BIOLOGIA PLANTARUM 52 (1): 149-152, 2008

BRIEF COMMUNICATION

Limited variability of CTG/CAG repeats in *Lycopersicon* nuclear DNA

D. DIMITROVA[1], O. GEORGIEV[2], C. VALKOVA[3], B. ATANASSOVA[4]* and L. KARAGYOZOV[5]

Institute of Biophysics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str. Bl. 21, 1113 Sofia, Bulgaria[1]
Institut für Molekularbiologie II, Universität Zurich, Winterthurer Str. 190, CH-8057 Zurich, Switzerland[2]
Fritz Liepmann Institut für Altersforschung, Beutenberger Str. 11, D-07745 Jena, Germany[3]
Institute of Genetics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria[4]
Friedrich-Schiller Universität Jena, Drackendörfer Str. 1, D-07747 Jena, Germany[5]

Abstract

Seven clones containing $(CTG)_n/(CAG)_n$ repeats ($n \geq 4$) were isolated by screening *Lycopersicon esculentum* genomic DNA. Four of the clones contained more than one simple sequence repeat (SSR). The SSRs were analyzed in several *L. esculentum* cultivars after polymerase chain reaction (PCR) amplification. No length variations were observed, suggesting considerable locus stability. Five clones are from transcribed regions, which might explain the lack of cultivar variations. However the conservation of CTG repeats was limited as differences in some transcribed loci were registered between *L. pennellii* and other *Lycopersicon* species. It is noted that in *Lycopersicon* trinucleotide repeat variation might be used for species identification.

Additional key words: tomato, transcribed DNA, trinucleotide simple sequence repeats.

Genomes of all eukaryotes contain regions of DNA, which are composed of short units, mostly 2, 3 or 4 nucleotides long, repeated several times one after another. These simple sequence repeats (SSRs), also known as microsatellites, frequently occur as highly dispersed elements, evenly spread throughout the genome. As a rule SSRs are highly polymorphic as the number of their units is subject to variation. Due to these features, and to the simplicity of allele identification, SSRs now are widely used as genetic markers in animals. The usefulness of SSRs based markers in plants is well documented (for a review see Rakoczy-Trojanowska and Bolibok 2004, Dikshit *et al*. 2007). Approaches are also developed for retrieval of microsatellite markers from selected regions of the plant genome (Požárková *et al*. 2002).

Most SSRs occur in non-coding regions, predominantly outside genes and in introns. Trinucleotide repeats, however, should be singled out as in animals and plants they are present also in exons (Toth *et al.* 2000, Cardle *et al*. 2000). During the last decade a number of investigations of microsatellite polymorphism and genomic distribution were performed in tomato. It was shown that many SSRs in tomato have biased distribution and are present mainly in the centromere regions (Areshchenkova and Ganal 1999, Yang *et al*. 2005). This suggests that identification of markers covering the entire genome by screening of random DNA clones would be highly ineffective.

Areshchenkova and Ganal (2002) made attempts to achieve more uniform distribution of microsatellite markers for genomic mapping purposes. One approach was to screen genomic libraries enriched for single-copy sequences. However, all markers isolated in this way also mapped into centromeric regions. Another approach was to develop tomato microsatellite markers by exploiting the tomato EST database. These markers were located in euchromatic regions, contained predominantly AT repeats, and some of them were polymorphic in a set of *L. esculentum* cultivars. He *et al*. (2003) also searched the GenBank database and characterized more microsatellite markers, mostly from expressed sequences. Frary *et al.* (2005) later employed the same strategy and now more than 100 markers derived from ESTs are placed on the genetic map of tomato.

Received 12 December 2006, *accepted* 15 June 2007.

Abbreviations: EST - expressed sequence tags; PCR - polymerase chain reaction; SDS - sodium dodecyl sulfate; 1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7.0; SSR - simple sequence repeat.

Acknowledgments: This work has been supported by grant CRP/BUL95-02 to L.K. from ICGEB, Trieste, Italy.

