DE GRUYTER OPEN

*Agriculture (Poľnohospodárstvo), 61, 2015 (4): 149−156*

DOI: 10.1515/agri-2015-0020

# *IN SILICO* RETRIEVING OF OPIUM POPPY (*PAPAVER SOMNIFERUM* L.) MICROSATELLITES

VERONIKA MASÁROVÁ[1], DANIEL MIHÁLIK[2,3], JÁN KRAIC[2,3*]

[1]Constantine the Philosopher University in Nitra, Slovak Republic
[2]National Agricultural and Food Centre – Research Institute of Plant Production, Piešťany, Slovak Republic
[3]University of SS. Cyril and Methodius in Trnava, Slovak Republic

MASÁROVÁ, V. – MIHÁLIK, D. – KRAIC, J.: *In silico* retrieving of opium poppy (*Papaver somniferum* L.) microsatellites. Agriculture (Poľnohospodárstvo), vol. *61*, 2015, no. 4, pp. 149–156.

Repetitive tandem sequences were retrieved within nucleotide sequences of opium poppy (*Papaver somniferum* L.) genomic DNA available in the GenBank® database. Altogether 538 different microsatellites with the desired length characteristics of tandem repeats have been identified within 450 sequences of opium poppy DNA available in the database. The most frequented were mononucleotide repeats (246); nevertheless, 44 dinucleotide, 148 trinucleotide, 62 tetranucleotide, 28 pentanucleotide and 5 hexanucleotide tandem repeats have also been found. The most abundant were trinucleotide motifs (27.50%), and the most abundant motifs within each group of tandem repeats were TA/AT, TTC/GAA, GGTT/AACC and TTTTA/TAAAA. Five hexanucleotide repeats contained four different motifs.

Key words: opium poppy, microsatellite, tandem repeat, *in silico*, GenBank®

The opium poppy (*Papaver somniferum* L.) belongs to the family *Papaveraceae* and is the most important member of this family from the agricultural point of view. This crop is cultivated for the production of not only edible seeds and oil but also pharmaceutically important alkaloids, especially morphine, codeine, narcotine, thebaine and papaverine. Opium poppy might also be regarded as one of the most important renewable resource for the extraction of pharmaceutical alkaloids used as narcotics, analgesics, and relaxants (Şelale *et al*. 2013).

The genus *Papaver* includes about 110 species occurred naturally mainly in the northern hemisphere, outside the tropics. Some of them are considered weeds (*Papaver dubium*, *Papaver rhoeas*, *Papaver rumelicum*), and others are grown as ornamental plants (*Papaver pseudo-orientale*, *Papaver alpinum*, *Papaver burseri*, *Papaver bracteatum*). The *P. somniferum* L. has been improved by breeding programmes, nevertheless, running only in a limited number of countries. Its improvement targets to three fundamentally different goals. The first is the development of cultivars producing high amounts of alkaloids (morphine, thebaine, codeine), and these are cultivated mainly in Australia (Tasmania), Turkey, Hungary, France, Spain, Poland and India. The second breeding goal is the opposite, that is, to develop low-morphine cultivars producing very low or zero amount of morphine in heads and straw. These cultivars are grown mainly in Poland and Austria.

Mgr. Veronika Masárová, Department of Botany and Genetics, Faculty of Natural Sciences, Constantine the Philosopher University, Tr. A. Hlinku 1, 949 74 Nitra, Slovak Republic
Mgr. Daniel Mihálik, PhD., Prof. RNDr. Ján Kraic, PhD. (*Corresponding author), Research Institute of Plant Production, National Agricultural and Food Centre, Bratislavská cesta 122, 921 68 Piešťany, Slovak Republic. E-mail: mihalik@vurv.sk, kraic@vurv.sk
Mgr. Daniel Mihálik, PhD., Prof. RNDr. Ján Kraic, PhD., Department of Biotechnology, Faculty of Natural Sciences, University of SS. Cyril and Methodius, Nám. J. Herdu 2, 917 01 Trnava, Slovak Republic. E-mail: daniel.mihalik@ucm.sk, jan.kraic@ucm.sk

The third goal of opium poppy breeding targets to creation of cultivars with high yield of edible seeds of blue, greyish, white, or ochry colour and seeds are the main product for direct consumption and the food industry. These cultivars contain low or medium level of morphine, but only in dry heads. The leading country in the production of poppy seeds is the Czech Republic.

