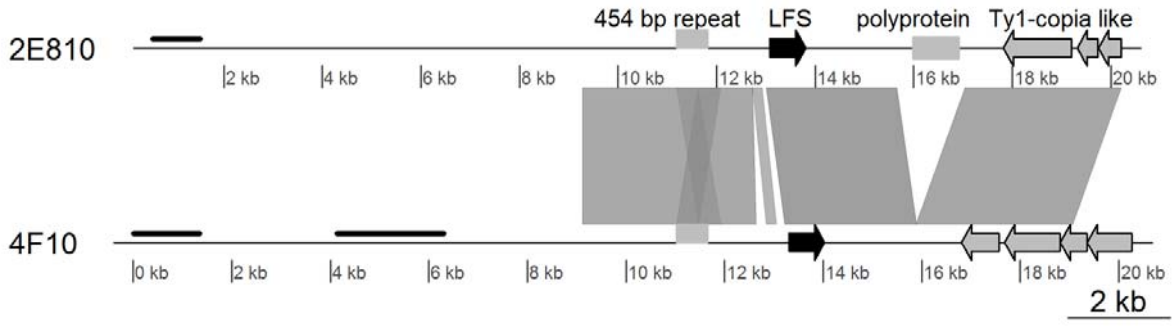


**Figure 4**

non-synonymous substitution 203A>T conditioning the mutation V50D in the translation; two synonymous variants (441T>C and 474C>T) in the coding region and two variants in the 3'UTR (596A>G and 654T>C).

**Chromosomal localizations of BAC clones:** The Dig-labeled BAC clones, 2E8/10 and 4F10/155 were hybridized to the mitotic metaphase chromosome preparation of *A. cepa*. By using the Cot-100 fraction to block the repetitive sequence hybridization, both BAC clones generated distinct signals from a single pair of somatic metaphase chromosomes (Figure 5A). The karyotype analysis revealed that signals were located on the proximal region of the long arm of chromosome 5 (Figure 5B). The identity of this chromosome was established on the basis of its chromosome size and the position of its centromere (Relative chromosome length 12.7±1.0, Centromere index 48.7±0.7; de Vries 1990). Each BAC's location from the centromere was measured (20 chromosomes from 10 metaphases per each BAC) and their relative positions were estimated. The position of 2E8/10 was 0.31±0.03 and the position of 4F10/155 was 0.32±0.03 (Figure 5C). No statistically significant difference was found between the position of hybridization signals in the two BAC clones (Student *t*-test, td = 0.195, tst = 2.04, n = 40, $P = 0.05$).