* Author for correspondence; fax +359 2 978 55 16, e-mail: bistra_a@yahoo.com

The aim of our work was to screen genomic DNA clones for genetic markers, presumably distributed evenly throughout the genome. We focused our attention in particular on the trinucleotide CTG/CAG repeats, which might be present also in exons.

The following species were used for DNA isolation: *Lycopersicon esculentum* Mill. (cultivars VF36, B-317, Start 24, Ideal, Mercurii, Druzhba, and Spartak), *L. esculentum* Mill. var. *cerasiforme* (Dun) A. Gray (WVA 106), *L. pimpinellifolium* (L.) Mill. (LA 121), *L. pennellii* Correll (LA 716). The plants were grown in greenhouse (day/night temperature of 16/27 °C, maximum irradiance 1 000 μmol m^{-2} s^{-1}, air humidity 65 - 75 %)

For isolation of nuclei and DNA young expanded leaves from 2-month-old plants were harvested. Nuclei were isolated by modification of a published procedure (Hewish and Burgoyne 1973). The crude nuclear pellet was suspended in 1 % SDS, 100 mM Tris-HCl (pH 8.0), 20 mM EDTA. The DNA solution was cleared and potassium acetate (0.5 volumes, 3 M, pH 5.5) was added. The resulting precipitate was removed and DNA was purified on a *Qiagen* (Hilden, Germany) column.

Previous estimates have shown that in the tomato genome most abundant are the (CT)n and (CA)n SSRs, which occur on average once per 1.2 Mb (Broun and Tanksley 1996). We enriched the tomato genomic fragments for the rare (CTG)n repeats by hybridization to immobilized oligonucleotides as described earlier (Karagyozov *et al.* 1993). Briefly, DNA from *L. esculentum* cv. Start 24 was digested with Taq I restriction endonuclease and the fragments were ligated to an adapter, formed by the oligonucleotides: RX 24 (AGCACTCTGCAGCCTCTAGATCTC) and RX 11 (CGGAGATCTAG). The genomic fragments were amplified (primer RX24), heat denatured and hybridized (0.35 cm^3 5× SSC, 5 % SDS, 50 mM sodium phosphate, pH 7.0, 50 °C, 16 h) to 10 pmol $(CTG)_{10}$ oligonucleotides bound to a nylon membrane piece (2 mm). The membrane was washed and heated in 0.5 % SDS. The detached strands were amplified (primer RX 24), digested with Xba I restriction endonuclease and cloned. Clones were screened by hybridization with 5'-end labeled $(CTG)_{10}$ and the positive clones were sequenced (both strands). Sequence data for seven genomic clones with (CTG)n repeats ($n \geq 4$) were deposited with the *EMBL/ GenBank* data libraries (see Table 1 for Accession numbers). The microsatellite in Dol_1 contains an uninterrupted stretch of $(TGC)_{18}$, so it is one of the longest CAG/CTG repeat registered in plants. It should be noted that four of the clones have additional simple sequence repeats. In clone Dol_3 a hexanucleotide repeat is present, the other clones carry additional trinucleotide repeats. No mononucleotide ($n \geq 12$) or dinucleotide ($n \geq 6$) repeats are present in cloned DNA.

Primers for the amplification of the SSRs were selected (Table 1) and genomic DNA from different sources was amplified (System 3 buffer, *Boehringer*, Ingelheim, Germany, 2.25 mM $MgCl_2$ plus detergents). PCR products were compared by denaturing DNA electrophoresis in polyacrylamide gels. Results showed a marked stability of the CTG repeats as no length polymorphism was observed with DNA from all *L. esculentum* cultivars tested (Table 1).

This result was quite unexpected. The triplet repeats, and in particular the CTG/CAG repeat, are unusually susceptible to genetic change, conclusion supported by investigations on the mammalian genome and also by model experiments with DNA constructs integrated into *Saccharomyces* or in *E. coli* (Schweitzer *et al*. 2001, Pearson *et al*. 2005, Bichara *et al.* 2006). In view of these data the limited CTG repeat variability, which we observed in tomato, required further clarification.

