Molecular Genetics and Genomics
https://doi.org/10.1007/s00438-018-1442-5

ORIGINAL ARTICLE

Microsatellite analysis and marker development in garlic: distribution in EST sequence, genetic diversity analysis, and marker transferability across Alliaceae

Karina Barboza[1] · Vanesa Beretta[1] · Perla C. Kozub[2] · Cecilia Salinas[1] · Mauro M. Morgenfeld[3] · Claudio R. Galmarini[1,4] · Pablo F. Cavagnaro[1,4]

Received: 28 March 2018 / Accepted: 25 April 2018

Abstract

Allium vegetables, such as garlic and onion, have understudied genomes and limited molecular resources, hindering advances in genetic research and breeding of these species. In this study, we characterized and compared the simple sequence repeats (SSR) landscape in the transcriptomes of garlic and related *Allium* (*A. cepa, A. fistulosum, and A. tuberosum*) and non-*Allium* monocot species. In addition, 110 SSR markers were developed from garlic ESTs, and they were characterized—along with 112 previously developed SSRs—at various levels, including transferability across Alliaceae species, and their usefulness for genetic diversity analysis. Among the *Allium* species analyzed, garlic ESTs had the highest overall SSR density, the lowest frequency of trinucleotides, and the highest of di- and tetranucleotides. When compared to more distantly related monocots, outside the Asparagales order, it was evident that ESTs of *Allium* species shared major commonalities with regards to SSR density, frequency distribution, sequence motifs, and GC content. A significant fraction of the SSR markers were successfully transferred across *Allium* species, including crops for which no SSR markers have been developed yet, such as leek, shallot, chives, and elephant garlic. Diversity analysis of garlic cultivars with selected SSRs revealed 36 alleles, with 2–5 alleles/locus, and PIC = 0.38. Cluster analysis grouped the accessions according to their flowering behavior, botanical variety, and ecophysiological characteristics. Results from this study contribute to the characterization of *Allium* transcriptomes. The new SSR markers developed, along with the data from the polymorphism and transferability analyses, will aid in assisting genetic research and breeding in garlic and other *Allium*.

Keywords Garlic · *Allium* · Microsatellite markers · SSRs · Cross-amplification · Genetic diversity

Introduction

Allium is the largest genus of the Alliaceae family and includes several economically important vegetables species, including garlic (*A. sativum*) and onion (*A. cepa*). *Allium* vegetables are cultivated and consumed worldwide, mainly as food condiments, due to their characteristic flavor.

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00438-018-1442-5) contains supplementary material, which is available to authorized users.

✉ Pablo F. Cavagnaro
pablocavagnaro@hotmail.com

[1] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto Nacional de Tecnología Agropecuaria (INTA)-E.E.A. La Consulta, Mendoza, Argentina

[2] Facultad de Ciencias Agrarias (FCA), Instituto de Biología Agrícola de Mendoza (IBAM), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Cuyo, Mendoza, Argentina

[3] Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Instituto de Investigaciones en Ingeniería Genética y Biología Molecular "Dr. Héctor N. Torres" (INGEBI-CONICET), Universidad de Buenos Aires, Buenos Aires, Argentina

[4] Facultad de Ciencias Agrarias, Instituto de Horticultura, Universidad Nacional de Cuyo, Mendoza, Argentina

Published online: 28 April 2018

In addition, *Allium*s are also consumed for their numerous health-enhancing properties (reviewed by Block 2010). Despite their economic, culinary, and medicinal significance, *Allium* vegetables have understudied genomes and limited molecular resources, hindering advances in genetic research and breeding of these species. Thus, progress towards the characterization of non-genic or expressed regions (e.g., transcriptome) of *Allium* genomes, as well as the development of new molecular resources, such as robust and informative molecular markers (e.g., SSRs), are important for the genetic improvement of vegetables in this taxonomic group.

Most *Allium* species have complex and extremely large genomes, as compared to other eukaryotes (Arumuganathan and Earle 1991). For example, the nuclear genomes of onion (16.4 Gbp per 1C) and garlic (15.9 Gbp), both diploid ($2n = 2x = 16$), are comparable in size to hexaploid wheat, and ~ 34 and 6 times larger than rice and maize, respectively (Arumuganathan and Earle 1991). Although large-scale genomic sequence data are not available for *Allium* species, data from sequencing of BAC clones, linkage mapping, cytogenetic studies, and biochemical analyses suggest that the onion genome has low GC content (Kirk et al. 1970), and is characterized by an abundance of intrachromosomal tandem duplications (Jones and Rees 1968; Ranjekar et al. 1978; King et al. 1998) and long tracks of moderately repetitive sequences interspersed with single copy regions (Stack and Comings 1979; Pearce et al. 1996). Genome-scale data regarding the content and genome distribution of other types of repetitive DNA, such as simple sequence repeats (SSRs), have not been reported in *Allium*.

Expressed sequenced tag (EST) data are commonly used for developing sequence-based molecular markers, such as SNPs and SSRs. Among them, SSRs are frequent in occurrence, widely distributed in the genome, highly polymorphic, somatically stable, and often inherited in a codominant Mendelian manner (Kalia et al. 2011). The robustness and amount of information derived from microsatellite analysis, together with the ease by which they can be developed (if sequence information is available); make them ideal markers for linkage mapping, varietal fingerprinting, population structure, and genetic diversity studies (Varshney et al. 2005). In addition, SSRs have proven useful for marker-assisted selection in several crop species, especially when the markers reside close to or inside the genes responsible for a phenotypic trait (Varshney et al. 2005).

SSR markers—particularly those developed from EST sequence—are also valuable because of their high level of transferability to related species (Varshney et al. 2005; Cavagnaro et al. 2011; Huang et al. 2014), allowing their use as anchor markers for comparative genetic mapping (Bodénès et al. 2012; Cavagnaro et al. 2014) and evolutionary studies (Xu et al. 2004). Because EST-derived SSRs (ESSRs) originate from transcribed regions of the genome, they are likely to be more conserved and serve as potential markers across a broader taxonomic range than other non-genic markers. SSR markers developed in *Allium* vegetables are rather limited, as compared to other crop species, and largely insufficient for exploring the complex and large genomes of these genus members. To date, a total of 312 SSR markers have been reported in garlic (Lee et al. 2011; Cunha et al. 2012; Ma et al. 2009; Chand et al. 2015; Ipek et al. 2015; Liu et al. 2015), 203 SSRs in onion (Fischer and Bachman 2000; Kuhl et al. 2004; Martin et al. 2005; Baldwin et al. 2012), 682 in bunching onion (*A. fistulosum*) (Tsukazaki et al. 2008, 2015; Yang et al. 2015), and 94 in Chinese chives (*A. tuberosum*) (Tang et al. 2017), whereas no SSR markers have been developed to date for leek (*A. ampeloprasum* var. *porrum*), chives (*A. schoenoprasum*), shallot (*Allium cepa* var. *aggregatum*), and elephant garlic (*A. ampeloprasum* var. *ampeloprasum*).

Recent publications on the sequencing of garlic transcriptomes from various organs and tissues (Sun et al. 2012; Kamenetsky et al. 2015; Liu et al. 2015) represent potential sources of ample sequence data for developing SSR markers for garlic and—if transferable within the genus—other *Allium* species. In addition, they provide an opportunity for characterizing the microsatellite landscape in the garlic transcriptome and compared it with that of other *Allium*, as well as with other more distantly related monocots. It has been hypothesized that genomic resources developed for economically important or model species in the Poales order (e.g., rice, sugarcane, and wheat) may be transferable to the Asparagales, which include the Alliaceae and Asparagaceae families, since Asparagales is a monophyletic order sister to the lineage carrying the Poales (Kuhl et al. 2004). Although this hypothesis was indeed tested—by comparing homologous EST sequences of onion, asparagus, and rice—and refuted in the same study, reporting strong differences in ESTs of the Asparagales and Poales (Kuhl et al. 2004), detailed comparisons of the SSR composition in transcript sequences of these taxonomic groups have not been reported.

The objectives of this study were to: (1) characterize and compare the type, frequency, and distribution of simple sequence repeats in the transcriptomes of garlic and related *Allium* (onion, bunching onion, and Chinese chives) and non-*Allium* species (asparagus, rice, sugarcane, and ginger); (2) to develop SSR markers for garlic and evaluate their transferability across *Allium* vegetables species; (3) to evaluate markers polymorphism in garlic and onion accessions as an indicator of their potential usefulness for fingerprinting and diversity analysis; and (4) to use selected markers for assessing genetic diversity in a collection of garlic cultivars.

Materials and methods

Source and processing of sequence data

To explore the microsatellite landscape in *Allium* species, we characterized and compared the type, frequency, and distribution of these repeats in the transcriptomes of garlic, onion, bunching onion, and Chinese chives. Asparagus (*Asparagus officinalis*), belonging to the Asparagales order, was included as the closest taxa outside *Allium* with abundant sequence data available. Similarly, EST sequence from ginger (*Zingiber officinale*), rice (*Oryza sativa*), and sugarcane (*Saccharum officinarum*) was also analyzed for comparison purposes.

Transcript sequence data used were collected from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) and the PlantGDB-assembled unique transcripts (PUT, http://www.plantgdb.org/prj/ESTCluster/progress.php) databases on June 6, 2017. In the case of bunching onion, Chinese chives, and asparagus, next-generation sequencing (NGS) reads were available and, thus, sequences were downloaded and processed—to minimize redundancy—prior to the analysis. For this, the program CAP3 (Huang et al. 1999) was used, with default parameters, to assemble sequence reads into contigs and singletons, resulting in 32.1, 82.2, and 29 Mbp of non-redundant sequence data for bunching onion, Chinese chives, and asparagus, respectively. EST sequence of garlic and onion was processed in the manner, resulting in 29.4 and 43.1 Mbp of non-redundant sequence, respectively. Transcript sequences from ginger (12 Mbp), rice (26.1 Mbp), and sugarcane (93.1 Mbp) were downloaded and analyzed for SSRs directly, as these sequence sets, obtained from the PUT database, represent non-redundant sequences. Information on sequence source (databases), sequence ID, type of transcript sequence, and total sequence length analyzed for each plant species is presented in online resource Table S1.