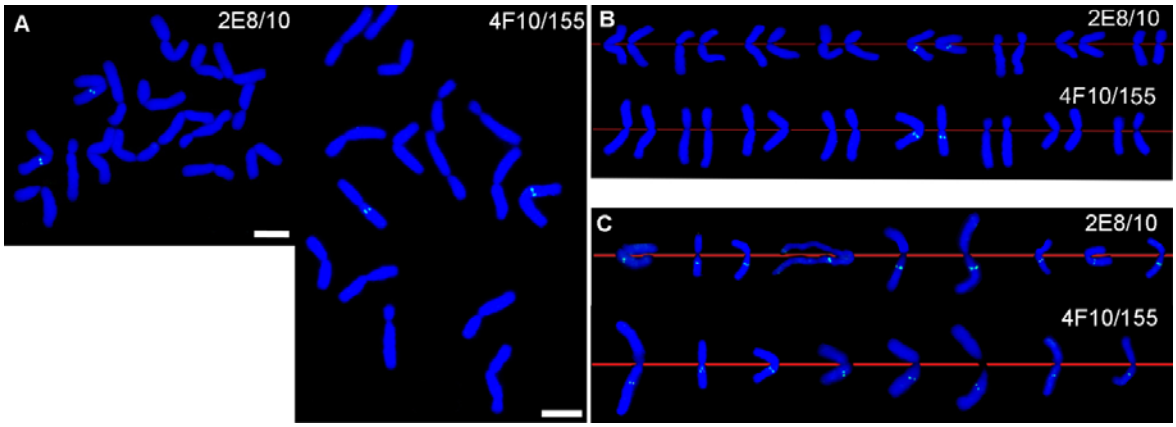


**Figure 5**

**Sequence variation in LFS amplicons and transcripts:** Mapping leaf transcript 454 reads from the doubled haploid line 'CUDH2150' to *A. cepa* LFS cDNA sequences (Genbank accessions AB094593, AB089203) with stringent criteria revealed 119 matching reads containing three variants at intermediate frequencies (Table 2). The same variants were also observed in 'Nasik Red' and formed two haplotypes. In a previous study (McCallum *et al.* 2008) we surveyed allelic variation at multiple SSR loci and failed to detect any heterozygosity in 'CUDH2150', suggesting that the observed variants are transcripts from duplicated LFS genes. BLAST searches of Genbank EST division revealed that in addition to ESTs with the same haplotype as AB094593 ('haplotype 1'), three onion ESTs exhibited the alternate haplotype ('haplotype 2') at these three sites (accession numbers CF451348, FK935151, FK936343).

Table 2 Counts of 454 cDNA reads from '*CUDH2150*' and *'Nasik Red*' cDNA 454 cDNA sequencing classified by variants.

| Site[a] | AB089203 | Variant | Total Depth | | Variant frequency % | |
|---|---|---|---|---|---|---|
| | | | *CUDH2150* | *Nasik Red* | *CUDH2150* | *Nasik Red* |
| 439 | T | C | 106 | 37 | 41 | 49 |
| 594 | A | G | 75 | 39 | 48 | 23 |
| 652 | T | C | 62 | 31 | 50 | 26 |

[a] Site positions are relative to Genbank Acc AB089203.

Sanger sequencing of LFS amplicons from a range of diverse onion germplasm and *A. roylei* revealed these sequence variants as well as rarer ones (Figure 6). The most commonly observed haplotype was that matching AB094593 ('haplotype 1'), and this was amplified from all onion populations surveyed. Trees based on alignment of cDNA and genomic sequences (Figure 7) revealed clear clustering of onion sequences into two groups corresponding to haplotypes 1 and 2 observed in 454 sequencing and an outgroup containing other Allium species. The variant