It is well documented that triplet repeats are present in exons, coding mostly for hydrophilic amino acid, particularly glutamine (Katti *et al*. 2000). In this case, microsatellite expansions or contractions may alter unfavorably the plant phenotype and would be selected against. Therefore we tested the possible involvement of SSR sites in transcription. Total RNA was isolated from *L. esculentum* leaves (1 g) according to Chomczynski and Sacchi (1987). RNA was treated with DNase (RNase free, *Stratagene*, USA) and reverse transcriptase (RT)-PCR was carried out using *M-MLV RT* (*Gibco-BRL*, Paisley, UK) and $(dN)_6$ oligonucleotides. Amplification of the cDNA was performed using specific primers (Table 1) and the RT-PCR products were size fractionated, transferred to filter and hybridized with labeled $(CTG)_{10}$ to confirm specificity. Results showed that specific products were synthesized using Dol_11 and Dol_15pr primers.

Thereupon we searched *GenBank* for expressed sequences that are homologous to the CTG clones. Result confirmed our findings and showed that additionally clones Dol_P4, Dol_4 and Dol_14 also contain regions, homologous to *L. esculentum* and *S. tuberosum* EST sequences (Table 1). The proportion of expressed clones is large, five out of seven. This parallels the predominant location of CTG/CAG repeats in exons, which is observed in the mammalian genome (O'Donovan and Guy 1997).

Earlier reports showed that SSR variations in *Lycopersicon* are not very frequent between cultivars, but are abundant between species (Smulders *et al*. 1997). So, we compared amplification products of DNA from several *Lycopersicon* species (*L. esculentum, L. esculentum* var. *cerasiforme, L. pimpinellifolium* and *L. pennellii*). Only *L. pennellii* PCR products showed in some instances size differences with products from other tomato species (Table 1).

The results show that inter-species differences in PCR product lengths do not depend directly on the transcription (or translation) of the triplet repeat. The SSR in clones Dol_P4 and Dol_15pr are transcribed and conserved between *Lycopersicon* species, however, the SSRs in clones Dol_4, Dol_11 and Dol_14 are transcribed and variable. Clone Dol_3 (variable) codes for a hypothetical protein with 138 amino acids (assigned *GenBank* accession AAB96938).

Table 1. Characteristics of clones containing CTG/CAG repeats isolated from *L. esculentum* genomic library. Amplification and polymorphism of the sites with CTG/CAG repeats. (CTG)n/(CAG)n = (TGC)n/(GCA)n = (GCT)n/(AGC)n.

Clone name (size) Accession	Position of SSR in the clone	SSR sequence	SSR amplification primers (position)	Species variance	Homology to EST (region)
Dol_1 (437 bp) AF040998	148-228	$(tgc)_{18}c(tgc)_3c(tgc)_3$	101-123 303-282	no	none
Dol_P4 (330 bp) G36970	203-220 230-244	$(gctgaa)_3$ $(gct)_5$	120-142 297-272	no	(1-330) *S. tuberosum* CK717254
Dol_15pr (250 bp) G36974	202-225 235-258 267-296 340-354	$(gct)_4(gtt)_4$ $(gct)_3(gtt)_5$ $(tgc)_4atctgt(tgc)_4$ $(ctg)_5$	65-82 214-193	no	(3-250) *L. esculentum* BI924965
Dol_3 (626 bp) AF041411	244-267	$(agc)_3atg(agc)_4$	38-61 288-266	yes	none Codes hypothetical protein AAB96938
Dol_4 (443 bp) AF042094	33-50 77-91	$(gct)_6$ $(tgc)_5$	154-177 407-384	yes	(1-47) and (150-443) *S. tuberosum* CK267379
Dol_11 (248 bp) G36971	61-75 210-224	$(cag)_5$ $(cca)_5$	12-29 232-213	yes	(1-248) *L. esculentum* BI925909
Dol_14 (450 bp) G36972	148-165	$(cag)_6$	35-57 125-104	yes	(39-450) *L. esculentum* AI772553

It should be noted that in all loci (Dol_3, Dol_4, Dol_11 and Dol_14) which manifest PCR length variations between species, the microsatellite repeats are clustered (see Table 1). This suggests that PCR genomic screening in regions with neighboring trinucleotide repeats might be useful for finding differences in *Lycopersicon*.

