



Geliş(Received) :07/02/2018
Kabul(Accepted) :13/03/2018

Research Article
DOI:10.30708/mantar.391682

High-Throughput Genomic Simple Sequence Repeat (SSR) Marker Development and Construction of a High Resolution Physical Map in Button Mushroom (*Agaricus bisporus*) Genome

Ali Tefrik UNCU*¹, Ayşe Özgür UNCU²

*Corresponding author: atuncu@konya.edu.tr

¹Necmettin Erbakan University, Department of Molecular Biology & Genetics, Meram, Konya, 42090, TURKEY

²Necmettin Erbakan University, Department of Biotechnology, Meram, Konya, 42090, TURKEY

Abstract: Button mushroom [*Agaricus bisporus* (J.E. Lange) Imbach], is both a model organism and a worldwide cultivated edible mushroom species. Despite its scientific and economic relevance, efforts to develop molecular markers in *A. bisporus* genome are scarce and relatively recent. The present research reports the development of 3943 novel simple sequence repeat (SSR) markers specific to the *A. bisporus* genome. The marker set is well-distributed among the 13 *A. bisporus* chromosomes, saturating the entire genome with sequence-specific markers. The large number of novel, *A. bisporus* specific DNA markers introduced in the present work constitute a valuable resource for molecular genetic research in button mushroom, including strain identification and protection, and gene/QTL (Quantitative Trait Locus) mapping.

Key words: Molecular genetics, Bioinformatics, Sequence-specific markers, DNA markers

Kültür Mantarı (*Agaricus bisporus*) Genomuna Özel Yüksek Miktarda Basit Dizi Tekrarı (SSR) Markörünün Geliştirilmesi ve Yüksek Çözünürlükte Fiziksel Genom Haritasının Oluşturulması

Öz: Kültür mantarı [*Agaricus bisporus* (J.E. Lange) Imbach], bir model organizma olmasının yanı sıra dünya genelinde yetiştirilen bir yenilebilir mantar türüdür. Bilimsel ve ekonomik önemine rağmen *A. bisporus* genomuna özel moleküler markör geliştirme çalışmaları ancak yakın zamanda gerçekleştirilmeye başlanmış olup, oldukça kısıtlı sayıdadır. Bu çalışma kapsamında *A. bisporus* genomuna özel 3943 adet yeni basit dizi tekrarı (SSR) markörü geliştirilmiştir. Geliştirilen markörler *A. bisporus* kromozomlarına dengeli şekilde dağılmakta olup, genomun sekans-spesifik markörler ile doyurulmasını sağlamaktadır. Çalışma kapsamında tanıtılmakta olan yüksek sayıdaki *A. bisporus*'a özel yeni DNA markörü, bu türde gerçekleştirilecek moleküler genetik çalışmalar için kıymetli bir kaynak teşkil edecektir.

Anahtar kelimeler: Moleküler genetik, Biyoinformatik, Sekans-spesifik markörler, DNA markörleri

Introduction

Button mushroom [*Agaricus bisporus* (J.E. Lange) Imbach], also referred to as common mushroom, is the most commonly cultivated edible mushroom species (Barroso et al., 2000; Foulongne-Oriol et al., 2012). In addition, it is the model fungus species for fungal adaptation, persistence and growth in humic-rich

environments (Morin et al., 2012; Heneghan et al., 2016). Despite its role as a model organism and economic value due to its widespread cultivation as the predominant edible mushroom in European countries and USA, molecular genetic research in *A. bisporus* has been quite limited. While gene and QTL (Quantitative Trait Loci) mapping became the standard approach in modern plant and



