Genome-wide identification and annotation of microsatellite markers in white truffle (*Tuber magnatum*)

Beyaz trüf mantarında (*Tuber magnatum*) mikrosatelit markörlerinin tüm genom düzeyinde tanımlanması ve anotasyonu

Ayse Ozgur UNCU[1] , Ali Tevfik UNCU[2]

[1]Necmettin Erbakan University, Faculty of Science, Department of Biotechnology, 42090, Konya, Turkey
[2]Necmettin Erbakan University, Faculty of Science, Department of Molecular Biology and Genetics, 42090, Konya, Turkey

Corresponding author (*Sorumlu yazar*): A. O. Uncu, e-mail (*e-posta*): auncu@konya.edu.tr
Author(s) e-mail (*Yazar(lar) e-posta*): atuncu@konya.edu.tr

ABSTRACT

The white truffle *Tuber magnatum* Pico is an ectomycorrhizal fungus, which produces the most economically valuable edible ascomata, referred to as truffles. Highly appreciated sensory properties and exceptional market prices make truffle fungi an important focus of research. Accordingly, there is increasing interest to characterize truffle aroma volatiles, fully understand their reproduction and, characterize their genomes, analyze their population structure and molecular genetic diversity. While production of truffle species was confined before to natural woodlands, development of host seedling inoculation protocols in 1970s established truffle production as a managed agricultural activity. On the other hand, *T. magnatum* inoculated seedling production was achieved only recently, since morphological criteria were insufficient for the precise identification of *T. magnatum* mycorrhizas without DNA-based identification protocols. The present research describes the development and functional annotation of novel microsatellite markers in *T. magnatum* genome. A bioinformatics workflow was applied in order to mine for microsatellite sequences in *T. magnatum* genome assembly, convert the loci to PCR markers, map marker amplification on genomic sequences and calculate allele sizes. As a result, 11 189 microsatellite markers specific to *T. magnatum* were developed. Annotation analysis was performed for marker sequences, resulting in 3377 marker loci matching with *T. magnatum* proteins. Thus, these markers have high potential to tag genes of interest in functional genomic analyses. As a result of the present work, a large database of *T. magnatum* specific DNA markers was introduced as a useful genomic resource for the valuable ectomycorrhizal fungus species.

ÖZ

Ektomikoriza mantarı *Tuber magnatum* Pico, en yüksek ekonomik değere sahip trüf mantarı türüdür. Trüf mantarlarının talep gören aroma özellikleri ve yüksek fiyatlar ile pazarlanmaları bu mantar türlerini önemli bir araştırma konusu haline getirmekte, trüf mantarı aroma bileşenleri, bu mantarların üreme mekanizmaları, trüf mantarı genomlarının karakterizasyonu, ve genetik çeşitlilik ve popülasyon yapılarının belirlenmesi konularında çalışmalar gerçekleştirilmektedir. Trüf mantarları önceden doğal ormanlık arazilerden toplanır iken, 1970'lerde fidan inokülasyon protokollerinin geliştirilmesi ile trüf mantarı yetiştiriciliği sistematik bir tarımsal aktivite haline gelmiştir. Ancak en değerli trüf mantarı olan *T. magnatum* ile fidan inokülasyonu, mikorizanın DNA-temelli protokoller olmaksızın yalnızca morfolojik yöntemler ile tanımlanmasındaki sorunlardan ötürü yakın zamanda uygulamaya geçirilebilmiştir. Bu çalışma kapsamında *T. magnatum* genomuna özel mikrosatelit markörleri geliştirilmiş ve fonksiyonel anotasyonları gerçekleştirilmiştir. Çalışma dahilinde biyoinformatik analiz yöntemleri kullanılarak *T. magnatum* genom sekansında mikrosatelit dizileri belirlenmiş, diziler PCR markörlerine çevrilmiş, markörler genom sekansında haritalanarak alel büyüklükleri hesaplanmıştır. Çalışma sonucunda *T.magnatum* türüne özel 11 189 yeni mikrosatelit markörü geliştirilmiştir. Anotasyon analizleri gerçekleştirilerek *T. magnatum* proteinleri ile eşleşen, bu sebeple fonksiyonel genomik çalışmaları için özellikle elverişli 3377 adet markör tanımlanmıştır. Çalışma sonucunda sunulan *T. magnatum*'a özel DNA markörü koleksiyonu bu değerli trüf mantarı türü için önemli bir genomik araç kaynağı oluşturmaktadır.