SSR identification and marker development

SSR motifs were detected using the program MISA (MIcroSAtellite identification tool) (Thiel et al. 2003). Perfect microsatellites with a basic motif of 2–7 nucleotides (nt) and a minimum length of 12 (for di to tetranucleotides), 15 (for pentanucleotide), 18 (for hexanucleotide), and 21 (for heptanucleotide) were recorded. The positions of the SSRs in garlic ESTs were recorded and primers were designed flanking the SSRs using Primer3 (v. 4.0.0) software (Untergasser et al. 2012). Primers were designed to generate amplicons of 200–500 bp in length (optimum 300 bp) with the following minimum, optimum, and maximum values for Primer3 parameters: primer length (bp): 18–22–24; Tm (°C): 50–55–60; GC content (%): 40–50–60. Other parameters used the program default values.

DNA extraction and PCR conditions

Genomic DNA was isolated from young leaf tissues as previously described (Murray and Thompson 1980). PCR reactions were performed in 20 µl final volume containing 9.6 µl water, 2 µl 10 × DNA polymerase buffer, 3 µl $MgCl_2$ (10 mM), 1.6 µl dNTPs (2.5 mM each), 0.6 µl of each primer at 5 µM, 0.2 µl Taq polymerase at 3 units/µl (TaqUBA), and 2.4 µl of genomic DNA (40–50 ng). Thermocyclers were programmed as follows: initial denaturation at 94 °C for 30 s, followed by 40 cycles of 94 °C for 45 s, appropriate annealing temperature for 30 s, and 72 °C for 1 min, and a final step at 72 °C for 5 min.

Marker transferability across Alliaceae

Marker transferability across Alliaceae was investigated for 110 garlic SSRs developed in this study, plus 24 and 88 additional SSRs developed previously from garlic genomic sequence (Cunha et al. 2012; Ma et al. 2009), and onion EST sequence (Kuhl et al. 2004), respectively (see online resource Table S6 for more information on these additional 112 markers developed in other studies). In total, 222 SSR primer pairs were tested in a transferability panel of 16 accessions, composed of 9 *Allium* species [4 garlic and 4 onion accessions, and one accession each of elephant garlic, *A. vineale*, bunching onion, Chinese chives, leek, shallot, and chives], and one accession of the non-*Allium* Alliaceae *Tulbaghia violacea*. These materials were obtained from the *Allium* germplasm bank at INTA La Consulta (Argentina), and from commercial seed companies. Further information about the accessions used is presented in online resource Table S2.

The SSR markers were amplified by PCR as described above and the amplicons were resolved by 3% agarose gel electrophoresis, carried out for 2 h at 120 V. The gels were stained with ethidium bromide for visualization of the PCR products photographed under UV light. Accessions that produced amplicons of approximately expected length (i.e., the expected size in the donor species ± 50 nt) were considered as successful PCR amplifications, whereas amplicons outside the expected-size range were considered non-specific amplifications and regarded as negative results. This criterion was used regardless of the number of bands observed (i.e., if one or more bands of expected size were present for an SSR marker, the PCR reaction was considered successful and specific). All the PCR reactions were repeated at least

once, to confirm positive results and/or detect false negatives due to technical issues.

Analysis of marker polymorphism in garlic and onion cultivars

To evaluate the potential utilization of the SSR markers for fingerprinting and diversity analysis in garlic and onion, all the SSR markers with positive amplifications in these species, as observed in the transferability assay, were further examined for polymorphism in a sample set of four garlic and four onion commercial cultivars. The SSR amplification products were resolved by denaturing polyacrylamide gel (6%) electrophoresis, run at 150 V, 60 W, and 40 mA for 3 h. The gels were stained with silver nitrate and photographed for later analysis. SSR allele sizes were estimated with a size ladder.

Garlic genetic diversity

Fourteen polymorphic SSR markers were selected for assessing genetic diversity in a collection of 21 Argentine garlic cultivars (online resource Table S3) and two other *Allium* species: *A. fistulosum* and *A. schoenoprasum*, used as putative outgroups. Among the 14 selected markers, 12 correspond to SSRs developed in this study and 2 SSRs, ACM086 and ACM066, were developed previously from onion ESTs (Kuhl et al. 2004). Primers for these markers were fluorescently labeled and used in PCR reactions with the same conditions as described above. The resulting amplicons were resolved—and allele sizes were estimated—by capillary electrophoresis in a 3130xl Genetic Analyzer (Applied Biosystems, California, USA) at the Biotechnology Institute of INTA Castelar, Buenos Aires, Argentina. For genotyping of the garlic cultivars, a matrix was constructed with data regarding the presence or absence of the different alleles for each SSR locus.

General statistics, such as number of alleles (N_A), expected (*He*), observed heterozygosities (*Ho*), and polymorphic index content (PIC), were calculated with the software Cervus (v. 3.0) (Kalinowski et al. 2007). Pairwise genetic similarity values among the taxa were estimated according to Jaccard's coefficient, using the software XLSTAT (v. 2016.05.33324) (XLSTAT 2016). To use the Jaccard coefficient, which requires the presence/absence information, the SSR data were transformed and coded as 1 and 0, indicating the presence and absence of an allele, respectively. The resulting matrix was used to construct a dendrogram depicting phenetic relations among the taxa, using the Unweighted Pair Group Method using Arithmetic means (UPGMA) procedure, with the aid of the software XLSTAT. In addition, a Principal Component Analysis (PCoA) was performed with the same marker data and accessions using the software XLSTAT (Addinsoft, 2017).

Results

Distribution of microsatellites in EST sequence of garlic and other species

We analyzed the distribution of perfect and compound microsatellites with ≥ 3 repeat units and a minimum total length of 12 nt in 29.4 Mbp of non-redundant garlic transcript sequences. For comparison purposes, the same analyses were performed in EST sequence of other *Allium* and non-*Allium* species (online resource Table S1). For consistency in estimating repeat frequencies, the SSR motifs presented here represent all variants of both strands of the DNA sequence (e.g., AC also includes CA and the reverse complements GT and TG). Unless otherwise stated, microsatellite content in DNA sequence was expressed as number of SSRs per Mbp or as relative frequencies (%) within a particular dataset.

A total of 5,742 SSRs were identified in garlic ESTs, resulting in a density of 226.6 SSRs/Mbp. Garlic had the highest SSR density of all *Allium* species, followed by Chinese chives (178.1 SSRs/Mbp), bunching onion (166.9 SSR/Mbp), and onion (167.1 SSR/Mbp) (Table 1). Interestingly, EST sequence of non-*Allium* monocots had, in average, more than twice the SSR density observed in *Allium*, with rice and sugarcane presenting the highest SSR content.

Relative distributions of SSR types in the eight species analyzed were, in general, comparable (Fig. 1). Trinucleotides were the most frequent SSRs in ESTs of garlic and all other species, followed by tetra and dinucleotides (with the exception of onion and sugarcane, in which hexanucleotides outnumbered dinucleotides). Despite this general trend in relative abundance of SSR types, substantial variation was observed for some SSR types, between garlic and the other species. Garlic ESTs had the lowest proportion of trinucleotides (41.4%) as compared to all other species, showing a range of 51.9–60% (Table 1). In addition, garlic ESTs had the highest frequency of di-, tetra-, and pentanucleotides.

The number of repeat units in microsatellites is often correlated with their mutation rate and level of polymorphism (Weber 1990; Wierdl et al. 1997; Cavagnaro et al. 2011) and—therefore—can affect the usefulness of these markers for fingerprinting and discriminating among closely related taxa. In the present study, we observed that the number of repeat units varied among the different SSR types and among species (Fig. 2). In transcript sequence of garlic and other *Allium*, for all repeat types (i.e., di- to heptanucleotide SSRs), their frequency decreased as the number of repeat units per SSR increased. Such decrease was very rapid in