animal breeding, mapping traits in fungal genomes is a relatively recent research objective (Foulongne-Oriol et al., 2012). Molecular markers are unquestionably the indispensable tools of modern breeding (Jones et al., 2009). Thus, an abundance of molecular markers specific for plant and animal genomes were developed and have been utilized for the improvement of cultivated crops and livestock. On the other hand, DNA markers specific to the *A. bisporus* genome are limited. Because linkage drag is a prominent problem that impairs breeding of improved button mushroom cultivars (Gao et al., 2016), saturating *A. bisporus* chromosomes with molecular markers would be highly useful to monitor the segregation of linked loci in high resolution and determine the desired recombinants. Yet, molecular breeding of white button mushroom toward improved yield, resistance to bacterial and fungal diseases, and technological properties (e.g. amenability to mechanical harvest) has not been properly established. For example, bacterial and fungal diseases are among major causes of yield loss, shortened shelf life and decreased quality in button mushroom cultivation (Samuels et al., 2002; Sokovic and Van Griensven, 2006; Berendsen et al., 2010). Major pathogens of button mushroom are listed as the bacterium *Pseudomonas tolaasii* Paine; the fungal species *Lecanicillium fungicola* (Preuss) Zare and W. Gams, and *Trichoderma harzianum* Rifai. On the other hand, there is lack of molecular genetic research to identify genes/QTLs that confer resistance to the above listed pathogens. Similarly, work that identified yield related molecular markers is extremely scarce. In addition to traits related to disease resistance and yield, it has to be noted that button mushroom is reported as a source of health beneficial dietary fibers and antioxidants (Jeong et al., 2010; Liu et al., 2013), therefore, traits regarding nutritional quality are also potential targets of breeding. Thus, molecular genetic research to dissect the genetic basis of nutritionally important traits in button mushroom is also required.

With the recent advances in high throughput DNA sequencing technologies, vast amount of genomic data is continuously being generated and deposited in publicly available databases (Uncu et al., 2015a), enabling *in silico* design of genome-specific molecular markers using bioinformatic approaches. Entire genomes can be surveyed for the high-throughput identification of molecular markers that represent the genomes of the relevant species in high-resolution. A draft genome

assembly of *A. bisporus* is available since 2012 (Morin et al., 2012), and the latest assembly that correctly represents the 13 chromosomes was recently published (Sonnenberg et al., 2016). Thus, it is now feasible to develop sequence-specific molecular markers and saturate the *A. bisporus* genome, which will greatly contribute to studies on chromosomal recombination and segregation patterns, as well as strain selection, conservation and molecular breeding efforts.

According to the common definition, SSRs (Simple Sequence Repeats) are iterations of short DNA motifs (Jones et al., 2009), where motif size ranges from one to eight bases. SSRs found wide applications in molecular genetics since they are highly polymorphic, reproducible and provide extensive genome coverage (Uncu et al., 2015b). In the present work, SSR (Simple Sequence Repeat) markers specific to the *A. bisporus* genome were developed via bioinformatics, by mining the recently released genome assembly for tandem repeat sequences. As a result, 4138 SSR loci were identified within the *A. bisporus* genome. In order to convert the SSR loci to markers, primers were designed that flank the repeat sequences. e-PCR was performed to confirm the amplifiability of the marker loci *in silico*, and it was feasible to further verify the e-PCR results with wet-lab experiments.

Material and Method

Simple sequence repeat identification

Chromosome assemblies of *A. bisporus* var. *bisporus* strain H39 (GenBank: CP015470.1 - CP015482.1) were retrieved from NCBI Genome Database (<https://www.ncbi.nlm.nih.gov/genome/>) as individual files in FASTA format. GMATA (Genome-wide Microsatellite Analyzing Tool Package) (Wang and Wang, 2016) software was utilized for mining the chromosome sequences for simple sequence repeats and, designing and mapping flanking amplification primers. Because GMATA requires Perl, R and Java running environments, the three programs were installed prior to running the software. For simple sequence repeat identification, FASTA files for individual chromosomes were imported to the 'SSR identification' module of the GMATA software. Parameters used for mining mononucleotide repeats were; Min-length (nt): 1, Max-length (nt): 1, Min. repeat-times: 8. Parameters for oligonucleotide repeat



identification were; Min-length (nt): 2, Max-length (nt): 8, Min. repeat-times: 5. SSR loci information files (.ssr) were generated individually for each chromosome that contained the SSR start & end positions, motifs and number of repeats.