References

Areshchenkova, T., Ganal, M.W.: Long tomato microsatellites are predominantly associated with centromeric regions. - Genome **42**: 536-544, 1999.

Areshchenkova, T., Ganal, M.W.: Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. - Theor. appl. Genet. **104**: 229-235, 2002.

Bichara, M., Wagner, J., Lambert, I.B.: Mechanisms of tandem repeat instability in bacteria. - Mutat. Res. **598**: 144-163, 2006.

Broun, P., Tanksley, S.D.: Characterization and genetic mapping of simple repeat sequences in the tomato genome. - Mol. Gen. Genet. **250**: 39-49, 1996.

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

Chomczynski, P., Sacchi, N.: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. - Anal. Biochem. **162**: 156-159, 1987.

Dikshit, H.K., Jhang, T., Singh N.K., Koundal, K.R., Bansal, K.C., Chandra, N., Tickoo, J.L., Sharma, T.R.: Genetic differentiation of *Vigna* species by RAPD, URP and SSR markers. - Biol. Plant. **51**: 451-457, 2007.

Frary, A., Xu, Y., Liu, J., Mitchell, S., Tedeschi, E., Tanksley, S.: Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. - Theor. appl. Genet. **111**: 291-312, 2005.

He, C., Poisa, V., Yu, K.: Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. - Theor. appl. Genet. **106**: 363-373, 2003.

Hewish, D.R., Burgoyne, L.A.: The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. - Biochem. biophys. Res. Commun. **52**: 504-510, 1973.

Karagyozov, L., Kalcheva, I.D., Chapman, V.: Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. - Nucl. Acids Res. **21**: 3911-3912, 1993.

Katti, M.V., Sami-Subbu, R., Ranjekar, P.K., Gupta, V.: Amino acid repeat patterns in protein sequences: their diversity and structural-functional implications. - Protein Sci. **9**: 1203-1209, 2000.

O'Donovan, M.C., Guy, C.: Bias in the genomic distribution of CAG and CTG trinucleotide repeats. - Amer. J. Med. Genet. **74**: 62-64, 1997.

Pearson, C.E., Edamura K.N., Cleary J.D.: Repeat instability: mechanisms of dynamic mutations. - Nat. Rev. Genet. **6**: 729-742, 2005.

Požárková, D., Koblížková, A., Román, B., Torres, A.M., Lucretti, S., Lysák, M., Doležel, J., Macas, J.: Development and characterization of microsatellite markers from chromosome 1-specific DNA libraries of *Vicia faba*. - Biol. Plant. **45**: 337-345, 2002.

Rakoczy-Trojanowska, M., Bolibok, H.: Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. - Cell mol. Biol. Lett. **9**: 221-

238, 2004.
Schweitzer, J.K., Reinke, S.S., Livingston, D.M.: Meiotic alterations in CAG repeat tracts. - Genetics **159**: 1861-1865, 2001.
Smulders, M.J.M., Bredemeijer, G., Rus-Kortekaas, W., Arens, P., Vosman, B.: Use of short SSR from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. - Theor. appl. Genet. **97**: 264-272, 1997.
Toth, G., Gaspari, Z., Jurka, J.: SSR in different eukaryotic genomes: survey and an analysis. - Genome Res **10**: 967-981, 2000.
Yang, T.-J., Lee, S., Chang, S.-B., Yu, Y., De Jong, H., Wing, R.A.: In-depth sequence analysis of the tomato chromosome 12 centromeric region: identification of a large CAA block and characterization of pericentromere retrotranposons. - Chromosoma **114**: 103-117, 2005.