Table 1 Distribution of SSRs in EST sequence of garlic and related species

	Garlic	Onion	Bunching onion	Chinese chives	Asparagus	Ginger	Rice	Sugarcane
No. of sequences analyzed	70,191	33,161	122,751	353,984	120,937	16,495	44,644	131,383
Total seq. length (Mbp)	29.4	43.1	32.1	82.2	29.0	12.04	26.1	93.1
GC content (%)	40.3	39.5	43.2	39.8	42.7	47.2	48.8	51.8
No. of sequences containing SSRs (%)	5742 (8)	5879 (17.7)	3673 (3)	11,029 (3.1)	8647 (7.2)	3055 (18.5)	12,813 (28.7)	31,132 (23.7)
SSR density (SSR/Mbp)	226.6	167.1	166.9	178.1	343.1	323.3	490.9	465.5
Total perfect SSRs (%)	6412 (96)	7040 (97.7)	5226 (97.7)	14,069 (96.1)	9458 (94.9)	3751 (96.4)	12,202 (95.2)	41,515 (95.8)
Frequency (%) of di/tri/tetra/penta/hexa/hepta-nt	14.9/40.8/28.3/8.8/6.3/0.9	5.9/53/26.4/6.8/7/0.9	6.1/60/21.4/5.9/5.9/0.8	12.6/51.9/26.1/4.4/4.5/0.5	9.6/57.7/19/6.8/6/0.9	10.1/55.9/21.2/5.2/6.4/1.2	9.6/52.5/26/6.7/5/0.3	6.3/58.5/20.2/6/8.6/0.3
Total compound SSRs (%)	249 (4)	163 (2.3)	124 (2.3)	578 (3.9)	507 (5.1)	141 (3.6)	611 (4.8)	1817 (4.2)
Most frequent dinucleotide motifs (%)[a]	AC (43.2), AT (30.5), AG (25.8)	AG (38.3), AT (30.8), AC (29.7)	AC (45), AT (28.4), AG (25.7)	AC (58.9), AT (21.5), AG (17.6)	AG (76.9), AT (15.4), AC	AG (51.1), AT (40.6)	AG (49.5), AC (29.7), AT (16.3)	AG (51.4), AC (20.1), AT (20.1)
Most frequent trinucleotide motifs (%)[a]	AAG (35), AGG (13), AAT (10)	AAG (31.2), AGG (12.1), AAT	AGG (21.4), AAG (15.2), ACG (13.1), AAC (10.4)	AAG (22.5), AGG (15.2), AAC (11), ACT (10.4)	AAG (23.3), CCG (17.7), AGG (12.1)	CCG (25.6), AGG (18.7), AAG (16.1)	CCG (33.8), AGG (13.9), AGC (12.4), ACG (11)	CCG (46.6), AGC (12.2), AGG (11.6), ACG (10.2)
Most frequent tetranucleotide motifs (%)[a]	AAAT (23.8), AAAG (23.7), AAAC (10)	AAAT (21.2), AAAG (18.9), AAAC (12)	AAAG (26.7), AAAT (15.3), AAAC (10.1)	AAAT (23.2), AAAG (15.4), AAAC	AAAG (14.6), AAAT (13.9), AAAC (13.2)	AAAT (17), AAAG (11), AATT (10.8)	AGCT (12.1), AAAT, AAAC, AAAG	AGCT, CCCG, AGGG, AAAG
Most frequent pentanucleotide motifs (%)[a]	AAAAT (32), AAAAG (27), AAAAC	AAAAT (26.2), AAAAG (18.9), AAAAC	AAAAG (34.4), AAAAT (25.8), AGGGGG	AAAAT (26), AAAAG	AAAAT (18.7), AAAAG (12.8), AAAAC (11.8)	AAAAG (13.8), AAAAT (10.8), AAAAC	AGAGG, AAAAG	AGGGG, AGAGG, CCGCG
Most frequent hexanucleotide motifs (%)[a]	AAAAAG (18), AAA AAT (16)	AAAAAT (10.1), AAAAAG, AAA AAC	AACTCC (18.4), ACG GGT (10.1)	AAGCTG (16.2), AAC TCC	AAAAAG, AACCCT, AAAAAT	AAGAGG, AGGCGG	AAGCTC, AGGCGG, CCGCGG	AAGCCC, CCGCGG
Most frequent hepta-nucleotide motifs (%)[a]	AAAAAAG (39), AAA AAAT (23)	AAAAAAG, AAA AAAT	AAAAAAT (63.4), AAGGAGG (24.4)	AAAACGG (16.4), AAA AAAG	AAAAACG (40.9), AAAAAAT, AAA AAAC	AAACAGC (11.1), AAG TCAT, AATACTG	AAAAAAT, AGC GGCG	AAACCCT, AGGAGGG

A minimum of six repeat units (r.u.) for dinucleotides, four r.u. for trinucleotides, and three r.u. for tetra-, penta-, hexa-, and heptanucleotides were used as parameters for searching for microsatellites in EST sequences

[a]SSR motif considering complementary (e.g., "AAG" includes AAG + CTT motifs); SSR motifs within each SSR type class are listed in decreasing order of occurrence, and motifs occurring at a rate of at least 10% of the total within each SSR type class are indicated with their occurrence as percentage in parentheses

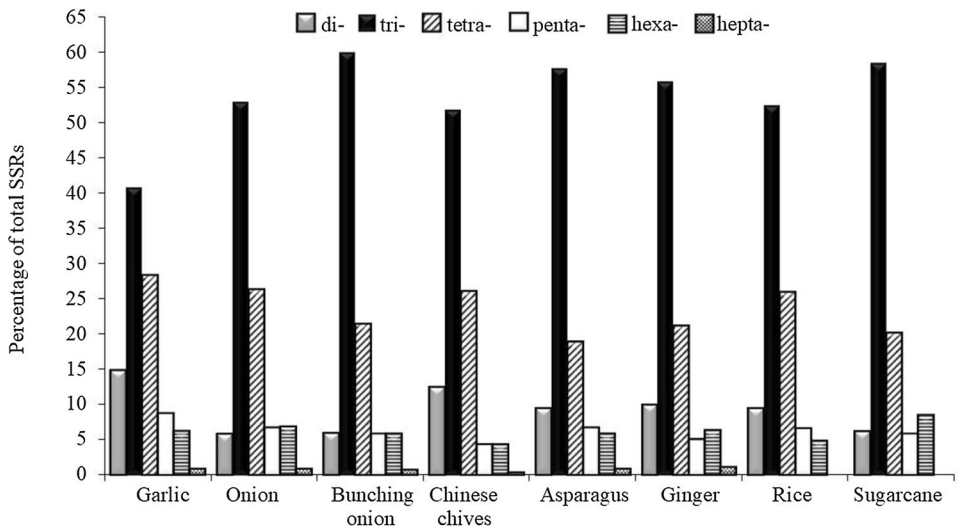

Fig. 1 Relative frequency (%) of SSR types in the transcriptomes of garlic and other monocot species

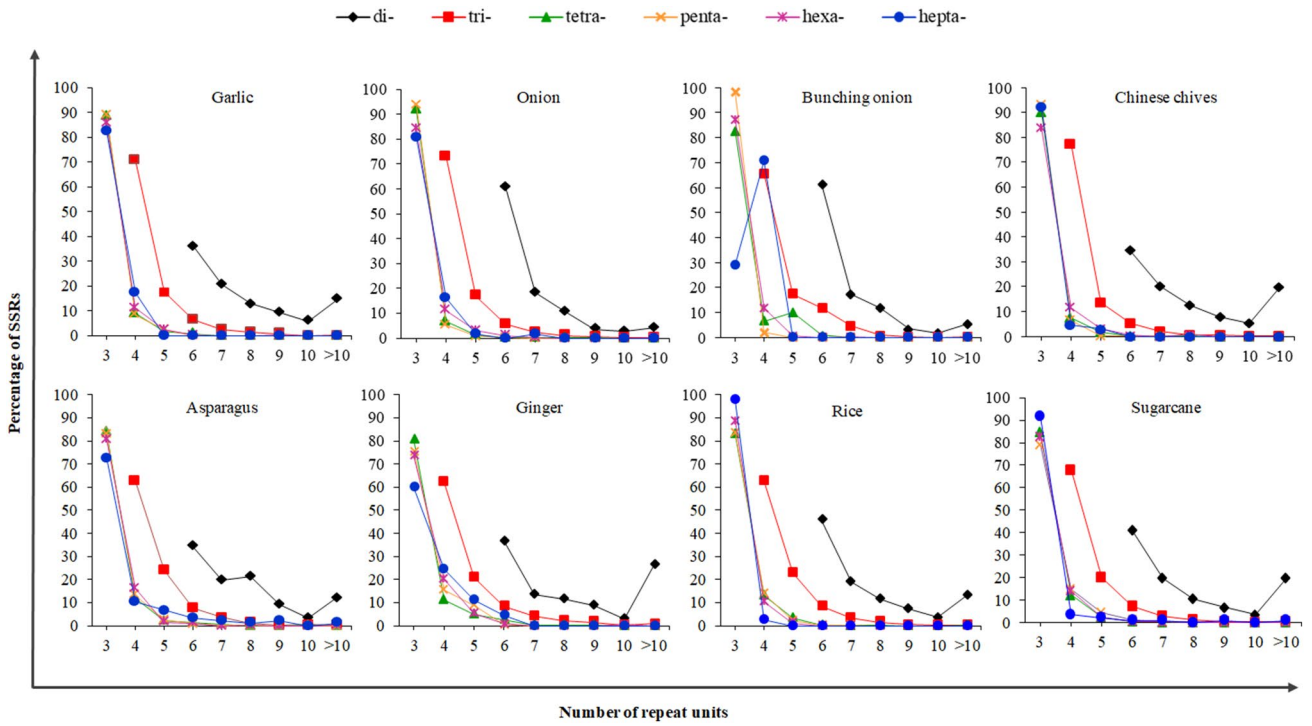

Fig. 2 Relative frequency (%) of SSR types by number of repeat units, in EST sequences of garlic and other monocot species. The search criterium for di-, tri-, and tetranucleotide SSRs was a minimum length of 12 nt, whereas, for pentanucleotides–heptanucleotides, a minimum of three repeat units were considered

tetra- to heptanucleotides, with SSRs having three repeats (the shortest for these SSR types) accounting for 80.6–98% of the total SSRs, whereas the shortest trinucleotides (with four repeats) accounted for 65.4–77.5% of the total trinucleotides. A more gradual decrease was observed in dinucleotides with the shortest class (six repeats) representing 34.4–61.2% of the total dinucleotides (Fig. 2). Overall, these results indicate that the SSR landscape of *Allium* EST sequence is mainly composed of short SSRs with a few repeats.