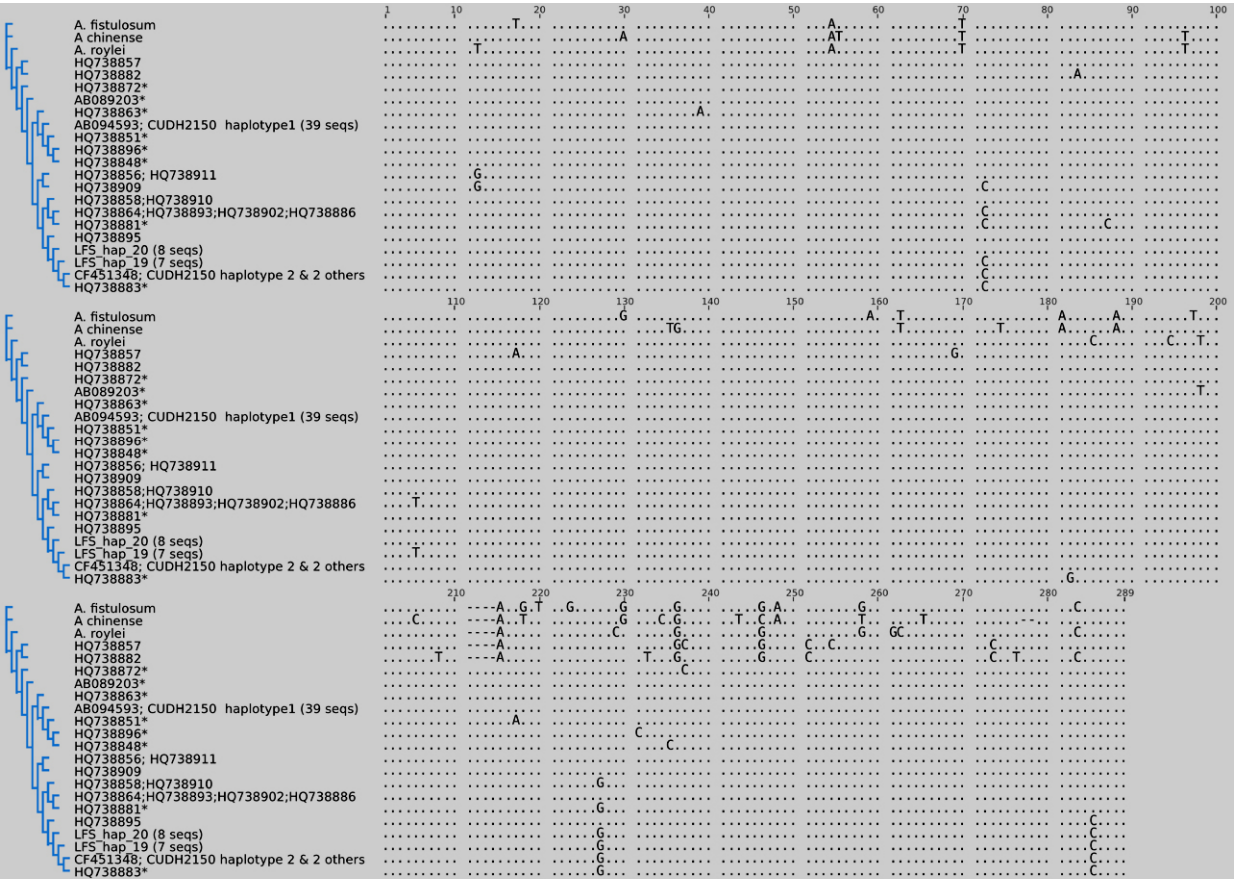


**Figure 6**

distinguishing AB089203 from other LFS sequences (A570T) was not observed in any other reads suggesting this is a PCR or sequencing error. Four singleton reads with single base differences to the haplotype 1 sequence were observed, one of which (HQ738863) conditioned a non synonymous M > I mutation. By contrast, several well-supported haplotypes were observed within the 'haplotype 2'

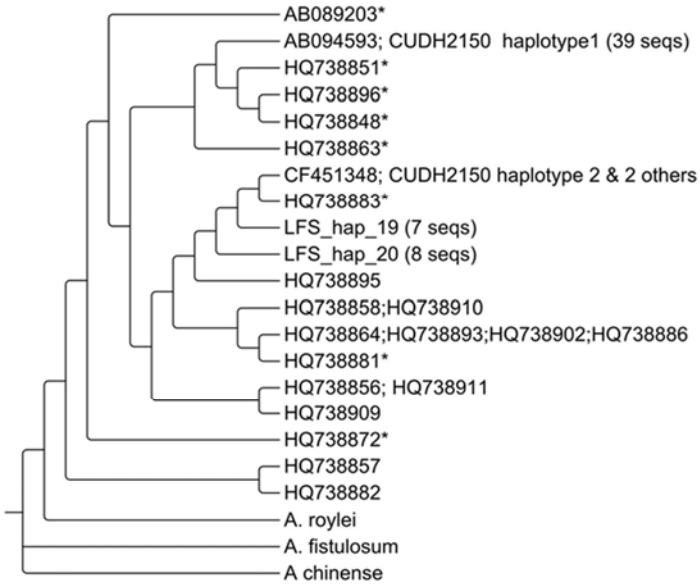


**Figure 7**

group corresponding to BAC 2E8/10. And a non-synonymous variant (HQ738883, V > G at 166) was also found in the group. Average nucleotide diversity was higher in the haplotype 2 group ($\pi$ = 0.0042) than in haplotype 1($\pi$ = 0.0028).

The out-group containing related Allium species also contained two sequences amplified from *A. cepa* 'Faridpuri' and 'AC43' (HQ738857, HQ738882,). Notably, these sequences exhibit a deletion corresponding to bases 578–581 of AB089203 and other variants in the 3' UTR that observed in LFS cDNA sequences from *A. roylei, A. chinense* and *A. fistulosum.* HQ738882 contains a premature stop (G450A) within the putative coding region. These may represent rare or ancestral alleles but could also be products amplified from other LFS loci.

A test of the hypothesis of selective neutrality in this sequence region in the two paralog groups was tested using Tajimas' D. This revealed a significant deviation from neutral expectations for the haplotype 1 group (D=−1.91; N=55 sequences; $P < 0.05$) but not the haplotype 2 group (D=−0.23; N=21 sequences; $P > 0.10$).

