

مقاله پژوهشی

آنالیز بیوانفورماتیکی ژنوم کلروپلاستی گونه‌های دیپلوئید *G.longicalyx* & *G.stocksii* و مقایسه آن با گونه‌های تتراپلوئید پنبه

فرشید طلعت^{۱*}، ماهرخ غیبی^۲

^۱ استادیار، 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 کارشناس ارشد زیست فناوری (بیوتکنولوژی)، دانشگاه آزاد اسلامی واحد ارومیه، ایران

* نویسنده مسئول مکاتبات: f.talat@areeo.ac.ir

تاریخ پذیرش: دی ۱۴۰۱

تاریخ دریافت: خرداد ۱۴۰۱

DOI:10.30495/JDB.2023.1959300.1305

<https://dorl.net/dor/20.1001.1.2008692.1401.15.1.4.8>

چکیده

پنبه (*Gossypium*) شامل ۵ گونه تتراپلوئید ($2n=4x=52$) و بیش از ۴۵ گونه دیپلوئید ($2n=2x=26$) است. گونه‌های تتراپلوئید پنبه از طریق هیبریدیزاسیون بین یک ژنوم A- و یک گونه ژنوم D- شکل گرفته‌اند. آنالیز نقشه‌های ژنتیکی نشان داد که ژنوم‌های دیپلوئید A- و D- سطح بالایی از همثانی با ژنوم‌های At- و Dt- در پنبه تتراپلوئید دارند. دو گونه دیپلوئید (*G.longicalyx* و *G.stocksii*) متعلق به ژنوم D در گیاه پنبه می‌باشند که نقش بسیار مهمی به عنوان منبع زیستی در برنامه‌های تحقیقاتی و اصلاحی پنبه دارد. این مطالعه به منظور مقایسه توالی ژنوم کلروپلاستی، تعداد نوکلئوتیدها، نواحی بین‌ژنی، توالیهای تکراری و آنالیز کاربرد کدونی بین دو گونه دیپلوئید (*G.longicalyx* و *G.stocksii*) و دو گونه تتراپلوئید پنبه (*G.hirsutum* و *G. barbadense*) صورت گرفت. ژنوم کامل کلروپلاست *G.longicalyx* دارای ۲۴۸، ۱۶۰ جفت باز است که بلندتر از *G.stocksii* و کوتاهتر از دو گونه تتراپلوئید بوده، ژنوم کامل کلروپلاست *G.stocksii* دارای ۱۵۹، ۰۳۹ جفت باز است که کوتاه‌تر از *G.longicalyx* و دو گونه تتراپلوئید بوده همچنین در هر دو گونه دو ناحیه تک نسخه‌ای توسط دو ناحیه تکرار معکوس جدا شده‌اند. ژن‌های کد شده توسط کلروپلاست *G.longicalyx* و *G.stocksii* شامل ۷۹ ژن کدکننده پروتئین، ۴ ژن rRNA، ۳۰ ژن tRNA، ۱۱۳ ژن تک و ۲۰ ژن مضاعف در IR قرار گرفته‌اند. در دو گونه دیپلوئید مورد بررسی ۵ ژن با عملکرد ناشناخته ژن‌های ycf کشف شدند که برای گیاهان ضروری در نظر گرفته شده و در میان گونه‌ها بسیار محافظت شده هستند.

کلیدواژه‌ها: آنالیز ساختاری، بیوانفورماتیک، کلروپلاست پنبه، *G.stocksii*، *G.longicalyx*

مقدمه

سال به میزان زیادی از آن استفاده گردیده است. پنبه ابتدا برای استفاده از فیبر آن در صنعت نساجی و بعد برای روغن دانه کشت می‌گردید [۱]. قدمت مطالعه ژنوم کلروپلاست به سال ۱۹۵۰ بر

پنبه یک گیاه بوته‌ای از خانواده پنیرکیان است. این گیاه از اولین گیاهانی است که بوسیله انسان کشت گردیده و حدود ۴۰۰۰

nephrolepis، توالی‌های IR بزرگی در نقاط وارونگی واقع شده‌اند. ژنوم کلروپلاست *A.nephrolepis* با درخت نراد کره‌ای که نزدیک به گونه مربوطه است یکسان می‌باشد. گزارشات در مطالعه توالی کامل کلروپلاست ۵ گونه *Epimedium* حاکی از آن است که رونویسی ژن *trnQ-UUG* در ۵ ژنوم کلروپلاست *Epimedium* که در دیگر دولپه‌ها پیدا نشد در ناحیه LSC اتفاق افتاده است و توالی نوکلئوتیدی هر کپی در ناحیه LSC در هر ۵ گونه یکسان بود [۱۱-۱۲]. طلعت و همکاران (۱۳) اعلام کردند ژنوم کلروپلاست جامع‌ترین ژنوم مورد مطالعه گیاه است و بدلیل ساختار ملکولی منحصر به فرد و وراثت تک والدی ویژگی‌های زیادی برای تجزیه و تحلیل‌های تکاملی را دارا می‌باشد. با بررسی ژنوم کلروپلاست *Gossypium thurberi* توسط طلعت و وانگ [۱۰] اعلام کردند از بین ژن‌های پلاستییدی فقط ۱۸ ژن پیدا شد که ۱ تا ۲ اینترون داشتند که در مقایسه با دو آوتراپلوئید زراعی *ips18* تنها ژن دونسخه‌ای در *G.thurberi* بود. با وجود سطح زیاد حفاظتی SSRهای ژنوم کلروپلاست این SSRها بعلت بازده بالای آن‌ها در مقابله با SSRهای ژنومی برای تجزیه تنوع ژنتیکی مفید هستند. میزان کم GC یک مشخصه مهم ژنوم پلاستییدی است که بعد از اندوسیمپوتیک بوسیله ویرایش و همانندسازی DNA انجام شده است. طلعت و همکاران [۱۳] اعلام کردند که گسترش روز افزون حجم عظیم داده‌های ژنومی و نیاز به ذخیره، بازیابی و تجزیه و تحلیل داده‌ها موجب پیدایش علم بیوانفورماتیک گردیده است. ژنوم توالی‌یابی شده کامل کلروپلاست یک منبع غنی از اطلاعات است که می‌تواند پاسخ بسیاری از سوالات مربوط به مطالعات فیلوژنتیکی را حل کند. این تحقیق برای شناسایی نواحی اینترونی در ژن‌های ژنوم کلروپلاست دو گونه دیپلوئید، سازماندهی توالی‌های تکراری و مقایسه بین ژنوم‌های موجود، شناسایی پلی‌مورفیسم در بین گونه‌های مختلف پنبه، مطالعه ساختار ژنوم، محتوای ژن، سازماندهی توالی‌های تکراری و کدون یوسیح مورد مطالعه قرار گرفت.