Marker design and mapping

'Marker designing' module of GMATA was utilized in order to design PCR markers for the identified SSR loci. For each chromosome, assembly files in FASTA format and output files generated by the 'SSR identification' module (.ssr) were imported to the module as the 'Sequence file' and the 'SSR loci file', respectively. Marker design parameters were adjusted as: Min. amplicon size: 100 bp, Max amplicon size: 300 bp, Optimal annealing T_m: 60 °C, Flanking sequence length: 400 bp, Max template length: 2000 bp. Two output files with .mk and .sts extensions were generated for each chromosome, which contained left and right primer sequences, calculated melting temperatures, primer locations and anticipated product sizes. 'e-mapping' module of the GMATA software was utilized to run the e-PCR algorithm (Schuler, 1997). The algorithm conducted *in silico* PCR, generating the amplicons and mapping the markers. For each chromosome, the FASTA file that contains the entire chromosome sequence and the .sts file that contains primer sequences were imported as the 'Sequence file' and the 'Marker file', respectively. Default parameters of GMATA were used for e-mapping. Marker position information for each chromosome was obtained in .emap format. MapChart 2.3 computer program (Voorrips 2002) was used to visualize the physical map of the *A. bisporus* genome constructed based on the absolute positions of mapped simple sequence repeat loci.

DNA extraction, simple sequence repeat amplification and genotyping

Fresh fruiting bodies of *A. bisporus* were prepared for DNA extraction as described by Loftus et al. (1988), with minor modifications. DNA was extracted using a CTAB protocol. Briefly, fruiting bodies were cut into coarse pieces and ground in liquid nitrogen using a mortar and pestle. Ground tissue (200 mg) was taken to a 2 ml centrifuge tube, mixed with 800 µL of CTAB buffer [100 mM Tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), 1.4 M NaCl, 2% (w/v) CTAB, 1% PVP] and incubated at 65 °C for 1 h. Following incubation, 600 µL of chloroform:isoamyl alcohol

(24:1) was added and the samples were centrifuged at 14,000 rpm for 10 min at room temperature. DNA in the supernatant phase was precipitated by incubating at 4 °C for 60 min with 200 µL of isopropanol. DNA pellets were obtained by centrifugation at 4 °C for 10 min at 14,000 rpm. Pellets were washed with 100 µL of 70% ethanol, then dried, and suspended in 100 µL of Tris-EDTA buffer (pH 8.0).

SSR marker fragments were amplified in 20 µL reaction mixtures that contained 1× Colorless GoTaq Flexi PCR buffer (Promega Corp., Madison, WI, USA), 1.5 mM MgCl₂, 0.25 mM of each deoxyribonucleotide triphosphate (dNTP) (Promega Corp.), 1 U GoTaq G2 Flexi DNA Polymerase (Promega Corp.), 0.25 µM of each primer, and 10 ng of DNA as template. Thermal cycling conditions consisted of one cycle of initial denaturation for 30 s at 98 °C, followed by 40 cycles of 98 °C for 10 s, 58 °C for 20 s, 72 °C for 30 s, with a final extension step of 2 min at 72 °C.

Amplified marker fragments were subjected to capillary electrophoresis analysis. PCR amplicons were directly loaded to the Qiaxcel Advanced capillary electrophoresis system (Qiagen, Valencia, CA) and run using a Qiaxcel DNA High Resolution Kit (Qiagen). QX DNA Size Marker 25–500 bp, v2.0 (Qiagen), was used as the size standard, and QX Alignment Marker 15 bp/600 bp (Qiagen) was used for size standard alignment. The high-resolution run method OM800 was applied with a sample injection time of 10 s. Marker amplicons were visualized using the QIAxcel ScreenGel Software (Qiagen).