**Existence of distinctive LFS haplotypes in the FF+5A:** The A594G and T652C variants distinguishing the two haplotype groups condition *Nsi*I and *Sph*I restriction sites respectively and we used this to test for existence of distinctive haplotypes in the FF+5A. This revealed both *Nsi*I and *Sph*I digestion of the LFS product amplified from this line but no evidence for double digestion (Figure 8). This confirms that the two putative paralogous loci are both located on onion chromosome 5.

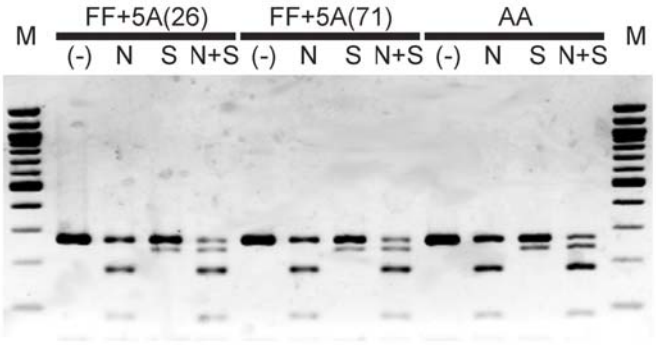


**Figure 8**

## DISCUSSION

By the use of multiple approaches this study has determined that LFS in onion is transcribed from at least two loci and that they are localized on chromosome 5. Southern blotting of eight LFS bearing clones revealed two distinct *Eco*RI RFLP patterns but assignments using alien addition lines and mapping analysis in an inter-specific cross placed the gene only on chromosome 5. Furthermore, BAC-FISH study showed co-localization of these BACs. Under the optical limit of a 1.4 numerical aperture for a conventional microscope objective, a maximum distance of only 0.2 µm can be resolved. De Jong *et al.* (1999) reported that mitotic metaphase FISH can resolve 4–5 Mbp in a tomato. The level of metaphase chromosome condensation in Allium is about five times higher

than in tomato metaphase chromosomes (Khrustaleva and Kik 2001). Thus, 2E8/10 and 4F10/155 clones can be distant from each other up to 25 Mbp and still be located in the same position on the mitotic metaphase chromosome. Because onion is highly heterozygous, a possible interpretation of these results is heterozygosity at the LFS locus. This possibility was eliminated through transcriptome sequencing in a doubled haploid genotype, which also revealed two distinct sequence variants. This confirms that LFS is transcribed from two distinct loci.

The observation of LFS-like sequences in onion closely resembling those from related Allium species may provide some further scope to study the evolution of LFS in Allium. These sequences may be rare or ancestral alleles at either of the major loci but could also be cross-amplified from other LFS loci. It is possible that such sequences could represent ancestral copies of the gene family, and indicate that gene duplication is still ongoing in onion genome. The LFS gene is small and intronless. In plants, small and intronless genes show that they are frequently plant- or lineage-specific (Jain *et al*. 2008). It is therefore plausible that LFS is such an example of a novel, genus-specific gene family that has arisen recently and provided a strong selective advantage. A survey by PCR across diverse germplasm suggested lower nucleotide diversity in haplotype group 1 in the region surveyed compared to the haplotype group 2 corresponding to BAC 2E8/10 locus. This is suggestive of possible differences in selective constraints but wider sampling of nucleotide diversity is required to test the role of purifying selection and gene conversion on these loci.

Gene duplication is known to be an important source of evolutionary innovation and adaptation (Des Marais and Rauscher 2008). Mapping studies have suggested extensive duplication in onion. King *et al.* (1998) reported that 21% of 91 cDNA probes of RFLP detected more than one segregating RFLP, of which 53% were unlinked, 47% were linked. These findings indicated gene