مواد و روش‌ها

— دانلود توالی کامل ژنوم

می‌گردد زمانی که زیست‌شناسان گیاهی برای اولین بار کشف کردند که کلروپلاست دارای DNA مستقل است، کروموزوم‌های پلاستییدی بطور نسبی تعداد ژن کم و سطح حفاظت بالایی دارند [۲-۴]. تنها ۴۴ ژن رمزکننده پروتئین یافت شده است که بین ۱۵ ژنوم کلروپلاست به نمایندگی همه اجداد اصلی موجودات فتوسنتزی مشترک باشد. ظاهراً در طول تکامل ژنوم پلاستییدی تعداد کمی از ژن‌ها بدست آمده است اما اکثریت تغییرات محتوای ژن نشان‌دهنده فقدان ژن است که برخی از آن‌ها بطور مستقل در اجداد مختلف از دست رفته‌اند [۵]. بیوانفورماتیک استفاده از تکنیک‌های محاسباتی برای تجزیه و تحلیل اطلاعات وابسته به زیست‌شناسی ملکولی در مقیاس بزرگ است. در حال حاضر بیوانفورماتیک بعنوان یک رشته در زیست‌شناسی ملکولی مطرح است و شامل طیف گسترده موضوع‌ها از زیست‌شناسی ساختاری و ژنومیک برای مطالعه بیان ژن است [۱]. تجزیه و تحلیل در بیوانفورماتیک عمدتاً بر سه نوع مجموعه داده بزرگ موجود در زیست‌شناسی ملکولی شامل ساختار ماکروملکول‌ها، توالی‌های ژنوم و نتایج حاصل از آزمایش‌های ژنتیک کاربردی اطلاعات اضافی مانند مقاله‌های علمی و روابط داده‌ها از مسیرهای متابولیکی، درخت تاکسونومی و شبکه‌های تعامل پروتئین-پروتئین تمرکز می‌کند. بیوانفورماتیک طیف گسترده‌ای از تکنیک‌های محاسباتی شامل توالی و تطابق ساختار، طراحی پایگاه داده و داده کاوی، ماکروملکول هندسی، ترکیب درخت فیلوژنتیکی، پیش‌بینی ساختار و عملکرد پروتئین، یافتن ژن و خوشه‌بندی بیان داده به کار می‌رود [۶]. مطالعه ترجیح کدونی فاکتوری تعیین کننده در بیان ژن و عملکرد سلولی است که مستقیماً در فرایندسازی RNA، ترجمه پروتئین و ساختارسازی پروتئین موثر است. ترجیح کدونی اشاره به منشا، الگوی جهشی و تکامل گونه‌ها یا ژن‌ها دارد [۷]. ترجیح کدونی می‌تواند تحت تاثیر ترکیب نوکلئوتیدی ژنوم (ژنوم‌های غنی از AT تمایل به استفاده از کدون‌های منتهی به A یا T دارند) قرارگیرد [۸]. Choudhury و همکاران [۹] اعلام کردند تجزیه و تحلیل الگوی کدون یوسیح برای فهم سازماندهی ژنتیکی و ملکولی یک ژن مهم است. کدون یوسیح همچنین به بیان ژن‌های هترولوگ و طراحی پرایمر و ژن مصنوعی کمک می‌کند [۱۰]. براساس مطالعات پیشین در ژنوم کلروپلاست گیاه *Abies*

شده از مرکز ملی اطلاعات بیوتکنولوژی با رنگ مربوط به گروه خود هایلایت گردید. پس از مشخص شدن موقعیت همه ژن‌ها روی توالی ژنوم مربوطه اطلاعات مربوطه به هر ژن از سایت مرکز ملی اطلاعات بیوتکنولوژی و از صفحه بانک ژن مربوط به هر ژنوم از قسمت توالی کدکننده پروتئین بدست آمد و در کنار ژن مربوطه نوشته شد. اینترون‌های مربوط به ژن‌های حاوی اینترون نیز با رنگ مجزا مشخص شدند. در نتیجه موقعیت ژن‌ها نواحی بین‌ژنی و همچنین ژن‌هایی که دارای هم‌پوشانی هستند نیز مشخص گردید. علاوه بر این طول نواحی رمزکننده (CDS)، (rRNA، tRNA) و نواحی غیر رمزکننده (اینترون و IGS) برای هر چهار ژنوم بطور جداگانه بدست آمده است. درصد چهار نوکلئوتید A، T، C و G و همچنین درصد AT و GC همه قسمت‌های ژنوم‌ها اعم از پروتئین، اینترون، نواحی بین‌ژنی و غیره جهت تکمیل جدول محتوای GC از طریق نرم‌افزار Visual Bioinformatics (V 2.1.0) بدست آورده شد. به بخش other tools وارد شده و توالی ژنوم کلروپلاست گونه مربوطه وارد گردید. روی گزینه start کلیک کرده در نتیجه درصد هر ۴ نوکلئوتید به همراه درصد AT و درصد GC در قسمت پایین صفحه ظاهر شدند. از نرم‌افزار آنالین Nucleic Acid Sequence Massager برای حذف اضافات، علامت‌ها و فاصله‌هایی که در توالی فستا وجود داشت استفاده شد.

نتایج

– محتوای GC

محتوا و نتایج GC برای چهار گونه مورد بررسی در این تحقیق در جدول‌های ۱، ۲، ۳ و ۴ آورده شده است. طبق این نتایج، محتوای GC در ژنوم کلروپلاست *G.longicalyx* ۲/۳۷ درصد است که مشابه به دیگر گونه‌ها مانند ۲/۳۷ درصد در *G.stocksii*، ۲/۳۷ درصد در *G.barbadense* و ۲/۳۷ درصد در *G.hirsutum* می‌باشد. هر دو نواحی کدینگ و غیرکدینگ در محتوای GC برای *G.longicalyx* بترتیب با ۳۳/۱ درصد و ۳۳/۰۸ درصد که *G.stocksii* است. ناحیه کدینگ در محتوای GC در *G.stocksii*، ۳۳/۱ درصد و ناحیه غیرکدینگ ۳۳/۰۸ درصد است. ناحیه کدینگ در محتوای GC در *G.barbadense*، ۴۰/۳ و ناحیه غیرکدینگ ۳۳/۱ درصد

توالی کامل ژنوم کلروپلاست دو گونه دیپلوئید پنبه (*G.stocksii* و *G.longicalyx*) با دو گونه تتراپلوئید (*G.hirsutum* و *G.barbadense*) در این پژوهش مورد مقایسه قرار گرفت. توالی کامل ژنوم کلروپلاست *G.stocksii*، *G.longicalyx*، *G.hirsutum* و *G.barbadense* بترتیب با شماره دسترسی‌های NC_023217.1، NC_023216.1، NC_008641.1 و HQ_901197.1 از NCBI برداشت گردید.

– نرم افزارهای مورد استفاده

با استفاده از نرم‌افزار Fast PCR توالی‌های کامل ژنتیکی به فرمت فستا تبدیل شده است. نقشه ساختار ژنوم کلروپلاست و توزیع ژن با استفاده از نرم‌افزار OGDRAW(V1.2) (<http://ogdraw.mpimpgolm.mpg.de>) با وارد کردن شماره دسترسی ژنوم هر گونه رسم گردید. شماره دسترسی ژنوم هر چهار گونه در قسمت specify an accession number وارد شده است. پس از کلیک بر روی گزینه submit sequence صفحه بعد ظاهر شده و ویژگی‌های نقشه ژنومی از جمله فتوسیستم I و II، ORF، RNAهای ریبوزومی، اینترون‌ها و ... انتخاب گردید. همچنین نوع فایل گرافیکی خروجی و کیفیت آن نیز تعیین گردید و بعد از انتخاب گزینه create map نقشه ژنومی مورد نظر ایجاد و ذخیره گردید. برای تفسیر و حاشیه‌نویسی ژن‌های موجود در توالی‌های ژنوم‌ها از نرم‌افزار DOGMA (<http://dogma.cccb.utexas.edu>) استفاده شد. این نرم‌افزار اسم ژن‌های ژنوم هر کدام از گونه‌های مورد بررسی را داده و ژن‌های مربوط به گروه‌های مختلف ژنی با توجه به رنگ آن‌ها تشخیص داده شد. شناسایی موقعیت تک تک ژن‌ها در هر چهار ژنوم کلروپلاست مورد بررسی با توجه به اطلاعاتی که در بانک ژن مربوط به هر ژنوم در سایت مرکز ملی اطلاعات بیوتکنولوژی وجود دارد با استفاده از تعیین محل دقیق و تعداد دقیق نوکلئوتید هر ژن در توالی و مشخص کردن آگزون، اینترون و سایر نواحی ژنوم به صورت هایلایت شده و دستی بدون استفاده از نرم افزار انجام گردید. ابتدا توالی کل هر کدام از ژنوم‌های کلروپلاست مورد بررسی از سایت مرکز ملی اطلاعات بیوتکنولوژی برداشت و ذخیره شد. سپس برای هر گروه ژنی رنگ مجزا انتخاب شد و موقعیت هریک از ژن‌ها روی توالی برداشت

می‌باشد. ناحیه کدینگ در محتوای GC در *G.hirsutum*، ۴۰/۴ و ناحیه غیرکدینگ ۳۳/۲ درصد است. تغییرات در محتوای GC در میان چهار ناحیه مختلف در DNA در *G.longicalyx*، *G.stocksii*، *G.barbadense* و *G.hirsutum* مشاهده شد و در چهار گونه IR غنی‌ترین بود بترتیب با درصدهای (۴۲/۹۵)، (۴۲/۹۸)، (۴۳/۱) و (۴۳/۵) مفروض است که ژن‌های ریبوزومی (*tRNA4.5*, *tRNA5*, *tRNA16*, *tRNA23*) و نواحی کدینگ برای محتوای GC بالا در IR مسئول هستند. توزیع محتوای GC هر ناحیه با دیگر گونه‌ها مشابه است. در

محتوای GC (۵۵/۴۸ درصد) و ژن‌های پروتئین پایین‌ترین محتوای GC (۳۹/۱۴ درصد) را دارند. مشابه با آن در *G.stocksii*، *G.barbadense* و *G.hirsutum* در ناحیه کدینگ ژن‌های *tRNA* بترتیب با درصدهای ۵۵/۴، ۵۵/۵ و ۳۸/۳، ۳۹/۱۴ و ژن‌های پروتئین بترتیب با ۳۸/۳ و ۳۸/۲ درصد پایین‌ترین محتوای GC را دارند. نواحی غیرکدینگ یک تکامل سریع را تجربه کردند و بنابراین ناحیه غیرکدینگ در GC از ناحیه کدینگ غنی‌تر است.