Although breeding programmes are directed to different goals, all need as much as available genetic variation within the *P. somniferum* L. and related germplasm along with effective tools for genetic diversity evaluation. However, the majority of commercially used cultivars contain low genetic variation (Acharya *et al*. 2009) and related high similarity in phenotype traits. Another aspect of opium poppy exploitation is its high specificity and economical significance especially related with the production of drugs derived from alkaloids. Owing to this aspect, the breeding, development and releasing of new cultivars as well as their cultivation are under strict legal control. Control over the handling with opium poppy cultivars and especially poppy straw containing alkaloids necessitates effective tools for distinguishing between cultivars. Commonly, it is based on the description of plant phenotype by morphological and agronomical traits and characteristics. Nevertheless, this approach cannot clearly differentiate cultivars from each other. Therefore, DNA marker systems such as the RAPD (random amplified polymorphic DNA), ISSR (inter-simple sequence repeat) and AFLP (amplified fragment length polymorphisms) (Saunders *et al*. 2001; Dittbrenner *et al*. 2008; Hari *et al*. 2009; Parmaksiz & Özcan 2011; Güçlü *et al*. 2014) have been applied for genetic diversity studies in opium poppy. Also SSR (simple sequence repeat) and EST-SSR (expressed sequence tag derived simple sequence repeat) markers for opium poppy have been developed recently (Lee *et al*. 2010; Lee *et al*. 2011; Şelale *et al*. 2013; Çelik *et al*. 2014). Microsatellites (SSRs) are tandemly repeated nucleotide motifs ranging usually up to six nucleotides within the DNA molecule (Litt & Luty 1989). Their advantages are hypervariability, co-dominant inheritance, multi-allelic character, high abundance and extensive genome coverage within genomes, especially eukaryotic. Moreover, their assays are simple, reproducible and convenient for polymerase chain reaction (PCR)-based technologies, automation and high-throughput genotyping (Gupta & Varshney 2000). SSRs are usually located outside DNA-coding sequences. The EST-SSRs reveal variation in microsatellites located in expressed and regulatory regions of the genome (Suárez *et al*. 2000) and usually exhibit transferability between different genus (Gupta *et al*. 2003). Screening of SSRs directly from genomic libraries or microsatellite-enriched libraries is laborious, time consuming and expensive. Another alternative is the *in silico* mining for tandem repeats within available databases of DNA nucleotide sequences followed by designing and testing primers within flanking regions of individual microsatellite loci (Sharma *et al*. 2007). This approach is more effective by using available specific tandem repeats mining software tools (Thiel *et al*. 2003; Tang *et al*. 2008; Sarmah *et al*. 2012; Wang *et al*. 2013), especially in species with low level of polymorphism.

The aim of this study was to retrieve microsatellites within all DNA sequences of opium poppy (*P. somniferum* L.) available in public database of nucleotide sequences and to evaluate their abundance and composition.

## MATERIAL AND METHODS

Tandem repeats were searched within DNA sequences deposited and available in the GenBank® database (http://www.ncbi.nlm.nih.gov/genbank, Benson *et al*. 2013). The software SSRLocator (da Maia *et al*. 2008) was used for searching tandem repeats. The minimum number of repeats required for searching mono-, di-, tri-, tetra-, penta- and hexanucleotide were 10, 6, 4, 3, 3 and 3, respectively.

## RESULTS AND DISCUSSION

The DNA variation in opium poppy is limited because of the moderate genetic diversity existing within the genus *Papaver*, especially within released cultivars and cultivated genotypes. The special status of opium poppy as producer of drug alkaloids decreases interest of scientists, breeders and growers because of the serious limitations and

restrictions. It is reflected by weak interest in opium poppy research in comparison to other crops and also by the relatively low number of nucleotide sequences of opium poppy genome deposited in the GenBank® database. This decreases the chance for retrieving tandem repeats *in silico*. Therefore, the possibility to differentiate opium poppy cultivars, lines and genetic resource accessions using polymorphism in revealed microsatellites is also limited. In total, 450 nucleotide sequences of opium poppy (*P. somniferum* L.) DNA were available in the GenBank® database at the time of analysis (May 25, 2015). Altogether 538 SSRs containing mono-, di-, tri-, tetra-, penta- and hexanucleotide tandem repeat motifs were identified (Figure 1).

Mononucleotide tandem repeats are generally prevalent over others within eukaryotic genomes (Sharma *et al*. 2007) and also the most abundant in opium poppy in our analysis. Their frequencies were A/T (56.18%), T/A (42.63%) and C/G (1.19%). The most frequented were mononucleotides containing 10 tandem repeats (Figure 2). After excluding mononucleotide repeats, the most frequent were trinucleotide repeats with a frequency of 27.50%, followed by tetranucleotide repeats (11.52%). Trinucleotide repeats were also the most abundant within the genomic DNA (49%) (Çelik *et al*. 2014) and ESTs (38.8%) detected in opium poppy (Şelale *et al*. 2013).