1. Introduction

Tuber species are ectomycorrhizal fungi which produce economically highly valuable, edible ascomata as a result of symbiotic interactions with plant roots (Rubini et al. 2004; 2005; Rizzello et al. 2012). The ascomata, referred to as truffles, produce a plethora of aroma volatiles. The volatile compounds are proposed to serve for diverse purposes such as facilitating communication of the fungi with host plant roots, inhibiting the growth of non-host plant species (Splivallo et al. 2007a) and aiding the dispersal of ascospores by attracting insects and mammals (Splivallo et al. 2012). The aroma volatiles are also highly appreciated by humans such that certain *Tuber* species, namely *Tuber magnatum* Pico and *Tuber melanosporum* Vittad., are traded at exceptionally high market prices. As a result, efforts were concentrated on moving the production of the highly demanded *Tuber* species from natural woodlands to managed truffle orchards. Accordingly, the geography of truffle cultivation is expanding and seedling inoculation protocols are commercially applied worldwide (Murat 2015). For example, the black truffle *T. melanosporum*, which is native to southern Europe (Martin et al. 2010), is now commercially cultivated in Australia in significant amounts (Murat 2015). In parallel with the efforts to establish truffle production as an agricultural activity, truffle fungi became an important focus of research in order to characterize and understand truffle aroma production (Splivallo et al. 2007a; 2007b; 2011; 2012), fully dissect reproduction attributes (Rubini et al. 2010a; 2010b; Zampieri et al. 2012), and explore *Tuber* genes, genomes and molecular genetic diversity (Mello et al. 2005; Rubini et al. 2005; Bertini et al. 2006; Martin et al. 2010; Bolchi et al. 2011; Murat et al. 2011; Linde and Selmes 2012; Payen et al. 2015). To date, majority of work that aimed at the molecular genetic characterization of ectomycorrhizal *Tuber* species involved the utilization of generic markers rather than genome-specific marker systems. Therefore, restriction-based typing or sequencing of the ITS (Internal transcribed spacer region of ribosomal DNA) locus is the most commonly employed genotyping approach to identify and characterize molecular genetic diversity in truffles (Amicucci et al. 1996; Paolocci et al. 1999; Murat et al. 2005; Bertini et al. 2006; Mello et al. 2015).

Among the highly demanded *Tuber* species, *T. magnatum*, also known as Piedmont white truffle, is of special interest as this species produces the most appreciated, hence, the most expensive fruit bodies which are also referred to as "vegetable gold" (Mello et al. 2005). In addition to the distinguished sensory properties of the fruit bodies, narrow geographical distribution of *T. magnatum* in Europe covering only Italy and Balkans definitely contributes to the reputation of this truffle species. The aim of the present research was to develop novel sequence-based markers specific to the *T. magnatum* genome. Toward this aim, a bioinformatic approach was employed in order to mine for microsatellites in the *T. magnatum* genome assembly and design PCR markers. In addition, marker loci that reside in protein coding genomic regions were annotated, producing useful information for the utilization of the newly designed markers in functional genomic analyses.

2. Materials and Methods

2.1. Identification of microsatellite loci in T. magnatum genome assembly

The genome assembly of *T. magnatum*, comprising 192.78 megabases (Mb) of sequence was accessed at NCBI (National Center for Biotechnology Information) database (NCBI 2018a) under the accession GCA_003182015.1. The genome assembly was analyzed for microsatellite loci using the GMATA (Genome-wide Microsatellite Analyzing Tool Package) software (Wang and Wang 2016). For microsatellite mining, 'SSR identification' module of the software was utilized with the assembly file in FASTA format as the input file. Microsatellite mining was carried out with the following parameters: Min-length (nt): 2, Max-length (nt): 6, Min. repeat-times: 5. The output file with .ssr extension contained position information, repeat motif and number of repetitions for each microsatellite and the .seq file contained the sequences that flanked the identified microsatellites.