duplication has occurred with high frequency in onion genome, the ratio of the linked duplicated RFLP were higher than other plants. Our results demonstrate that LFS genes exist as functional linked duplicated genes , implying a low possibility for obtaining null mutants through mutant screening. However, BAC sequence analysis and PCR experiment to distinguish two haplotypes indicated that 2E8/10 was 'haplotype 2' and 4F10/155 was 'haplotype 1', and showed a large difference in upper flanking 1.2 kbp region of LFS between the BACs. These observations of variation in the promoter regions suggest that there may be differential expression between the paralogs. Kim *et al*. (2005) speculated similarly in pink onion. More detailed surveys of paralog-specific expression might reveal mutants with lower LFS activity. The identification of the complete genomic sequence surrounding the two paralogs will now enable wider surveys for natural mutants by deep sequencing, like the variant (HQ738882) possessing premature stop we identified in 'Faridpuri'. Using such genotypes as a material for further mutant induction or breeding, we may achieve production of LFS null onion. Non-synonymous variants found in this paper, HQ738883 (V > G) may have its enzymatic activity equivalent to that of wild-type LFS (Masamura et al. 2012) Because a deletion C-terminal nine amino acids (160-169) had no effect on the enzymatic activity (Masamura *et al*. 2012) and it remains unclear whether another non-synonymous variant HQ738863 (M > I) has enzymatic activity. However, LFS would catalyze an intra-molecular $H^+$ substitution reaction, thus this variant which changes between non-polar amino acids might retain the enzymatic activity same as that of wild type onion LFS. From same reason, we speculate that the variant (V50D) found in 4F10/155 has a low potential for loss of the enzymatic activity.

It has been reported that a well-supported QTL affecting bulb dry matter and a putative sucrose transporter gene are located on chromosome 5 (Galmarini *et al.* 2001; Martin *et al.* 2005;

Masuzaki *et al.* 2007; McCallum *et al.* 2007). Dry matter is an important trait affecting firmness and storability of onion bulb, and therefore survival under human or natural selection. It is notable that LFS is located also in this region, sufficiently close that we would expect some effects on LFS through linkage drag. Consequently this region of chromosome 5 could be an important domestication region in the onion genome and it would be fruitful for further functional and crop evolutionary studies.

These findings provide a highly relevant case study of a duplication leading to multiple functional loci encoding a gene with important adaptive trait in onion.

## REFERENCES


Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Arumuganathan K., E. D. Earle, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9:** 208–218.

Braun D. M., T. L. Slewinski, 2009 Genetic control of carbon partitioning in grasses: roles of sucrose transporters and tie-dyed loci in phloem loading. Plant Physiol. **149:** 71–81.

Bennett M. D., I. J. Leitch, 1995 Nuclear DNA amounts in angiosperms. Ann. Bot. **76:** 113–176.

Des Marais D., and M. D. Rausher, 2008 Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature **454:** 762–765.

De Vries J. N., 1990 Onion chromosome nomenclature and homology relationships—workshop report. Euphytica **49:** 1–3

Drummond A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper *et al.*, 2011 Geneious v5.4, Available from http://www.geneious.com/

Eady C.C., T. Kamoi, M. Kato, G. Porter Noel, S. Davis *et al.*, 2008 Silencing onion lachrymatory factor synthase causes a significant change in the sulfur secondary metabolite profile. Plant Physiology **147:** 2096–2106.

Galmarini C. R., I. L. Goldman, M. J. Havey, 2001 Genetic analyses of correlated solids, flavor, and health-enhancing traits in onion (*Allium cepa* L.). Mol. Genet. Genomics. **265:** 543–551.

Guy L., J. R. Kultima, S. G. E. Andersson, 2010 genoPlotR: comparative gene and genome visualization in R. Bioinformatics **26 (18):** 2334–2335

Guindon S., O. Gascuel, 2003 A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:** 696–704.

Van Heusden A. W., J. W. van Ooijen, R. Vrielink-van Ginkel, W. H. J. Verbeek, W. A. Wietsm *et al.*, 2000a A genetic map of an interspecific cross in Allium based on amplified fragment length polymorphism (AFLP$^{TM}$) markers. Theor. Appl. Genet. **100**: 118–126.

Van Heusden A. W., M. Shigyo, Y. Tashiro, R. Vrielink-van Ginkel, C. Kik, 2000b AFLP linkage group assignment to the chromosomes of *Allium cepa* L. via monosomic addition lines. Theor. Appl. Genet.**100:** 480–486.

Imai S., N. Tsuge, M. Tomotake, Y. Nagatome, H. Sawada *et al.*, 2002 An onion enzyme that makes the eyes water. Nature **419**: 685.

Imai S., N. Tsuge, M. Tomotake, N. Masamura, H. Sawada *et al.*, 2005 The production of a lachrymatory factor in onion is mediated by a novel enzyme, pp. 123–126 in Sulfur transport

and assimilation in plants in the post genomic era, edited by K. Saito, L. J. DeKok, I. Stulen, M. J. Hawkesfold, E. Schnug *et al.*, Backhuys Publishers, Leiden, The Netherlands.

