# Comparative and bioinformatics analysis of the chloroplast genomes of three varieties of Strawberry: *F. vesca*, *F. virginiana* and *F. chiloensis*

**Farshid Talat [1*], Nesa Ghaffari [2], Abolhasan Faraji [2], Mehdi Badri Anarjan [3]**
[1] Research Division of Horticultural and Crop Sciences, West Azarbaijan Agricultural and Natural Resources Research and Education Center (AREEO), Urmia, Iran
[2] Department of Horticultural Sciences, Saba Institute of Higher Education, Urmia, Iran
[3] Department of Agronomy and Plant Breeding, Faculty of Agriculture, Urmia University, Urmia, Iran

## Abstract
Strawberry belongs to the genus *Fragaria* and family Rosaceae. Nowadays, the species *chiloensis*, *virginia*, and hybrid between them, namely, *ananasae*, are cultivated. This perennial herb has a small genome that is prone to the transfer of *Fragaria ananasae* cultivated strawberry genes and other important economic plants from the Rosaceae family. Present study was conducted to study and compare the complete chloroplast sequence of *Fragaria vesca*, analyses of its genome structure, gene content and organization, repeated sequences, codon usage and comparison with two cultivated octaploid sequenced *Fragaria* species *(F. chiloensis* and *F. virginiana). F. vesca* chloroplast (cp) genome is 155691 bp in length with conserved quadripartite structure. Single copy region of cp genome is separated by the two inverted regions. The large single copy region is 85561 bp, and the small single copy region is 20204 bp whereas the inverted repeat is 24963 bp each. The plastidic genome has 130 single genes and 18 duplicated genes located in inverted repeats. The singletons encode 85 proteins, 8 ribosomal RNA genes and 37 transfer RNA genes. Amongst all plastidic genes only 18 genes appeared to have 1-2 introns and when compared with DNA of two cultivated octaploeids, *rps18* was the only duplicated gene in *F. vesca* and *rps 12, rpl 2* and *trn I-GAU* are genes which are located in inverted repeats. Despite the high level of conservation in chloroplast SSRs, these are useful in analysis of genetic diversity due to their greater efficiency as opposed to genomic SSRs.
**Keywords:** Bioinformatics, Genome, Strawberry's chloroplast, Structural analysis

* Corresponding Author: f.talat@areeo.ac.ir

Copyright©2019, University of Isfahan. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/BY-NC-ND/4.0), which permits others to download this work and share it with others as long as they credit it, but they cannot change it in any way or use it commercially.

زیست‌شناسی گیاهی ایران، سال یازدهم، شمارهٔ چهلم، تابستان ۱۳۹۸، صفحهٔ ۷۶-۶۳
تاریخ دریافت مقاله: ۱۳۹۷/۰۸/۰۹
تاریخ بازنگری: ۱۳۹۸/۰۴/۳۰
تاریخ پذیرش نهایی: ۱۳۹۸/۰۵/۰۸

# مقایسه و تحلیل بیوانفورماتیک ژنوم کامل کلروپلاست سه رقم توت‌فرنگی *F. vesca* و *F. virginiana ،F. chiloensis*

**فرشید طلعت [۱*]، نسا غفاری [۲]، ابوالحسن فجری [۲]، مهدی بدری انرجان [۳]**

[۱] بخش تحقیقات علوم زراعی و باغی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی آذربایجان غربی، سازمان تحقیقات، آموزش و ترویج کشاورزی، ارومیه، ایران

[۲] گروه علوم باغبانی، دانشکدهٔ کشاورزی، مؤسسهٔ آموزش عالی صبا، ارومیه، ایران

[۳] گروه زراعت و اصلاح نباتات، دانشکدهٔ کشاورزی، دانشگاه ارومیه، ارومیه، ایران

## چکیده

توت‌فرنگی متعلق به جنس *Fragaria* و خانوادهٔ Rosaceae است. در حال حاضر، از رقم‌های شیلنسیس (*Chiloensis*)، ویرجینیانا (*Virginiana*) و هیبرید بین این دو یعنی آناناس (*Ananasae*) استفادهٔ زراعی می‌شود. این گیاه علفی چندساله، ژنوم کوچک متمایل به انتقال ژن توت‌فرنگی کشت‌شدهٔ *Fragaria ananasae* و دیگر گیاهان مهم اقتصادی از خانوادهٔ Rosaceae را دارد. پژوهش حاضر، برای بررسی و مقایسهٔ توالی کامل کلروپلاستی، تحلیل ساختار ژنومی، سازمان‌دهی و محتوای ژن، توالی‌های تکراری و تمایل کدونی توت‌فرنگی جنگلی با دو رقم اکتاپلوئید کشت‌شدهٔ *F. chiloensis* و *F. virginiana* انجام شد. ژنوم کلروپلاستی *F. vesca* ساختمان حفاظت‌شدهٔ چهار قسمتی با ۱۵۵۶۹۱ جفت‌باز طول دارد. منطقهٔ تکرار کپی کوچک با دو منطقهٔ تکرار معکوس جدا شده است. طول منطقهٔ تک‌نسخه‌ای بزرگ، ۸۵۵۶۱؛ منطقهٔ تک‌نسخه‌ای کوچک، ۲۰۲۰۴ و مناطق تکرار معکوس، هردو طولی معادل ۲۴۹۶۳ جفت‌باز داشتند. ژنوم کلروپلاستی، ۱۳۰ ژن دارد که ۱۸ ژن در منطقهٔ تکرار معکوس قرار دارند. همچنین، میان این ژن‌ها هشت ژن rRNA، ۳۷ ژن tRNA و ۸۵ ژن کدکنندهٔ پروتئین وجود داشتند. به نظر می‌رسد میان همهٔ ژن‌های پلاستیدی، تنها ۱۸ ژن، ۱ تا ۲ اینترون دارند و هنگامی‌که با DNA کلروپلاستی دو رقم اکتاپلوئید مقایسه می‌شوند، ژن *rps18* تنها ژن مضاعف‌شده است و ژن‌های *trn I-GAU*، *rpl 2* و *rps 12* در منطقهٔ تکرار معکوس وجود دارند و باوجود حفاظت زیاد از نشانگرهای SSR در ژنوم کلروپلاستی و با توجه به بهره‌وری بیشتر خود نسبت به SSRهای ژنومی، در تحلیل تنوع ژنتیکی مفیدتر هستند.

**واژه‌های کلیدی:** بیوانفورماتیک، تحلیل ساختاری، ژنوم، کلروپلاست توت‌فرنگی

[*] نگارندهٔ مسئول: نشانی پست الکترونیک: f.talat@areeo.ac.ir، شمارهٔ تماس: ۰۴۴۳۲۷۲۲۲۲۰

Copyright©2019, University of Isfahan. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/BY-NC-ND/4.0), which permits others to download this work and share it with others as long as they credit it, but they cannot change it in any way or use it commercially.