Results and Discussion

Simple sequence repeat identification

A. bisporus genome assembly is comprised of 30.8 megabases (Mb) of sequence distributed over the 13 chromosomes (Sonnenberg et al., 2016). In the present work, simple sequence repeat mining in the *A. bisporus* genome resulted in the identification of 4138 SSR loci, corresponding to an average marker density of 7.31 kilobases (kb) per marker interval (Table 1). The highest number of SSRs (439 SSRs) was identified in chromosome 1 whereas the lowest number of repeat loci (233 SSRs) was identified in chromosome 10. When individual chromosomes were evaluated for simple sequence repeat frequency, chromosome 13 displayed the highest SSR frequency of 5.16 kb/marker interval and chromosome 2 displayed the lowest SSR density of 9.13



kb between adjacent markers (Table 1). Physical map of the *A. bisporus* genome, constructed based on the absolute positions of the SSR loci along the 13 chromosomes, is displayed in Figure 1. Data regarding the physical map (simple sequence repeat motifs, amplification primers for each motif and absolute marker positions on each chromosome) are provided in Excel format as 'Electronic Supplementary Material'. Figure 1 displays the homogeneity of marker distribution along physical chromosomes.

There is a single report on genome wide simple sequence repeat identification in button mushroom, where the previous, unrepaired version of the *A. bisporus* genome assembly was surveyed for simple sequence repeat loci (Foulongne-Oriol et al., 2013). While some chromosomal segments were mislocated in the assembly and later corrected (Sonnenberg et al., 2016), the study provided valuable information regarding repeat abundance in the *A. bisporus* genome. On the other hand, SSR loci were not converted to PCR markers in the study. Foulongne-Oriol et al. (2013) reported an SSR density of 7.4 kb/marker interval with 4062 microsatellite loci, which indicate similar overall SSR abundance values with the present study. In addition, in concordance with the findings of the present research, Foulongne-Oriol et al. (2013) reported the highest and lowest SSR densities for chromosomes 13 and 2, respectively.

In the present research, among the 4138 SSR loci identified in the *A. bisporus* genome, mononucleotide repeats were predominant and represented 49.78% of the total number of SSRs (Table 2). Because mononucleotide repeats do not exert a functional constraint on most genomic sequences (introns and intergenic sequences), it is likely that SSR searches in genomes will yield mononucleotide SSRs as the most abundant motif type (Gu et al., 2010). In contrast, when SSR searches are restricted to coding sequences, it is reasonable to anticipate trinucleotide repeats as the predominant SSR type, since repeats that cause frameshift mutations are likely to get eliminated (Celik et al., 2014). According to the results of our SSR survey, dinucleotide motifs constituted the second most abundant simple sequence repeat type in the *A. bisporus* genome, corresponding to 30.52% of the identified SSRs (Table 2). The sum of mono-, di- and

trinucleotide SSRs was 4071, representing 98.38% of the total number of SSRs. Thus, simple sequence repeats in the *A. bisporus* genome were mainly iterations of one to three nucleotides and repeats of motifs longer than trinucleotide represented only 1.62% of the identified SSRs (Table 2). Our results were in agreement with Foulongne-Oriol et al. (2013), who reported a similar pattern of SSR motif abundance in the *A. bisporus* genome. In agreement with the relevant literature regarding genome wide SSR surveys (Uncu et al. 2015b), majority of the mononucleotide repeats were stretches of A/T nucleotides and constituted 46.45% of the identified SSRs. The second most abundant repeat motif was TC/GA with a relative abundance of 8.72% among *A. bisporus* genomic SSRs (Table 3).