A comparative analysis of the main SSR motifs found in ESTs of garlic and other species is presented in Table 1 and Fig. 3 (a complete and detailed analysis of all SSR motifs found in each species can be found in Supplemental Table S4). Among dinucleotide SSRs, AC, AT, and AG were the most abundant motifs in all the species analyzed. With

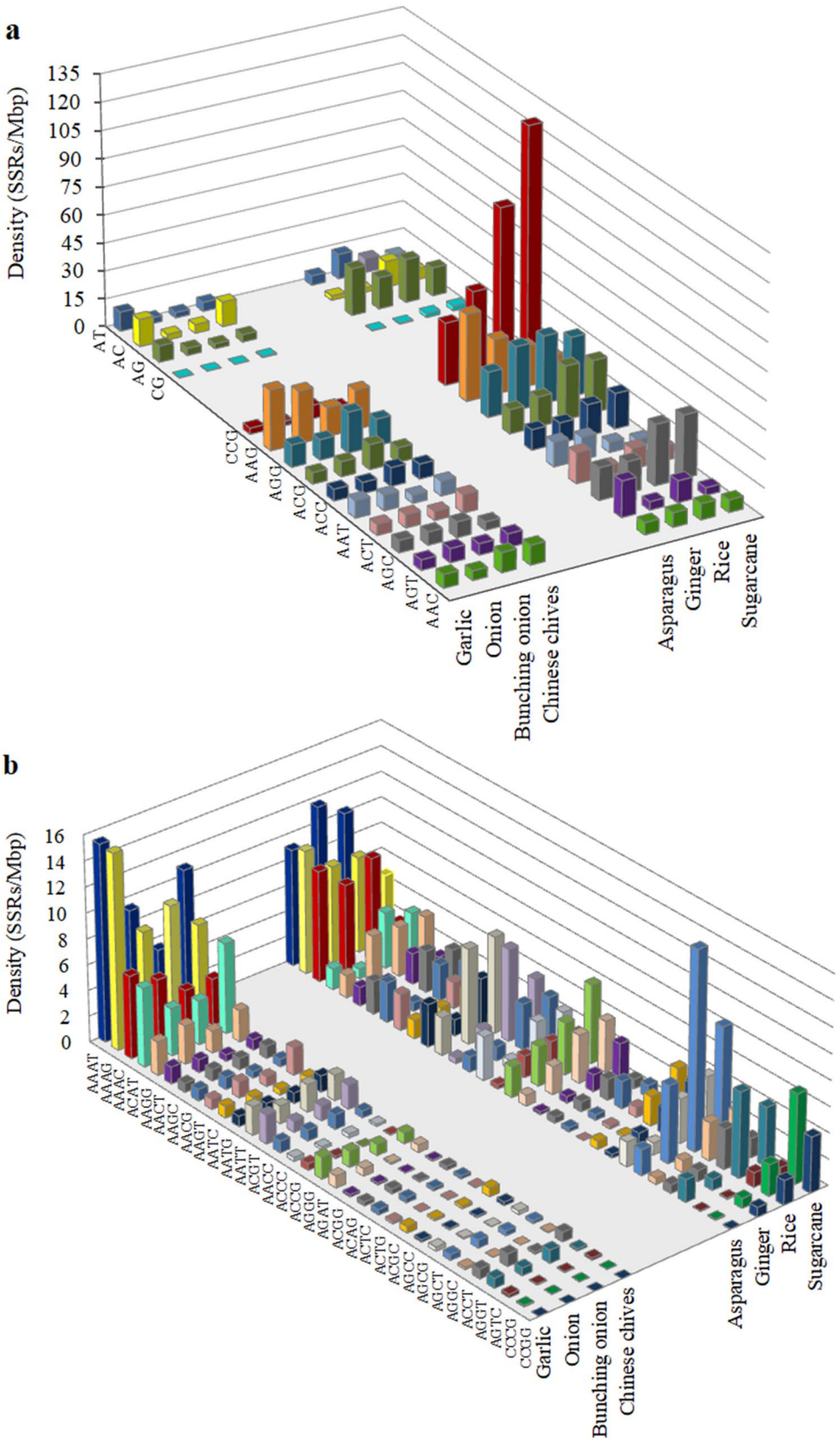

Fig. 3 Distribution of di- (**a**), tri- (**a**), and tetranucleotide motifs (**b**) in EST sequence of garlic and related species. Frequency values are expressed as number of repeats per million base pairs of sequence

the exception of onion, AC was the most frequent dinucleotide motif in *Allium*, whereas AG predominated in dinucleotides of onion and all non-*Allium* species. In garlic ESTs, these three motifs represented ~ 99.5% of the dinucleotide SSRs.

Garlic and onion had very similar relative frequencies (%) for the main trinucleotides, with AAG (31–35%), AGG (12–13%), and AAT (9–10%) being the most frequent motifs. Besides garlic and onion, AAG was also the most frequent individual motif in Chinese chives and asparagus, whereas AGG predominated in bunching onion. Motifs of CCG, AGG, and AGC were most frequent in ginger, rice, and sugarcane. These differences in the main trinucleotide motifs reflect the broad variation observed for GC-rich trinucleotides content among the taxa. Thus, in *Allium* species and asparagus, GC-rich motifs accounted for 36 to 59% of the total trinucleotides, with a mean of 45%, whereas GC-rich trinucleotides were present at much higher frequencies in the non-Asparagales species: ginger (68%), rice (78%), and sugarcane (87%).

GC-poor motifs of AAAT, AAAG, and AAAC predominated in tetranucleotide repeats of all *Allium* species and asparagus, representing these three motifs 41.7–57.5% of the total tetranucleotides found in these taxa. As observed for trinucleotides, garlic, and onion also had comparable relative frequencies for the main tetranucleotide motifs. In the non-Asparagales species rice and sugarcane, AGCT was the main motif, whereas other GC-rich motifs were particularly frequent in sugarcane.

Although broad variation for pentanucleotide motifs was found (e.g., 62 different pentanucleotides were identified in garlic ESTs), two GC-poor motifs (AAAAT and AAAAG), largely predominated in the transcripts of all *Allium* species, asparagus and ginger. Together, these two motifs represented 24.6% (in ginger) to nearly 60% (in garlic and bunching onion) of the total EST-pentanucleotides. AAAAC motifs were also relatively frequent in these species. As observed for tri- and tetranucleotides, sugarcane revealed a high frequency of GC-rich pentanucleotides, with predominance of AGGGG, AGAGG, and CCGCG motifs. The latter species did not share any of its major pentanucleotide motifs with those found in *Allium*, asparagus, and ginger (Table 1).

AT-rich hexanucleotide motifs of AAAAAG and AAA AAT predominated in transcripts of garlic, onion, and asparagus, whereas AAGCTG and AACTCC were the most frequent hexanucleotides in bunching onion and Chinese chives, respectively. In the other species, no clear predominance of a particular motif was observed.

Heptanucleotides were the most underrepresented repeat types in EST data from all the species (Fig. 1). GC-poor motifs of AAAAAAG and AAAAAAT largely predominated in garlic, together accounting for 62% of the total heptanucleotides, whereas AAAAAAT represented 63% of the total EST-heptanucleotides in bunching onion. These motifs were also frequent in other *Allium* and asparagus. As observed for other repeat types, sugarcane had the highest frequency of GC-rich heptanucleotide motifs (Table 1).

Development of SSR marker from garlic EST sequence

Primer pairs for 110 ESSR loci were designed. These SSR loci included 11 dinucleotides, 49 trinucleotides, 9 tetranucleotides, 10 pentanucleotides, and 21 hexanucleotides, whereas the remaining 10 markers were compound SSRs. Further information on these 110 SSR markers, including primer sequence, annealing temperature, repeat motif, and its position in template sequence, expected amplicon length, and the template DNA sequence carrying the SSR (for developing alternative primers if desired) is presented in online resource Table S5.

SSR marker transferability across Alliaceae

The 110 newly developed garlic ESSR markers were tested across 16 Alliaceae accessions for a total of 1760 primer/accession combinations. Of these, 857 (~ 49%) produced fragments within the expected range. Combinations that produced fragments outside the expected-size range (i.e., > 50 bp larger or smaller than the original garlic sequence) were considered non-specific amplifications and regarded as negative results. This range was arbitrarily selected to simplify the analysis, especially in the cases where more than two bands were amplified.

The number of garlic SSR markers that produced amplicons of expected size in garlic accessions (i.e., the success rate) ranged from 68 (62%) to 76 (69%), with a mean value of 73 (66%). The potential transferability of these garlic ESSRs across other Alliaceae varied widely (Fig. 4a). In accessions of onion and shallot, both belonging to *A. cepa*, the number of successful markers ranged from 48 to 62 (44–56%). A similar transfer rate was observed for leek, with 52 (47%) positive markers. The remaining *Allium* taxa and *Tulbaghia violacea* had the lowest transfer rates, presenting approximately half the number of positive markers (34–42 SSRs, 31–38%) of garlic, the SSR donor species. Despite the broad variation found in SSR transfer rates among the accessions, 18 markers produced expected-size amplicons in all *Allium* accessions.

To increase the potential number of useful SSRs in garlic and other *Allium* crops, 112 additional SSRs reported in the previous studies were also evaluated in our panel of Alliaceae accessions. Among these, 88 markers were developed from onion ESTs (ESSR) (Kuhl et al. 2004) and 24 were from garlic genomic sequence (GSSR) (Ma et al. 2009; Cunha et al. 2012) (see online resource Table S6 for