**مقدمه**

با پیشرفت علم ژنتیک، ژن، عامل اصلی در برنامه‌ریزی عملکرد سلول و به‌دنبال آن تنظیم ویژگی‌های موجودزنده شناخته شد. به این ترتیب، تمایل برای شناخت هرچه بیشتر ژن‌ها برای توجیه پدیده‌های زیستی و بهبود زندگی انسان که پیچیده‌ترین موجود است، به‌ظور چشمگیری افزایش یافت؛ به طوری که در چند دههٔ اخیر، تجهیرات مورد نیاز در تحقیقات مولکولی به‌طور گسترده‌ای افزایش یافته‌اند و در حال حاضر، پژوهش‌های مولکولی، جزء بررسی‌های رایج آزمایشگاه‌های زیستی هستند. پروژه‌های تعیین توالی ژنوم موجودات مختلف از پروژه‌های بسیار رایج به شمار می‌روند (Khrustalev and Barkovsky, 2009; Saski *et al*., 2011). بیوانفورماتیک، شاخه‌ای از علم است که از علوم محاسباتی بهره می‌برد تا اطلاعات زیست‌شناسی - مولکولی، توالی‌های DNA یا پروتئین‌ها را در رایانه ذخیره و با ابزارهای محاسباتی و الگوریتم‌های قدرتمند ریاضی تحلیل کند. توالی‌های DNA مربوط به موجودات‌زندهٔ مختلف برای دسترسی سریع و مقایسهٔ آنها با یکدیگر در پایگاه‌های داده ذخیره می‌شوند. برای بهینه‌کردن تولید پروتئین در سیستم‌های بیانی، تحلیل الگوی کدون‌های مترادف و تعیین کدون‌های بهینه بسیار مهم هستند. در این راستا، کلروپلاست به‌دلیل اندازهٔ ژنوم کوچک و تعداد کپی‌برداری زیاد اهمیت ویژه‌ای دارد؛ از این رو به تحلیل الگوی کدون‌های مترادف و تعیین کدون‌های بهینه در کلروپلاست گیاهان با روش‌های بیوانفورماتیک پرداخته می‌شود (Persson, 2000; Hirao *et al*., 2008; Perrow, 2011). توت‌فرنگی از خانوادهٔ Rosaceae و جنس *Fragaria* است و گونه‌های متعددی دارد. تعدادکروموزوم پایه در این گیاه $X$=۷ و تنوع در سطح پلوئیدی در آن زیاد است. در پژوهش حاضر، ژنوم کامل کلروپلاستی *F. virginiana* , *F. chiloensis* و مشهورترین گونهٔ این جنس، *F. vesca*، بررسی شدند. پژوهش حاضر، با هدف بررسی و مقایسهٔ توالی کامل کلروپلاست *F. vesca*، *F. chiloensis* و *F.virginiana* میوهٔ توت‌فرنگی، تحلیل ساختار ژنوم، محتوای ژنوم‌و توالی‌های تکراری، کاربرد کدون و مقایسهٔ بین ژنوم‌های موجود انجام شد. فرضیه‌های بررسی حاضر، شامل ژنوم کلروپلاست یعنی هدف اصلی در پژوهش‌های بیوتکنولوژی گیاهان، شناسایی ژن‌های دخیل در این ژنوم و نقش آنها در سیستم ژنتیکی گیاه است که کمک مؤثری به بهبود تولیدات مرتبط با سیستم فتوسنتزی گیاه خواهد کرد. همچنین شناسایی نقش ژن‌های موجود در ژنوم کلروپلاست و توانایی این ژن‌ها در بیان بیشینه از اهداف بررسی حاضر است که تولید واکسن‌های گیاهی را ممکن خواهد کرد. در چند دههٔ گذشته، پیشرفت در زیست‌شناسی مولکولی و تجهیزات لازم برای بررسی‌ها در این زمینه، افزایش سریع تعیین توالی ژنوم بسیاری از گونه‌های موجودات را باعث شد. در حال حاضر، توالی ژنوم بسیاری از موجودات ساده مانند باکتری‌ها تا موجودات بسیار پیشرفته مانند یوکاریوت‌های پیچیده

شناسایی شده است. پروژهٔ شناسایی ژنوم انسان در سال ۱۹۹۰ آغاز شد و در سال ۲۰۰۳ پایان یافت و اکنون اطلاعات کامل مربوط به توالی هر ۲۴ کروموزوم انسان موجود است (Cordenunsi *et al.*, 2003; Debnath and Teixeira da Silva, 2007; Diekmann *et al.*, 2009). گسترش روزافزون حجم زیاد داده‌های ژنومی و نیاز به ذخیره، بازیابی و تحیل مناسب این داده‌ها پیدایش علم بیوانفورماتیک را موجب شد. این دانش نوظهور به‌صورت دانشی بین‌رشته‌ای تلاش می‌کند با روش‌های موجود در علوم رایانه‌ای، ریاضیات، ژنتیک، شیمی، فیزیک و علوم مرتبط دیگر، مسائل مختلف زیست‌شناختی را که معمولاً در سطح مولکولی هستند حل کند. تلاش‌های پژوهشی اصلی در این رشته عبارتند از: تطابق توالی، کشف ژن، گردآوری ژنوم، تنظیم ساختار پروتئینی، پیش‌بینی ساختارهای دوم و سوم پروتئین، پیش‌بینی بیان ژن و تعاملات پروتئین – پروتئین و مدل‌سازی (Kim *et al.*, 2009). Talat و Wang (۲۰۱۵) تحلیل بیوانفورماتیک ژنوم کلروپلاست پنبهٔ دیپلوئید وحشی و دو گونهٔ آلوتتراپلوئید را انجام دادند. نتایج بررسی آنها نشان دادند ژنوم کلروپلاست *G. thurberi*، ۱۶۰۲۶۴ جفت‌باز طول دارد و با ساختاری چهارگانه، محافظت شده است. ناحیه‌ای از نسخهٔ تکی ژنوم کلروپلاستی از دو رشتهٔ معکوس جدا شده است. قطعهٔ بزرگ‌تر، ۸۸۷۳۷ جفت‌باز و قطعهٔ کوچک‌تر ۲۰۲۷۱ جفت‌باز دارد؛ در حالی که نواحی تکرار معکوس هرکدام ۲۵۶۲۸ جفت‌باز طول داشتند.

## مواد و روش‌ها

**توالی‌یابی:** ژنوم کامل کلروپلاست سه رقم توت‌فرنگی *F.chiloensis*، *F.vesca* و *F.virginiana* به‌ترتیب با شماره دسترسی‌های NC_019601.2، NC_019601.1 و NC_015206.1 از پایگاه NCBI با فرمت FASTA دانلود شدند.

**تفسیر ژن:** تفسیر ژن به‌طور سنتی بسیار خسته‌کننده و مستعد خطا است. همچنین، تفسیر ژن در بانک ژنی ازنظر نام ژن سازگار نیست و معمولاً خواندن‌ها (ycfs یا ORF) به‌روز نیستند. در گذشته، بسیاری از توالی‌های ژنوم کلروپلاستی با انجام BLAST N و BLAST X در بانک ژنی مشروح می‌شدند؛ اما در حال حاضر، با نرم‌افزار DOGMA بسیاری از این مشکلات، کاهش یافته و به پایان رسیده‌اند؛ بنابراین در پژوهش حاضر از نرم‌افزار DOGMA سال ۲۰۱۷ استفاده شد.

**نقشه و ساختار ژنوم:** نقشهٔ دایره‌ای‌شکل ژنوم کامل کلروپلاستی برای مشاهدهٔ ساختار و توزیع ژن هر سه رقم بررسی‌شده با ارجاع به تفسیرهای گذشته و دادن شمارهٔ دسترسی موجود هر ژنوم با نرم‌افزار OGDRAW سال ۲۰۱۷ ایجاد شد.