Marker design and experimental validation

Marker design through GMATA software was performed on the 4138 recently identified SSR loci. Simple sequence repeats that were located on chromosomal ends were not suitable for primer design, therefore eliminated by the marker design algorithm. As a result, a total of 3943 primer pairs that amplify SSR fragments from the *A. bisporus* genome were generated. Marker distribution over the 13 *A. bisporus* chromosomes is displayed in Table 1. Amplifiability of the newly developed markers was verified *in silico* by e-mapping. Primer sequences, repeat motifs and genomic locations of the entire set of markers (3943 markers) are provided in Excel format as 'Electronic Supplementary Material'. Amplifiability of a total of 39 randomly selected marker loci (three SSR markers per chromosome) was experimentally evaluated in order to confirm the results obtained by e-mapping. SSR fragments were visualized in high resolution with a capillary electrophoresis system. Capillary electropherograms displaying a representative set of six SSR markers is provided in Figure 2. As a result, the entire set of tested SSR markers yielded amplification products, as expected. In addition, the experimentally obtained product sizes matched those calculated by the e-mapping algorithm. Thus, it was valuable to experimentally evaluate the newly designed markers and confirm the results obtained by primer design and e-mapping.

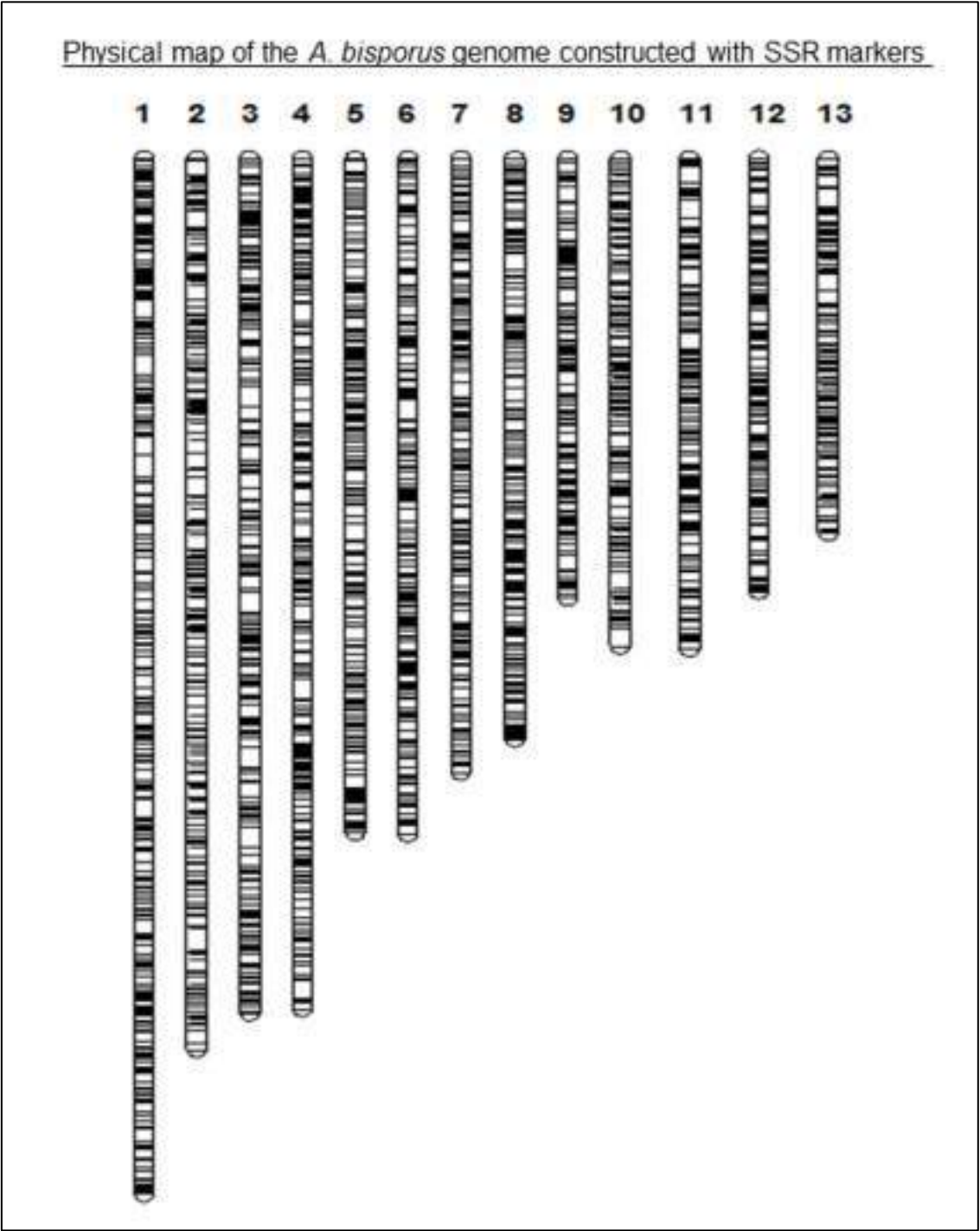


Figure 1. Physical map of the *A. bisporus* genome constructed with SSR markers. Numbers refer to *A. bisporus* chromosomes. Black bars represent simple sequence repeat loci. Primer and genomic location information for each SSR locus is provided in 'Electronic Supplementary Material'.

Table 1. Simple sequence repeat (SSR) identification statistics in *A. bisporus* genome.