جدول ۱: محتوای GC برای *G.longicalyx*

	Coding Region			Non Coding Region			Complete Genome	LSC	SSC	IR	
	Protein	tRNA	rRNA	Total	IGS	Intron	Total				
Length	۸۳۷۴۰	۲۷۹۵	۹۰۵۰	۹۵۵۸۵	۴۷۳۸۳	۱۷۲۸۰	۶۴۶۶۳	۱۶۰۲۴۸	۸۸۷۴۸	۲۰/۲۳۲	۲۵۶۳۴
Proportion	۵۲/۲۵	۱/۷۴	۵/۶۴	۵۹/۶۴	۲۹/۵۶	۱۰/۷۸	۴۰/۳۵	۱۰۰/۰۰	۵۵/۳۸	۱۲/۶۲	۱۵/۹۹
T%	۳۱/۰۲	۲۲/۱۹	۲۲/۳۰	۲۷/۱	۳۰/۳۱	۳۲/۲۸	۳۳/۷۳	۳۱/۸	۳۳/۱۷	۳۴/۴۶	۲۸/۵۰
A%	۲۹/۸۱	۲۴/۵۹	۲۲/۱۹	۲۶/۲	۲۸/۹۵	۳۰/۹۷	۳۳/۱۹	۳۰/۸	۳۱/۵۹	۳۳/۹۲	۵۶/۲۸
C%	۲۰/۰۳	۲۶/۱۲	۲۷/۷۵	۱۵/۷	۲۰/۵۶	۱۹/۱۲	۱۶/۹۰	۱۹/۱	۱۸/۱۹	۱۶/۵۴	۲۰/۶۲
G%	۱۸/۴۳	۲۷/۱۵	۲۷/۶۹	۱۷/۴	۱۹/۷۵	۱۷/۶۳	۱۶/۱۸	۱۸/۳	۱۷/۰۵	۱۷/۰۹	۲۲/۳۳
A+T%	۶۰/۸۶	۴۸/۰۹	۴۴/۵۲	۵۳/۳	۵۹/۶۵	۶۳/۲۵	۶۶/۹۲	۶۲/۷	۶۴/۸۱	۶۸/۳۸	۵۷/۰۵
C+G%	۳۹/۱۴	۵۱/۹۱	۵۵/۴۸	۱/۳۳	۴۰/۳۵	۳۶/۷۵	۳۳/۰۸	۳۷/۲	۳۵/۱۹	۳۱/۶۲	۴۲/۹۵

جدول ۲: محتوای GC برای *G.stocksii*

	Coding Region			Non Coding Region			Complete Genome	LSC	SSC	IR	
	Protein	tRNA	rRNA	Total	IGS	Intron	Total				
Length	۸۳۰۷۵	۲۷۳۵	۹۰۵۰	۹۴۸۶۰	۴۵۵۴۷	۱۸۶۳۲	۶۴۱۷۹	۱۵۹۰۳۹	۸۸۷۶۲	۱۹۰۹۳	۲۵۶۹۲
Proportion	۵۹/۶۴	۱/۷۱	۵/۶۹	۵۹/۶۴	۲۸/۶۳	۱۱/۷۱	۴۰/۳۵	۱۰۰/۰۰	۵۵/۸۱	۱۲/۰۰	۱۶/۰۹
T%	۳۱/۰۲	۲۲/۱۹	۲۲/۳۰	۲۷/۱	۳۰/۲۸	۳۲/۲۹	۳۳/۷۴	۳۱/۸	۳۳/۲۱	۳۴/۴۵	۲۸/۴۸
A%	۲۹/۸۴	۲۴/۵۸	۲۲/۱۹	۲۶/۳	۲۸/۹۵	۳۰/۹۳	۳۳/۱۸	۳۰/۸	۳۲/۰۲	۳۳/۹۳	۲۸/۵۸
C%	۲۰/۷۰	۲۶/۱۶	۲۷/۷۵	۱۵/۶	۲۰/۵۸	۱۹/۱۲	۱۶/۹۰	۱۹/۱	۱۸/۱۷	۱۶/۵۵	۲۰/۶۳
G%	۱۸/۴۴	۲۷/۱۶	۲۷/۶۹	۱۷/۴	۱۹/۷۵	۱۷/۶۳	۱۶/۱۸	۱۸/۳	۱۷/۰۸	۱۵/۰۵	۲۲/۳۱
A+T%	۶۰/۸۶	۴۸/۰۹	۴۴/۵۲	۵۳/۳	۵۹/۶۵	۶۳/۲۵	۶۶/۹۲	۶۲/۳	۶۴/۵۲	۶۸/۳۲	۵۷/۰۲
C+G%	۳۹/۱۴	۵۱/۹۱	۵۵/۴۸	۳۳/۱	۴۰/۳۵	۳۶/۷۵	۳۳/۰۸	۳۷/۴	۳۵/۴۸	۳۱/۶۸	۴۲/۹۸

جدول ۳: محتوای GC برای *G.barbadense*

	Coding Region			Non Coding Region			Complete Genome	LSC	SSC	IR	
	Protein	tRNA	rRNA	Total	IGS	Intron	Total				
Length	۷۸۶۷۵	۲۷۹۱	۹۰۵۰	۹۰۵۱۶	۴۸۵۵۶	۲۱۲۴۵	۶۹۸۰۱	۱۶۰۳۱۷	۸۸۸۹۷	۲۰۰۳۶	۲۵۶۹۲
Proportion	۴۹/۰۷	۱/۷۴	۵/۶۴	۵۶/۶۶	۳۰/۲۸	۱۳/۲۷	۴۳/۵۳	۱۰۰/۰۰	۵۵/۴۵	۱۲/۴۹	۱۶/۰۳
T%	۳۱/۵۰	۲۳/۳	۲۲/۳	۳۰/۴	۳۴/۳	۳۲/۰۰	۳۳/۷	۳۱/۸	۳۳/۱	۳۴/۴	۲۸/۵
A%	۳۰/۲	۲۳/۹	۲۲/۳	۲۹/۳	۳۴/۱	۳۱/۲	۳۳/۲	۳۰/۹	۳۱/۷	۳۳/۸	۲۸/۵

	Coding	Region	-	-	Non	Coding	Region	-	-	-	-
	Protein	tRNA	rRNA	Total	IGS	Intron	Total	Complete Genome	LSC	SSC	IR
C%	۱۹/۶	۲۶/۷	۲۷/۷	۲۰/۶	۱۶/۰	۱۹/۰	۱۶/۸	۱۹/۰	۱۸/۱	۱۶/۶	۲۱/۶
G%	۱۸/۷	۲۶/۱	۲۷/۷	۱۹/۷	۱۵/۶	۱۷/۸	۱۶/۲	۱۸/۲	۱۷/۱	۱۵/۳	۲۱/۶
A+T%	۶۱/۷	۴۷/۷	۴۴/۶	۵۹/۷	۶۸/۴	۶۳/۲	۶۶/۹	۶۲/۳	۶۴/۸	۶۸/۱	۵۶/۹
C+G%	۳۸/۳	۵۲/۸	۵۵/۲	۴۰/۳	۳۱/۶	۳۶/۸	۳۳/۱	۳۷/۲	۳۵/۲	۳۱/۹	۴۳/۱

