Research in Plant Biology, 2(6): 01-09, 2012 ISSN : 2231-5101
www.resplantbiol.com

Regular Article

In silico EST-SSRs Analysis in UniGene of *Quercus robur* L.

Ertugrul Filiz[1*], Ibrahim Koc[2] and F. Cigdem Sakinoglu[1]

[1]Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, Duzce, Turkey
[2]Department of Molecular Biology and Genetics, Gebze Institute of Technology, Kocaeli, Turkey
*Corresponding author E-mail: ertugrulfiliz@gmail.com

Pedunculate oak (*Quercus robur* L.) is one of the most important tree components of Europe's forest ecosystems, possessing both ecological and economical value. Development of genomic resources, such as genetic markers, is needed to support gene conservation and tree improvement activities. Experimental methods to develop SSR markers are laborious, time consuming and expensive, while in silico approaches have become a practicable and inexpensive alternative in genetic studies. The aim of this study was to characterize simple sequence repeat (EST-SSR) markers and functional annotation of SSR containing sequences in *Q. robur* unigene sequences. 7170 unigene sequences (5147.315 kb) of *Q. robur* were downloaded from National Center for Biotechnology Information (NCBI). A total of 475 (6.62 %) unigene sequences containing 525 SSRs (microsatellites) were identified by using MISA software. The average frequency of microsatellites was found, on average, one in every 9.8 kb of sequence. The analysis revealed that tri-nucleotide repeats (42.6%) were most abundant followed by di-nucleotide (36.9%), hexa-nucleotide (11.8%), penta-nucleotide (4.9%) and tetra-nucleotide repeats (3.8%), respectively. Flanking sequences of the 525 SSRs generated 500 primers (95.2%) with forward and reverse strands by using Primer3 software. Gene based SSR markers can be used for studies of genetic diversity, population genetics, genetic mapping, gene tagging and more. Large numbers of unigenes containing SSRs (77.4%), annotations were available 46.75% of which were predicted, 23.91% were hypothetical, 8.83% were putative and 20.51% belonged to other protein types. Only 22.5% sequence could not assign to any specific protein class.

Key words: EST-SSR, *in silico* analysis, oak, *Quercus robur*, unigene.

Introduction

Q. robur is a member of the white oak section (*Lepidobalanus*) and distributed throughout Europe from central Spain to the Urals (Barreneche et al., 1998). It is also known as pedunculate oak, European oak or common oak and it is the most widespread oak species in Europe. This tree supports important environmental services such as carbon sequestration, water cycle, reservoir of biodiversity and economic contribution. *Quercus robur* L. is a diploid species (2n=2x=24) but in some cases B type chromosomes have been observed (Besendorfer et al., 1996). The physical genome size of *Q. robur* (1.88 pg/2C) is larger than other several woody angiosperms such as *Populus*, *Eucalyptus*, *Acacia*, *Fraxinus* (Zoldos et al., 1998). Various types of molecular markers have

Received: 27.9.2012; Revised: 31.10.2012; Accepted: 2.11.2012

been used to study genetic architecture in *Q. robur* including random amplified polymorphic DNA (RAPD) markers (Barreneche et al., 1998), amplified fragment length polymorphisms (AFLP) markers (Saintagne et al. 2004), and simple sequence repeats (SSR) markers (Barreneche et al., 2004; Neophytou et al., 2010).