An increasing number of tandem repeats in dinucleotide microsatellites reduced their abundance (Figure 3). The longest dinucleotide microsatellite contained 23 tandem repeats. The most abundant dinucleotide repeats were TA/AT (47.72% of all dinucleotides), followed by AT/TA (34.09%) (Figure 3). This result is different from the others. The AT/TA was the most frequented within the EST-SSRs (Şelale *et al*. 2013) and also genomic DNA (50.4%) of opium poppy (Çelik *et al*. 2014). Morgante and Olivieri (1993) revealed that AT/TA motif was by far the most frequented dinucleotide repeat in 34 analysed plant species. The poly$(AT)_n$ repeat was the most abundant and polymorphic class of SSRs also in the rice genome (Temnykh *et al*. 2001) and in organellar genomes of rice, wheat, maize and sorghum (Rajendrakumar *et al*. 2008).

The most frequented trinucleotide SSRs had four tandem repeats and also repeats consisted of five and six tandem repeats were relatively frequented (Figure 4). The longest trinucleotide microsatellite had 12 tandem repeats. The predominant trinucleotide motif was TTC/GAA with the frequency of 20.95% (Figure 5). It is consistent with the frequency of the same tandem motif within the EST-SSRs

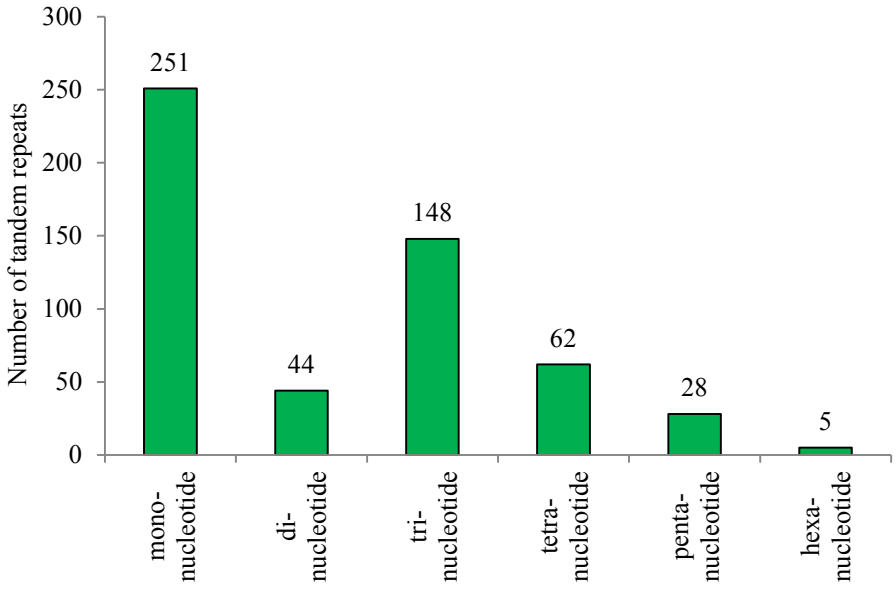


Figure 1. Simple sequence repeat motifs identified in 450 nucleotide sequences of opium poppy