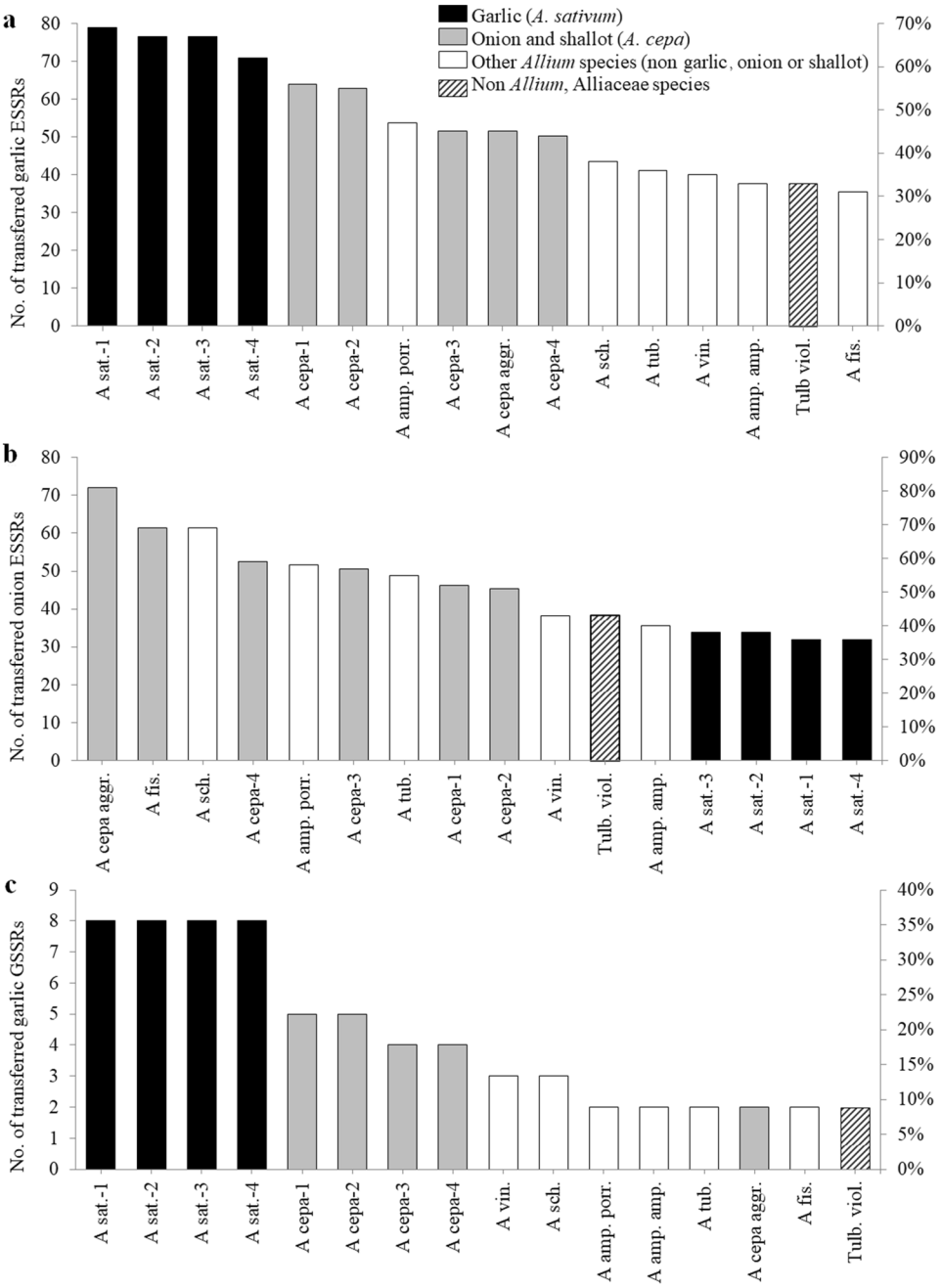

Fig. 4 Transferability of SSR markers developed from garlic (**a**) and onion (**b**) ESTs and garlic genomic sequences (**c**) across Alliaceae species. Number and percentage of SSR markers that successfully generated PCR products of nearly expected size in 16 Alliaceae accessions, including 4 accessions each of garlic and onion, 7 other *Allium* species, and 1 non-*Allium* Alliaceae species. *A sat.* = *A. sativum* (1. Rubí, 2. Nieve, 3. Morado, 4. Castaño); *A. cepa* (1. Refinta 20, 2. Morada, 3. Navideña, 4. Valcatorce); *A amp. porr.* = *A. ampeloprasum* var. *porrum; A cepa aggr.*= *A. cepa* var. *aggregatum; A sch.* = *A. schoenoprasum; A fis.* = *A. fistulosum; A tub.* = *A. tuberosum; A vin.* = *A. vineale; A amp. amp.* = *A. ampeloprasum* var. *ampeloprasum; Tulb. viol.* = *Tulbaghia violacea*

detailed information on these 112 SSRs). Success rate of these onion ESSRs in our onion accessions ranged from 46 to 52 (52–59%), whereas the garlic GSSRs obtained from the literature revealed low rates of success in our garlic materials (33.3%).

Marker transferability of both sets of markers to other Alliaceae taxa varied broadly. Transfer rates for onion ESSRs ranged from 36%, in garlic cultivar Castaño INTA, to 81% in shallot (Fig. 4b). In general, higher transfer rates were obtained for close relatives of the donor species, like shallot (81%), (belonging to the same *A. cepa* species), bunching onion (*A. fistulosum*) (69%), and chives (*A. schoenoprasum*) (69%), whereas garlic accessions—more distantly related to the SSR donor species—had the lowest marker transfer rates (36–38%). For the garlic GSSR markers developed by Ma et al. (2009) and Cunha et al. (2012), low cross-species transferability was observed, with a ranged of 8–21% (Fig. 4c).

Altogether, considering the three sets of markers evaluated, 108–116 (49–52%) and 104–113 (47–51%) SSRs amplified fragments of expected size in garlic and onion accessions, respectively. Among the other Alliaceae, 122 markers (55%) were successfully transferred to shallot, 106 (48%) to chives, 105 (47%) to leek, 97 (44%) to bunching

onion, 90 (41%) to Chinese chives, 76 (34%) to Tubalghia, and 73 (33%) to elephant garlic. A total of 33 SSR markers produced expected-size products across all the *Allium* accessions. The performance of each SSR marker across the 16 Alliaceae accessions is presented in online resource Table S7.

SSR polymorphism in garlic and onion accessions

Based on the results from the transferability analysis, 134 SSR markers (76 garlic ESSRs developed in this study, and 50 onion ESSRs and 8 garlic GSSRs from the literature) that showed positive amplifications in garlic and/or onion, were further evaluated for polymorphism in accessions of both species by denaturing polyacrylamide gel electrophoresis. Data on the performance and number of alleles observed for each of these markers in garlic and onion accessions are presented in online resource Table S8.

After excluding the markers did not yield amplicons in either garlic (20 SSRs) or onion (24 SSRs), 114 and 110 positive SSRs remained in each species, respectively. Of these, 42 markers (37%) were polymorphic in garlic accessions and revealed 2–5 alleles/locus, whereas 23 SSRs (21%) were polymorphic in onion accessions, exhibiting 2–5 alleles/locus.

Genetic diversity in garlic cultivars

Fourteen selected SSRs were used to assess marker polymorphism and genetic diversity in a collection of 21 garlic cultivars. A total of 36 alleles were revealed (2.6 alleles/SSR), with each SSR presenting 2–4 alleles/locus. Observed (*Ho*) and expected heterozygosities (*He*) ranged from 0 to 0.95 and 0.29 to 0.67, respectively. The markers discriminatory power, as estimated by the polymorphic index content (PIC), varied from 0.24 to 0.61 with a mean 0.38 (Table 2).

Pair-wise genetic similarities (GS) among the garlic accessions ranged from 0.38 to 1.00 (Jaccard coefficient), with an average GS among all the accessions of 0.70. The lowest GS values (0.38) were between 'Castaño', a bolting garlic cultivar of Russian origin belonging to the *ophioscorodon* botanical variety, and the cultivars 'INCO 283', 'Nieve' and 'Plata', all non-bolting garlics (var. *sativum*) from the Mediterranean region. Conversely, three groups of 3, 4, and 2 accessions, respectively, had maximum genetic similarities (GS = 1.00) among their group members.

A dendrogram was constructed based on these pair-wise GS values to depict relations among the garlic accessions (Fig. 5). Comparisons of the cophenetic values, calculated from the Jaccard GS matrix demonstrated a correlation of 0.992, indicating that the dendrogram was a very good representation of the data in the GS matrix (Sokal and Rohlf 1962). The outgroup species *A. fistulosum* and *A. schoenoprasum*, were clearly separated from all the garlic materials and formed the bottom branch of the tree, indicating a coherent genetic clustering of the taxa.

Table 2 General diversity statistics for 14 SSR loci evaluated in 21 garlic accessions

Locus	N_A	*Ho*	*He*	PIC
AsESSR-004	2	0.00	0.51	0.37
AsESSR-012	3	0.91	0.53	0.41
AsESSR-014	2	0.57	0.42	0.33
AsESSR-027	3	0.52	0.41	0.34
AsESSR-030	3	0.52	0.41	0.34
AsESSR-038	2	0.62	0.44	0.34
AsESSR-078	2	0.57	0.42	0.33
AsESSR-083	4	0.62	0.51	0.42
AsESSR-091	3	0.91	0.67	0.61
AsESSR-102	2	0.43	0.35	0.28
AsESSR-103	2	0.33	0.29	0.24
AsESSR-107	2	0.71	0.47	0.35
ACM066	3	0.95	0.65	0.60
ACM086	3	0.52	0.43	0.35
Total	36			
Mean	2.6	0.58	0.47	0.38

N_A number of alleles, *Ho* observed heterozygosity, *He* expected heterozygosity, *PIC* polymorphic index content

By clustering the accessions with more than 70% similarity, four groups were revealed (Fig. 5). The accessions were clustered according to their flowering behavior, botanical variety, and ecophysiological classification (Burba 1997). Group A included ten non-bolting "white" garlics (var. *sativum*) of ecophysiological group (EG) III. Group B was composed of 8 bolting "red" garlics (var. *sativum*) of EG IVa. Group C included two bolting garlics (var. *pekinense*) of EG II. Group D, composed of a single accession and the only *ophioscorodon* botanical variety in the analysis (a brown type garlic of EG IVb), was the outermost branch among the garlic materials.

In addition, a Principal Component Analysis (PCoA) was performed using the same marker data and accessions, and the results were presented in online resource Figure S1. Results from PCoA and UPGMA analyses were fully concordant, and separated the garlic accessions based on the same phenotypic and ecophysiological parameters.

