بررسی ESTهای ثبت شده در بانک ژن برای تشخیص و شناسایی EST-SSR در انگور

عبدالکریم زارعی*[۱]، عزیز ابراهیمی[۲]

[۱*] گروه بیوتکنولوژی، دانشکده کشاورزی، دانشگاه جهرم، جهرم، ایران

[۲] گروه جنگلداری و منابع طبیعی، دانشگاه پوردو، وست لافایات، ایالات متحده آمریکا

*نویسنده مسئول: zarei@jahrom.ac.ir

چکیده

توالی‌های نشانه بیان شونده (ESTs) منابع مهمی برای جنبه‌های مختلف مطالعات ژنومی از قبیل کشف ژن، بیان ژن، ژنومیکس مقایسه‌ای، و توسعه نشانگرهای مولکولی می‌باشند. در مطالعه حاضر، ESTهای گیاه انگور (*Vitis vinifera*) ثبت شده در بانک ژن NCBI مورد ارزیابی قرار گرفتند تا توالی‌های تکراری موجود در آن‌ها مورد بررسی قرار گیرد تا بتوان نشانگرهای EST-SSR جدیدی را شناسایی کرد. در مجموع ۵۱۱۶۹۰ توالی مورد بررسی قرار گرفت که از بین آن‌ها ۹۲۳۶ توالی (میانگین ۱/۸٪ از توالی‌های ثبت شده) حداقل دارای یک توالی تکراری بودند. از بین موتیف‌های مختلفی که مورد آنالیز قرار گرفت، تکرارهای دو نوکلئوتیدی با دارا بودن بیش از ۴۵٪ توالی‌های تکراری، فراوان‌ترین تکرار بودند و بعد از آن به ترتیب تکرارهای سه نوکلئوتیدی (۳۳٪) و چهار نوکلئوتیدی (۲۲٪) در رتبه‌های بعدی قرار داشتند. در بین تکرارهای دو نوکلئوتیدی هم توالی AG/GA/CT/TC با حدود ۴۶٪ از این توالی‌ها، فراوان‌ترین توالی بود. بر اساس توالی‌های تکراری ساده شناسایی شده، آغازگرهای مناسب برای تکثیر برخی از EST-SSRهای عملکردی طراحی گردید. ارزیابی این آغازگرها روی برخی از ارقام انگور کشور در حال بررسی می‌باشد.

***کلمات کلیدی:* انگور، نشانگرهای مولکولی، EST، SSR**

مقدمه

انگور (*Vitis vinifera*) یکی از مهم‌ترین گیاهان باغی موجود در دنیا می‌باشد. تولید سالیانه این میوه بیش از ۶۳ میلیون تن بوده، که بیشترین تولید در بین میوه‌های مناطق معتدله می‌باشد و بعد از مرکبات و موز بیشترین تولید میوه دنیا را به خود اختصاص داده است. به علت اهمیت اقتصادی این میوه، مطالعات ژنتیکی بسیار گسترده‌ای روی آن انجام گرفته است و اولین درخت میوه‌ای بود که ژنوم آن به‌طور کامل توالی یابی گردید (Velasco *et al.*, 2007). بعلاوه مطالعات بیان ژن متنوعی در مورد این میوه انجام گرفته است و توالی‌های بسیار زیادی مربوط به رونوشت‌های بافت‌های مختلف و مراحل نموی مختلف این گیاه در بانک داده‌های ژنومی ثبت گردیده است.

توالی‌های تکراری ساده (SSRs)، تکرارهای ۱ تا ۷ نوکلئوتیدی در ژنوم می‌باشند که به‌طور گسترده در موجودات یوکاریوتی پخش‌شده‌اند و تنوع طول بسیار زیادی از خود نشان می‌دهند، به همین علت اهمیت زیادی به‌عنوان نشانگرهای مولکولی و بررسی تنوع ژنوم موجودات مختلف دارا می‌باشند. این نوع توالی به طرق مختلف گروه‌بندی می‌شوند: از جمله تعداد واحدهای تکراری (یک، دو، سه، چهار، پنج و شش نوکلئوتیدی)، نحوهٔ قرار گرفتن نوکلئوتیدها در واحدهای تکراری (ساده: فقط از یک نوع واحد تکراری تشکیل شده، مرکب: دارای بیش از یک نوع واحد تکراری، و ناکامل: که دارای واحدهای غیر تکراری در بین واحدهای تکراری می‌باشند) و محل قرار گرفتن در ژنوم: (هسته‌ای، کلروپلاستی و میتوکندریایی). روش سنتی شناسایی این توالی‌ها بر مبنای ساخت کتابخانه‌های ژنومی و هیبریداسیون آن‌ها با توالی‌های تکراری و سپس کلون کردن و توالی یابی انواعی که هیبرید می‌شدند، بود (Edwards *et al.*, 1996). توسعه نشانگرهای SSR با این روش بسیار هزینه‌بر و وقت‌گیر می‌باشد. به علت افزایش مطالعات روی ترانسکریپتوم