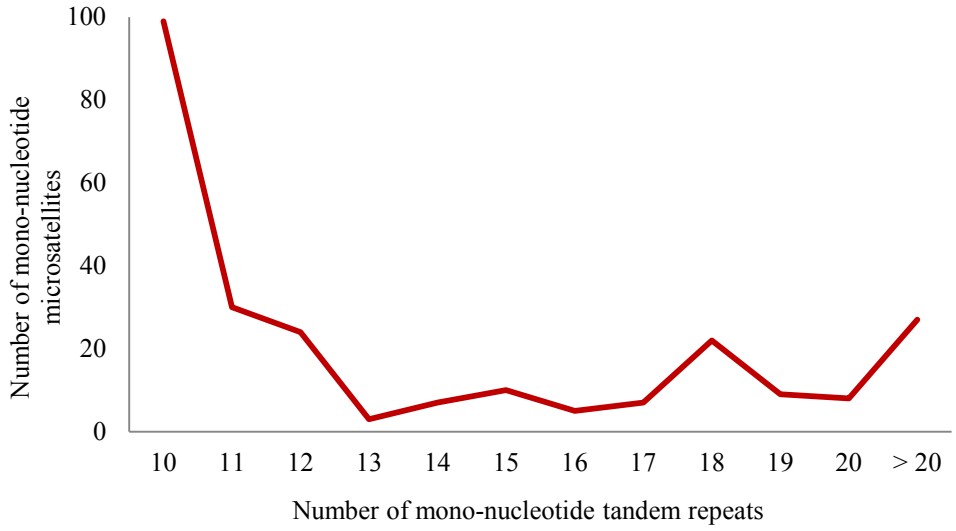


Figure 2. Frequency of length variants of mononucleotide microsatellites

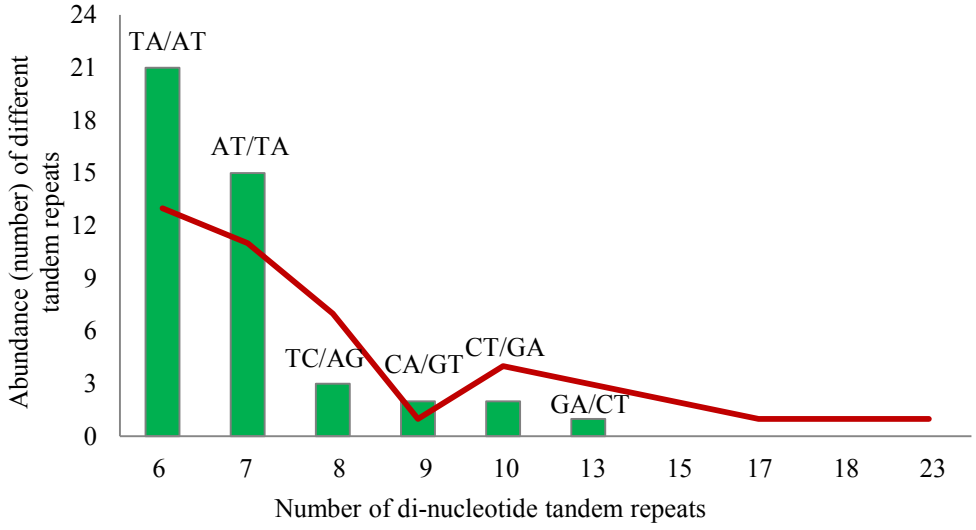


Figure 3. Frequency of length (line) and sequence variants (columns) of different dinucleotide motifs

in opium poppy (Şelale *et al*. 2013) but it does not agree with the frequency of trinucleotide repeats in genomic opium poppy DNA where the most frequented was AAG/TTC (Çelik *et al*. 2014). This is also in contrast with the results of studied 34 plant species where the most frequented trinucleotide was TAT/ATA (Morgante & Olivieri 1993), whereas in our analysis, its frequency was very low (2.7%).

Altogether 62 tetranucleotide microsatellites have been retrieved within the available DNA sequences of opium poppy DNA. Only two length variants were detected, either three or four tetranucleotide tandem repeats. Twenty-three different tetranucleotide motifs have been identified and the most frequented (12.90%) was GGTT/AACC (Figure 6). In contrast, the most frequented tetranucleotide tandem repeat motif identified in the database of opium poppy EST sequences as well as in genomic DNA was AAAT/TTTA (Şelale *et al*. 2013; Çelik *et al*. 2014).

Altogether 28 pentanucleotide tandem repeats with 11 different compositions were retrieved. Number of tandem repeats was three, only one contained four tandem repeats – $(CTTTT/GAAAA)_4$. The predominant (25%) pentanucleotide tandem repeat was TTTTA/AAAAT (Figure 7). Also, this result is different from those detected by Şelale *et al*. (2013) and Çelik *et al*. (2014). They identified the most frequented pentanucleotide tandem repeat AAATA/TTTAT within both the opium poppy database of EST sequences as well as genomic DNA, respectively.

Altogether only 5 hexanucleotide repeats have been identified within 450 nucleotide sequences of opium poppy genomic DNA. They contained only four motifs: $(TCAGTT)_3$, $(TTTATA)_3$, $(AATCAA)_3$ and $(CATCTC)_3$. The most frequented hexanucleotide tandem repeat in opium poppy genomic DNA detected by Çelik *et al*. (2014) was AAAAAT/TTTTTA, not found in our study.