Jakše J., J. D. F. Meyer, G. Suzuki, J. McCallum, F. Cheung *et al.*, 2008 Pilot sequencing of onion genomic DNA reveals fragments of transposable elements, low gene densities, and significant gene enrichment after methyl filtration. Mol. Genet. Genomics **280:** 287–292.

Jain M., P. Khurana, A. K. Tyagi, J. P. Khurana, 2008 Genome-wide analysis of intronless genes in rice and Arabidopsis. Funct. Integr. Genomics **8:** 69–78.

Jiang J., B. S. Gill, G-L. Wang, P. C. Ronald, D. C. Ward, 1995 Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc. Natl. Acad. Sci. USA **92:** 4487–4491.

Khrustaleva L.I., and C. Kik, 2001 Localization of single-copy T-DNA insertion in transgenic shallots (*Allium cepa*) by using ultra-sensitive FISH with tyramide signal amplification. The Plant Journal **25:** 699–707.

Khrustaleva L.I., P. E. de Melo, A. W. van Heusden, C. Kik, 2005 The integration of recombination and physical maps in a large-genome monocots using haploid genome analysis in a trihybrid Allium population. Genetics **169:** 1673–1685.

Kim S., K-S. Yoo, L. M. Pike, 2005 Development of a codominant PCR-based marker for allelic selection of the pink trait in onion (*Allium cepa*), based on the insertion mutation in the promoter of the anthocyanidin synthase gene. Theor. Appl. Genet. 110**:** 573–578

King J. J., J. M. Bradeen, O. Bark, J. A. McCallum, M. J. Havey, 1998 A low-density genetic map of onion reveals a role for tandem duplication in the evolution of the extremely large diploid genome. Theor. Appl. Genet **96:** 52–62.

Koornneef M., C. Alonso-Blanco, D. Vreugdenhil, 2004 Naturally occurring genetic variation in Arabidopsis thaliana. Annu. Rev. Plant Biol. **55:** 141–172.

Kuhl J.C., F. Cheung, Q. Yuan, W. Martin, Y. Zewdie *et al.*, 2004 A unique set of 11,008 onion expressed sequence tags reveals expressed sequence and genomic differences between the monocot orders Asparagales and Poales. The Plant Cell **16:** 114–125.

Lamb J. C., T. T. Danilova, M. J. Bauer, J. M. Meyer, J. J. Holland *et al.*, 2007 Single-gene detection and karyotyping using small-target fluorescence *in situ* hybridization on maize somatic chromosomes. Genetics **175:** 1047–1058.

Librado P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:** 1451–1452

Martin W. J., J. McCallum, M. Shigyo, J. Jakše, J. C. Kuhl *et al.*, 2005 Genetic mapping of expressed sequences in onion and in silico comparisons with rice show scant colinearity. Mol. Genet. Genomics **274:** 197–204.

Masamura N., W. Ohashi, N. Tsuge, S. Imai, A. Ishii-Nakamura *et al.*, 2012 Identification of amino acid residues essential for onion lachrymatory factor synthase activity. Biosci. Biotechnol. Biochem. **76:** in press.

Masuzaki S., M. Shigyo, N. Yamauchi, 2006 Direct comparison between genomic constitution and flavonoid contents in *Allium* multiple alien addition lines reveals chromosomal locations of genes related to biosynthesis from dihydorokaempferol to quercetin glucosides in scaly leaf of shallot (*Allium cepa* L.). Theor. Appl. Genet. **112:** 607–617.

Masuzaki S., S. Yaguchi, N. Yamauchi, M. Shigyo, 2007 Morphological characterisation of multiple alien addition lines of *Allium* reveals the chromosomal locations of gene(s) related to bulb formation in *Allium cepa* L. J. Hortic. Sci. Biotechnol. **82:** 393–396.

McCallum J, A. Clarke, M. Pither-Joyce, M. Shaw, R. Butler *et al.*, 2006 Genetic mapping of a major gene affecting onion bulb fructan content. Theor. Appl. Genet. **112:** 958–967.