**ساختار ژنوم کلروپلاستی و تحلیل توالی:** Codon Usage اشاره به تفاوت در تعداد تکرار کدون مترادف در برنامه‌نویسی DNA دارد. کدون، متشکل از سه نوکلئوتید است که آمینواسید ویژه‌ای در زنجیرهٔ پلی‌پپتید یا را رمز می‌کند یا کدون خاتمهٔ ترجمه است که هیچ آمینواسیدی رمز نمی‌کند. نسبت کدون‌های مترادف به کدون‌های مختلف در هر نمونهٔ ژن (RSCU) با نرم‌افزار Codon W سال ۲۰۱۸ به‌صورت برنامهٔ آنلاین محاسبه شد. Codon

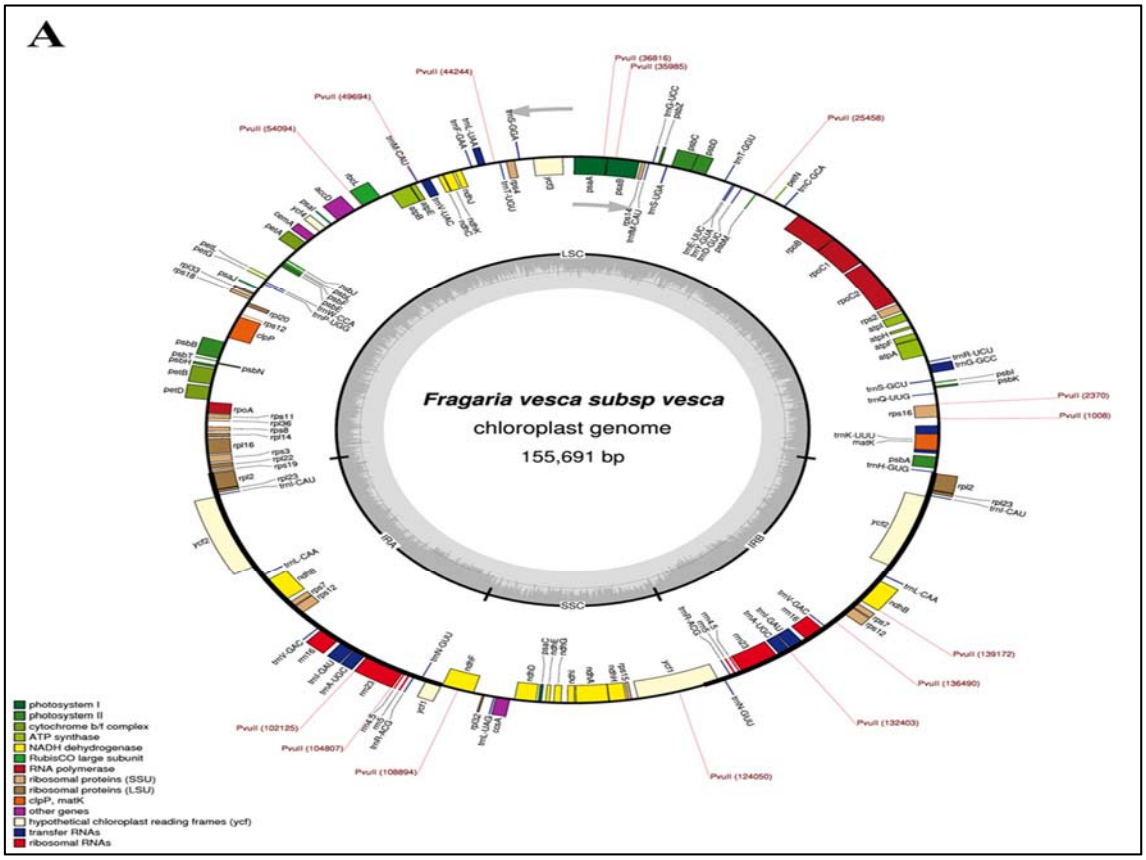
W برنامهٔ طراحی‌شدهٔ ساده‌ای برای تحلیل چندمتغیرهٔ کدون و آمینواسید است که برای محاسبهٔ شاخص استاندارد کدون استفاده می‌شود.

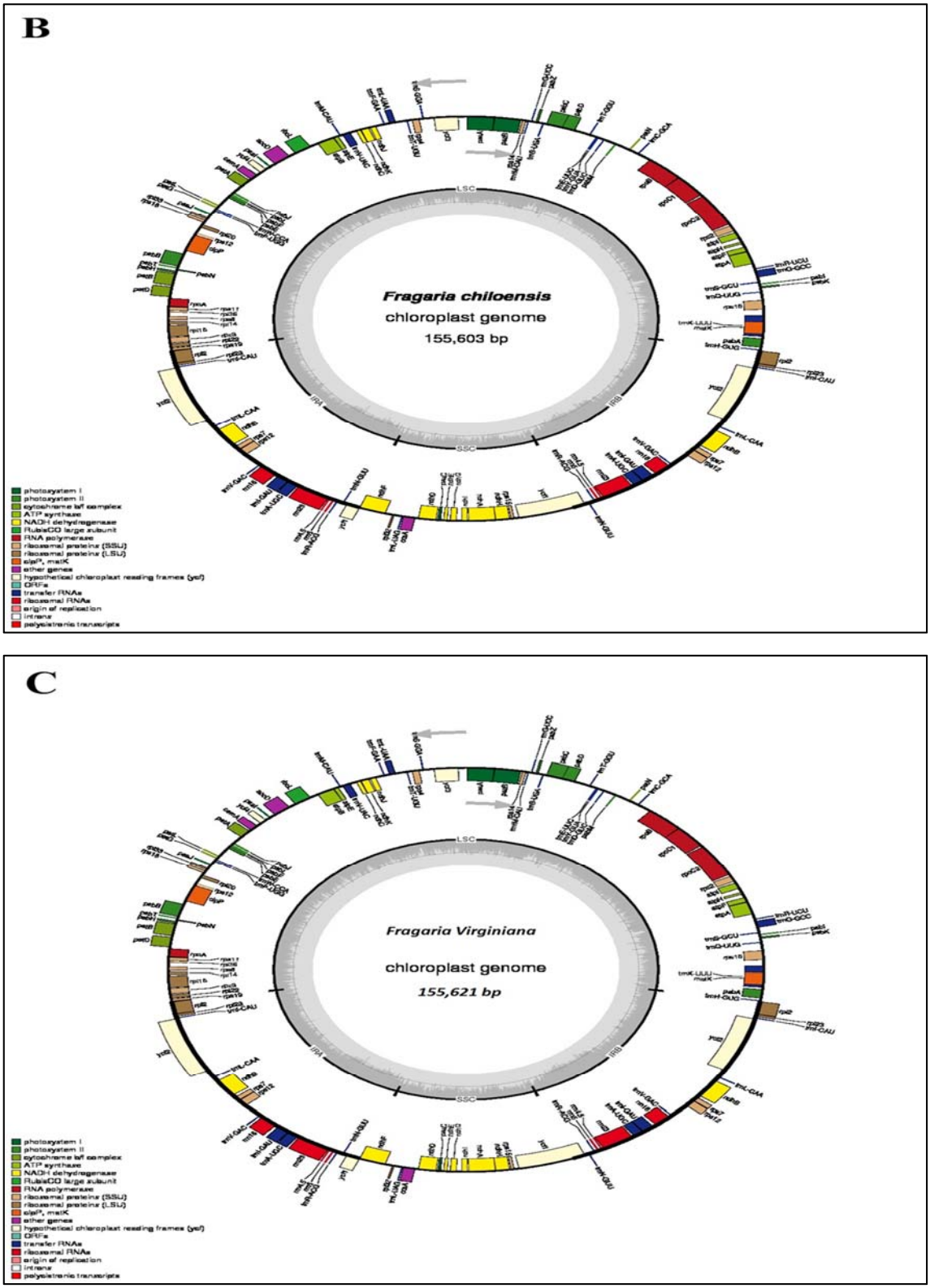
**محتوای GC:** محتوای GC عاملی بسیار مهم در ژنوم کلروپلاستی است که بر ثبات ژنوم تأثیر می‌گذارد. محتوا و درصد GC موجود در سه رقم یادشده با داشتن توالی ژن‌های مربوطه از پایگاه NCBI و فایل بانک ژنی استخراج شد و با قراردادن توالی مربوطه در نرم‌افزار Visual Bioinformatics سال ۲۰۱۷، درصد مربوط به هر منطقه به دست آمد.