The best confrontation of results obtained in our *in silico* analysis could be with the first and the only one published report about the development of genomic SSR markers in opium poppy published by Çelik *et al*. (2014). They sequenced the opium poppy genomic DNA using the pyrosequencing technology and examined the DNA sequences for the presence of SSRs. Tri- and tetranucleotide repeats were the most abundant, accounting for 49.0% and 27.9% of all SSRs, respectively. Only this finding corresponds with us, all others were different. The most abundant di-, tri-, tetra-, penta- and hexanucleotide repeats in their study were AT/TA, AAG/TT, AAAT/TTTA, AAATA/TTTAT and AAAAAT/TTTTTA, respectively. However, both the studies on the opium poppy genomic DNA (Çelik *et al*. 2014) and the genic DNA (Şelale *et al*. 2013) revealed microsatellites useful for diversity analysis

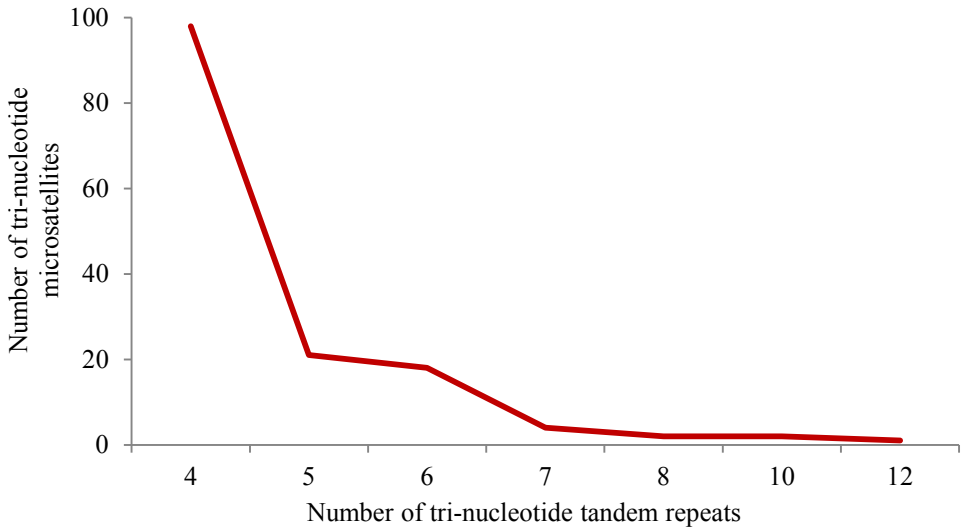


Figure 4. Frequency of length variants of trinucleotide tandem repeats

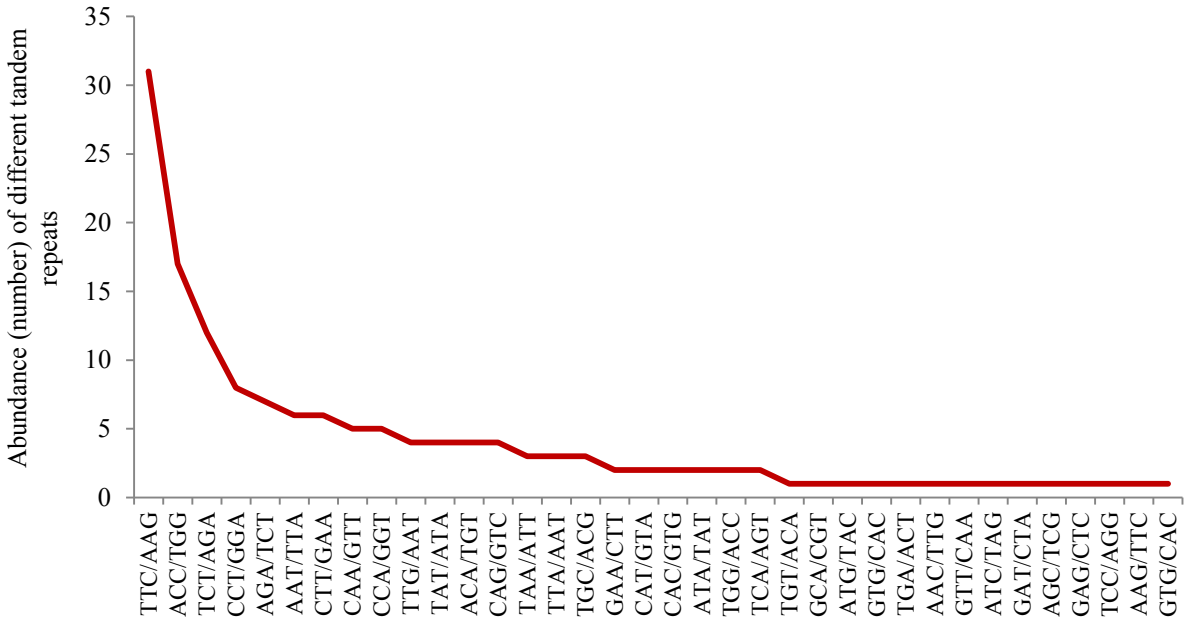


Figure 5. Frequency of sequence variants of different trinucleotide motifs

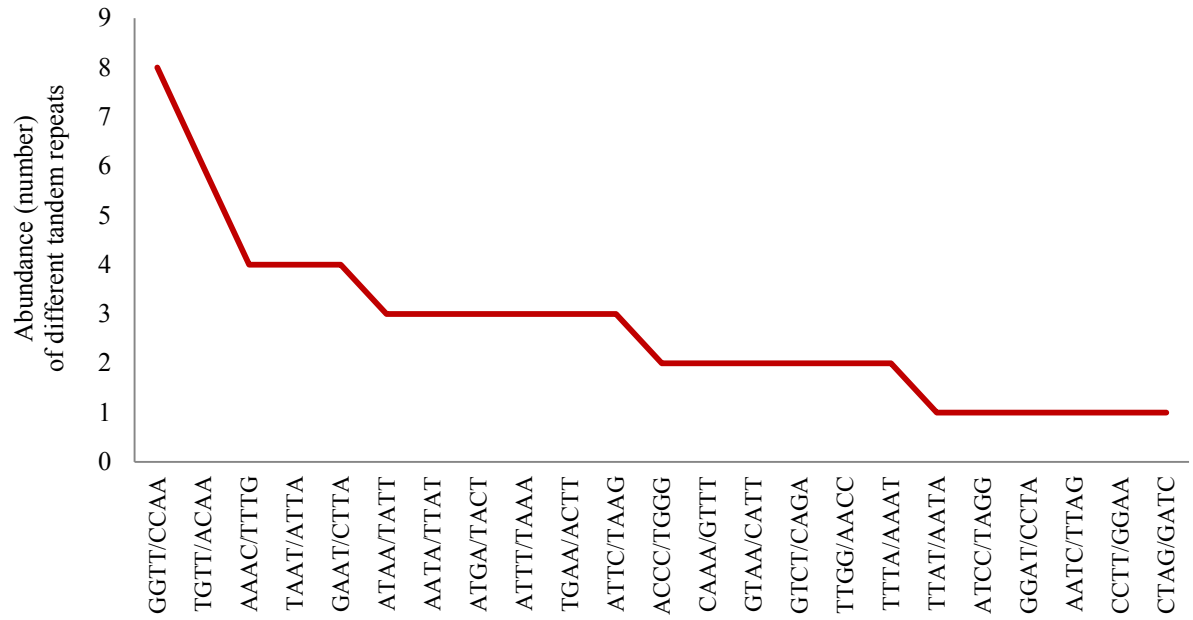


Figure 6. Frequency of sequence variants of different tetranucleotide tandem repeats

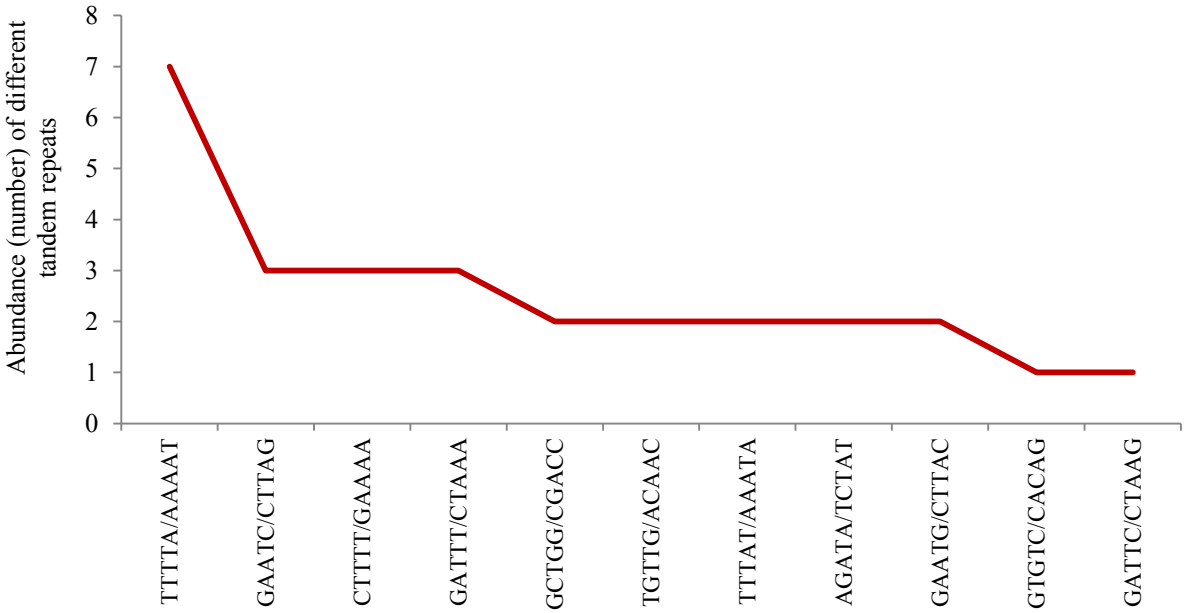


Figure 7. Frequency of sequence variants of different pentanucleotide tandem repeats