Chromosome	Located SSRs	SSR density (kb)	SSR marker ^a
1	439	8.61047836	422
2	355	9.126760563	346
3	387	8.087855297	370
4	393	7.938931298	377
5	310	8	299
6	332	7.469879518	317
7	285	7.824561404	270
8	328	6.585365854	311
9	268	6.044776119	249
10	233	7.682403433	222
11	274	6.496350365	260
12	262	6.030534351	253
13	273	5.164835165	247

^aNumber of SSR markers designed per chromosome.

Table 2. Simple sequence repeat (SSR) types in *A. bisporus* genome.

Motif length	Number of occurrences	Frequency (%)
Mononucleotide	2060	49.78
Dinucleotide	1263	30.52
Trinucleotide	748	18.08
Tetranucleotide	21	0.51
Pentanucleotide	7	0.17
Hexanucleotide	25	0.6
Heptanucleotide	3	0.07
Octanucleotide	11	0.27
Total	4138	100

Table 3. Most abundant SSR motifs.

SSR motif	Number of SSRs	Motif frequency ^a (%)
A/T	1922	46.45
TC/GA	361	8.72
CT/AG	233	5.63
AT	209	5.05

^aOnly motifs with a frequency ≥ 0.05 are listed.



The first tandem repeat sequence was cloned from the *A. bisporus* genome in 2000 (Barroso et al., 2000), followed by the development of a set of 33 SSR markers (Foulongne-Oriol et al 2009). The high abundance of simple sequence repeats in the *A. bisporus* genome was displayed later by Foulongne-Oriol et al. (2013), yet, the work did not involve marker design but introduced repeat bearing loci as candidates for future marker development. Thus, a database of *A. bisporus* specific markers with high

genome coverage has not been reported before the present study. As a result of the present work, it was feasible to develop a total of 3943 novel, sequence-specific SSR markers that provide a high-resolution coverage of the *A. bisporus* genome. Absolute genomic positions and primer sequences of the 3943 novel SSR markers are available for researchers who conduct molecular genetic research in the important mushroom species.