2.2. Marker design and mapping

The output file (.ssr) generated by the 'SSR identification' module and the genome assembly file in FASTA format were used as input files in order to design PCR markers that flank microsatellite sequences. Processing parameters of GMATA for marker design were set as: Min. amplicon size: 100 bp, Max amplicon size: 300 bp, Optimal annealing Tm: 60 °C, Flanking sequence length: 400 bp, Max template length: 2000 bp. The output files generated by the marker design process were utilized as the input files together with the genome assembly sequence to run an e-PCR algorithm (Schuler 1997). The algorithm mapped the newly designed markers on *T. magnatum* genome assembly sequences. High-stringency parameters were applied for e-Mapping, with maximum number of mismatches and indels both set as 0. Amplification profiles, including the allele sizes and the number of loci targeted by the PCR primers, were obtained for each marker as a result of e-Mapping.

2.3. Functional annotation of microsatellite markers

Identified microsatellite loci were functionally annotated using the Blast2GO Functional Annotation and Genomics Tool. Identical Protein Groups Database of *T. magnatum* (NCBI 2018b), which consists of peptide sequences grouped based on common identity, was used to establish a local reference protein database within the Blast2GO platform. The *T. magnatum* genomic sequences that harbor the identified microsatellite loci were merged as a single FASTA file in order to perform the annotations. Blastx algorithm within Blast2GO was operated in order to translate *T. magnatum* DNA sequences, align sequences with the local reference protein database, determine statistically significant sequence identities and create protein descriptions for *T. magnatum* genomic sequences.

3. Results and Discussion

3.1. Microsatellite identification, marker design and mapping

Continuous development of novel computational genomic tools customized for different genome analyses purposes enables to perform detailed genomic analyses in any species with sufficient quantity of accessible genomic sequences. Despite the fact that *T. magnatum* is the most appreciated and valuable truffle fungus species, a genome assembly was made available only recently in 2018. Yet, the assembly constitutes the most useful resource in order to establish a collection of sequence-specific markers. In the course of the present work, *T. magnatum* genome assembly comprising 192.78 Mb of sequence was mined for microsatellites, resulting in the identification of 12 223 candidate loci for the design of PCR

markers. The overall frequency of microsatellite abundance in the *T. magnatum* genome assembly corresponded to 15.7 kilobases (kb) marker interval^{-1}. Iterations of dinucleotide repeats were the most abundant repeat loci (7048 dinucleotide repeats), representing 57.6% of all identified microsatellites (Table 1). Trinucleotide microsatellites were the second most abundant repeat type (3833 trinucleotide repeats) and represented 31.4% of the overall pool of microsatellites (Table 1). Genome-wide microsatellite markers were developed before in *T. melanosporum*, the ectomycorrhizal fungus that produces the most appreciated black Perigord truffles (Murat et al. 2011). According to the results of their study, the authors reported 5325 microsatellites (iterations of two to six nucleotide motifs) in an assembled genome sequence of 124.95 Mb, corresponding to an overall frequency of 23.5 kb marker interval^{-1}. Thus, the microsatellite density detected in the *T. magnatum* genome as a result of the present study (15.7 kb marker interval^{-1}) was higher than the density reported for *T. melanosporum* genome (Murat et al. 2011). On the other hand, because the number and relative abundance of detected loci depend on the applied repeat search parameters, a more reliable comparison requires that the abundance of microsatellites identified using identical parameters are compared. Accordingly, dinucleotide repeats were excluded from the comparison as they were identified with different search parameters in the two studies and the overall frequencies of the sum of remaining microsatellite types were compared. As a result, a marker density value of 37.3 kb marker interval^{-1} was obtained for microsatellites of tri-to hexanucleotides in the *T. magnatum* genome, which was a similar value calculated for the *T. melanosporum* genome (38.4 kb marker interval^{-1}). Thus, the results imply that microsatellite abundance in the two *Tuber* genomes is comparable.