Discussion

Frequency and distribution of SSRs in EST sequence of garlic and related species

In the present study, a detailed characterization of microsatellite repeats in EST sequence of four economically

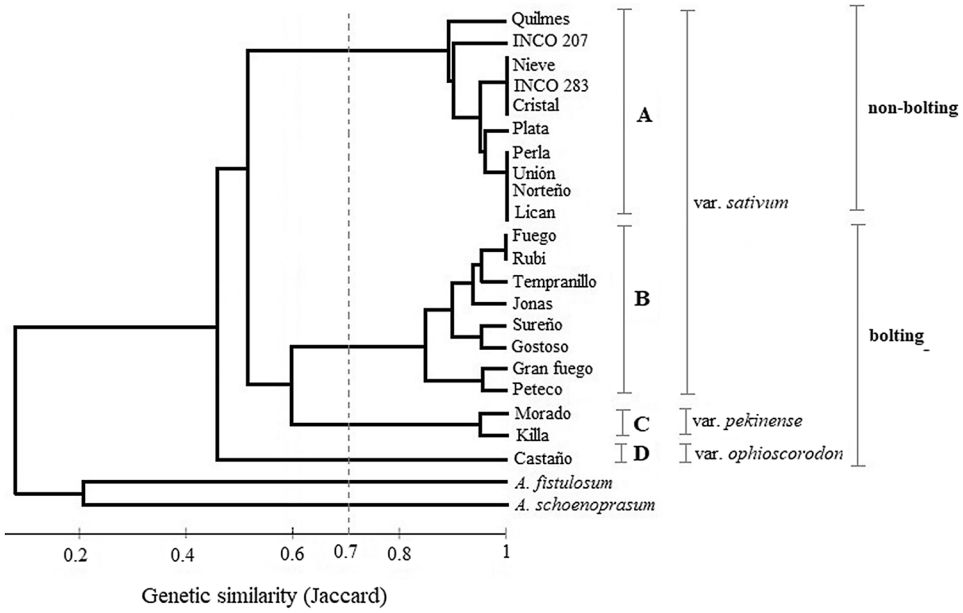

Fig. 5 UPGMA dendrogram demonstrating genetic relationships among 21 garlic cultivars and two *Allium* outgroup species. The dendrogram is based on the Jaccard coefficient of genetic similarity (GS). Letters A–D indicate the groups revealed by clustering the garlic accessions at GS > 0.7. Groups A, B, C, and D, belong to the ecophysiological group III, IVa, II, and IVb, respectively, according to the classification of Burba (1997)

important *Allium* species and four other reference monocot species was performed. To our knowledge, this is the first report to explore and compare the SSR landscape across *Allium* taxa, as well as between *Allium* members, asparagus (belonging to the Asparagaceae family which is phylogenetically closest to Alliaceae) and other more distantly related monocots. Our analysis revealed that SSR density in EST sequence of *Allium* was rather uniform, showing a range of 166.9–226.6 SSRs/Mbp, being garlic the most SSR dense *Allium* species. In addition, garlic ESTs had the lowest frequency of trinucleotides and the highest of di- and tetranucleotides among *Allium* and non-*Allium* species, whereas bunching onion was the highest in trinucleotide repeats (Fig. 1). Besides these differences in their relative abundance of SSR types (i.e., dinucleotides, trinucleotides, etc.), transcript sequences of the *Allium* species analyzed shared major commonalities with regard to SSRs, as evidenced by their similar frequency distributions for the main sequence motifs within each SSR type class (Table 1).

The mean SSR density in *Allium* ESTs (184.7 SSRs/Mbp) was much lower than in asparagus (343 SSRs/Mbp) and less than 40% of the SSR density in the Poales species rice (490.9 SSRs/Mbp) and sugarcane (465.5 SSRs/Mbp). In addition to the overall SSR density, strong differences were also found between *Allium*s and the non-Asparagales species regarding the major SSR motifs found in each SSR type class. As expected, the distribution and sequence motifs of microsatellite repeats in *Allium* differed less from their counterparts in asparagus (both genera belong to the Asparagales order) than compared to SSRs in ginger (Zingiberales), rice, and sugarcane (Poales) (Table 1). In the latter, a much higher frequency of GC-rich motifs was found. This likely reflects the differences in the overall GC content in EST sequence of *Allium* species (40.7% in average) and asparagus (42.7%), compared to the non-Asparagales species (mean = 49.3%, range = 47.2–51.8%).

Altogether, these data indicate that *Allium* members share common features in their EST-SSR composition which may be exploited—within *Allium*—for transferring SSR markers from economically important vegetables crops, like garlic and onion, to less resourceful taxa. Conversely, the important differences observed in EST sequence composition, SSR density, SSR motifs, and SSR frequency distributions, between *Allium* members (and asparagus) and rice, sugarcane and ginger, suggest that molecular and genomic resources (including molecular markers) from non-Asparagales model species will not be transferable to *Allium*. Coincidently with our results, Kuhl et al. (2004) compared EST sequences of onion, asparagus and rice, reporting major differences between the Asparagales and Poales for codon usage, GC content, and GC distribution. Furthermore, they concluded that Asparagales were more similar to eudicots than to the Poales for these characteristics.

Transfer success of garlic and onion SSRs across Alliaceae

The availability of SSR loci for economically important species has increased interest in primer transferability to related taxa, especially for species in which molecular resources are limited. In Alliaceae, SSRs have been developed for major crop species such as garlic, onion, and bunching onion. Results from this study indicate that a significant fraction of the garlic ESSRs developed by us, as well as onion ESSRs (Kuhl et al. 2004) and garlic GSSRs (Ma et al. 2009) developed in the previous studies, transfer successfully across

Alliaceae. Altogether, considering the three sets of markers evaluated, locus amplification success was detected in 73–122 markers across nine Alliaceae species, including economically important crops like garlic (110–116 SSRs), onion (104–113 SSRs), bunching onion (97 SSRs), leek (105 SSRs), and Chinese chives (90 SSRs). Successfully transferred markers may be particularly useful in orphan *Allium* crops for which no SSR markers have been developed yet, such as leek, shallot (122 SSRs), chives (106 SSRs), and elephant garlic (73 SSRs), as well as in the non-vegetable Alliaceae *A. vineale* (80 SSRs) and *Tulbaghia violacea* (76 SSRs). Prospects of a broader utilization of these markers beyond garlic and onion include their application in population and conservation studies as well as for linkage mapping and assisting breeding in *Allium* crop species.

It must be noted that when using SSR markers across distantly related species, the amplification of PCR product does not necessarily imply locus conservation, since homoplasy (i.e., convergence in size of non-homologous amplicons) may occur. Thus, verification of the PCR product identity by sequencing has been suggested, especially when working across genera or higher taxonomic ranks (Rossetto et al. 2000). However, according to Arnold et al. (2002), verification through sequencing may not be necessary if working within the same genus as the SSR donor species (i.e., the species from which de SSR markers were developed). This suggests that the SSR markers developed and evaluated in the present study may be safely used within *Allium*, the genus that includes the majority of Alliaceae vegetable crops species. Furthermore, the *Allium* vegetables used in this work represent a subset of closely related species within the genus *Allium*, as they belong to two closely related monophyletic sister clades within the genus *Allium*, namely the subgenera 'Cepa' and 'Allium' (Friesen et al. 2006). In addition, ad hoc analysis of amplicon size variation for 10 SSR loci in garlic and three other *Allium* species (bunching onion, chives, and leek) revealed that, for all the markers evaluated, amplicon size variation (in number of nucleotides) corresponded to a multiple number of the basic SSR motif length (e.g., for a dinucleotide SSR marker, amplicon size varied by multiples of two nucleotides; for a trinucleotide SSR, amplicons size varied by multiples of 3 nucleotides; etc.), suggesting that size variation was due to variation in the number of repeats of the basic SSR motif and, therefore, that specific SSR products were amplified in these *Allium* species. Altogether, these data, and the fact that the SSR markers developed in this work derive from more-conserved EST sequences (as compared to SSRs developed from genomic sequence), strongly suggest that they may be safely used for the *Allium* vegetables evaluated herein.

Success rate across Alliaceae was comparable for garlic ESSRs (31–69%) and onion ESSRs (36–81%). Conversely, success rate was much lower for garlic GSSRs (8–21%). These results are in agreement with the higher transferability of ESSRs compared to GSSRs reported previously for many plant taxa (reviewed by Varshney et al. 2005). In *Allium*, Tsukazaki et al. (2008) observed higher success rate in onion ESSRs (75%) compared to GSSRs from the same species (43%) when transferred and used for linkage mapping in bunching onion. The higher transferability observed, generally, for ESSRs is likely due to the higher degree of sequence conservation in ESTs across related taxa, as compared to non-transcribed sequences (Varshney et al. 2005).

Of the 222 SSR markers evaluated for cross-species transferability, 33 SSRs (15%) successfully amplified across all *Allium* taxa. These markers may be particularly useful for comparative mapping among the main *Allium* crop species. Although they are insufficient for the construction of full genetic maps, they may be used as reference markers for anchoring *Allium* maps. This approach was successfully applied—by the use of common SSRs—for anchoring linkage groups of a bunching onion (*A. fistulosum*) map to bulb onion (*A. cepa*) chromosomes (Tsukazaki et al. 2008). The need for more common markers for successful comparative mapping across *Allium* species has been repeatedly stated (Mallor et al. 2014; Tsukazaki et al. 2008). Our results indicating transferable SSRs among specific *Allium* species (see online resource Table S7) and—in particular—the 33 SSRs fully transferable across all *Allium*s, contribute to addressing this issue.

SSR polymorphism in garlic and onion accessions

The potential usefulness of SSR markers for diversity studies in Alliaceae will depend, to a great extent, on the possibility that markers successfully amplify across different species and on the ability of the marker to detect polymorphism among the taxa. Although the proportion of polymorphic markers found in garlic and onion may seem rather low, as compared to other studies using ESSRs from *Allium* [48 and 63% of polymorphic markers were reported for garlic (Ipek et al. 2015) and onion ESSRs (Khul et al. 2004), respectively, using both studies a screening panel of 8 accessions], it is, however, important to bear in mind that only four accessions of each species were used in our preliminary screening of polymorphic markers. Thus, if more accessions, or more diverse accessions, are used in a screening panel, more polymorphic markers may be revealed.