**اینترون ها:** اینترون‌ها بخش‌هایی از ژنوم هستند که رونویسی از روی آنها انجام می‌شود؛ اما ترجمه نمی‌شوند. برای مقایسهٔ مقدار اینترون‌ها در ژنوم سه گونهٔ بررسی‌شده از نرم‌افزار آنلاین DOGMA سال ۲۰۱۷ استفاده شد.

## نتایج

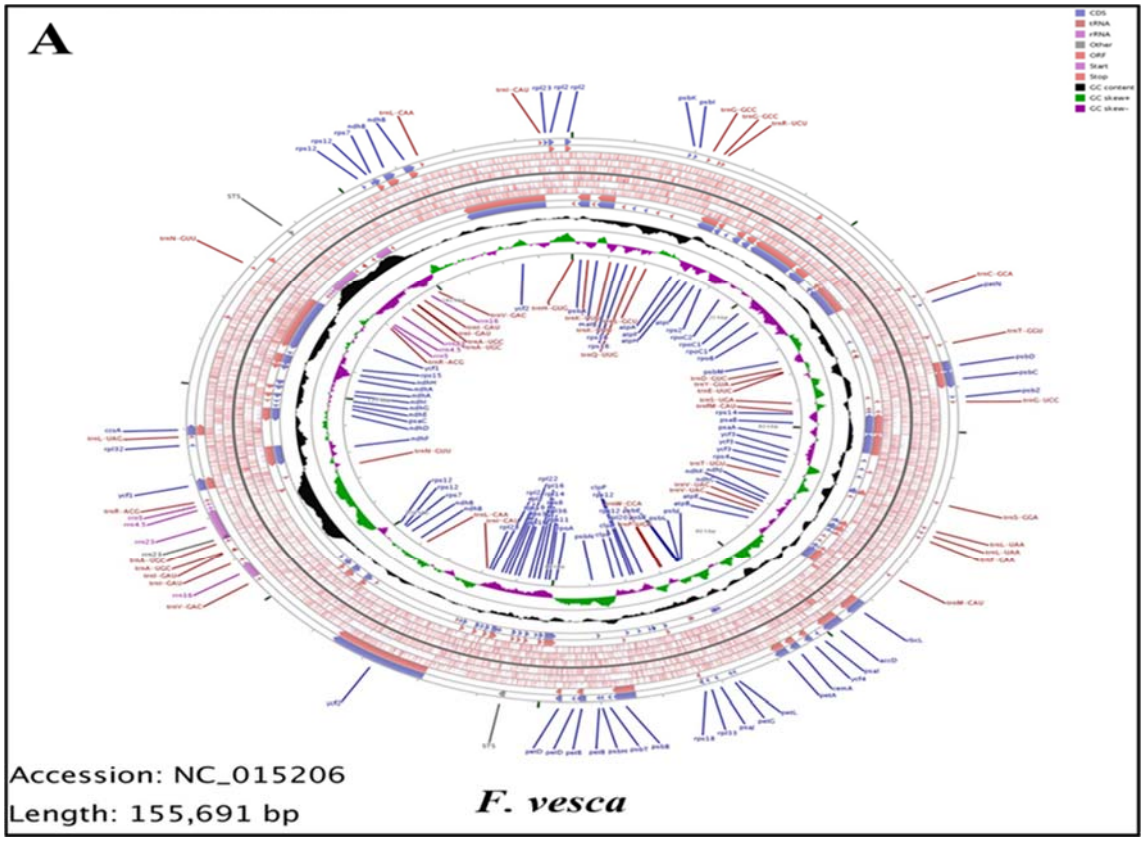
**ساختار و تفسیر ژنوم کلروپلاستی:** ساختار ژنوم کلروپلاستی سه گونه در شکل ۱ آورده شده است و نشان می‌دهد مجموع ژنوم کامل دایره‌ای‌شکل *F. vesca*، *F. chiloensis* و *F. virginiana* به‌ترتیب، طول‌هایی معادل ۱۵۵۶۹۱، ۱۵۵۶۰۳ و ۱۵۵۶۲۱ جفت‌باز دارند که شامل مناطق تکرار بزرگ (LSC)، مناطق تکرار کوچک (SSC)، دو منطقهٔ تکرار معکوس (IR) و ساختار چهارجزئی هستند.