انگور، تعداد زیادی از توالی‌های بیان شونده در بانک‌های ژن از قبیل NCBI ثبت گردیده است که می‌توان از آن‌ها به‌عنوان منبع بهتری برای شناسایی SSR های این گیاه استفاده کرد. بانک داده‌های ESTs دارای توالی‌هایی می‌باشند که دارای واحدهای تکراری می‌باشند (Chen *et al.*, 2006). پیشرفت‌های اخیر در روش‌های مبتنی بر کامپیوتر محققان را قادر کرده است تا توالی‌های EST را در بانک‌های داده‌های ژنومی مانند NCBI بررسی کرده و بتوانند EST-SSRs و یا SSRهای مبتنی بر ژن را شناسایی کنند. این روش شناسایی EST-SSR نسبت به روش سنتی بسیار آسان‌تر بوده و می‌توان با هزینه کمتر و کارایی بیشتر توالی‌های تکراری ساده را در گیاهان مختلف شناسایی کرده و در نهایت به توسعه نشانگرهای EST-SSRs اقدام کرد (Wang *et al.*, 2012). استفاده از این نشانگرهای مزایای متعددی نسبت به نشانگرهای SSR دارد از جمله کارایی بالای این نشانگرها برای انتخاب به کمک نشانگر و قابلیت انتقال‌پذیری[1] یعنی استفاده از این نشانگرها برای گونه‌های نزدیک از نظر ژنتیکی را دارا می‌باشد.

در مطالعه حاضر ESTهای موجود در بانک ژن NCBI مورد آنالیز قرار گرفت تا فراوانی و توزیع توالی‌های تکراری ساده در ESTهای ثبت شده در مورد انگور گونه *vinifera* ارزیابی گردد و نشانگرهای EST-SSR مناسب برای مطالعات بعدی این گونه شناسایی و توسعه یابد.

مواد و روش‌ها

در مجموع ۵۱۱۶۹۰ توالی EST مربوط به *V. vinifera* از بانک ژن NCBI در تاریخ ۱۰ دسامبر ۲۰۱۶ گرفته شد. در ابتدا توالی‌های انتهایی و دم پلی A (در صورت وجود) از داده‌ها خام توسط نرم‌افزار EST-Trimmer ((http://pgrc.ipk-gatersleben.de/misa/download/est_trimmer.pl)) حذف گردید و رونوشت‌های هم‌پوشان شناسایی شدند. شاخص تعداد تکرارها برای شناسایی نوع نوکلئوتیدها به شرح زیر بکار برده شد؛ تکرارهای دو نوکلئوتیدی ۸ تا ۴۰ بار، تکرارهای سه نوکلئوتیدی ۷ تا ۳۰ بار، و تکرارهای چهار نوکلئوتیدی ۶ تا ۲۰ بار. فقط تکرارهای کامل گزارش شدند و تکرارهای مرکب (دو یا تعداد بیشتری توالی تکراری متفاوت کنار هم) در نظر گرفته نشدند. بعلاوه حداقل توالی ۱۵ باز غیر تکراری در دو سمت توالی تکراری مدنظر قرار گرفت تا امکان طراحی آغازگر را فراهم کند. خوانش های دارای کانتیگ SSR سرهم بندی شدند و خوانش های تکه‌تکه شده از یک مکان ژنی با استفاده از نرم‌افزار CAP3 بر مبنای p ۹۵ به یک کانتیگ تکی تبدیل شدند (Huang *et al.*, 1999). برای شناسایی نواحی دارای پیچیدگی کمتر، توالی‌های حاصل با سطح یک dustmasker آنالیز شدند (Morgulis *et al.*, 2006) و آغازگرهای مکمل نواحی دو سمت قسمت تکراری توسط نرم‌افزار Primer3 نسخه ۵،۳،۲ طراحی شدند (Untergasser *et al.*, 2012).