Genetic diversity and relatedness in garlic cultivars

Genetic diversity analysis of Argentine garlics using 14 SSR markers revealed a mean PIC value of 0.38, with a range of 0.24–0.61. This level of polymorphism is lower than the mean PIC of 0.51 (range 0.19–0.78) obtained by Cunha et al. (2012) using ten SSR markers in a collection of 75

garlic accessions from Brazil, and substantially lower than the mean PIC of 0.72 (range 0.65–0.80) reported by Chen et al. (2014) using eight SSRs in 39 accessions of different Asian origins, as well as the PIC of 0.62 (range 0.19–0.83) reported by Ma et al. (2009) with eight SSRs in 90 accessions from 11 countries.

Variation in mean PIC values across genetic diversity studies may be due to various factors, including the number and type of SSR markers used (i.e., ESSRs or GSRs), the number of accessions, and the actual genetic variation in the germplasm collection analyzed. The lower polymorphism found in the present study, as compared to the previous reports, may be partially due to the fact that we used ESSRs instead of GSSRs, as used by Ma et al. (2009), Cunha et al. (2012), and Chen et al. (2014). Because ESSRs reside in transcribed sequences, they are, in general, more conserved and—therefore—less polymorphic than GSSRs (Varshney et al. 2005). However, this factor does not explain our lower polymorphism as compared to the mean PIC value of 0.60 obtained by Ipek et al. (2015) using 26 ESSRs in 31 garlic accessions. In this case, the differences in the level of polymorphism may reside in the garlic accessions analyzed in both studies. The garlic materials analyzed by Ipek et al. (2015) were obtained from the diverse collection of the USDA (Pullman, Washington, USA) and they were intentionally selected—in order to maximize genetic variation in the sampled set—to represent ten different phylogenetic groups identified in a previous AFLP-based diversity analysis of garlic (Ipek et al. 2003); whereas, in the present study, we mainly used Argentine commercial cultivars, more than half of which were developed from local garlic populations originally introduced from Spain (see online resource Table S2). Thus, the genetic variation found in this study, as estimated by the mean PIC value, reflects the real genetic diversity in Argentine garlic cultivars. Other estimated parameters, such as observed (*Ho*) and expected heterozygosity (*He*), and number of alleles per SSR (N_A), reflected similar levels of genetic variation, as compared to other published studies (data not presented).

This is the first report on the analysis of Argentine garlics using microsatellite markers. Cluster analysis depicting genetic relationship among 21 garlic accessions—19 of which are commercial cultivars—revealed four clearly differentiated clusters of accessions, with members in each cluster sharing ≥ 70% genetic similarity (Fig. 5). No association between genetic clustering and geographical origin, pungency level, thiosultinates content, and health-enhancing properties (antiplatelet activity) was found (Cavagnaro et al. 2005; Gonzalez et al. 2009). Instead, the accessions were clustered according to their flowering behavior, botanical variety, and ecophysiological characteristics (Burba 1997). A first cluster (A) included only "white" non-bolting cultivars that have medium requirements of cold temperatures and photoperiod for bulbification, and medium length of postharvest conservation. Within this cluster, two subgroups of three and four accessions each could not be resolved with the SSR loci analyzed (i.e., the accessions had GS = 1.00). Despite the fact that these accessions are genetically closely related, this is likely due to an insufficient resolution of these SSR markers to discriminate among these materials, rather than to lack of genetic variation among the cultivars, since they present distinct morphological characteristics (Burba 1997) and differentiation among some of them was achieved previously using AFLP markers (Garcia-Lampasona et al. 2012). Cluster B was exclusively composed of "red" bolting garlics, characterized by their high bulbification requirements in low temperatures and photoperiod, and their long postharvest conservation. Cluster C was composed of two bolting "purple" type garlics, characterized by their very low temperature and photoperiod requirements for bulbification, presenting early harvests, and short dormancy and postharvest conservation. Presumably, cultivar 'Killa INTA' is a non-purple garlic mutation derived from a 'Morado INTA' population (Burba, personal communication). Cluster D included a single bolting "brown" cultivar ('Castaño INTA') and the only garlic material belonging to the *ophioscorodon* botanical variety, characterized by its very high bulbification requirements and longest dormancy and postharvest storage conservation. Morphologically, this cultivar is easily distinguished from the rest by its few, but very large cloves covered by brown outer dry scales.

Two previous studies have analyzed genetic diversity in Argentine garlics using AFLP markers (Garcia-Lampasona et al. 2003, 2012). Coincidently with our results using SSR markers, both studies of Garcia Lampasona et al. reported a similar association between genetic clustering and the accessions flowering behavior and ecophysiological characteristics and lack of association between genetic clustering and geographical origin of the accessions. Similar findings (i.e., genetic clustering according to flowering phenotype but not to geographical origin) were reported using different marker systems in garlic germplasm collections from other countries, including USA (Ipek et al. 2003, 2015; Volk et al. 2004), Czech Republic (Ovesná et al. 2015), Chile (Paredes et al. 2008), Brazil (Morales et al. 2013), Nepal (Panthee et al. 2006), and Australia (Bradley et al. 1996). It must be noted, however, that a few studies have found association between genetic clusters and geographic origin of the accessions (Ma et al. 2009; Zhao et al. 2011).

Analysis of 14 ESSR loci differentiated most of the garlic materials evaluated, and allowed to depict coherent genetic relationship among the Argentine cultivars, based on their morphological and ecophysiological characteristics. Despite of this, a few cultivars could not be resolved with these markers. Thus, to adequately discriminate among genotypes in large garlic collections, more ESSRs

should be evaluated, ideally in combination with GSSRs, which are generally more polymorphic than the formers (Varshney et al. 2005). However, the scarce availability of garlic genomic sequence has seriously limited the development of GSSRs in this species (to date, only 287 genomic sequences totaling 82 kbp are available at NCBI), with only 66 garlic GSSR markers reported so far, and all of them developed from SSR-enriched libraries constructed from genomic DNA (Ma et al. 2009; Lee et al. 2011; Cunha et al. 2012). The combined use of some of these GSSRs with ESSRs developed in the present study may provide sufficient resolution for fingerprinting and characterizing genetic diversity in large sample sets, such as the garlic germplasm bank at INTA La Consulta (Argentina). The latter represents a broad and diverse collection of ~ 200 accessions from multiple geographic origins, and serves as the main genetic resource for garlic breeding programs in Argentina.

In the present study, a comparative characterization of the SSR composition in EST sequence of the four economically most important *Allium* species and four non-*Allium* monocots was performed. These data revealed most striking differences between the Asparagales (*Allium* and *Asparagus*) and non-Asparagales taxa (rice, sugarcane and ginger), although distinct features were also noted across *Allium* species. These data contribute to the general knowledge on *Allium* transcriptomes and, particularly, with regards to their composition in these repeats. In addition, 110 new SSR markers were developed in this study from garlic ESTs, and they were characterized—along with 122 other garlic and onion SSRs developed previously—for their level of polymorphism and cross-species transferability in Alliaceae. Results from the polymorphism evaluation allowed the identification of polymorphic SSRs suitable for genetic analyses in garlic and onion germplasm. A significant fraction of the SSR markers evaluated were successfully transferred across *Allium* species, including crops for which no SSR markers have been developed yet, such as leek (105 SSRs), shallot (122 SSRs), chives (106 SSRs), and elephant garlic (73 SSRs). Finally, 14 selected SSR markers were used for estimating genetic diversity and relatedness among Argentine garlic cultivars, revealing a moderate level of SSR diversity in these materials, and genetic clustering according to the accessions flowering behavior (i.e., bolting or non-bolting), botanical variety, and ecophysiological characteristics. The new SSR markers developed, along with the data from the polymorphism and transferability analyses, will hopefully aid in assisting genetic research and breeding in garlic and other *Allium*.

Acknowledgements The authors acknowledge Jose L. Burba (INTA La Consulta, Argentina) for providing the garlic materials used in this work and for the fruitful discussion on the garlic diversity analysis.

Funding This work was supported by the Agencia Nacional de Promoción Científica y Técnica (ANPCyT) through the grant PICT-PRH-69. Karina Barboza, Vanesa Beretta, Perla C Kozub, and Cecilia Salinas have fellowships granted by the Argentine National Scientific and Technical Research Council (CONICET).

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Research involving human and animal participants This article does not contain any studies with human participants or animals performed by any of the authors.