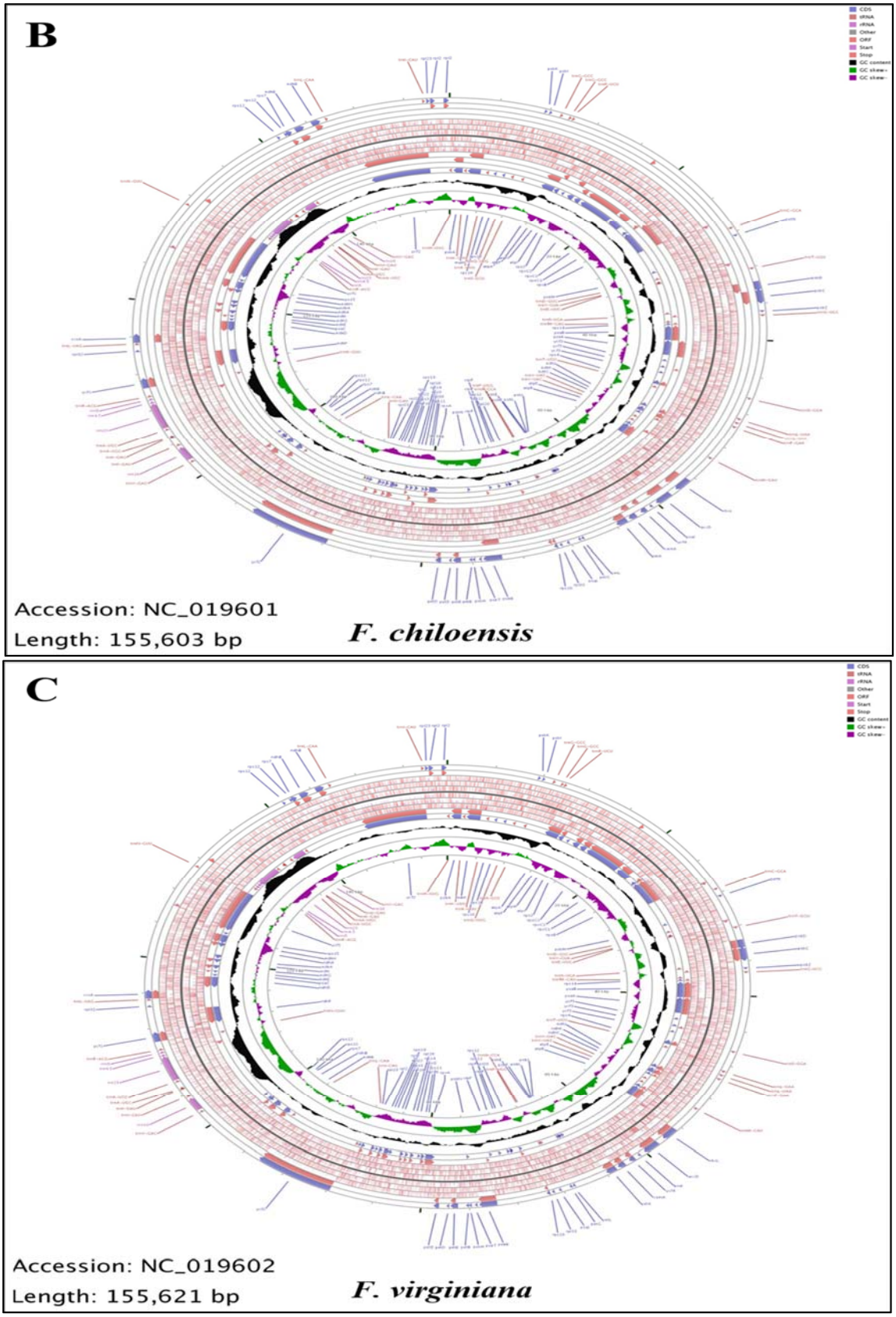





شکل ۱- نقشهٔ دایره‌ای‌شکل ساختار ژنوم کلروپلاستی ۳ گونهٔ *F. vesca* ssp. *vesca* (A)، *F. chiloensis* (B) و *F. virginiana* (C) و ۱۳۰ ژن شناسایی‌شده در ژنوم کلروپلاست- ژن‌هایی که در خارج از حلقه نشان داده شده‌اند، در خلاف جهت حرکت عقربه‌های ساعت و ژن‌هایی که در داخل حلقه هستند، در جهت حرکت عقربه‌های ساعت رونویسی می‌شوند.

**محتوای GC:** محتوای GC عاملی بسیار مهم در ژنوم کلروپلاستی است که بر ثبات ژنوم اثر می‌گذارد. در زیست‌شناسی مولکولی و ژنتیک، محتوای GC یا گوانین - سیتوزین، درصدی از پایگاه‌های نیتروژنی در مولکول DNA است که در مقابل آدنین و تیمین قرار دارد که ممکن است به قطعه‌ای از DNA یا RNA یا کل ژنوم مربوط باشد. تنوع در محتوای GC، میان چهار منطقهٔ مختلف در ژنوم کلروپلاستی سه رقم بررسی‌شده در شکل ۲ نشان داده شده است. مقدار محتوای GC برحسب درصد در هریک از مناطق برنامه‌نویسی‌شده که شامل پروتئین‌ها و ژن‌های tRNA و rRNA هستند، مناطق غیررمزگذاری‌شده که شامل مناطق بین‌ژنی و اینترون‌ها، کل توالی ژنوم کلروپلاستی و مناطق تک‌نسخه‌ای و تکرار معکوس هستند که توالی هرکدام به‌صورت دستی از پایگاه NCBI دانلود شد و همچنین با برنامهٔ آنلاین CP Gavas توالی‌های دردسترس rRNA، tRNA و پروتئین به دست آمدند. در هر سه رقم بررسی‌شده، نتایج مشابه به دست آمدند و داده‌های به‌دست‌آمده، با CG view روی یک نقشهٔ ژنومی دایره‌ای‌شکل برای هر سه رقم نمایش داده شدند. علاوه‌بر محتوای GC و انحراف GC، ژن‌های CDS، ORF، rRNA و tRNAها و کدون‌های شروع و پایان بر نقشه نمایان شدند. با توجه به نتایج، در منطقهٔ رمزگذاری‌شده، ژن‌های rRNA بیشترین و ژن‌های پروتئینی، کمترین مقدار GC را داشتند. مناطق غیرکدکننده، تحول سریع تجربه می‌کنند؛ درنتیجه، این مناطق نسبت به مناطق دیگر، GC و Coding غنی‌تری دارند.