and differentiation within cultivars, landraces, and accessions. Even though the other motifs of tandem repeats were the most abundant in our study in comparison to the two mentioned earlier, it is reasonable to assume that microsatellites that are able to generate polymorphism between commercial opium poppy cultivars will be found.

## CONCLUSIONS

The amount of genomic DNA sequences of opium poppy (*P. somniferum* L.) available in public database is limited. Only 450 DNA sequences of opium poppy were available at the time of this study. Altogether 538 microsatellites have been amongst them. In addition to predominant mononucleotide repeats, microsatellites containing di-, tri-, tetra-, penta- and hexanucleotide tandem repeats were also identified. The most abundant were trinucleotide motifs and the most frequent motifs within each group of tandem repeats were TA/AT, TTC/GAA, GGTT/AACC and TTTTA/TAAAA. The *in silico* analysis is probably the simplest way to obtain microsatellite sequences for following designing of relevant primers and also for differentiation studies in the opium poppy.

***Acknowledgements.*** This study was supported by the grants UGA VII/16/2014 from the Constantine the Philosopher University in Nitra and ITMS 26210120039 Systems biology for protection, reproduction and use of plant resources of Slovakia from the Operational Program Research and Development co-financed from the European Union Fund for Regional Development.

## REFERENCES

ACHARYA, H. – SHARMA, V. 2009. Molecular characterization of opium poppy (*Papaver somniferum*) germplasm. In *American Journal of Infection Diseases*, vol. *5*, pp. 148–153.

BENSON, D.A. – CAVANAUGH, M. – CLARK, K. – KARSCH-MIZRACHI, I. – LIPMAN, D.J. – OSTEL, J. – SAYERS E.W. 2013. GenBank. In *Nucleic Acids Research*, vol. *41*, pp. D36–D42. DOI: 10.1093/nar/gks1195.

ÇELIK, I. – GÜLTEKIN, V. – ALLMER, J. – DOĞANLAR, S. – FRARY, A. 2014. Development of genomic simple sequence repeat markers in opium poppy by next-generation sequencing. In *Molecular Breeding*, vol. *34*, pp. 323–334. DOI: 10.1007/s11032-014-0036-0.

DA MAIA, L.C. – PALMIERI, D.A. – DE SOUZA, V.Q. – KOPP, M.M. – DE CARVALHO, F.I.F. – DE OLIVEIRA, A.C. 2008. SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. In *Internation-*

*al Journal of Plant Genomics*, Article ID 412696. DOI:10.1155/2008/412696.

DITTBRENNER, A. – LOHWASSER, U. – MOCK, H.P. – BÖRNER, A. 2008. Molecular and phytochemical studies of Papaver somniferum in the context of intraspecific classification. In GROENDIJK-WILDERS, N. – ALEXANDER, C. – VAN DEN BERG, R.G. – HETTERSCHEID, W.L.A. (Eds). *Proceedings of the 5th International Symposium on the Taxonomy of Cultivated Plants*. Wageningen : Netherlands, October 15 – 19, 2007. (Series: Acta Horticulturae, Vol. 799) Leuven: ISHS (2008) 81–88. DOI: 10.17660/ActaHortic.2008.799.9.

GUPTA, P.K. – RUSTGI, S. – SHARMA, S. – SINGH, S. – KUMAR, N. – BALYAN, H.S. 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. In *Molecular Genetics and Genomics*, vol. *270*, pp. 315–323. DOI: 10.1007/s00438-003-0921-4.

GUPTA, P.K. – VARSHNEY, R.K. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. In *Euphytica*, vol. *113*, pp. 163–185. DOI: 10.1023/A:1003910819967.

GÜÇLÜ, G.B. – GÜRKÖK, T. – KOYUNCU, M. – ARSLAN, N. – PARMAKSIZ, İ. 2014. Genetic characterization of Turkish commercial opium poppy (*Papaver somniferum* L.) cultivars using ISSR and SSR markers. In *Journal of New Results in Science*, vol. *7*, pp. 48–57.

HARI, S. – SHARMA, A. – SHARMA, V. 2009. Molecular characterization of opium poppy (*Papaver somniferum*) germplasm. In *American Journal of Infection Diseases*, vol. *2*, pp. 155–160.