Microsatellites or simple sequence repeats (SSRs) are stretches of DNA consisting of tandemly repeated short units of 1–6 base pairs in length. SSRs are useful in plant research because of their variations, co-dominant inheritance, relative abundance, multi-allelic nature, extensive genome coverage, high reproducibility and ease of detection (Singh et al., 2011; Powel et al., 1996). Recently, expressed sequence tags (ESTs), genes or cDNA clones can be downloaded from various public biological databases and by using bioinformatic tools for identification of SSR regions referred as EST-SSR or genic microsatellite. EST is short strand of DNA which is a part of a cDNA molecule and can act as identifier of a gene. As they locate in or near coding DNA, they should be more conserved than genomic sequences, making them more significant, inexpensive and cross-species transferability than traditional anonymous SSRs (Chagne et al., 2004; Liewlaksaneeyanawin et al., 2004; Gutierrez et al., 2005; Pashley et al., 2006). Exploiting the publicly available genomic resources (EST databases) for the development of SSRs could be feasible option for obtaining high-quality nuclear markers (Gupta et al., 2003; Bhat et al., 2005). In addition, SSR regions have some variations in repeat units of SSRs and coding region and may affect gene expression or function such as gene transcription and/or translation, inactivate or activate genes or truncate protein, gene silencing or transcription slippage etc. Thus, microsatellites are utilized effectively for genetic mapping, functional diversity, transferability and comparative mapping studies (Varshney et al., 2005).

The present study focuses on for mining UniGene sequences for determination of microsatellite in *Q. robur*, which helps the development of SSR markers and annotates SSR containing sequences.

Material and method

Retrieval of UniGene sequences and Identification of EST-SSRs

UniGene (http://www.ncbi.nlm.nih.gov/UniGene/) is a database which contains sequences of well-characterized genes as well as hundreds of thousands novel EST sequences. A total of 7170 UniGene sequences (Csi.seq.uniq) of *Q. robur* were downloaded from the NCBI (ftp://ftp.ncbi.nih.gov/repository/UniGene/Quercus_robur/) for microsatellite discovery and annotation. The unigene sequences were mined for microsatellites using a program MISA (MIcroSAtellite) (http://pgrc.ipk-gatersleben.de/misa/) identification tool written in the Perl scripting language. The minimum motif repeat size were set to 10 for dinucleotide, 6 for trinucleotide, 5 for tetranucleotide, 4 for pentanucleotide and 4 for hexanucleotide in locating the microsatellites. The analysis of SSRs was done on the basis of their types (di-three-tetra-penta-hexa-nucleotide), number of repeats, percentage frequency of occurrence of each SSR motif and their distribution in the sequence (Gupta et al., 2010a).

Primer designing for SSRs

A pair of primers flanking each SSR was designed using Primer3 software available at www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi. The micro-satellites containing unigenes were used for designing primers pairs. The primers were designed from the flanking sequences having microsatellite repeats using PRIMER3 software (Whitehead Institute, USA). In the present study, default parameters of the PRIMER3 were taken for primer designing.

Functional Annotation of SSR containing sequences

Functional annotation of all the SSR containing sequences was determined on the basis of 75% similarity against nonredundant (nr) protein database entries. It was performed using program BLAST (Basic Local Alignment Search Tool) and BLASTX, available at NCBI (http://www.ncbi.nlm.nih.gov/blast). The resulting proteins obtained during similarity search by BLASTX program were classified into their respective classes. The most significant matches (EXP <1e−4) for each sequence were recorded.

Results

SSR mining from EST sequences

In the present study, 7170 UniGene sequences of *Q. robur* available at NCBI (http://www.ncbi.nlm.nih.gov/unigene) were searched for simple sequence repeats with a minimum length of 18 bp. A total of 525 SSRs detected from 5147.315 kb of data screened, excluding Poly A and Poly T. Depending upon the length of the repeat unit itself (2–6 bp), the lengths of SSRs varied from 18 to 30 bp, respectively.

Table 1. Summary of EST-SSR mining of UniGene sequences for *Q. robur*

Parameters	Values
Total number of sequences examined:	7170
Total size of examined sequences (bp):	5147315
Total number of identified SSRs:	525
Number of SSR containing sequences:	475
Number of sequences containing more than 1 SSR:	46
Number of SSRs present in compound formation:	24
Repeat type	
Dinucleotide	194
Trinucleotide	223
Tetranucleotide	20
Pentanucleotide	26
Hexanucleotide	62

Microsatellites frequency with different repeat types

A subset of 475 sequences contains 525 SSRs in 7170 UniGene sequences, indicating that merely 6.62% of sequences contained SSRs, which represent average density of one SSR per 9.8 kb. Fig. 1a shows the frequencies of SSRs with di-, tri-, tetra-, penta- and hexanucleotide repeat units. The most frequent repeat type found within the UniGene sequences of *Q. robur* were tri-nucleotide repeats (42.6%) followed by di-nucleotide (36.9%), hexa-nucleotide (11.8%), penta-nucleotide (4.9%) and tetra-nucleotide repeats (3.8%), respectively. The observed frequency of different repeat types comprising the SSRs is summarized in Table 1.