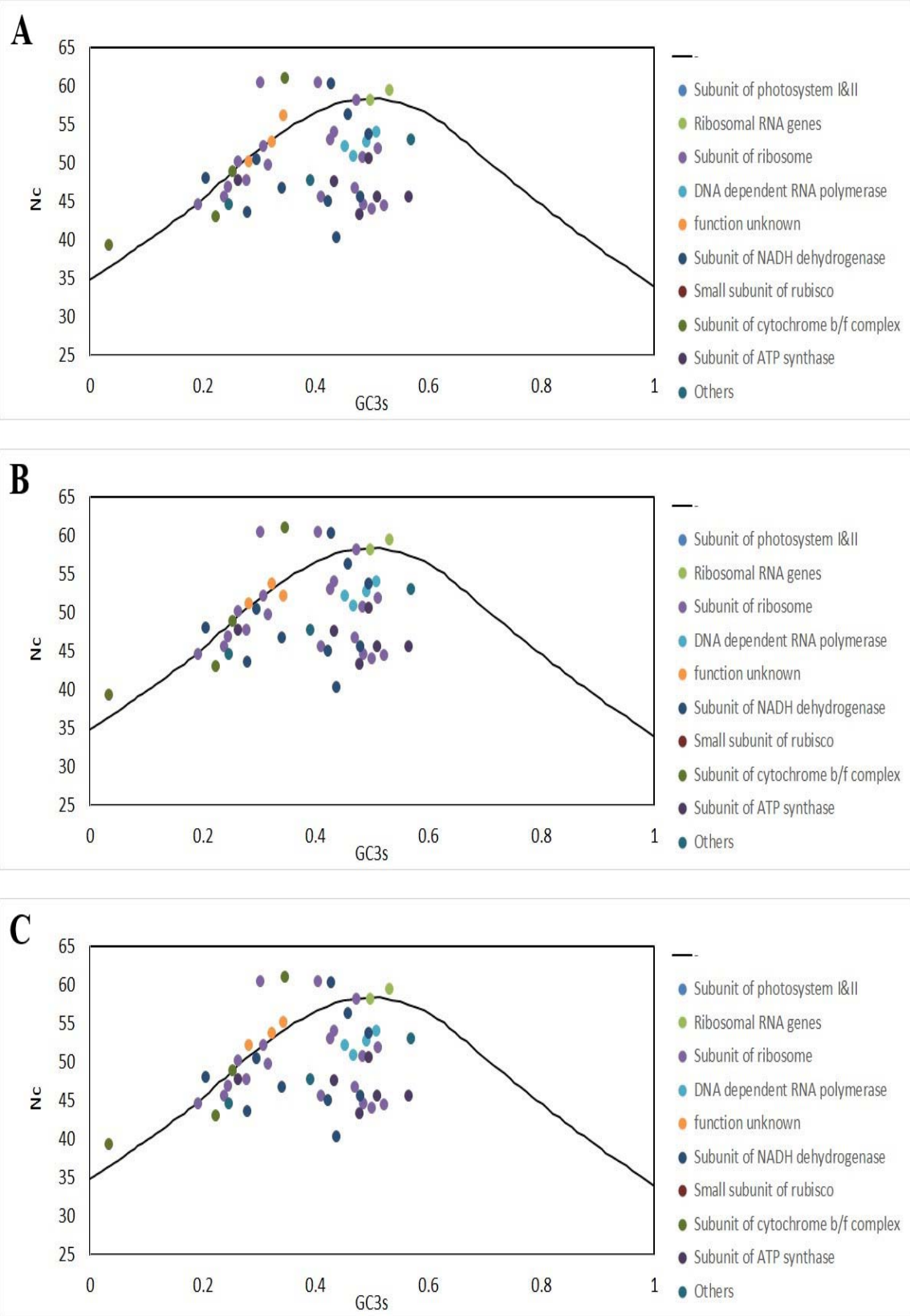



شکل ۲- نقشهٔ دایره‌ای نشان‌دهندهٔ مقدار GC، جایگاه و فراوانی ژن‌های tRNA و rRNA در ۳ گونهٔ *F. vesca* ssp. *vesca* (A)، *F. chiloensis* (B) و *F. virginiana* (C)

**نمودار توزیع NC:** نمودار توزیعNC که شاخص‌های ENC و GC3S در آن دخیل هستند برای هر سه ژنوم کلروپلاست بررسی‌شده در شکل ۳ نشان داده شده است. از آنجا که نتایج مربوط به این تحلیل برای هرسه ژنوم، تقریباً یکسان به دست آمدند، ژنوم کلروپلاست *Fragaria vesca* مرجع قرار داده شد و توضیحات، براساس آن ارائه می‌شوند. مقادیر ENC از ۳۸/۷۲ تا ۶۱/۷۲ متغیر هستند که میانگین ۵۰/۲۲ و انحراف استاندارد ۵/۵۶ دارند. ناهمگونی کاربرد کدون، بیشتر به‌دلیل تنوع مقادیر GC3 است که از ۱۷ تا ۵۸ درصد متغیر است و میانگین ۴۱ درصد و انحراف استاندارد ۱۰ درصد دارد. اگر GC3 به‌طور کامل بر انحراف کاربرد کدون تأثیر بگذارد، مقادیر NC روی نمودار، بین GC3S وNC پیش‌بینی می‌شود. تعداد درخورتوجهی از نقاط، روی نمودار قرار گرفته‌اند و به‌سمت ناحیه‌ای هستند که ازنظرGC فقیر است که این از ترکیب نوکلئوتیدی شدید منشاء می‌گیرد؛ اما موضوع جالب توجه این است که بیشتر نقاط با مقادیر NC کم در زیر نمودار پیش‌بینی‌شده قرار گرفته‌اند. این نتایج نشان می‌دهند برخی ژن‌ها در *Fragaria vesca* کاربرد کدونی مستقل از ترکیب نوکلوتیدی مشتق از فشار جهشی و متأثر از عوامل دیگر دارند که مستقل از محدودیت‌های ترکیبی هستند.

**اینترون:** اینترون‌ها بخش‌هایی از ژنوم هستندکه در فرایند رونویسی رمزهایDNA، رونویسی از روی آنها انجام می‌شود؛ ولی در فرایند پیرایش برداشته می‌شوند؛ از این رو در فرایند رمزخوانی، پروتئین‌ها از روی آنها ساخته نمی‌شوند. به همین دلیل، به آنها اینترون گفته می‌شود. اینترون‌ها در چیدمان DNA در میان رمزهای ژن و در چیدمان RNA پس از رونویسی رمزهای DNA دیده می‌شوند. این بخش‌ها در بسیاری از موجودات و حتی در برخی ویروس‌ها نیز دیده می‌شوند. اندازهٔ آنها ممکن است بسیار بزرگ باشد.

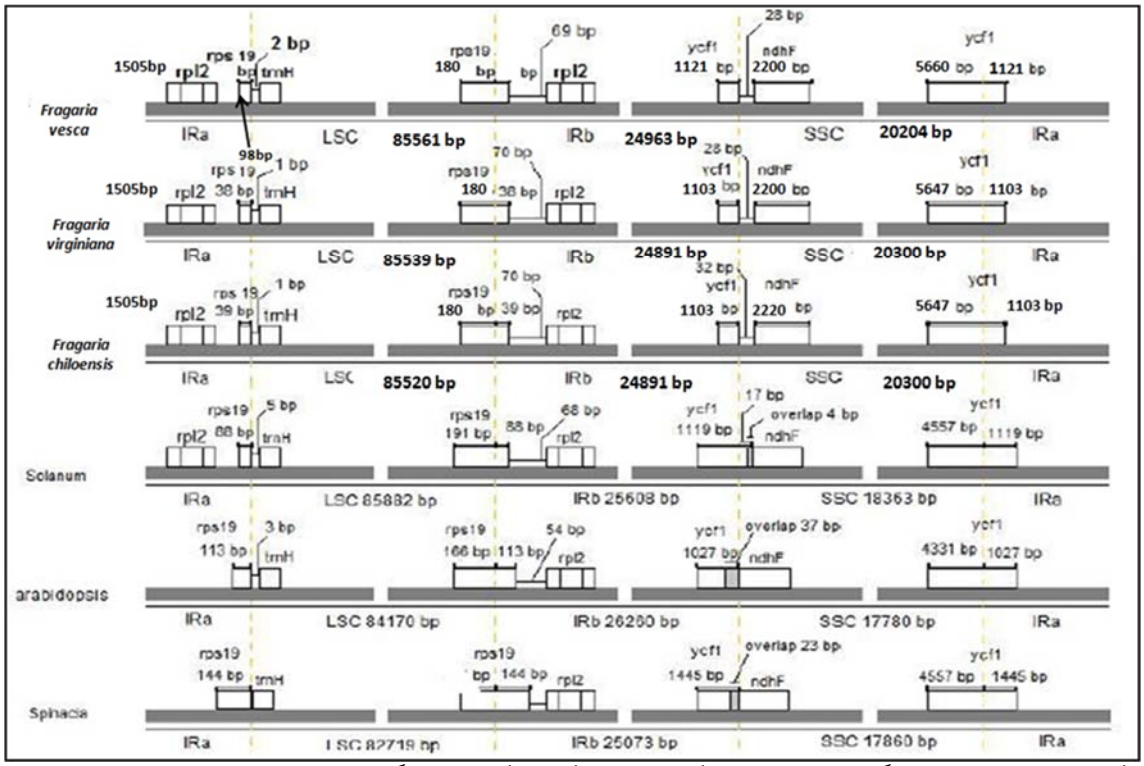
**فراوانی ژن‌ها در مناطق تک‌نسخه‌ای وتکرار معکوس:** در تحول کلروپلاست پیداکردن مناطق مناسب بین‌ژنی برای یکپارچه‌کردن کارآمد ترانس‌ژن‌ها بسیار مهم است. در گوجه‌فرنگی و سیب‌زمینی، پژوهشگران از rps 12، trnV، accD، rpcL، trnG، trnfM و مناطق بین‌ژنی rRNA16s-ORF 70B برای ادغام مناطق بین‌ژنی تنباکو استفاده کرده‌اند (Kim *et al.*, 2009)؛ بنابراین، یک نسخهٔ کامل از کپی دوم ژن rps 19 و یک کپی ناقص از rpl 22 در مجاورت trnH وجود دارد. این تکراری است که به احتمال زیاد با توجه به گسترش IR B در محل اتصال LSC، فرایندی مشترک در ژنوم کلروپلاست است (Talat and Wang, 2015). در پژوهش حاضر، مطابق با شکل ۴، روش قرارگیری و طول مربوط به هر ژن در محل اتصال LSC/IR, SSC/IR با سه گونهٔ مرجع نشان داده شده‌اند.