LEE, E.J. – HWANG, I.K. – KIM, N.Y. – LEE, K.L. – HAN, M.S. – LEE, Y.H. – KIM, M.Y. – YANG, M.S. 2010. An assessment of the utility of universal and specific genetic markers for opium poppy identification. In *Journal of Forensic Sciences*, vol. *55*, pp. 1002–1008. DOI: 10.1111/j.1556-4029.2010.01423.x.

LEE, E.J. – JIN, G.J. – LEE, K.L. – HAN, M.S. – LEE, Y.H. – YANG, M.S. 2011. Exploiting expressed sequence tag database for the development of gene-derived simple sequence repeat markers in opium poppy (*Papaver somniferum* L.) for forensic applications. In *Journal of Forensic Sciences*, vol. *56*, pp. 1131–1135. DOI: 10.1111/j.1556-4029.2011.01810.x.

LITT, M. – LUTY, J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. In *American Journal of Human Genetics*, vol. *44*, pp. 388–396.

MORGANTE, M. – OLIVIERI, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. In *Plant Journal*, vol. *3*, pp. 175–182.

PARMAKSIZ, I. – OZCAN, S. 2011. Morphological, chemical, and molecular analyses of Turkish Papaver accessions (Sect Oxytona). In *Turkish Journal of Botany*, vol. *35*, pp. 1–16. DOI: 10.3906/bot-1003-39.

RAJENDRAKUMAR, P. – BISWAL, A.K. – BALACHANDRAN, S.M. – SUNDARAM, R.M. 2008. *In silico* analysis of microsatellites in organellar genomes of major cereals for understanding their phylogenetic relationships. In *In Silico Biology*, vol. *8*, pp. 87–104.

SARMAH, R. – SAHU, J. – DEHURY, B. – SARMA, K. – SAHOO, S. – SAHU, M. – BAROOAH, M. – SEN, P. – MODI, M.K. 2012. ESMP: A high-throughput computational pipeline for mining SSR markers from ESTs. In *Bioinformation*, vol. *8*, pp. 206–208. DOI: 10.6026/97320630008206.

SAUNDERS, J.A. – PEDRONI, M.J. – PENROSE, L.D.J. – FIST, A.J. 2001. AFLP analysis of opium poppy. In *Crop Science*, vol. *41*, pp. 1596–1601.

ŞELALE, H. – ÇELIK, I. – GÜLTEKIN, V. – ALLMER, J. – DOĞANLAR, S. – FRARY, A. 2013. Development of EST – SSR markers for diversity and breeding studies in opium poppy. In *Plant Breeding*, vol. *132*, pp. 344–351. DOI: 10.1111/pbr.12059.

SHARMA, P.C. – GROVER, A. – KAHL, G. 2007. Mining microsatellites in eukaryotic genomes. In *Trends in Biotechnology*, vol. *25*, pp. 490–498. DOI:10.1016/j.tibtech.2007.07.013.

SUÁREZ, M.C. – BERNAL, A. – GUTIÉRREZ, J. – TOHME, J. – FREGENE, M. 2000. Developing expressed sequence tags (ESTs) from polymorphic transcript-derived fragments (TDFs) in cassava (*Manihot esculenta* Crantz). In *Genome*, vol. *43*, pp. 62–67. DOI: 10.1139/g99-073.

TANG, J. – BALDWIN, S.J. – JACOBS, J.M.E. – VAN DER LINDEN, C.G. – VOORRIPS, R.E. – LEUNISSEN, J.A.M. – VAN ECK, H. – VOSMAN, B. 2008. Large-scale identification of polymorphic microsatellites using an in silico approach. In *BMC Bioinformatics*, vol. *9*, pp. 374. DOI:10.1186/1471-2105-9-374.

TEMNYKH, S. – DECLERCK, G. – LUKASHOVA, A. – LIPOVICH, L. – CARTINHOUR, S. – MCCOUCH, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. In *Genome Research*, vol. *11*, pp. 1441–1452. DOI: 10.1101/gr.184001.

THIEL, T. – MICHALEK, W. – VARSHNEY, R.K. – GRANER, A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). In *Theoretical and Applied Genetics*, vol. *106*, pp. 411–422. DOI: 10.1007/s00122-002-1031-0.

Wang, X. – LU, P. – LUO, Z. 2013. GMATo: A novel tool for the identification and analysis of microsatellites in large genomes. In *Bioinformation*, vol. *9*, pp. 541–544. DOI: 10.6026/97320630009541.

*Received: November 12, 2015*