Identified SSRs belongs to nine different types of di-nucleotide repeats, 48 different types of tri-nucleotide repeats, 17 different types of tetra-nucleotide repeats and 23 different types of penta- nucleotide and 62 different types of hexa-nucleotide repeats, respectively (Table 2). The most abundant repeat motifs were $(TC)_n$, $(CT)_n$ and $(AG)_n$ in di-nucleotide, $(AAG)_n$, $(TTC)_n$ and $(TCT)_n$ in tri-nucleotide, $(TAAA)_n$, $(TCTT)_n$ and $(TTCT)_n$ in tetra-nucleotide, $(AAAAG)_n$, $(TTTTA)_n$ and $(TTTTG)_n$ in penta-nucleotide, $(AACAAA)_n$ in hexa-nucleotide repeats.

SSRs were categorized into two groups based on length of SSR tracts and their potential as informative genetic markers: Class I SSRs contain perfect repeats ≥20 nucleotides in length and Class II contain perfect repeats >10 nucleotides and <20 nucleotides in length (SINGH et al. 2011). Out of 525 SSRs, 302 repeats were categorized as Class I SSRs (57.6%) and 223 repeats were categorized as Class II SSRs (42.4%) (Fig. 1b). 36.9% of dinucleotide repeats were Class I SSRs, followed by hexanucleotide (11.8%) pentanucleotide (4.9%), and tetranucleotide repeats (3.8%). All di-, tetra- , penta - and hexanucleotide repeats were Class I SSRs while only trinucleotide repeats were Class II SSRs.

The estimated frequency of Class I SSRs was one SSR per 17 kb, whereas the frequency of Class II was one SSR per 23.1 kb.

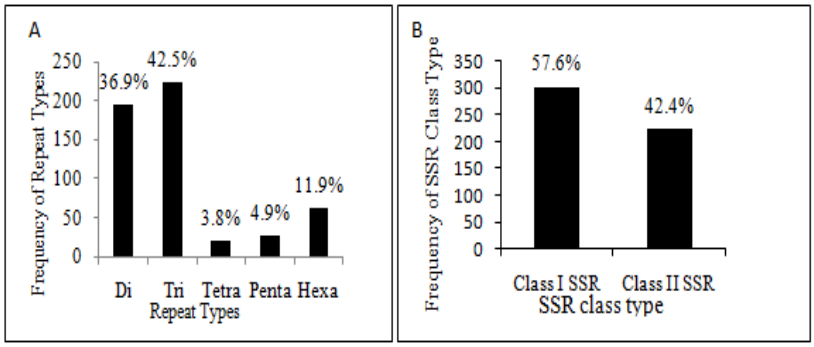

Figure 1. (a) Frequency distribution of different repeat types, (b) Frequency of SSR class

Table 2. Different types of di-, tri-. tetra-, penta- and hexa-nucleotide repeat motifs