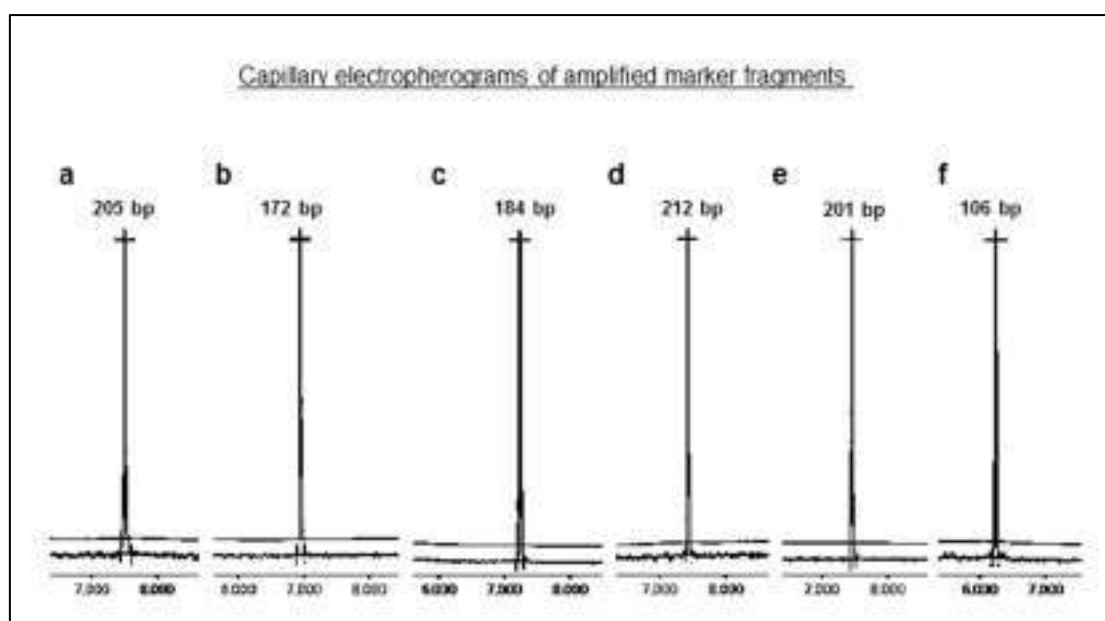


Figure 2. Capillary electropherograms displaying the amplification profiles of six *A. bisporus* SSR markers. Marker information is as follows: a, Chromosome 1, MK111, mononucleotide SSR (A/T); b, Chromosome 1, MK193, dinucleotide SSR (AT); c, Chromosome 5, MK139, mononucleotide SSR (A/T); d, Chromosome 5, MK72, mononucleotide SSR (A/T); e, Chromosome 5, MK50, trinucleotide SSR (ATG/CAT); f, Chromosome 12, MK14, dinucleotide SSR (TG/CA).

References

- Celik I, Gultekin V., Allmer J., Doganlar S., Frary A., Development of genomic simple sequence repeat markers in opium poppy by next-generation sequencing, *Molecular Breeding*, 34(2), 323-334 (2014).
- Barroso G., Sonnenberg A. S. M., Van Griensven L. J. L. D., Labarère J., *Molecular cloning of a widely distributed microsatellite core sequence from the cultivated mushroom Agaricus bisporus*, *Fungal Genetics and Biology*, 31, 115-123 (2000).
- Berendsen R. L., Baars J. J. P., Kalkhove S. I. C., Lugones L. G., Wösten H. A. B., Bakker P. A. H. M., *Pathogen profile Lecanicillium fungicola: causal agent of dry bubble disease in white-button mushroom*, *Molecular Plant Pathology*, 11(5), 585-595 (2010).
- Foulongne-Oriol M., Spataro C., Savoie J. M., *Novel microsatellite markers suitable for genetic studies in the white button mushroom Agaricus bisporus*, *Applied Microbiology and Biotechnology*, 84, 1125-1135 (2009).
- Foulongne-Oriol M., Rodier A., Rousseau T., Savoie J. M., *QTL mapping of yield-related components and oligogenic control of the cap colour in the button mushroom Agaricus bisporus*, *Applied and Environmental Microbiology*, 78(7), 2422-2434 (2012).
- Foulongne-Oriol M., Murat C., Castanera R., Ramírez L., Sonnenberg A. S. M., *Genome-wide survey of repetitive DNA elements in the button mushroom Agaricus bisporus*, *Fungal Genetics and Biology*, 55, 6-21 (2013).
- Gao W., Baars J. J. P., Maliepaard C., Visser R. G. F., Zhang J., Sonnenberg A. S. M., *Multi-trait QTL analysis for agronomic and quality characters of Agaricus bisporus (button mushrooms)*, *AMB Express*, 6(1), 67 (2016).