Table 1. Statistics of microsatellite identification in *T. magnatum* genome assembly.

Motif type	Number of occurences	Relative abundance (%)
Dinucleotide	7048	57.6
Trinucleotide	3833	31.4
Tetranucleotide	961	7.9
Pentanucleotide	273	2.2
Hexanucleotide	108	0.9
TOTAL	12 223	100

T. magnatum genomic sequences harboring microsatellite loci were used for designing PCR markers. The high-throughput marker design process yielded a total of 11 189 PCR primers that amplify microsatellite sequences. Primer sequences, amplified microsatellite motifs and number of motif iterations are provided for each marker in Supplementary Material. Following the marker design process, an e-Mapping algorithm was operated in order to determine the amplification profiles of the markers. Allele sizes determined for each marker and amplification patterns (single/multilocus) are provided in Supplementary Material.

Marker sequences were subjected to annotation analysis in order to identify markers that reside in protein coding sequences. An e-value threshold of 1.00E-05 was applied to determine statistically significant matches between *T. magnatum* proteins and genomic sequences that harbor microsatellite markers. As a result, 3377 sequence descriptions were obtained for the 11 189 microsatellite markers based on matching protein sequences. Thus, 30.2% of the microsatellite markers introduced in the present work represent protein coding genomic sequences. Therefore, these markers would be

especially useful genomic tools for functional genomic analyses in *T. magnatum*. Protein annotations of the marker loci are available in Supplementary Material. e-value and similarity distributions of statistically significant matches are displayed in Figure 1.
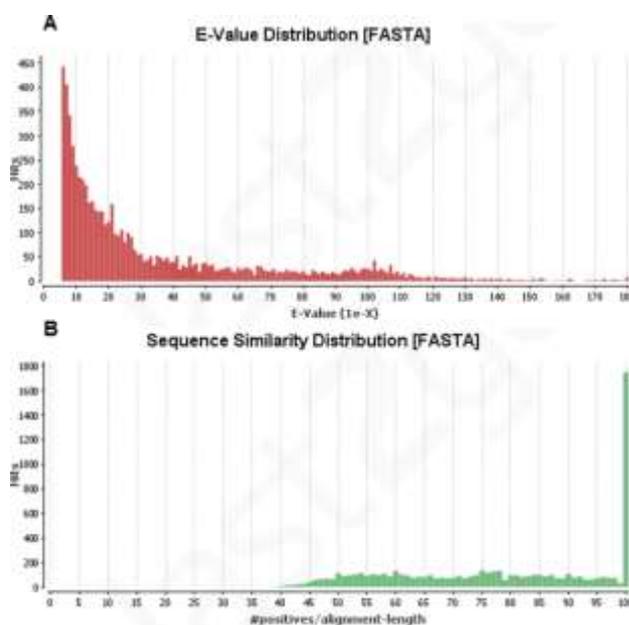

Figure 1. e-value and similarity distribution of marker sequence-peptide matches obtained by annotation analysis. A; e-value distribution of matches between microsatellite loci and *T. magnatum* proteins. e-value for each annotated marker locus is available in Supplementary Material. B; Similarity rate distribution of matches between microsatellite loci and *T. magnatum* proteins.

As a result of annotation analyses, almost one third of the newly designed markers (30.2%) corresponded to coding sequences. Regarding microsatellite abundance in fungal genomes, Murat et al. (2011) indicates that microsatellite density does not correlate with genome size. This is indeed related to the fact that on the contrary to a common assumption; microsatellites do not cluster at repetitive, noncoding DNA but tend to occur at single copy, coding regions of the genome (Morgante et al. 2002). Therefore, it is reasonable to expect that genome-wide microsatellite development studies yield high numbers of markers that represent protein coding sequences.