Repeats	Motifs
Di-nucleotide	AC, AG, AT, CA, CT, GA,TA, TC, TG
Tri-nucleotide	AAC, AAG, AAT, ACA, ACC, AGA, AGC, AGG, AGT, ATA, ATC, ATG, ATT, CAA, CAC, CAG, CAT, CCA, CCG, CGG, CGT, CTC, CTG, CTT, GAA, GAC, GAG, GAT, GCC, GCG, GCT, GGA, GGT, GTA, GTG, GTT, TAA, TAT, TCA, TCC, TCT, TGA, TGC, TGG, TGT, TTA, TTC, TTG
Tetra-nucleotide	AAAC, AAAG, AAGA, AGAA, ATAA, ATAC, CTTT, GTTT, TAAA, TATC, TATT, TCTT, TTAT, TTCT, TTTA, TTTC, TTTG
Penta-nucleotide	AAAAG, AAACC, AAATA, AAATT, AACCA, AACTC, AATCC, ACCCG, AGGGT, ATGCT, ATTAT, CAAAC, CATTA, GAACT, GAGCA, GTATG, GTGTT, TATAT, TCTCT, TTTGG, TTTTA, TTTTC, TTTTG
Hexa-nucleotide	AAAAAG, AAACAA, AAATTC, AACAAA, AACCCC, AAGCCA, ACTCCC, AGAAAA, AGAACC, AGCCCA, AGCTCA, ATAGCA, ATCACC, ATGAAC, ATTTTT, CAAACA, CAAACG, CAAATC, CAACAG, CACAAC, CACCAG, CAGAAA, CAGCAA, CAGCCA, CAGCCT, CATGGA, CCAAAG, CCACCT, CCCGTT, CCTACT, CCTGCA, CGCCGT, CTCCTT, CTCTCC, CTTCAA, CTTCAT, GAAACC, GAGAAC, GAGAGT, GAGCAC, GATCCA, GCACCA, GCTATG, GGAATG, GGAGAT, GGTAGC, GGTGGA, GGTGTC, GGTGTT, GTACCA, TAGCAA, TCAGCC, TCGCCA, TCTCAA, TGGGTA, TTAGGG, TTTCTT, TTTGGT, TTTTAG, TTTTTA, TGGTTA, TAACTG

SSRs primer design

Out of 525 SSRs detected primers could be designed for 500 (95.2%) SSRs with only 25 (4.8%) remaining sequences did not produce any acceptable primers. These 500 SSRs for which primers were designed

include 155 di-, 254 tri-, 20 tetra- , 18 penta- and 53 hexa-nucleotide repeats. The details of the accession numbers of UniGene sequences of *Q. robur*, repeat motif of SSRs for which primer were designed, length of repeat unit, primer sequences, annealing temperature and product size are available as supplementary information.

Annotation of *Q. robur* sequences containing SSRs

475 SSR containing EST sequences were scanned and annotated against the non-redundant (nr) protein database to determine the function using BLASTX, available at NCBI (http://www.ncbi.nlm.nih.gov/blast). Out of 475 SSR containing sequences, annotations of which 1615 (46.75%) were predicted proteins, 826 (23.91%) were hypothetical proteins, 305 (8.83%) were putative proteins and 708 (20.51%) belonged to other protein types, were available for a large number 368 (77.4%) of sequences (Fig. 2). Only 107 (22.5%) sequences could not be assigned to any specific class due to the absence of a homology in the protein sequence database.

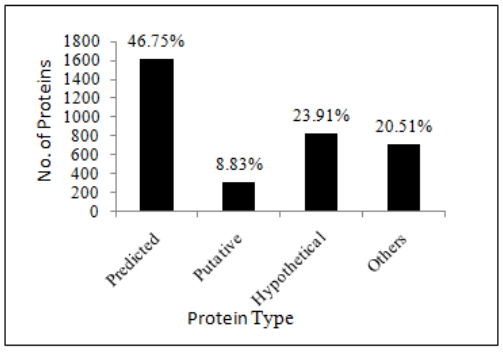

Figure 2. Distribution of SSR containing sequences according to the proteins encoding.

Distribution of tri-nucleotide SSRs and encoded amino acids

Plants are able to generate all 20 amino acids which are the building blocks of protein by themselves. Each trinucleotide motif codes a particular amino acid which has putative roles in biological activity of protein molecules. Out of a total of 223 trinucleotides, 11.21% trinucleotides SSRs encoded Serine, closely followed by Glutamic Acid (9.41%), Leucine (7.62%), Phenylalanine and Histidine each has equal contribution (7.17 %) (Fig. 3.).