- Gu T., Tan S., Gou X., Araki H., Tian D., Avoidance of long mononucleotide repeats in codon pair usage, *Genetics*, 186, 1077-1084 (2010).
- Heneghan M. N., Burns C., Costa A. M. S. B., Burton K. S., Challen M. P., Bailey M., Foster G. D., *Functional analysis of Agaricus bisporus serine proteinase 1 reveals roles in utilization of humic rich substrates and adaptation to the leaf-litter ecological niche*, *Environmental Microbiology*, 18(12), 4687-4696 (2016).
- Jeong S. C., Jeong Y. T., Yang B. K., Islam R., Koyyalamudi S. R., Pang G., Cho K. Y., Song C. H., *White button mushroom (Agaricus bisporus) lowers blood glucose and cholesterol levels in diabetic and hypercholesterolemic rats*, *Nutrition Research*, 30, 49-56 (2010).
- Jones, N., Ougham H., Thomas H., Pasakinskiene I., *Markers and mapping revisited: Finding your gene*, *New Phytologist*, 183, 935-966 (2009).
- Liu J., Jia L., Kan J., Jin C. H., *In vitro and in vivo antioxidant activity of ethanolic extract of white button mushroom (Agaricus bisporus)*, *Food and Chemical Toxicology*, 51, 310-316 (2013).
- Loftus M.G., Moore D., Elliott T. J., *DNA polymorphisms in commercial and wild strains of the cultivated mushroom, Agaricus bisporus*, *Theoretical and Applied Genetics*, 76, 712-718 (1988).
- Morin E., Kohler A., Baker A. R., Foulongne-Oriol M., Lombard V., Nagy L. G., et al. *Genome sequence of the button mushroom Agaricus bisporus reveals mechanisms governing adaptation to a humic-rich ecological niche*, *Proceedings of the National Academy of Sciences*, 109(43), 17501-17506 (2012).
- Samuels G. J., Dodd S. L., Gams W., Castlebury L. A., Petrini O., *Trichoderma species associated with the green mold epidemic of commercially grown Agaricus bisporus*, *Mycologia*, 94(1), 146-170 (2002).
- Schuler G. D., *Sequence mapping by electronic PCR*, *Genome Research*, 7, 541-550 (1997).
- Sokovic M., Van Griensven L. J. L. D., *Antimicrobial activity of essential oils and their components against the three major pathogens of the cultivated button mushroom, Agaricus bisporus*, *European Journal of Plant Pathology*, 116, 211-224 (2006).
- Sonnenberg A. S. M., Gao W., Lavrijssen B., Hendrickx P., Sedaghat-Tellgerd N., Foulongne-Oriol M., et al. *A detailed analysis of the recombination landscape of the button mushroom Agaricus bisporus var. bisporus*, *Fungal Genetics and Biology*, 93, 35-45 (2016).
- Uncu A. O., Uncu A. T., Celik, İ., Doganlar S., Frary A., *A primer to molecular phylogenetic analysis in plants*, *Critical Reviews in Plant Sciences*, 34, 454-468 (2015a).
- Uncu A. O., Gultekin V., Allmer J., Frary A., Doganlar S., *Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in sesame*, *The Plant Genome*, 8, doi: 10.3835/plantgenome2014.11.0087 (2015b).
- Voorrips R. E., *MapChart: software for the graphical presentation of linkage maps and QTLs*, *Journal of Heredity*, 93, 77-78 (2002).
- Wang X., Wang L., *GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing*, *Frontiers in Plant Science*, 7, 1350 (2016).