جدول ۴: محتوای GC برای *G.hirsutum*

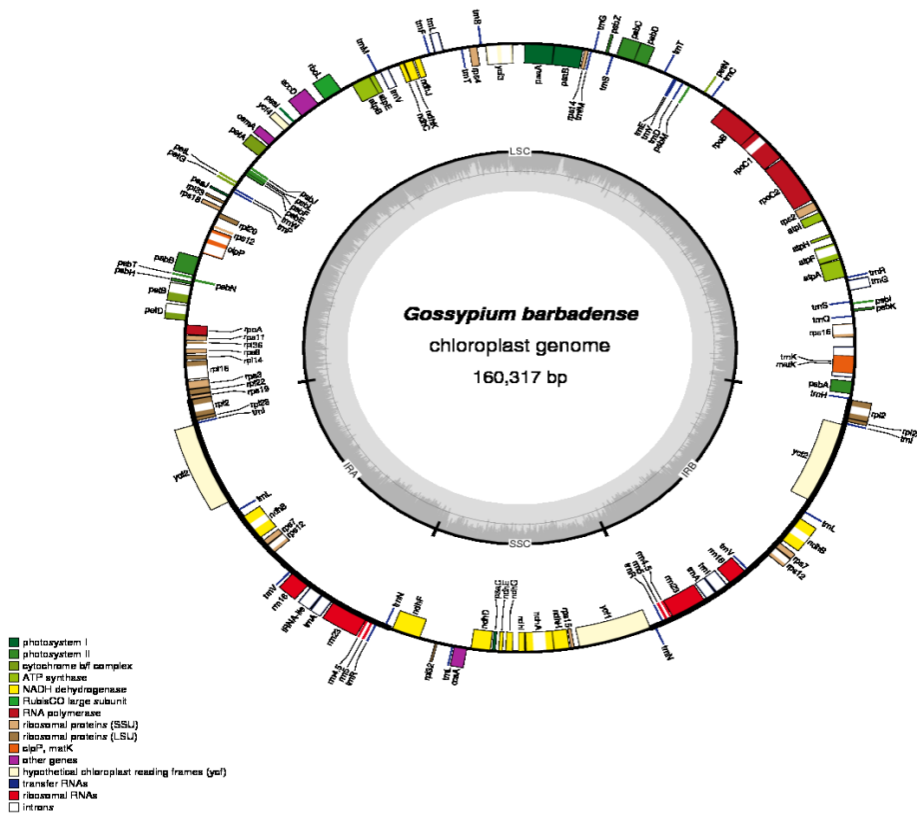
	Coding	Region	-	-	Non	Coding	Region	-	-	-	-
	Protein	tRNA	rRNA	Total	IGS	Intron	Total	Complete Genome	LSC	SSC	IR
Length	۷۵۸۳۱	۲۸۰۱	۹۰۴۸	۹۰۳۸۰	۴۸۸۹۸	۲۱۱۲۳	۶۹۹۲۱	۱۶۰۳۰۱	۸۸۸۶۹	۲۰۵۰۹	۲۵۴۶۵
Proportion	۴۸/۹۸	۱/۷۴	۵/۶۴	۵۶/۳۸	۳۰/۴۴	۱۳/۲۰	۴۳/۶۱	۱۰۰/۱۰۰	۵۵/۴۳	۱۲/۷۹	۱۵/۸۸
T%	۳۱/۴	۲۲/۸	۲۲/۳	۳۰/۲	۳۴/۴	۳۱/۹	۳۳/۷	۳۴/۷	۳۳/۲	۳۳/۹	۲۸/۵
A%	۳۰/۴	۲۳/۹	۲۲/۳	۲۹/۴	۳۳/۹	۳۱/۳	۳۳/۲	۳۱/۰	۳۱/۶	۳۴/۴	۲۸/۵
C%	۱۹/۲	۲۷/۱	۲۷/۷	۲۰/۳	۱۶/۰	۱۸/۹	۱۶/۹	۱۸/۸	۱۸/۱	۱۵/۱	۲۱/۵
G%	۱۹/۰	۲۶/۲	۲۷/۷	۲۰/۱	۱۵/۶	۱۷/۹	۱۶/۳	۱۸/۴	۱۷/۱	۱۶/۵	۲۱/۵
A+T%	۶۱/۸	۴۶/۷	۴۴/۵	۵۹/۶	۶۸/۴	۳۶/۲	۶۶/۸	۶۲/۸	۶۴/۸	۶۸/۳	۵۷/۰
C+G%	۳۸/۲	۵۳/۰	۵۵/۵	۴۰/۴	۳۱/۶	۳۶/۶	۳۳/۲	۳۷/۲	۳۵/۲	۳۱/۵	۴۳/۷

ژنوم کلروپلاست *G.hirsutum*

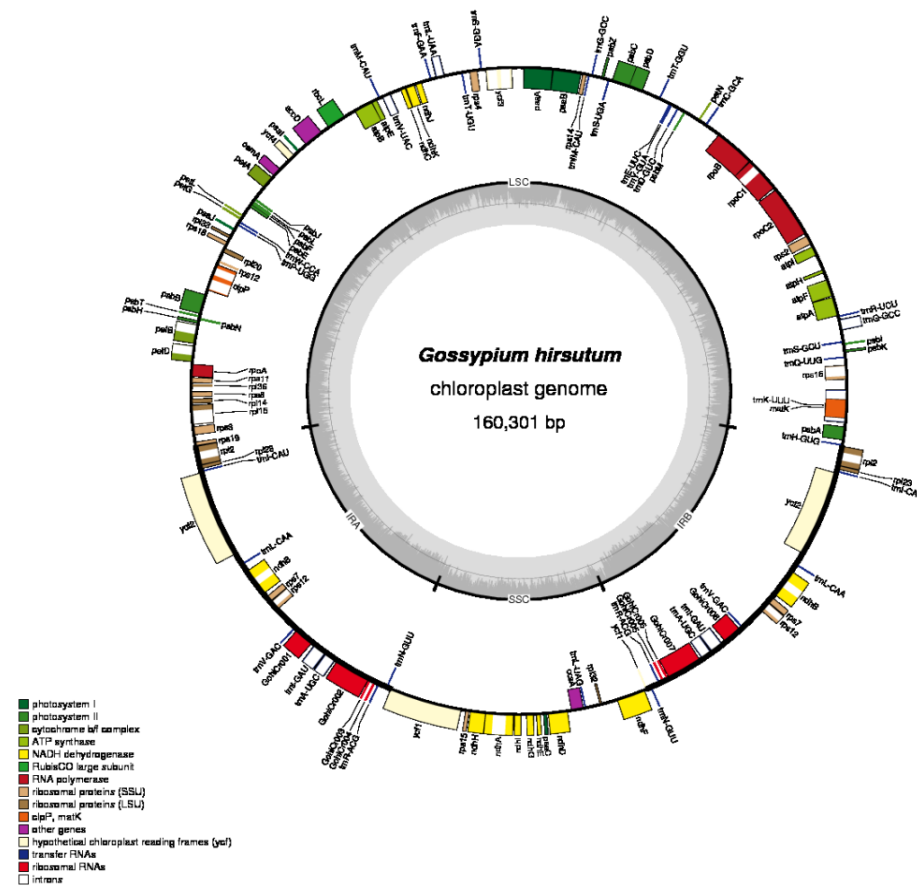
ژنوم کامل یک ملکول DNA حلقوی *G.hirsutum* دارای ۱۶۰۳۰۱ جفت باز است که بلندتر از *G.stocksii* و *G.longicalyx* ولی کوتاه تر از *G.barbadense* است. دو ناحیه کپی منفرد توسط دو ناحیه تکرار معکوس جدا شده اند. کل ژنوم این گونه آنالیز شده و در جدول ۴ آورده شده است. با توجه به نتایج حاصل از این جدول LSC و SSC بترتیب دارای ۸۸۸۶۹ و ۲۰۵۰۹ جفت باز و دو ناحیه تکرار معکوس (IR) که هر کدام دارای ۲۵۴۶۵ جفت باز هستند. ناحیه کدینگ دارای طول ۹۰۳۸۰ جفت باز بوده و ۵۶/۴۷ درصد از ژنوم کامل پیلستیدیک را شامل می شود که مشابه *G.barbadense* با ۵۶/۳۸ درصد درحالیکه کوچک تر از *G.longicalyx* با ۵۹/۶۴ و *G.stocksii* با ۵۹/۶۴ درصد است. ژنوم *G.barbadense* برای پروتئین ۴۹/۰۷ درصد، ژن های tRNA، ۱/۴۷ درصد و tRNA، ۵/۶۴ درصد را کد می کنند. طول ناحیه غیرکدینگ ۶۹۸۰۱ جفت باز و ۴۳/۵۳ درصد است. نسبت فضاهای بین ژنی و اینترون بترتیب ۱۳/۲۷ و ۳۰/۲۸ درصد هستند.