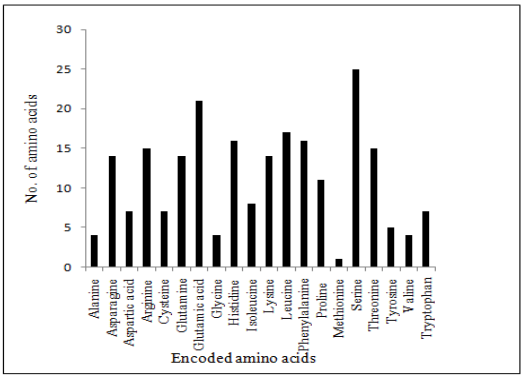

Figure 3. Distribution of encoded amino acids.

Nature of trinucleotide SSR encoded amino acids

The chemical and physical properties of the amino acids of proteins determine the biological activity of the protein. Trinucleotide SSR encoded amino acids were classified on the basis of their polar & nonpolar nature and hydrophilic & hydrophobic nature. Polar amino acids were more frequent (70.8 %) than nonpolar amino acids (29.2 %). Also, hydrophobic amino acids (57%) were more frequent than hydrophilic amino acids (43%). (Fig. 4.)

Discussion

Forest genetic resources are vital to adapting to changes in the environment for species. A better understanding of the diversity of these species is crucial for their sustainable use and conservation. In the present study, one of most important the forest tree species *Q. robur* UniGene sequences were used to mine for simple sequence repeats. Following SSRs were characterized; these SSRs were detected for primer designing. Also, SSRs containing sequences were annotated and tri-nucleotide SSRs were evaluated to determine encoded amino acids and nature of amino acids. While calculating the frequency of SSRs, mononucleotides were

not taken into account. 475 SSRs were identified based on the 7170 *Q. robur* Unigene sequences. The overall density of SSRs is one SSR per 9.8 kb.

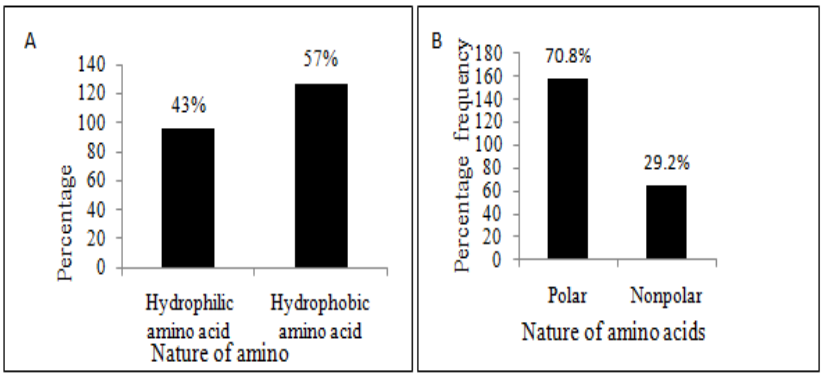

Figure 4. (a) Percentage frequency of hydrophilic & hydrophobic amino acids, (b) Percentage frequency of polar & non-polar amino acids.

The frequency of SSRs detected in this study was lower than earlier studies such as 1SSR/1.67 kb in wheat (Morgante et al., 2002), 1SSR/6 kb in *Arabidopsis* (Cardle et al., 2000), 1 SSR/1.3 kb in *S. lycopersicum* (Gupta et al., 2010a), 1SSR/10.4 kb in *G. barbadense* (Yuanda et al., 2010), 1SSR/0.7 kb, 1SSR/1.67 kb, 1SSR/0.22 kb and 1SSR/3.5 kb in four different species including the sequences of major palms like coconut, arecanut , oil palm and date palm respectively (Palliyarakkal et al., 2011), 1SSR/1.77 kb in *R. communis* (Qiu et al., 2010), 1SSR/6.0 kb averaged over six different species including barley, maize, rice, rye, sorghum and wheat (Varshney et al., 2002). Morgante et al. (2002) reported that whole genome SSR frequency was inversely related to the genome size in four plant species. These results suggest that large genome size of *Q. robur* may affect SSRs density in EST sequences. However, the density of SSRs in *Q. robur* (1SSR/9.8 kb) is higher than densities found in 1SSR/12.92 kb in *C. sinensis* (Shanker et al., 2007), 1SSR/21.2 kb in *F. graminearum* (Singh et al., 2011), 1SSR/14.73 kb in *C. longa* (Joshi et al., 2010), 1SSR/56.6 kb in loblolly pine and 1SSR/42.9 kb in spruce (Bérubé et al., 2007). Particularly, we observed that *Q. robur* has higher SSRs density than the other forest tree species as loblolly pine and spruce. This result may be affected by SSRs related to gene loss, gene duplications, etc. in evolutionary history of *Q. robur*.