شکل ۳- مقادیر GC3S در مقابل NC در ژنوم کلروپلاستی ۳ گونهٔ *F. vesca* ssp. *vesca* (A)، *F. virginiana* (B) و *F. chiloensis* (C)

جدول ۱- تفاوت در طول و جایگاه اینترون‌ها در سه رقم بررسی‌شده

| *F. virginiana* | *F. chiloensis* | *F. vesca* | اینترون |
|---|---|---|---|
| ۵۳۶ | ۵۳۶/۷۲۹ | ۵۳۶/۲۳۰ | Rps12 |
| ۳۱ | ۳۵ | ۳۵ | trnK-UUU |
| ۸۵۶ | ۸۵۶ | ۸۶۹ | Rps16 |
| ۶۹۲ | ۶۹۲ | ۶۹۲ | trnG-GCC |
| ۷۵۶ | ۷۵۶ | ۷۵۶ | Rpo C1 |
| ۸۰۵/۷۲۱ | ۸۰۵/۷۲۱ | ۸۰۵/۷۲۳ | ycf 3 |
| ۴۲۰ | ۴۲۰ | ۴۲۰ | trnL-UUU |
| ۵۹۸ | ۵۹۸ | ۵۹۸ | trnV-UAC |
| ۶۶۰/۸۲۵ | ۶۶۱/۸۲۵ | ۶۵۷/۸۲۵ | clp P |
| ۹۹۱ | ۹۹۱ | ۹۹۱ | rpl 16 |
| ۶۷۶ | ۶۷۶ | ۶۷۶ | ndh B |
| ۸۵۲ | ۹۵۲ | ۹۰۵ | trnI- GAU |
| ۸۱۱ | ۸۱۱ | ۸۱۱ | rtnA- UGC |
| ۱۱۸۹ | ۱۱۸۹ | ۱۱۸۸ | ndh A |
| ۷۷۰ | ۷۷۰ | ۷۷۰ | pet B |
| ۷۱۴ | ۷۱۴ | ۷۱۴ | pet D |
| ۶۸۱ | ۶۸۱ | ۶۸۱ | rpl 2 |
| ۶۷۶ | ۶۷۶ | ۶۷۶ | ndh B |




شکل۴- نقشهٔ تعیین محل قرارگیری ژن‌ها بر مناطق تک‌نسخه‌ای و تکرار معکوس از سه گونهٔ مرجع با ژنوم‌های بررسی‌شده- ردیف‌های ۴، ۵ و ۶ از منبع Talat و Wang (۲۰۱۵) اقتباس شده‌اند.

## بحث

براساس نتایج به‌دست‌آمده، ژنوم کلروپلاستی *F. chiloensis*، *F. virginiana* و *F. vesca* ساختار چهاروجهی دارد. مجموع ژنوم کامل دایره‌ای‌شکل *F.chiloensis* با طولی برابر ۱۵۵۶۰۳ جفت باز کوتاه‌تر از دو رقم *F. virginiana* (۱۵۵۶۲۱) و *F. vesva* (۱۵۵۶۹۱) است. در هرسه رقم، همهٔ مناطق تکرار کپی کوچک با دو منطقهٔ تکرار معکوس از هم جدا شده‌اند که با توجه به تنوع ژنوم کلروپلاستی بین این ارقام، بسیار شبیه به یکدیگر بودند. هنوز اطلاعات بسیار اندکی دربارهٔ میزان تنوع بین مناطق رمزگذاری‌نشدهٔ ژنوم کلروپلاستی وجود دارد و در بررسی حاضر، در ارزیابی نسبی سطح تنوع بین همهٔ مناطق غیررمزگذار منطقهٔ کوچک کپی ژنوم کلروپلاستی به‌طور مستقیم چند مورد استثناء دیده شد که ژنوم کلروپلاستی، دو تکرار معکوس دارد که هرکدام به‌طور تقریبی ۲۵ کیلو باز دارد و تصویر آینه‌ای از یکدیگر به‌صورت مکمل ژن هستند. با توجه به تحلیل کل ژنوم در رقم‌های *F. chiloensis*، *F. virginiana* و *F. vesca*، طول مناطق تکرار بزرگ آنها به‌ترتیب، ۸۵۵۶۱، ۸۵۵۳۹، ۸۵۵۲۰ جفت‌باز، طول منطقهٔ تکرار کپی کوچک آنها به‌ترتیب، ۲۰۲۰۴، ۲۰۳۰۰ و ۲۰۳۰۰ جفت‌باز و طول مناطق تکرار معکوس آنها به‌ترتیب ۲۴۹۶۳، ۲۴۸۹۱ و ۲۴۸۹۱ جفت‌باز بود که نشان‌دهندهٔ شباهت توالی بسیار زیاد بین این ارقام است. همچنین برای رقم‌های یادشده، طول منطقهٔ برنامه‌نویسی‌شده به‌ترتیب ۸۳۹۲۵، ۸۳۷۹۵ و ۸۳۷۹۵ جفت‌باز و طول منطقهٔ برنامه‌نویسی‌نشده به‌ترتیب ۶۶۰۴۴، ۶۵۷۸۴ و ۶۵۷۸۹ جفت‌باز به دست آمدکه نشان‌دهندهٔ تفاوت بسیار جزئی در توالی است و دو رقم شیلنسیس و ویرجینیانا طولی مشابه درمنطقهٔ برنامه‌نویسی دارند. ژنوم کلروپلاستی سه رقم، ۴۴/۷ درصد از پروتئین را رمزگذاری می‌کند. همچنین نسبت میان مناطق بین‌ژنی و اینترون‌ها مشاهده شد؛ بدین ترتیب که در *F. vesca* طول منطقهٔ بین‌ژنی و اینترون به‌ترتیب ۳۰/۵۳ و ۱۲/۰۷ درصد، در *F. virginiana* به‌ترتیب ۳۰/۵۲ و ۱۲/۰۷ درصد و برای *F. chiloensis* به‌ترتیب ۳۰/۵۱ و ۱۲/۰۷ درصد به دست آمد. محتوای GC در رقم‌های ویرجینیانا و شیلنسیس، ۳۶/۴۵ درصد و در رقم *F. vesca*، ۳۷/۲۱ درصد بود که بسیار نزدیک به مقدار جهانی آن (۳۶/۷۳ درصد) است. در رقم‌های *F. chiloensis*، *F. virginiana* و *F. vesca*، محتوای GC در منطقهٔ برنامه‌نویسی‌شده، به‌ترتیب برابر با ۳۷/۰۲، ۳۷/۰۲ و ۳۶/۰۹ درصد و در منطقهٔ برنامه‌نویسی‌نشده به‌ترتیب برابر با ۳۱/۹۷، ۳۱/۹۵ و ۳۱/۹۳ درصد بود که نشان می‌دهد هردو منطقه مقدار اندکی GC دارند. دلیل این مسئله، وجود ژن‌های ریبوزومی و مناطق رمزگذاری‌شده در منطقهٔ IR برای داشتن مقدار زیاد GC است. مناطق غیررمزگذار، تحولی سریع را تجربه می‌کنند؛ درنتیجه، این مناطق نسبت به مناطق رمزگذار، GC غنی‌تری دارند. محتوای GC یکی از ویژگی‌های مهم ژنوم است که با تعدادی از جایگاه‌های اتصال micro RNA همبسته است. محتوای GC در سمت 5'UTR و 3'UTR متفاوت است. همچنین محتوایGC اثری بر RNAi دارد که دلیل آن همبستگی زیاد بین جایگاه دسترسی زیاد RNAi و