References

Arnold C, Rossetto M, McNally J, Henry RJ (2002) The application of SSRs characterized for grape (*Vitis vinifera*) to conservation studies in Vitaceae. Am J Bot 89:22–28

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Report 9:208–218

Baldwin S, Pither-Joyce M, Wright K, Chen L, McCallum J (2012) Development of robust genomic simple sequence repeat markers for estimation of genetic diversity within and among bulb onion (*Allium cepa* L.) populations. Mol Breed 30:1401–1411

Block E (2010) Garlic and other *Alliums*: the lore and the science. Royal Society of Chemistry, Cambridge

Bodénès C, Chancerel E, Gailing O, Vendramin GG, Bagnoli F, Durand J, Goicoechea PG, Soliani C, Villani F, Mattioni C, Koelewijn HP, Murat F, Salse J, Roussel G, Boury C, Alberto F, Kremer A, Plomion C (2012) Comparative mapping in the Fagaceae and beyond using EST-SSRs. BMC Plant Biol 12:153

Bradley KF, Rieger MA, Collins GG (1996) Classification of Australian garlic cultivars by DNA fingerprinting. Aust J Exp Agric 36:613–618

Burba JL (1997) Panorama mundial y nacional de poblaciones y cultivares de ajo, posibilidades de adaptación. In: Burba JL (ed) 50 Temas sobre La producción de ajo, INTA EEA La Consulta, Mendoza-Argentina, vol 2, pp 11–31

Cavagnaro PF, Camargo A, Piccolo RJ, García-Lampasona S, Burba JL, Masuelli RW (2005) Resistance to *Penicillium hirsutum* Dierckx in garlic accessions. Eur J Plant Pathol 112:195–199

Cavagnaro PF, Chung SM, Manin S, Yildiz M, Ali A, Alessandro MS, Iorizzo M, Senalik DA, Simon PW (2011) Microsatellite isolation and marker development in carrot-genomic distribution, linkage mapping, genetic diversity analysis and marker transferability across Apiaceae. BMC Genomics 12:1

Cavagnaro PF, Iorizzo M, Yildiz M, Senalik D, Parsons J, Ellison S, Simon PW (2014) A gene-derived SNP-based high resolution linkage map of carrot including the location of QTL conditioning root and leaf anthocyanin pigmentation. BMC Genomics 15:1118

Chand SK, Nanda S, Rout E, Joshi RK (2015) Mining, characterization and validation of EST derived microsatellites from the transcriptome database of *Allium sativum* L. Bioinformation 11:145

Chen S, Chen W, Shen X, Yang Y, Qi F, Liu Y, Meng H (2014) Analysis of the genetic diversity of garlic (*Allium sativum* L.) by simple sequence repeat and inter simple sequence repeat analysis and agro-morphological traits. Biochem Syst Ecol 55:260–267

Cunha CP, Hoogerheide ES, Zucchi MI, Monteiro M, Pinheiro JB (2012) New microsatellite markers for garlic, *Allium sativum* (Alliaceae). Am J Bot 9:e17–e19

Fischer D, Bachmann K (2000) Onion microsatellites for germplasm analysis and their use in assessing intra-and interspecific relatedness within the subgenus Rhizirideum. Theor Appl Genet 101:153–164

Friesen N, Fritsch RM, Blattner FR (2006) Phylogeny and new intrageneric classification of *Allium* (Alliaceae) based on nuclear ribosomal DNA ITS sequences. Aliso 22:372–395

García-Lampasona S, Martínez L, Burba J (2003) Genetic analysis of a garlic (*Allium sativum* L.) germplasm collection from Argentina. Euphytica 132:115

García-Lampasona S, Asprelli P, Burba JL (2012) Genetic analysis of a garlic (*Allium sativum* L.) germplasm collection from Argentina. Sci Hortic 138:183–189

González RE, Soto VC, Sance MM, Camargo AB, Galmarini CR (2009) Variability of solids, organosulfur compounds, pungency and health-enhancing traits in garlic (*Allium sativum* L.) cultivars belonging to different ecophysiological groups. J Agric Food Chem 57:10282–10288

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Huang D, Zhang Y, Jin M, Li H, Song Z, Wang Y, Chen J (2014) Characterization and high cross-species transferability of microsatellite markers from the floral transcriptome of *Aspidistra saxicola* (Asparagaceae). Mol Ecol Resour 14:569–577

Ipek M, Ipek A, Simon PW (2003) Comparison of AFLPs, RAPD markers and isozymes for diversity assessment of garlic and detection of putative duplicates in germplasm collections. J Am Soc Hortic Sci 128:246–252

Ipek M, Sahin N, Ipek A, Cansev A, Simon PW (2015) Development and validation of new SSR markers from expressed regions in the garlic genome. Sci Agric 72:41–46

Jones R, Rees H (1968) Nuclear DNA variation in *Allium*. Heredity 23:591–605

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309–334

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kamenetsky R, Faigenboim A, Mayer E, Michael T, Gershberg Ch, Kimhi S, Esquira I, Shalom S, Eshe D, Rabinowitch HD, Sherman A (2015) Integrated transcriptome catalogue and organ-specific profiling of gene expression in fertile garlic (*Allium sativum* L.). BMC Genom 16:12

King JJ, Bradeen JM, Bark O, McCallum JA, Havey MJ (1998) A low-density genetic map of onion reveals a role for tandem duplication in the evolution of an extremely large diploid genome. Theor Appl Genet 96:52–62

Kirk JT, Rees H, Evans G (1970) Base composition of nuclear DNA within the genus *Allium*. Heredity 25:507–512

Kuhl JC, Cheung F, Yuan Q, Martin W, Zewdie Y, McCallum J, Catanach A, Rutherford P, Sink KC, Jenderek M, Prince JP, Town CD, Havey MJ (2004) 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 16:114–125

Lee GA, Kwon SJ, Park YJ, Lee MC, Kim HH, Lee JS, Lee SY, Gwag JG, Kima CK, Ma KH (2011) Cross-amplification of SSR markers developed from *Allium sativum* to other *Allium species*. Sci Hortic 128:401–407

Liu T, Zeng L, Zhu S, Chen X, Tang Q, Mei S, Tang S (2015) Large-scale development of expressed sequence tag-derived simple sequence repeat markers by deep transcriptome sequencing in garlic (*Allium sativum* L.). Mol Breed 35:204

Ma KH, Kwag JG, Zhao W, Dixit A, Lee GA, Kim HH, Chung IM, Kim NS, Lee JS, Ji JJ, Kim TS, Park YJ (2009) Isolation and characteristics of eight novel polymorphic microsatellite loci from the genome of garlic (*Allium sativum* L.). Sci Hortic 122:355–361

Mallor C, Arnedo-Andrés MS, Garcés-Claver A (2014) Assessing the genetic diversity of Spanish *Allium cepa* landraces for onion breeding using microsatellite markers. Sci Hortic 170:24–31

Martin WJ, McCallum J, Shigyo M, Jakse J, Kuhl JC, Yamane N, Pither-Joyce M, Gokce AF, Sink KC, Town CD, Havey MJ (2005) Genetic mapping of expressed sequences in onion and *in silico* comparisons with rice show scant colinearity. Mol Genet Genomics 274:197

Morales RG, Resende JT, Resende FV, Delatorre CA, Figueiredo AS, Da-Silva PR (2013) Genetic divergence among Brazilian garlic cultivars based on morphological characters and AFLP markers. Genet Mol Res 12:270–281

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326

Ovesná J, Mitrová K, Kučera L (2015) Garlic (*A. sativum* L.) alliinase gene family polymorphism reflects bolting types and cysteine sulphoxides content. BMC Genet 16–53

Panthee DR, Kc RB, Regmi HN, Subedi PP, Bhattarai S, Dhakal J (2006) Diversity analysis of garlic (*Allium sativum* L.) germplasms available in Nepal based on morphological characters. Genet Resour Crop Evol 53:205–212

Paredes M, Becerra V, Gonzalez MI (2008) Low genetic diversity among garlic (*Allium sativum* L.) accessions detected using random amplified polymorphic DNA (RAPD). Chilean J Agric Res 68:3–12

Pearce SR, Pich U, Harrison G, Flavell AJ, Heslop-Harrison JS, Schubert I, Kumar A (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

Ranjekar PK, Pallotta D, Lafontaine JG (1978) Analysis of plant genomes V Comparative study of molecular properties of DNAs of seven *Allium* species. Biochem Genet 16:957–970

Rossetto MF, Harriss CL, Mclauchlan A, Henry RJ, Baverstock PR, Lee LS (2000) Interspecific amplification of tea tree (*Melaleuca alternifolia*-Myrtaceae) microsatellite loci: potential implications for conservation studies. Aust J Bot 48:367–373

Sokal RR, Rohlf FJ (1962) The comparison of dendrograms by objective methods. Taxon 11:33–40

Stack SM, Comings DE (1979) The chromosomes and DNA of *Allium cepa*. Chromosoma 70:161

Sun X, Zhou S, Meng F, Liu S (2012) De novo assembly and characterization of the garlic (*Allium sativum*) bud transcriptome by Illumina sequencing. Plant Cell Rep 31:1823–1828

Tang Q, Yi L, Yuan X, Li F (2017) Large-scale development, characterization, and cross-amplification of EST-SSR markers in Chinese chive. Genet Mol Res 17:gmr16039861

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Tsukazaki H, Yamashita KI, Yaguchi S, Masuzaki S, Fukuoka H, Yonemaru J, Kanamori H, Kono I, Hang TTM, Shigyo M, Kojima A, Wako T (2008) Construction of SSR-based chromosome map in bunching onion (*Allium fistulosum*). Theor Appl Genet 117:1213–1223

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, Hamada M (2015) Development of transcriptome shotgun assembly-derived markers in bunching onion (*Allium fistulosum*). Mol Breed 35:55

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res 40:e115

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Volk GM, Henk AD, Richards CM (2004) Genetic diversity among U.S. garlic clones as detected using AFLP methods. J Amer Soc Hort Sci 129:559–569

Weber JL (1990) Informativeness of human $(dC–dA)_n \cdot (dG–dT)_n$ polymorphisms. Genomics 7:524–530

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769–779

XLSTAT (2016) Data analysis and Statistical Solution for Microsoft Excel. Addinsoft, Paris

Xu Y, Ma RC, Xie H, Liu JT, Cao MQ (2004) Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. Genome 47:1091–1104

Yang L, Wen C, Zhao H, Liu Q, Yang J, Liu L, Wang Y (2015) Development of polymorphic genic SSR markers by transcriptome sequencing in the welsh onion (*Allium fistulosum* L.). Appl Sci 5:1050–1063

Zhao WG, Chung JW, Lee GA, Ma KH, Kim HH, Kim KT, Chung IM, Lee JK, Kim NS, Kim SM, Park YJ (2011) Molecular genetic diversity and population structure of a selected core set in garlic and its relatives using novel SSR markers. Plant Breed 130:46–54