Microsatellite markers are widely utilized in forensic genetics, genome mapping, gene identification and population genetics studies. While microsatellite markers are abundant for a wide range of organisms, genome-specific microsatellite development studies are rare for fungal species (Dutech et al. 2007). As a result, molecular genetic analyses in truffle fungi generally involve the utilization of random marker systems (Dutech et al. 2007) or typing/sequencing the universal ITS locus (Amicucci et al. 1996; Paolocci et al. 1999; Murat et al. 2005; Bertini et al. 2006; Mello et al. 2015). However, it should be noted that conserved, universal sequences, such as the ribosomal ITS, serve well to discriminate among species, but are not ideal targets to explore the genetic diversity at the intraspecific level or characterize the genomic constitution. Reliable analysis of molecular genetic diversity and genome structure requires that sufficient number of data points is available to cover the genome in high resolution. Development

of genome-specific microsatellite markers using microsatellite-enriched genomic libraries is an inefficient process, yielding limited number of markers representing only the motifs defined by the capture probes (Uncu et al. 2015). On the other hand, advances in DNA sequencing technologies and development of computational genomics tools revolutionized the process, enabling high-throughput, genome-wide identification and design of novel species-specific markers. In the course of the present study, mining for microsatellites in *T. magnatum* genome assembly with bioinformatics resulted in the development of more than 10 000 novel genomic markers. Moreover, it was feasible to functionally annotate the marker loci and assign peptide definitions for markers that correspond to protein coding sequences. The novel microsatellite markers introduced in the present work would be beneficial tools for the precise identification of *T. magnatum* ectomycorrhiza and, genetic diversity assessment, population genetics and gene identification studies.

References

Amicucci A, Rossi I, Potenza L, Zambonelli A, Agostini D, Palma F, Stocchi V (1996) Identification of ectomycorrhizae from *Tuber* species by RFLP analysis of the ITS region. Biotechnology Letters 18: 821-826.

Bertini L, Rossi I, Zambonelli A, Amicucci A, Sacchi A, Cecchini M, Gregori G, Stocchi V (2006) Molecular identification of *Tuber magnatum* ectomycorrhizae in the field. Microbiology Research 161: 59-64.

Bolchi A, Ruotolo R, Marchini G, Vurro E, Sanita di Toppi L, Kohler A, Tisserant E, Martin F, Ottonello S (2011) Genome-wide inventory of metal homeostasis-related gene products including a functional phytochelatin synthase in the hypogeous mycorrhizal fungus *Tuber melanosporum*. Fungal Genetics and Biology 48: 573-584.

Dutech C, Enjalbert J, Fournier E, Delmotte F, Barres B, Carlier J, Tharreau D, Giraud T (2007) Challenges of microsatellite isolation in fungi. Fungal Genetics and Biology 44: 933-949.

Linde CC, Selmes H (2012) Genetic diversity and mating type distribution of *Tuber melanosporum* and their significance to truffle cultivation in artificially planted truffieres in Australia. Applied and Environmental Microbiology 78: 6534-6539.

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O et al. (2010) Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature Letters 464: 1033-1038.

Mello A, Murat C, Vizzini A, Gavazza V, Bonfante P (2005) *Tuber magnatum* Pico, a species of limited geographical distribution: its genetic diversity inside and outside a truffle ground. Environmental Microbiology 7: 55-65.

Mello A, Lumini E, Napoli C, Bianciotto V, Bonfante P (2015) Arbuscular mycorrhizal fungal diversity in the *Tuber melanosporum* brule. Fungal Biology 119: 518-527.

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genetics 30: 194-200.

Murat C, Vizzini A, Bonfante P, Mello A (2005) Morphological and molecular typing of the below-ground fungal community in a natural *Tuber magnatum* truffle-ground. FEMS Microbiology Letters 245: 307-313.