نتایج و بحث

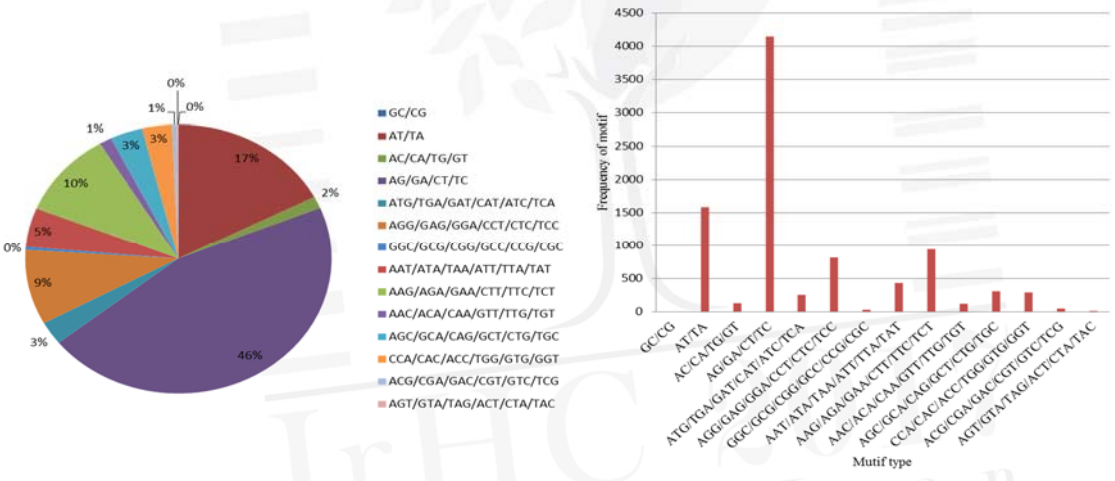
نتایج نشان داد از مجموع ۵۱۱۶۹۰ توالی EST مربوط به بافت‌های مختلف انگور گونه *vinifera* موجود در بانک ژن ۹۴۷۴ قطعه دارای توالی تکراری بودند (جدول ۱). از بین توالی‌های آنالیز شده، ۹۲/۵٪ از آن‌ها تشکیل کانتیگ دادند که بیانگر این است که بیشتر توالی‌های آنالیز شده هم‌پوشان بودند، درحالی‌که ۷/۵٪ از توالی‌ها منحصر به فرد بوده و در مورد آن‌ها توالی هم‌پوشانی مشاهده نشد. بعد از سر هم بندی، گروه‌های غیر تکراری ESTها سرهم شدند و کانتیگ ها و اسکلت‌ها را تشکیل دادند و به‌عنوان توالی‌های EST سرهم شده معرفی شدند. در مجموع ۷۳/۴۸٪ کاهش در غیر تکراری‌ها دیده شد. به عبارت دیگر تعداد توالی‌های EST قبل از جستجو برای نشانگرهای ریز ماهواره‌ای به این میزان کاهش یافت. این نتایج نشان دهنده میزان بالای هم‌پوشانی موجود در توالی‌های رونوشت متعلق به این ژنوم را نشان می‌دهد.