همبستگی منفی با فعالیت RNAi است. با تفسیر ژنوم کلروپلاستی سه رقم بحث‌شده، درمجموع برای هریک از رقم‌ها ۱۳۰ ژن کلروپلاستی یافت شدند که مطابق با عمل ژن آنها به چهار گروه طبقه‌بندی می‌شوند که عبارتند از: ژن‌های مرتبط با خود جایگزینی، ژن‌های مرتبط با فتوسنتز، ژن‌های مرتبط با اعمال دیگر و ژن‌های با اعمال ناشناخته. میان این ژن‌ها، هشت ژن rRNA، ۳۷ ژن tRNA و ۸۵ ژن کدکنندهٔ پروتئین در هرسه رقم هستند.که بین اینها هفت ژن tRNA و چهار ژن rRNA در منطقهٔ تکرار معکوس نمایان شدند. ژن rRNA به‌طور مرتب در ۵، ۴، ۲۳، ۱۶ و ۵ و srRNA در هردو طرف IR تنظیم شده است. *ycf 1* تقریباً در همهٔ توالی‌های پلاستیدی تا به امروز وجود داشته است و پیش‌بینی می‌شود در برخی از مسیرهای اساسی در متابولیسم سلولی یا در خدمت برخی از اعمال ساختمانی برای محافظت پلاستید باشد. همچنین فرض شده است *ycf 1* مسئولیت کدگذاری غشای درونی کلروپلاست را بر عهده دارد؛ با این حال عمل *ycf 1* به‌طور دقیق مشخص نشده است؛ ولی به نظر می‌رسد برای بقای گیاه ضروری باشد. در توالی ژنوم کلروپلاستی، این ژن معمولاً در دهانهٔ مرز مناطق IR و SSC قرار دارد. با توجه به جایگاه مشترک ژن *ycf 1* در ژنوم کلروپلاستی یک کپی از آن در مرز بین IRA/SSC و یک کپی کوتاه از شبه‌ژن *ycf 1* در مرز بین IR b/SSC در رقم *F. vesca* و در رقم‌های شیلنسیس و ویرجینیانا به‌ترتیب با ۱۱۰۳ و ۵۶۴۷ جفت‌باز به دست آمد. در این ارقام، جایگاه شبه‌ژن‌ها در محل اتصال LSC/SSC و IR قرار دارد.

## جمع‌بندی

بررسی حاضر، زمینه‌ای برای پژوهش‌های پروتئومیکس است. پروتئومیکس، علم بررسی گستردهٔ پروتئوم شامل مطالعهٔ بیان، تغییرات پس از ترجمه و بررسی برهم‌کنش پروتئین‌ها با سایر مولکول‌ها است. تحلیل جامع پروتئین‌ها با هدف بررسی تنوع ژنتیکی، بررسی تفاوت‌ها و پاسخ به تنش‌ها موضوع اصلی علم بین‌رشته‌ای پروتئومیکس است. پروتئومیکس با بهره‌گیری از علوم شیمی پروتئین‌ها، بیوانفورماتیک و زیست‌شناسی به شناسایی، کمیت‌سنجی، بررسی تغییرات پس از ترجمه و برهم‌کنش‌های پروتئین‌ها می‌پردازد. هدف پروتئومیکس، شناسایی پروتئین‌های جدید براساس نقش، وظیفه و بررسی روش بیان آنها در سیستم‌های تنظیمی است. در حال حاضر، بخش اصلی فعالیت‌های پروتئومیکس، جداکردن پروتئین‌های بافت‌ها با الکتروفورز دوبعدی و شناسایی متعاقب آنها با طیف‌سنجی جرمی است. نتایج حاصل از پژوهش حاضر و بررسی‌های مشابه، با در اختیار قرار دادن اطلاعات تکمیلی از توالی‌یابی‌های ژنی و متابولیت‌های سلولی، زمینهٔ بهره‌برداری را از این داده‌ها در تحلیل پروتئین‌های سنتزشده فراهم خواهند کرد. پژوهش حاضر، آغاز پروژهٔ ترانسکریپتومیکس و پروتئومیکس در گونه‌های بررسی‌شدهٔ دانش‌پژوهان در زمینهٔ علوم ژنومیکس است.

## سپاسگزاری

نگارندگان مقاله از مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان آذربایجان غربی و گروه علوم باغبانی دانشکدهٔ کشاورزی مؤسسهٔ آموزش عالی صبا به‌دلیل حمایت از انجام پژوهش حاضر سپاسگزاری می‌کنند.

## منابع


Cordenunsi, B. R., Nascimento, J. D. and Lajolo, F. M. (2003) Physico-chemical changes related to quality of five strawberry fruit cultivars during cool-storage. Food Chemistry 83(2): 167-173.

Debnath, S. C. and Teixeira da Silva, J. A. (2007) Strawberry culture *in vitro*: applications in genetic transformation and biotechnology. Fruit, Vegetable and Cereal Science and Biotechnology 1(1): 1-12.

Diekmann, K., Hodkinson, T. R., Wolfe, K. H., van den Bekerom, R., Dix, P. J. and Barth, S. (2009) Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* L.). DNA Research 16(3): 165-176.

Hirao, T., Watanabe, A., Kurita, M., Kondo, T. and Takata, K. (2008) Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biology 8: 70.

Khrustalev, V. V. and Barkovsky, E. V. (2009) Mutational pressure is a cause of inter-and intragenomic differences in GC-content of simplex and varicello viruses. Computational Biology and Chemistry 33(4): 295-302.

Kim, Y. K., Park, C. W. and Kim, K. J. (2009) Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthis saniculifolia*, and its evolutionary implications. Molecules and Cells 27(3): 365-381.

Perrow, C. (2011) Normal accidents: Living with high risk technologies-Updated edition. Princeton University Press, Princeton.

Persson, B. (2000) Bioinformatics in protein analysis. In: Proteomics in Functional Genomics (Eds. Jollès, P. and Jörnvall, H.) 215-231. Birkhäuser, Basel.

Saski, C., Lee, S. B., Daniell, H., Wood, T. C., Tomkins, J., Kim, H. G. and Jansen, R. K. (2005) Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. Plant Molecular Biology 59(2): 309-322.

Talat, F. and Wang, K. (2015) Comparative bioinformatics analysis of a wild diploid *Gossypium* with two cultivated allotetraploid species. Iranian Journal of Biotechnology 13(3): 47-56.