ژنوم کلروپلاست *G.barbadense*

ژنوم کامل یک ملکول DNA حلقوی *G.barbadense* دارای ۱۶۰۳۱۷ جفت باز است که بلندتر از *G.hirsutum*، *G.stocksii* و *G.longicalyx* است. دو ناحیه کپی منفرد توسط دو ناحیه تکرار معکوس جدا شده اند. کل ژنوم این گونه آنالیز شده و در جدول ۳ آورده شده است. با توجه به نتایج حاصل از این جدول LSC و SSC بترتیب دارای ۸۸۸۶۹ و ۲۰۵۰۹ جفت باز و دو ناحیه تکرار معکوس (IR) که هر کدام دارای ۲۵۴۶۵ جفت باز هستند. ناحیه کدینگ دارای طول ۹۰۳۸۰ جفت باز بوده و ۵۶/۳۸ درصد از ژنوم کامل پیلستیدیک را شامل می شود که مشابه *G.barbadense* با ۵۶/۴۶ درصد درحالیکه کوچک تر از *G.longicalyx* با ۵۹/۶۴ و *G.stocksii* با ۵۹/۶۴ درصد است. ژنوم *G.barbadense* برای پروتئین ۴۸/۹۸ درصد، ژن های tRNA، ۱/۷۴ درصد و tRNA، ۵/۶۴ درصد را کد می کنند. طول ناحیه غیرکدینگ ۶۹۹۲۱ جفت باز و ۴۳/۶۱ درصد است. نسبت فضاهای بین ژنی و اینترون بترتیب ۱۳/۲۰ و ۳۰/۴۴ درصد هستند.