[1]Transferability

نتایج این تحقیق نشان داد که موتیف‌های دی نوکلئوتیدی شایع‌ترین تکرار در بین انواع تکرارهای موجود بودند و بیش از ۴۶٪ از کل عوامل تکراری در رونوشت‌های انگور را تشکیل می‌دهند (تصویر ۱). بعد از نوکلئوتیدهای دوتایی، انواع سه‌تایی و چهارتایی با ۳۳٪ و ۲۲٪ عوامل تکراری به ترتیب در رده‌های بعدی وجود داشتند. نتایج این تحقیق مشابه با گزارشات مربوط به EST-SSR های شناسایی شده در مورد هسته‌دارها می‌باشد (Sorkhe *et al.*, 2016). این محققین گزارش کردند که تکرارهای دو نوکلئوتیدی بیشترین نوع تکرارهای مورد شناسایی در انواع مختلف میوه‌های هسته‌دار می‌باشد. برخلاف نتایج این پژوهش، در مورد زیتون گزارش شده که تکرارهای تک نوکلئوتیدی، با فاصله زیاد (بیش از ۷۷٪ تکرارهای شناسایی شده)، فراوان‌ترین نوع توالی تکراری در این گیاه می‌باشد و بعد از آن به ترتیب تکرارهای سه نوکلئوتیدی (۱۱٪/۸۴) و دو نوکلئوتیدی (۸٪/۶۲) قرار داشتند (Adawy *et al.*, 2013). همچنین گزارش شده است که تکرارهای سه نوکلئوتیدی، فراوان‌ترین نوع تکرارها در گیاه *Elymus sibiricus* L.، و بعد از آن به ترتیب تکرارهای دو و تک نوکلئوتیدی قرار داشتند (Zhou *et al.*, 2015). بنابراین می‌توان نتیجه گرفت که میزان و فراوانی نوع تکرارها فاکتوری وابسته به جنس و گونه گیاهی می‌باشد.

جدول ۱. خلاصه‌ای از اطلاعات مربوط به داده‌کاوی توالی‌های مربوط به رونوشت‌های ثبت شده در مورد *V. vinifera*

شاخص	تعداد
تعداد کل ESTs مورد بررسی	۵۱۱۶۹۰
تعداد کل SSRs شناسایی شده	۹۴۷۴
تعداد رونوشت‌هایی که حداقل دارای یک SSR بودند	۹۲۳۶
تعداد SSRهای مرکب	۱۴۴
تعداد SSRهایی که از آن‌ها امکان طراحی آغازگر وجود داشت	۴۹۳۳



تصویر ۱. فراوانی (سمت چپ) و درصد (سمت چپ) موتیف‌های مختلف مربوط به ESTs های گیاه انگور که تا تاریخ ۱۰ دسامبر ۲۰۱۶ در بانک داده‌ها ثبت شده است.

از بین تکرارهای دو نوکلئوتیدی، نوع AT/TA با میزان ۱۵۸۱ توالی، فراوان‌ترین تکرار بودند، در حالیکه تکرار GC/CG دارای کمترین تعداد رونوشت بودند. در حقیقت هیچ تکرار GC/CG در توالی‌های ثبت شده رونوشت‌های انگور مشاهده نشد. در این تحقیق تفاوت بسیار زیادی بین تکرارهای دو سه تایی مشاهده شد. از بین موتیف‌های مختلفی که ما در این پژوهش بررسی نمودیم، ۱۴۴ رونوشت دارای SSRهای مرکب بودند و دارای بیش از یک نوع موتیف بودند که در تهیه آغازگرهای ریزماهواره‌ای حذف شدند. فراوانی وقوع SSR با تعداد تکرارها در انواع دو، سه و

چهارتایی متفاوت بود. به‌عبارتی‌دیگر، بیشترین تعداد نشانگرهای ریزماهواره‌ای شناسایی شده در این تحقیق به ترتیب مربوط به تکرارهای دو، سه و چهار نوکلئوتیدی بود.

توالی‌های تکراری ساده که در مناطق کد کننده ژنوم قرار دارند نسبت به انواع مشابه در نواحی غیر کد شونده، از نظر تکاملی حفاظت شده تر بوده و تکرارپذیری بسیار بالایی داشته و انتقال‌پذیری بالایی بین گونه‌ها و جنس‌های مشابه دارا می‌باشند (Haq *et al.*, 2012). بنابراین این نوع از SSRها را می‌توان در مطالعات تنوع ژنتیکی، انتقال‌پذیری و بررسی چندشکلی افراد مورد استفاده قرار داد، و این امکان وجود دارد که به‌عنوان لنگر در مطالعات نقشه یابی مقایسه‌ای و تکاملی در گونه‌های گیاهی مختلف بکار برد (Varshey *et al.*, 2005). به عبارتی با توجه به طبیعت حفاظتی بیشتر بخش کد شونده ژنوم نسبت به قسمت غیر کد شونده و همچنین تائید حفظ ترتیب توالی ژن‌های مختلف در گونه‌های نزدیک[2] استفاده از EST-SSR های گونه‌های گیاهی که مطالعات ژنتیکی گسترده‌ای روی آن‌ها انجام گرفته و بعضاً انواعی مانند انگور که توالی کامل ژنوم آن‌ها موجود می‌باشد، برای استفاده در گونه‌های کمتر مطالعه شده، این امکان را فراهم می‌کند که با استفاده از این نشانگرها نقشه‌های ژنتیکی را با اطمینان بیشتری ترسیم نمود و از این نشانگرها به‌عنوان لنگر برای ترسیم نقشه این گونه‌ها استفاده نمود.