McCallum J., M. Pither-Joyce, M. Shae, F. Kenel, S. Davis *et al.*, 2007 Genetic mapping of sulfur assimilation genes reveals a QTL for onion bulb pungency. Theor. Appl. Genet. **114:** 815–822

McCallum J., S. Thomson, M. Pither-Joyce, F. Kenel, A. Clarke *et al.*, 2008 Genetic diversity analysis and single-nucleotide polymorphism marker development in cultivated bulb onion based on expressed sequence tag-simple sequence repeat markers. J. American Society for Horticultural Science **133:** 810–818.

van Ooijen R., 2006 JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen,: Netherlands.

Pearce S. R., U. Pich, G. Harrison, A. J. Flavell, J. S. Heslop-Harrison *et al.*, 1996 The Ty1-copia group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. Chromosome Res. **4:** 357–364.

Pijnacker L. P., and M. A. Ferwerda, 1984 Giemsa C-banding of potato chromosomes. Can. J. Genet. Cytol. **26:** 415–419.

Randle W. M., D. E. Kopsell, D. A. Kopsell, 2002 Sequentially reducing sulfate fertility during onion growth and development affects bulb flavor at harvest. HortScience **37:** 118–121.

1 Rice, P., I. Longden, A. Bleasby, 2000 EMBOSS: The European molecular biology open Software
2 suite. Trends in Genetics **16:** 276–277
3 Ricroch A., R. Yockteng, S. C. Brown, S. Nadot, 2005 Evolution of genome size across some
4 cultivated *Allium* species. Genome **48:** 511–520.
5 Rogers, S.O., and A. J. Bendich, 1988 Extraction of DNA from plant tissues, pp. 1–10 in Plant
6 Molecular Biology Manual A6 edited by S. B. Gelvin and R. A. Schilperoort. Kluwer
7 Academic Publishers, Dordrecht, The Netherlands..
8 Shigyo M., Y. Tashiro, S. Isshiki, S. Miyazaki, 1996 Establishment of a series of alien monosomic
9 addition lines of Japanese bunching onion (*Allium fistulosum* L.) with extra chromosomes
10 from shallot (*A. cepa* L. Aggregatum group). Genes Genet. Syst. **71:** 363–371.
11 Stack S. M., and D. E. Comings, 1979 The chromosomes and DNA of *Allium cepa*. Chromosoma
12 **70:** 161–181.
13 Staden R., 1996 The staden sequence analysis package. Molecular Biotechnology **5:** 233–241.
14 Suzuki G, N. Saito, G. S. Do, B. B. Seo, M. Yamamoto *et al.*, 2001 BAC FISH analysis in *Allium*
15 *cepa*. Genes Genet. Syst. **76:** 251–255.
16 Suzuki G, G. S. Do, Y. Mukai, 2002 Efficient storage and screening system for onion BAC clones.
17 Breeding Science **52:** 157–159.
18 Szinay D., S. B. Chang, L. Khrustaleva, S. Peters, E. Schijlen *et al.*, 2008 High-resolution
19 chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a
20 backbone for sequencing tomato chromosome 6. Plant J. **56:** 627–37.

Tsukazaki H, K., Yamashita, S. Yaguchi, S. Masuzaki, H. Fukuoka *et al.*, 2008 Construction of SSR-based chromosome map in bunching onion (*Allium fistulosum*). Theor. Appl. Genet. **117:** 1213–1223.

Yaguchi S., J. McCallum, M. Shaw, M. Pither-Joyce, S. Onodera *et al.*, 2008 Biochemical and genetic analysis of carbohydrate accumulation in *Allium cepa* L. Plant Cell Physiol. **49:** 730–739.

Zwick M. S., R. E. Hanson, T. D. McKnight, M. N. Islam-Faridi, D. M. Stelly *et al.*, 1997 A rapid procedure for the isolation of Cot-1 DNA from plants. Genome **40:** 138–142.

**Figure Legends:**

Figure 1

Amplification of LFS from AMALs. M, molecular size marker (100bp ladder); FF, *A. fistulosum*; 1A-8A, eight different *A. fistulosum* – shallot monosomic additions; AA, shallot

Figure 2

Genetic mapping of LFS heteroduplex marker LFS5 to chromsome 5 in the *A. cepa* x *A. roylei* population and alignment with the onion linkage map (BYG15 x AC43) of Martin et al. (2005). The scale denotes the recombination distance in Kosambi units. AFLP markers in the interspecific map have been deleted for clarity.