شکل ۱: ژنوم کلروپلاست *G. barbadense*



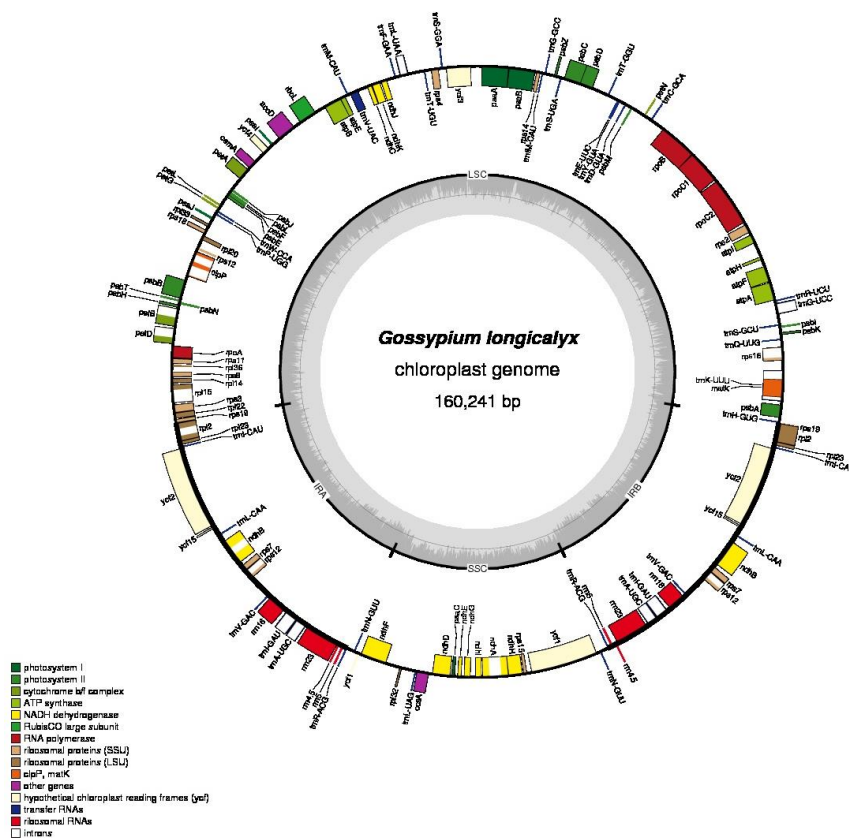
شکل ۲: ژنوم کلروپلاست *G. hirsutum*

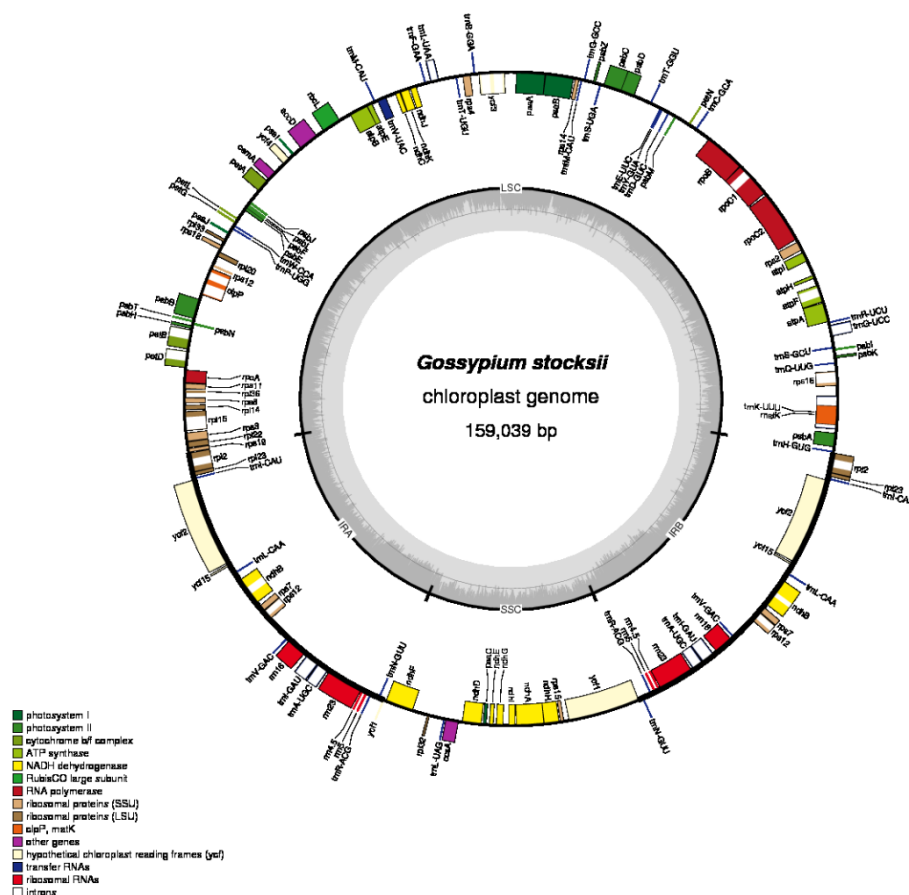
- ژنوم کلروپلاست *G. longicalyx*

ژنوم کامل یک مولکول DNA حلقوی *G. longicalyx* دارای ۱۶۰/۲۴۱ جفت باز است که بلندتر از *G. stocksii* و کوتاهتر از *G. hirsutum* و *G. barbadense* است. دو ناحیه کپی منفرد توسط دو ناحیه تکرار معکوس جدا شده‌اند. کل ژنوم این گونه آنالیز شده و در جدول ۱ آورده شده است. با توجه به نتایج حاصل از این جدول LSC و SSC بترتیب دارای ۸۸۷۴۸ و ۲۰۲۳۲ جفت باز و دو ناحیه تکرار معکوس (IR) که هر کدام دارای ۲۵۶۳۴ جفت باز هستند. ناحیه کدینگ دارای طول ۹۵۵۸۵ جفت باز بوده و ۵۹/۶۴ درصد از ژنوم کامل پیلاستیدیک را شامل می‌شود که مشابه *G. stocksii* با ۵۹/۶۴ درصد درحالی‌که بزرگ‌تر از *G. barbadense* با ۵۶/۴۶ و *G. hirsutum* با ۵۶/۳۸ درصد است. ژنوم *G. longicalyx* برای پروتئین ۵۲/۲۵ درصد، ژن‌های *tRNA* ۱/۷۴ درصد و *rRNA* ۵/۶۴ درصد را کد می‌کند. طول ناحیه غیرکدینگ ۴۶۶۳ جفت باز و ۴۰/۳۵ درصد است. نسبت فضاهای بین ژنی و اینترون بترتیب ۲۹/۵۶ و ۱۰/۷۸ درصد هستند.

- ژنوم کلروپلاست *G. stocksii*

ژنوم کامل یک مولکول DNA حلقوی *G. stocksii* دارای ۱۵۹/۰۳۹ جفت باز است که کوتاه‌تر از *G. longicalyx*، *G. hirsutum* و *G. barbadense* است. دو ناحیه کپی منفرد توسط دو ناحیه تکرار معکوس جدا شده‌اند. کل ژنوم این گونه آنالیز شده و در جدول ۲ آورده شده است. با توجه به نتایج حاصل از این جدول LSC و SSC بترتیب دارای ۸۸۷۶۲ و ۱۹۰۹۳ جفت باز و دو ناحیه تکرار معکوس (IR) که هر کدام دارای ۲۵۵۹۲ جفت باز هستند. ناحیه کدینگ دارای طول ۹۴۸۶۰ جفت باز بوده و ۵۹/۶۴ درصد از ژنوم کامل پیلاستیدیک را شامل می‌شود که مشابه *G. longicalyx* با ۵۹/۶۴ درصد درحالی‌که بزرگ‌تر از *G. barbadense* با ۵۶/۴۶ و *G. hirsutum* با ۵۶/۳۸ درصد است. ژنوم *G. stocksii* برای پروتئین ۵۹/۶۴ درصد، ژن‌های *tRNA* ۱/۷۱ درصد و *rRNA* ۵/۶۹ درصد را کد می‌کند.

شکل ۳: ژنوم کلروپلاست *G. longicalyx*

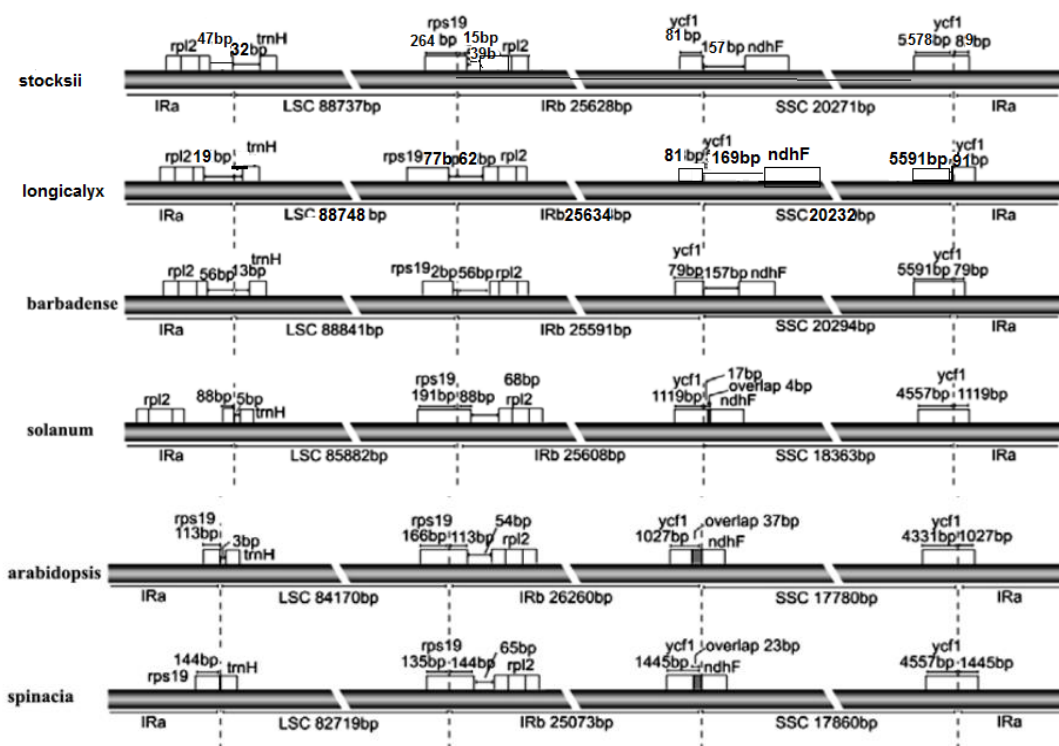


شکل ۴: ژنوم کلروپلاست *G.stocksii*

نشان داد که *rps19* در محل اتصال IRa و LSC دو تایی شد که این دو تایی شدن در گیاهان خیلی متداول است. در مرز IRb و *G.stocksii* SSC با *G.longicalyx* مشابه و در جهت مخالف است که دارای ۸۱ جفت باز از قطعه *ycf1* در مرز IRb است. *Spinacia*، *Arabidopsis* و *Solanum* هم‌پوشانی مشابهی از *ycf1* و *ndhf* در محل اتصال دارند. بزرگترین هم‌پوشانی در *Arabidopsis* با ۳۷ جفت باز و کوتاهترین در *Solanum* با ۱۷ جفت باز است. *ycf1* در محل اتصال SSC/IRb واقع شده و بنابراین *ycf1* در مرز IRb و SSC دو نسخه‌ای شد. در *spinacia*، *ycf1* بزرگترین دوپلیکیشن با ۱۴۴۵ جفت باز دارد، در *G.barbadense*، *ycf1* کوتاهترین دوپلیکیشن را با ۷۹ جفت باز دارد. *G.longicalyx* در جهت مخالف SSC است همانطور که با ۵ گونه دیگر مقایسه شد *ycf1* در محل اتصال IRb/SSC با ۸۱ جفت باز واقع شد.

– انبساط و انقباض IR

مرز بین دو IR/LSC و دو IR/SSC معمولاً در میان فضاهای گوناگون فرق می‌کند. انبساط و انقباض ناحیه IR اغلب تغییرات طولی بزرگی از ژنوم کلروپلاست در گروه‌های مختلف گیاهی ایجاد می‌کند. مقایسه دقیق مرزهای IR نشان می‌دهد که میزان وسیعی از انقباض و انبساط IR تکاملی متداول هست. بعنوان یک نتیجه، ژن‌های کاذب گوناگونی در ژن‌هایی مانند *rps19* و *ycf1* در مرز بین ناحیه SC و ناحیه IR ایجاد شدند. مرز IR معمولاً بین گونه‌ها مختلف است و انبساط و انقباض IR تا حدی که اندازه ژنوم اهمیت دارد مهم است. انبساط IR اغلب به اندازه بزرگتر ژنوم منجر می‌شود. معمولاً ژن‌های کاذب در اتصال IR و LSC/SSC قرار می‌گیرند. اختلاف محل‌های اتصال در میان *G.stocksii* و ۳ گونه دیگر آنالیز شده و در شکل ۵ آورده شده است. محل‌های اتصال IRb/LSC در *rps19* در *G.stocksii*، *Solanum lecopersium*، *G.barbadense*، *G.longicalyx* و *Arabidopsis thaliana* و *Spinacia oleracea* پیدا شد و