Many studies have suggested that the trinucleotide repeat is the predominant EST-SSR repeat type in most plants, followed by the dinucleotide and tetranucleotide repeat types (Kota et al., 2001). The SSRs identified contained 194 di, 223 tri-, 20 tetra-, 26 penta- and 62 hexanucleotides**.** In present study, the SSRs with tri-nucleotide repeats (42.6%) were most abundant in UniGene sequences of *Q. robur*. This is in agreement with the results of the previous studies on *R. communis* (Qiu et al., 2010), some cereal species (Varshney et al., 2002), *C. longa* (Joshi et al., 2010), *N. crassa* (Shanker et al., 2007), *T. aestivum* (Gupta et al., 2003), *O. basilicum* (Gupta et al., 2010b), *L. perenne* (Asp et al., 2007), *E. globulus* (Acuña et al., 2011), *Q. robur* (Durand et al., 2010). Many EST sequences have exonic regions consisting of codons and encode proteins. Such dominance of trinucleotide repeats over other repeats in coding regions may be explained on the basis of the selective risk of non-trimeric SSR variants in coding regions, possibly causing frame-shift mutations (Metzgar et al., 2000). For this reason, trinucleotide repeats are the most abundant SSR class found in ESTs. Among trinucleotide repeats, 48 different types of repeat motifs

were identified and the AAG & TTC repeat motif was predominant in *Q. robur* genome. This result is similar to in the earlier studies on *Gossypium barbadense* (Yuanda et al. 2010), *Curcuma longa* (Siju et al., 2010), loblolly pine (Liewlaksaneeyanawin et al., 2004) and *Q. robur* (Durand et al., 2010). Also, AAG and GGC motifs appeared common in *Arabidopsis* (Cardle et al., 2000), AAG is the second most abundant motif in loblolly pine (Bérubé et al., 2007), while it is the most abundant motif in *Q. robur* (Durand et al., 2010). Hexa-nucleotide repeats (11.8%) are more abundant than tetra- (3.8%) and penta-nucleotide repeats (4.9%) in ESTs of *Q. robur* in this study. This result is consistent with the earlier studies on *Q. robur* (Durand et al., 2010), major crops (Gao et al., 2003) and various eukaryotic genomes (Tóth et al., 2000). It was observed that out of all coded amino acids Serine (Ser) demonstrated the highest percentage of occurrence followed by Glutamic acid (Glu). Serine is important in biological processes in which it participates in the biosynthesis of purines and pyrimidines and also plays an important role in the catalytic function of many enzymes (Rai, 2002). Thus, our findings support this hypothesis that the abundance of serine residues may reflect importance of serine using in many proteins as structural and/or catalytic functions. Primer design is not an exact science, and a success rate of 60–90% amplification for both genomic and EST-SSRs has been reported in different studies (Varshney et al., 2005). In present study, the putative primers could be designed successfully for a very large number (500, 95.2%) of SSRs containing sequence and these primers can be used for a variety of purposes such as gene tagging, genetic mapping, functional diversity, comparative mapping, etc.