Murat C, Riccioni C, Belfiori B, Cichocki N, Labbe J, Morin E, Tisserant E, Paolocci F, Rubini A, Martin F (2011) Distribution and localization of microsatellites in Perigord black truffle genome and identification of new molecular markers. Fungal Genetics and Biology 48: 592-601.

Murat C (2015) Forty years of inoculating seedlings with truffle fungi: past and future perspectives. Mycorrhiza 25: 77-81.

NCBI (2018a) National Center for Biotechnology Information Database. https://www.ncbi.nlm.nih.gov/genome/?term=txid42249[orgn] Accessed 1 October 2018.

NCBI (2018b) National Center for Biotechnology Information Database. https://www.ncbi.nlm.nih.gov/ipg/?term=Tuber%20magnatum Accessed 9 October 2018.

Paolocci F, Rubini A, Granetti B, Arcioni S (1999) Rapid molecular approach for a reliable identification of *Tuber* spp. ectomycorrhizae. FEMS Microbiology Ecology 28: 23-30.

Payen T, Murat C, Gigant A, Morin E, De Mita S, Martin F (2015) A survey of genome-wide single nucleotide polymorphisms through genome resequencing in the Perigord black truffle (*Tuber melanosporum* Vittad.). Molecular Ecology Resources 15: 1243-1255.

Rizzello R, Zampieri E, Vizzini A, Autino A, Cresti M, Bonfante P, Mello A (2012) Authentication of prized white and black truffles in processed products using quantitative real-time PCR. Food Research International 48: 792-797.

Rubini A, Topini F, Riccioni C, Paolocci F, Arcioni S (2004) Isolation and characterization of polymorphic microsatellite loci in white truffle (*Tuber magnatum*). Molecular Ecology Notes 4: 116-118.

Rubini A, Paolocci F, Riccioni C, Vendramin GG, Arcioni S (2005) Genetic and phylogeographic structures of the symbiotic fungus *Tuber magnatum*. Applied and Environmental Microbiology 71: 6584-6589.

Rubini A, Belfiori B, Riccioni C, Tisserant E, Arcioni S, Martin F, Paolocci F (2010a) Isolation and characterization of MAT genes in the symbiotic ascomycete *Tuber melanosporum*. New Phytologist 189: 710-722.

Rubini A, Belfiori B, Riccioni C, Arcioni S, Martin F, Paolocci F (2010b) *Tuber melanosporum*: mating type distribution in a natural plantation and dynamics of strains of different mating types on the roots of nursery-inoculated host plants. New Phytologist 189: 723-735.

Schuler GD (1997) Sequence mapping by electronic PCR. Genome Research 7: 541-550.

Splivallo R, Novero M, Bertea C, Bossi S, Bonfante P (2007a) Truffle volatiles inhibit growth and induce an oxidative burst in *Arabidopsis thaliana*. New Phytologist 175: 417-424.

Splivallo R, Bossi S, Maffei M, Bonfante P (2007b) Discrimination of truffle fruiting body versus mycelial aromas by stir bar sorptive extraction. Phytochemistry 68: 2584-2598.

Splivallo R, Ottonello S, Mello A, Karlovsky P (2011) Truffle volatiles: from chemical ecology to aroma biosynthesis. New Phytologist 189: 688-699.

Splivallo R, Valdez N, Kirchoff N, Castiella Ona M, Schmidt JP, Feussner I, Karlovsky P (2012) Intraspecific genotypic variability determines concentrations of key truffle volatiles. New Phytologist 194: 823-835.

Uncu AO, Gultekin V, Allmer J, Frary A, Doganlar S (2015) Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in sesame. The Plant Genome, 8, doi: 10.3835/plantgenome2014.11.0087.

Wang X, Wang L (2016) GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. Frontiers in Plant Science 7: 1350.

Zampieri E, Rizzello R, Bonfante P, Mello A (2012) The detection of mating type genes of *Tuber melanosporum* in productive and non productive soils. Applied Soil Ecology 57: 9-15.