شکل ۵: انبساط و انقباض IR

ژن *trnL-UAA* به گروه I متعلق است. دو ژن *clpP* و *ycf3* و اینترون دارند. ژن *ips12* ژن جدا کننده بی نظیری است که آگزون انتهایی 5' در ناحیه تک نسخه بزرگ دور از آگزون سوم و دوم که مانند مضاعفها در ناحیه تکرار معکوس واقع شده اند و یک مکانیسم trans-splicing بین آگزون I و II برای تولید و رونویسی *ips12* بالغ نیاز دارد. محتوی A-T بالا در موقعیت کدون سوم به وضوح کدون یوسیج را به A یا T منعکس می کند، برای مثال کدون های دژنره شده دو تایی در موقعیت سوم فراوانی کدون یوسیج را از ۳۴-۲۱ درصد برای C یا G و ۷۹-۶۶ درصد برای A یا T را نشان می دهند. کدون های دژنره شده چهار تایی در موقعیت سوم مانند Ala, Thr, Pro, Ile, Val همچنین یک کدون یوسیج قوی به اندازه A یا T از ۶۶ به ۷۹ درصد نشان می دهد. برای کدون های دژنره شده شش تایی مانند Arg, Ser, Leu کدون یوسیج A-T برای هر دو موقعیت کدون سوم و اول مشاهده شد. فراوانی کدون های پایان همچنین به A یا T در هر دو موقعیت سوم و دوم نشانگر ۴۰ درصد برای TAA، ۳۴ درصد برای TGA و ۲۶ درصد برای TAG می باشد. همان نتایج در ژنوم کلروپلاست چندین گیاه عالی دیگر به خوبی ثبت شدند (۶). وجود یک تکرار معکوس بزرگ (IR) بطول ۱۲ تا ۵۰

ژن های کدکننده و اینترون در *G.longicalyx* و

G.stocksii

ژن های کد شده توسط کلروپلاست *G.longicalyx* و *G.stocksii* جدول ۵ آورده شده که در کل، ۷۹ ژن کدکننده پروتئین، ۴ ژن *tRNA*، ۳۰ ژن *tRNA*، ۱۱۳ ژن تک و ۲۰ ژن مضاعف در IR قرار گرفته اند. بر طبق عملکرد ژن همه ژن ها می توانند بعنوان ژن های سیستم ژنتیک عملکردی، سیستم فتوسنتتیک، بیوسنتز و بعضی با عملکرد ناشناخته طبقه بندی شوند. در *G.stocksii* و *G.longicalyx*، ۵ ژن با عملکرد ناشناخته (ژن های *ycf*) کشف شدند و برای گیاهان ضروری در نظر گرفته شده اند که در میان گونه ها بسیار محافظت شدند. ژن *ips12* یک اینترون دارد، (*ips12* بوسیله یک اینترون) بین ۲ قطعه با یک آگزون قرار گرفته در (5'-end) LSC و دیگری در انتهای 3' در IR قرار گرفتند. *Matk* دارای طول ۱/۵ کیلو جفت باز بوده و در اینترون *trnk-UUU* پیدا شد که تنها ژن واقع شده در یک اینترون است و *maturase k* را کد می کند که در این ژن هر دو قطعه حفظ شده است. نوزده ژن شامل یک یا دو اینترون هستند، پنج تا از این اینترون ها، *ndh B*، *rpl2*، *GAU*، *rps12*، *trnL-UAA* و *trnA-UGC* در ناحیه IR واقع شده اند. فقط اینترون

نسبتاً حفظ شده است. به‌هرحال بوسیله جهش‌های معکوس که می‌تواند بوسیله پدیده‌های بازسازی داخل مولکولی واسطه شود اندازه جایگزینی ژن‌های کلروپلاست با موقعیت ژن روی ژنوم مرتبط است، برای مثال ژن‌ها در ناحیه IR در یک مقیاس آهسته‌تر واگرا می‌شوند هم‌چنانکه با ژن‌های واقع شده در SSC و LSC مقایسه شد بیشتر نقاط تکامل که نشانگر فراوانی بالایی از جهش‌ها و جانمایی بازهاست روی فضاها بین ژنی که فاقد واحدهای رونویسی پلی‌سیسترونیک هستند مورد توجه قرار گرفت [۱۴].

کیلو باز یکی از ویژگی‌های ساختاری حفظ شده است و برای تغییرات طول ژنوم شرح داده شد. دو قطعه از IR بوسیله یک ناحیه تک نسخه‌ای بزرگ (LSC) و یک ناحیه تک نسخه‌ای کوچک (SSC) جدا شدند. ادغام حداکثر یا از دست دادن ناحیه IR تنها در *Pinus* و *Medicago* مشاهده شد. محتوی ژن و واحدهای پلی‌سیسترونیک ژنوم کلروپلاست هم‌چنین در میان گونه‌های گیاهان آوندی حفظ شدند که تنها استثنا گیاه انگلی غیرفتوسنتتیک *Epifagus virginiana* است که فاقد چندین ژن مرتبط با فتوسنتز است. دستورات ژنی ژنوم کلروپلاست هم‌چنین

جدول ۵: ژن‌های کدکننده در *G.stocksii* & *G.longicalyx*

Group	Gene Name
Subunit of Acetyl-CoA-carboxylase	<i>AccD</i>
Large subunit of rubisco	<i>RbcL</i>
Subunits of NADH-dehydrogenase	<i>ndhA*</i> , <i>ndhB*§</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>
Subunits of cytochrome b/f complex	<i>petA</i> , <i>petB*</i> , <i>petD*</i> , <i>petG</i> , <i>petL</i> , <i>petN</i> <i>ccsA</i>
Protein name subunits of photosystem I and II	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> , <i>psaB</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
DNA dependt RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1*</i> , <i>rpoC2</i>
Large subunit of ribosome	<i>rpl14</i> , <i>rpl16*</i> , <i>rpl2*§</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23§</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
Small subunit of ribosome	<i>rps11</i> , <i>rps12*§</i> , <i>rps14</i> , <i>rps15*</i> , <i>rps16*</i> , <i>rps18</i> , <i>rps19§</i> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7§</i> , <i>rps8</i>
Others	<i>cemA</i> , <i>clpP**</i> , <i>matK</i>
Function unknown	<i>ycf1§</i> , <i>ycf15§</i> , <i>ycf2§</i> , <i>ycf3**</i> , <i>ycf4</i>
ribosomal RNA gene	<i>rrn16§</i> , <i>rrn23§</i> , <i>rrn4.5§</i> , <i>rrn5§</i>
RNA genes transfer RNA gene	<i>trnA-UGC*§</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnFM-CAU</i> , <i>trnGUCC*</i> , <i>trnG-GCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU§</i> , <i>trnI-GAU*§</i> , <i>trnK-UUU*</i> , <i>trnLCAA§</i> , <i>trnL-UAA*</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU§</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG§</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnTUGU</i> , <i>trnV-GAC§</i> , <i>trnV-UAC*</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>

§نمایانگر ژن‌های موجود در ناحیه IR. * نمایانگر ژن‌های با ۱ ایترون و ** نمایانگر ژن‌های با یک ایترون می‌باشد. *ycf3*, *rpoC1* و *atpF* در *longicalyx* ایترون ندارد. *ndhA* در *stocksii* ایترون ندارد