Functional roles of microsatellites located near or within coding regions, are important for plants. SSRs may affect DNA replication (Field and Wills 1996) and also play an important role in regulation of gene activity (Sandaltzopoulos et al., 1995). Tri-nucleotide repeats were overrepresented in ORFs and encoded a biased set of amino acids (Young et al., 2000). In *Arabidopsis*, the TAIR10 release contains 27,416 protein coding genes, 4827 pseudogenes or transposable element genes and 1359 ncRNAs (33,602 genes in all, 41,671 gene models), which provides a good source for annotating sequences, but it is still inadequate. In current study, homologies of SSR containing sequences were analyzed in non-redundant (nr) protein database using BLASTX. Out of 475 SSRs containing sequences were annotated, only for 368 (77.4%) sequences were categorized into different classes of proteins (predicted, hypothetical, putative and others). In 107 (22.5%) SSRs containing sequences, no homology could be found because of large genome size that obvious function has not been assigned.

Conclusions

This study supports the utilization of in-silico approaches to analyze microsatellites (SSRs) from unigene sequences of plants. Unigene database provide a valuable resource for the development of SSR markers which are associated with transcribed gene regions in genome. Annotation of SSR containing sequences provides an opportunity to identify the functional diversity of amino acids. This study revealed the analysis of microsatellites in UniGenes of *Quercus robur*. A total of 525 non-redundant SSRs were detected in silico using UniGenes data and tri-nucleotide repeats were the most abundant type of repeats followed by di- and hexa-nucleotide repeats. 502 (95.2%) EST-SSR primers could be designed for 525 non-redundant SSRs that these EST-SSR primers should be used for genetic mapping, comparative mapping, functional diversity and population structure analysis, facilitating breeding of different oak species especially for *Q. robur*.

References

Acuña CV, Fernandez P, Villalba PV, García MN, Hopp HE, Poltri SNM. 2011. Discovery, validation, and in silico

functional characterization of EST-SSR markers in *Eucalyptus globules*. Tree Genet Genomes doi: 10.1007/s11295-011-0440-0.

Asp T, Frei UK, Didion T, Nielsen KK, Lübberstedt T. 2007. Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*. BMC Plant Biol 7**:** 36.

Barreneche T, Bodenes C, Lexer C. et al 1998. A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. Theor Appl Genet 97: 1090-1103.

Barreneche T, Casasoli M, Russell K, Akkak A, Meddour H, Plomion C, Villani F, Kremer A. 2004. Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). Theor Appl Genetic 108(3): 558–566.

Bérubé Y, Zhuang J, Rungis D, Ralph S, Bohlmann J, Ritland K. 2007. Characterization of EST-SSRs in loblolly pine and spruce. Tree Genet Genomes 3: 251–259.

Besendorfer V, Zoldos V, Peskan T, Krnik-Rasol M, Littway T, Papes D. 1996. Identification of potential cytogenetical and biochemical markers in bioindication of common oak forest. Phyton (Horn, Austria) Special issue 36: 139-146.

Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, Aggarwal RK 2005. Identification and characterization of expressed sequence tags-derived simple sequence repeats, markers from robusta coffee variety 'C X R' (an interspecific hybrid of *Coffea canephora* X *Coffea congensis*). Mol Ecol Notes 5: 80–83.

Cardle L, Ramsay L, Milborne D, Macaulay M, Marshall D, Waugh R. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156: 847–854.

Chagne D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT. et al. 2004. Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109: 1204–1214.

Durand J, Bodénès C, Chancerel E. et al. 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11: 570-582.

Field D, Wills C. 1996. Long, polymorphic microsatellite in simple organisms. Proc R Soc London B Biol Sci 263: 209–215.

Gao L, Tang J, Li H, Jia J. 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 12: 245–261.

Gupta S, Tripathi KP, Roy S, Sharma A. 2010a. Analysis of unigene derived microsatellite markers in family *Solanaceae*. Bioinformation 5(3): 113-121.

Gupta S, Shukla R, Roy S, Sen N, Sharma A. 2010b. In silico SSR and FDM analysis through EST sequences in *Ocimum basilicum*. Plant Omics J 3(4): 121-128.

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. 2003. EST-SSRs for transferability, polymorphism and genetic diversity in bread wheat. Mol Genet Genom 270: 315–323.