منابع

Adawy, S.S., Mokhtar, M.M., Alsamman, AM. and Sakr, M.M., 2013. Development of Annotated EST-SSR Database in Olive (*Olea europaea*). International Journal of Science and Research; 4, 9: 1063-1073.

Chen C., Zhou P., Choi Y.A., Huang S. and Gmitter F.G. 2006. Mining and characterizing microsatellites from citrus ESTs. Theoretical and Applied Genetics,112:1248–57.

Edwards, K.J., Barker, A., Daly, A., Jones, C. and Karp, A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants, BioTechniques, 20: 758–760.

Haq, S.U., Jain, R., Sharma, M., Kachhwaha, S. and Kothari, S.L. 2014. Identification and characterization of microsatellites in expressed sequence tags and their cross transferability in different plants, International Journal of Genomics; 2014, Article ID 863948, 12 P.

Huang, X. and Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Research, 9: 868–877. PMID: 10508846

Morgulis, A., Gertz, E.M., Schäffer, A.A. and Agarwala, R. 2006. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. Journal of Computational Biology, 13: 1028–1040. PMID: 16796549

Sorkheh, K., Prudencio. A.S., Ghebinejad, A., Kohei Dehkordi, M., Erogu, D., Rubio, M. and Martínez-Gómez, P. 2016. In silico search, characterization and validation of new EST-SSR markers in the genus *Prunus*. BMC Research Notes; 9, 366, DOI 10.1186/s13104-016-2143-y

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Research; 40: e115–e115. PMID: 22730293

Varshney, R.K., Graner, A. and Sorrells, M.E. 2005. Genic microsatellite markers: features and applications, Trends Biotechnology; 23: 48–55.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D.A., Cestaro, A., Pruss, D. et al. 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2: e1326.

Wang , L.H., Zhang, Y.X., Qi, X.Q., Gao, Y. and Zhang, X.R. 2012 . Development and characterization of 59 polymorphic cDNA-SSR markers for the edible oil crop *Sesamum indicum* (Pedaliaceae). American Journal of Botany; 99 : e394 – e398 .

Zhou, Q., Luo, D., Ma, L., Xie, W., Wang, Y., Wang, Y. and Liu, Z. 2015. Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus* L.) using Illumina sequencing. Scientific Reports, 6:20549, DOI: 10.1038/srep20549

[2] Gene synteny

IrHC 2017
Tehran-Iran

Analysis of Ests Deposited in Gene Bank for Identification of EST-Ssrs in Grapevine

Abdolkarim Zare[1], Aziz Ebrahimi[2]
[1]Department of Biotechnology, Agriculture College, Jahrom University, Jahrom, Iran
[2]Department of Forestry and Natural Resources, Purdue University, 715 State Street, West Lafayette, IN 47907, USA
**Corresponding author: zarei@jahrom.ac.ir*

Abstract

Expressed sequence tags (ESTs) are importance resource for different aspects of molecular studies including gene discovery, gene expression, comparative genomics as well as molecular marker development. In the present study, EST sequences of grapevine (*Vitis vinifera)* available in NCBI were exploited to estimate and analysis simple sequence repeats and to develop SSR molecular markers. A total of 511690 sequences were analyzed, out of which 9236 contained at least one SSR (on average 1.8 % of deposited sequences). From 9474 SSRs were found in our study, primer were developed for 4933 SSRs and these sequences showed to have potential to be used as molecular markers (52.06% of found SSRs). Among different motif we analyzed in our study, dinucleotide repeats (45%) were the most abundant followed by trinucleotide (33%) and tetranucleotide (22%) repeats, respectively. Among the dinucleotide repeats, the motifs of AG/GA/CT/TC were the most frequent (46%). Primers were designed for some of the functionally active EST-SSRs and evaluation of their polymorphism is underway in different grapevine cultivars.

Keyword: grapevine, express sequence tags, molecular markers, SSR