Quest-Ritson, C.: **Climbing Roses of the World**. - Timber Press, Portland - Cambridge 2003. 306 pp., 200 colour photographs. ISBN 0-88192-563-2.

The rhodologic book is rather extensive, whether it concerns works about wild roses and their taxonomy, nomenclature, morphology, physiology, fytogeography, *etc*., or literature dealing with cultivated garden roses that have accompanied mankind for thousands of years. With the exception of only a few works, *e.g*. Williams (1913), Thomas (1965), and Stevens (2003), there probably do not exist too many of such highly specialized publications on rhodology, devoted only to climbing roses, as this one written by Quest-Ritson.

In the 14 chapters he is partly trying to put some order into the maybe chaotic opinions concerning classification and origin of such an inorganic group as the "climbing roses", and he partly describes their history with all its peripeties he managed to research, and deduct, which is definitely a more successful part.

The term "climbing roses" is a purely horticultural, technical term, and to a certain extent a confusing one. Roses in their enormous diversity and polymorphism are not bushes, but leaning climbers most often. They lean on their support with the help of their thorns (in particular, the hook-like thorns) and short flower-bearing branchlets. It is possible to insert the equals sign between the terms "climber" and "climbing plant". Even our briar roses are leaning climbers, *i.e.* they are climbing roses as well as the majority of the wild and cultivated roses. Therefore, the horticultural term "climbing roses" identifies those (cultivated) roses that grow significantly longer sparsely branched shoots that can be tied to a trellis or to another kind of support soon - or they can be left to climb into treetops - or they can creep on the ground as ground-covering roses.

Quest-Ritson has searched the past of those cultivated roses that have been identified as the climbing ones and he acquainted the readers (or, to be more exact, students) of this book with potential former sources of the existing cultivars. He presents Chinese roses from the *Synstylae* section, including the remarkable and problematic gigantic rose, *Rosa gigantea* and the proper Chinese rose (*Rosa chinensis*). The European adherent-style roses could be a similar source: the deciduous *Rosa arvensis* and the evergreen *R. sempervirens.* The long, stoloniferous shoots are characteristic for the *Synstylae* section regardless of the continent from which they come. Therefore, another chapter is dedicated to the American roses from this group; the basis of this group forms the American prairie rose, *Rosa setigera.* However, the most essential influence on the climbing and multifloreous roses development has *Rosa multiflora* and its relative, *R. wichuraiana* (it is mentioned in the book under the frequent name, *R. wichurana).*

Attention has been paid not only to the early cultivated roses with a part of "east Asian blond", to noisettes and to the first tea roses (this name has nothing in common with the colour of the roses but with their scent!) - but to the recent climbing roses as well, *i.e*. to the descendants of the first patented rose in the world, the New Dawn cultivar (1930), to the world famous hybrids of the German firm Kordes and, in particular, to the newest, spontaneous "long-shooted" mutations, the so-called sports of common bed roses, polyanthes or tea roses, that are usually named after the original cultivar with the appellation "Climbing". The last listing includes a summary of large-flowered climbing roses listed according to the sequence of their breeding at the end (not only) of the 20th century and their introduction to the market by the most important rose growers in the whole world. Altogether about 2000 thousand cultivars (cultivar names and synonyms) are mentioned in this book in various connections and 70 botanical taxa on the level of species or interspecies or on the level of named primary hybrids.

Of course it was impossible to picture all the taxa mentioned; for that matter the book is neither an atlas nor an aid for determination of roses, but it is more a historically tuned genealogy of an artificial, nevertheless practical horticultural group of "climbing roses". It will serve theoretical rhodologists and similar specialists more than that to the broad public, who is interested more in the beauty and scent of roses than in climbing roses history and origin.

In any case this is a useful book.

V. VĚTVIČKA (*Praha*)