Gutierrez MV, Patto MCV, Huguet T, Cubero JI, Moreno MT, Torres AM 2005. Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. Theor Appl Genet 110: 1210–1217.

Joshi RK, Kuanar A, Mohanty S, Subudhi E, Nayak S. 2010. Mining and characterization of EST derived microsatellites in *Curcuma longa* L. Bioinformation 5(3). 128-131.

Kota R, Varshney RK, Thiel T. et al. 2001. Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum Vulgare* L.). Hereditas 135: 145−151.

Liewlaksaneeyanawin C, Ritland CE, Kassaby YA, Ritland K. 2004. Single-copy, species-transferable microsatellite

markers developed from loblolly pine ESTs. Theor Appl Genet 109: 361–369.

Metzgar D, Bytof J, Wills C. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10: 72–80.

Morgante M, Hanafey M, Powell W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet 30: 194–200.

Neophytou C, Aravanopoulos FA, Dounavi A. 2010. Detecting interspecific and geographic differentiation patterns in two interfertile oak species (Quercus petraea (Matt.) Liebl. and Q. robur L.) using small sets of microsatellite markers. Forest Ecology and Management 259: 2026–2035.

Palliyarakkal MK, Ramaswamy M, Vadivel A. 2011. Microsatellites in palm (*Arecaceae*) sequences. Bioinformation 7(7): 347- 351.

Parida SK, Yadava DK, Mohapatra T. 2010. Microsatellites in *Brassica* unigenes: relative abundance, marker design, and use in comparative physical mapping and genome analysis. Genome 53(1): 55-67.

Pashley CH, Ellis JR, Mccauley DE, Burke JM 2006. EST databases as a source for molecular markers: lessons from *Helianthus*. J Hered 97: 381–388.

Powell W, Machray GC, Provan J. 1996. Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1: 215-222.

Rai VK. 2002. Role of amino acids in plant responses to stresses. Biol Plantarum, 45(4): 481-487.

Qiu L, Yang C, Tian B, Yang JB, Liu A. 2010. Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). BMC Plant Biol 10: 278.

Saintagne C, Bodénès C, Barreneche T, Pot D, Plomion C, Kremer A. 2004. Distribution of genomic regions differentiating oak species assessed by QTL detection. Heredity 92(1): 20–30.

Sandaltzopoulos R, Mitchelmore C, Bonte E, Wall G, Becker PB 1995. Dual regulation of the *Drosophila* hsp26 promoter in vitro. Nucl Acids Res 23: 2479–2487.

Shanker A, Bhargava A, Bajpai R, Singh S, Srivastava S, Sharma V. 2007. Bioinformatically mined simple sequence repeats in UniGene of *Citrus sinensis*. Sci Hortic 113: 353–361.

Siju S, Dhanya K, Syamkumar S, Sasikumar B, Sheeja TE. 2010. Development, Characterization and Cross Species Amplification of Polymorphic Microsatellite Markers from Expressed Sequence Tags of Turmeric (*Curcuma longa* L.). Mol Biotechnol 44: 140.

Singh R, Sheoran S, Sharma S, Chatrath R. 2011. Analysis of simple sequence repeats (SSRs) dynamics in fungus *Fusarium graminearum*. Bioinformation 5(10): 402-404.

Tóth G, Gáspári Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10: 967–981.

Varshney RK, Graner A, Sorrells ME. 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23: 1.

Varshney RK, Thiel T, Stein N, Langridge P, Graner A. 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7: 537–546.

Young ET, Sloan JS, Van Riper K 2000. Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. Genetics 154:1053–1068.

Yuanda L, Caiping C, Lei W, Shaoyan L, Liang Z, Liangliang T, Junhong L, Tianzhen Z, Wangzhen G. 2010. Mining, characterization, and exploitation of EST-derived microsatellites in *Gossypium barbadense*. Crop Genetics 55: 1889−1893.

Zoldos V, Papes D, Brown SC, Panaud O, Siljak-Yakovlev S. 1998. Genome size and base composition of seven *Quercus* species: inter- and intra-population variation. Genome 41 (2): 162-168.