Figure 3

Southern blot analyses of eight LFS-containing BAC plasmid with LFS probe. HindIII and EcoRI indicate digestion with HindIII or EcoRI, respectively. M lane shows 1 kb ladder marker lane of the gel used to this southern blot.

Figure 4

Sequence similarity and annotation of LFS flanking regions of 2E8/10 and 410/155. Gray-scale shading denotes sequence similarity over 80% detected by BLASTN alignment of sequences. Black arrow indicates intronless LFS ORF. Gray side-bar denotes a 454 bp direct tandem repeat. Gray bar denotes insertion in 2E810 sequence with BLASTX homology to polyproteins and gray arrows

denote regions with BLASTX similarity to Ty1-copia like sequences. Black lines denote regions with BLASTN similarity to other onion BAC sequences.

Figure 5

BAC-FISH using two LFS bearing clones (2E8/10 and 4F10/155) as a probe. A) mapping of BAC clones on onion mitotic metaphase chromosomes: Bar indicated 10μm, B) FISH karyotypes and C) chromosome 5 extracted from a number of mitotic metaphases, which possess hybridization signal (green) in the proximal region of the long arm.

Figure 6

Alignment of cDNA sequences from onion and related Allium species with haplotypes observed in region corresponding to bases 370-658 of GB accession AB089203 in genomic PCR amplicons of LFS genes from diverse onion germplasm. Accession numbers marked by asterisk denotes haplotypes based on sequence variants only supported by a single read, which may represent PCR errors.

Figure 7

Cladogram of LFS sequences based on consensus maximum likelihood tree of LFS sequence alignments with 100 bootstraps

Figure 8

Restriction analysis of LFS PCR products from chromosome 5 AMALs. Lanes 1-4 FF+5A (plant number 26); Lanes 5-8 FF+5A (71); Lanes 9-12 AA ('Chiang Mai'). (-), N, S, and N+S indicate uncut PCR product, NsiI digested, SphI digest and double digest by NsiI + SphI. M shows 100 bp ladder marker. The two AMALs were the sibs obtained from a single cross.

**Supporting information**

## File S1 detailed experimental conditions of BAC-FISH

### Chromosome preparation

Young root meristems of bulb onion (*Allium cepa* cv. Khalcedon) were collected in water saturated α-bromnaphthalene (1:1000). The root meristems were kept in α-bromnaphthalene overnight at 4°. After this treatment the roots were fixed in methanol/acetic acid fixative (3:1) for at least 1h at room temperature. The roots were rinsed three times in distilled water and once in 10 mM sodium citrate buffer (pH 4.5) before transferring to an enzyme mix containing 0.1% pectolyase Y23 (Sigma-Aldrich, St. Louis, Missouri, USA), 0.1% cellulase RS (Yakult Pharmaceutical, Tokyo, Japan) and 0.1% cytohelicase (Bio Sepra, Idstein, Germany) in citrate buffer for 35–50 min at 37º. The roots were carefully transferred to water and then left on ice until further use. Somatic metaphase chromosome spreads were made according to Pijnacker and Ferwerda (1984) with slight modification. In brief, dissected meristemic tissue were placed in very clean grease-free slides and cell suspension were made with fine needles, then 60% acetic acid were dropped onto the cells for further maceration on a hot block at 42° with stirring for 2 min. 100 µl ethanol/acetic acid fixative (3:1) were added around droplet of acetic acid. Finally slides were rinsed briefly in 96% ethanol and were air dried.

### Pre-treatment procedure of slide

Slides were dried overnight at 37º, treated with RNAse (0.1 mg/ml) in 2x SSC for 1 h and then with pepsin (5 µg/ml) in 0.01 M HCl, after that treatment slides were incubate in 4% paraformaldehyde for

10 min. Slides were washed in 2x SSC three times for 5 min between each steps of treatment. Finally, slides were dehydrated in 70%, 90% and 100% ethanol for 3 min each and air dry.

**Hybridization condition**

The mixture was boiled for 5 min and subsequently placed on ice for 5 min. 40 ml of the mixture was administered to the chromosome preparations, covered with a coverslip (22 x 22 mm), and denaturated for 5 min at 80°. An 81% stringency washing was applied. The slides were washed in 2x SSC for 5 min at 42°, in 50% (v/v) formamide in 2x SSC twice for 7 min at 42°, in 2x SSC for 3 min at 42°.