جدول ۶: اینترون در *G.stocksii* & *G.longicalyx*

Intron	<i>G.longicalyx</i>	<i>G.stocksii</i>	<i>G.hirsutum</i>	<i>G.barbadense</i>	Sequence Identity (%)
<i>trnK-UUU</i>	۲۵۲۹	۲۵۲۷	۲۵۴۲	۲۵۳۵	۹۸/۳۷
<i>rps16</i>	۸۷۰	۸۸۰	۸۷۱	۸۷۰	۹۹/۲۴
<i>trnG-UCC</i>	۸۰۵	۸۸۰	۸۷۱	۸۷۰	۹۸/۲۹
<i>atpF</i>	-	۷۸۹	۸۰۴	۸۰۵	۹۸/۹۷
<i>rpoC1</i>	-	۷۴۶	۷۵۳	۷۵۳	۹۹/۴۲
<i>ycf3-2</i>	-	۷۵۸	۷۷۷	۷۷۷	۹۹/۶۱
<i>ycf3-1</i>	-	۷۵۸	۷۸۹	۷۸۹	۱۰۰
<i>trnL-UAA</i>	۵۹۲	۵۴۴	۵۷۵	۵۸۲	۹۸/۸۰
<i>trnV-UAC</i>	۶۰۶	۶۰۵	۶۱۸	۶۰۹	۹۷/۸۵
<i>rps12-3end</i>	۵۳۶	۵۳۶	۵۳۶	۵۳۶	۹۹/۹۴
<i>clpP2</i>	۹۰۸	۹۰۱	۸۹۱	۸۹۰	۹۹/۵۲
<i>clpP1</i>	۶۹۲	۶۷۲	۶۸۳	۶۷۹	۹۸/۸۳
<i>petB</i>	۷۶۰	۷۵۸	۷۶۰	۷۶۱	۹۹/۵۲
<i>petD</i>	۷۶۸	۷۶۰	۷۵۷	۷۵۴	۹۹/۸۷
<i>rpl16</i>	۱۱۲۳	۱۱۰۰	۱۱۴۰	۱۱۳۵	۹۹/۵۰
<i>rpl2</i>	۶۹۳	۶۸۷	۶۹۵	۶۸۸	۹۹/۵۲
<i>ndhB</i>	۶۸۹	۶۸۹	۶۸۳	۶۸۳	۹۹/۹۵
<i>trnI-GAU</i>	۹۵۵	۹۴۸	۹۵۴	۹۵۹	۹۹/۴۸
<i>trnA-UGC</i>	۷۹۷	۷۹۷	۷۹۷	۷۹۵	۹۹/۷۹
<i>ndhA</i>	۱۰۸۴	-	۱۰۷۶	۱۰۷۶	۹۹/۸۷

بحث

گزارش کردند که SSR در کلروپلاست رایج می‌باشد حتی اگر در ژنوم کلروپلاست بروز نکند. توالی‌های تکرار ساده در کلروپلاست‌ها برای مطالعات تنوع ژنتیکی مهم و قابل استفاده می‌باشند و در *Maritime pine* مورد استفاده قرار گرفته‌اند [۱۵]. در هر یک از ژنوم‌های کلروپلاست یک ژن با عملکرد ناشناخته بنام *ycf* شناسایی و بعنوان ژن‌های ضروری گیاهان در نظر گرفته شده که در میان گونه‌ها بشدت محافظت شده‌اند. با توجه به نتایج بدست آمده در ژنوم کلروپلاست دو ژن حائز اهمیت می‌باشند. یکی از آنها ژن *rps12* است و به دو قطعه که یکی با آگزون انتهایی ۵ واقع در LSC و دیگری با آگزون انتهایی ۳ واقع در IR (توسط یک اینترون) جدا می‌گردد که تنها ژن شکافته شده در تمام گونه‌ها می‌باشد، ژن دیگر *matk* که ۱/۵ کیلوباز طول داشته و در ناحیه اینترون *trnK/UUU* شناسایی شده است که تنها ژن واقع در اینترون بوده و پروتئین *maturase K* را رمزدهی می‌کند و دارای هر دو قطعه متغیر و محافظت شده بوده و در مطالعات

ژنوم کلروپلاست مولکول DNA حلقوی با ساختاری چهارجزی و واحدی چهار قسمتی (LSC، SSC) و دو ناحیه تکرار معکوس) می‌باشد که در این ژنوم نواحی کد کننده (*tRNA*، *rRNA*) و توالی‌های کد کننده پروتئینی) و نواحی غیر کد کننده (IGS) و اینترون) درصدی از کل ژنوم را به خود اختصاص می‌دهند که برای هر چهار گونه مورد مطالعه مشابه بود و در این پژوهش برای تفسیر نتایج یک گونه (*G.stocksii*) مرجع قرار داده شد. درصد نواحی کد کننده *tRNA*، *rRNA* و توالی‌های کد کننده پروتئینی بترتیب ۵/۶۹، ۱/۷۱ و ۵۹/۶۴ درصد گزارش شد که این نتایج مشابه با *M.esculenta* [۱۴]، *Panax* [۶] و *coffee* [۴] بود. ژنوم کلروپلاست یک ژنوم محافظت شده است و بنابراین بنظر می‌رسد فاقد توالی‌های تکراری ساده باشد درحالی‌که حضور SSR در ژنوم کلروپلاست بسیاری از گونه‌ها گزارش شده است. Kim و همکاران [۶]

- vascular plants. DNA research. 2004 Jan 1; 11(4): 247-61.
- [7] Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022 Jan; 49(1):539-565. doi: 10.1007/s11033-021-06749-4. Epub 2021 Nov 25. PMID: 34822069; PMCID: PMC8613526.
- [8] Subramanian K, Payne B, Feyertag F, Alvarez-Ponce D. The Codon Statistics Database: A Database of Codon Usage Bias, *Molecular Biology and Evolution*, Volume 39, Issue 8, August 2022, msac157, <https://doi.org/10.1093/molbev/msac157>
- [9] Choudhury MN, Uddin A, Chakraborty S. Nucleotide composition and codon usage bias of SRY gene. *Andrologia*. 2018 Feb; 50(1): e12787.
- [10] Soto-Blanco B, Spinosa HS, G'orniak SL, Neto JP. Gossypol factors antinutricionais da soja in Toxicologia Aplicada`a Medicina VeterIn`aria. 2008 Brazil. 545
- [11] Turmel M, Gagnon MC, O'Kelly CJ, Otis C, Lemieux C. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Molecular biology and evolution*. 2009 Mar 1; 26(3): 631-48.
- [12] Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W, Zhang W, Kim K, Lee SC, Yang TJ, Wang Y. The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Frontiers in plant science*. 2016 Mar 15; 7: 306.
- [13] Talat F, Wang K. CHLOROPLAST GENOME STUDY, NEW TOOL IN PLANT BIOTECHNOLOGY; GOSSYPIMUM SPP. AS A MODEL CROP. *Journal of Current Research in Science*. 2014 Nov 1; 2(6): 838.
- [14] Daniell H, Wurdack KJ, Kanagaraj A, Lee SB, Saski C, Jansen RK. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. *Theoretical and Applied Genetics*. 2008 Mar; 116(5): 723-37.
- [15] Reichhardt T. It's sink or swim as a tidal wave of data approaches. *Nature*. 1999 Jun 10; 399(6736): 517-20.
- فیلوژنتیکی می‌تواند مورد استفاده قرار گیرد. در هر چهار ژنوم مورد مطالعه ۱۸ ژن حاوی اینترون (۱ یا ۲ تایی) شناسایی شد که مشابه *Panax schinseng* بود [۶]. در ناحیه تکرار معکوس همه ژن‌های دو نسخه‌ای وجود دارند. دو ناحیه تکرار معکوس وجود دارد که طول یکسانی دارند. دو ناحیه تک نسخه‌ای بزرگ و تک نسخه‌ای کوچک بوسیله دو ناحیه IR_A و IR_B از هم جدا شده‌اند. بررسی نتایج روشن ساخت که ژنوم‌های کلروپلاست بدلیل اندازه کوچک و تعداد ژن‌های کمی که دارند بهترین گزینه برای بررسی روابط فیلوژنتیک بین گونه‌ها هستند. با استفاده از مطالعه ژنوم کلروپلاست گونه‌های موجود در یک خانواده به سادگی می‌توان پی برد که چه ژن‌هایی در طول تکامل از دست رفته‌اند و چه ژن‌هایی حفظ شده‌اند.

منابع

- [1] Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences*. 2002 Sep 17; 99(19): 12246-51.
- [2] Chan CX, Gross J, Yoon HS, Bhattacharya D. Plastid origin and evolution: new models provide insights into old problems. *Plant physiology*. 2011 Apr; 155(4): 1552-60.
- [3] Greiner S, Rauwolf UW, Meurer J, Herrmann RG. The role of plastids in plant speciation. *Molecular ecology*. 2011 Feb; 20(4): 671-91.
- [4] Talat F, Wang K. Comparative bioinformatics analysis of the chloroplast genomes of a wild diploid *Gossypium* and two cultivated allotetraploid species. *Iranian Journal of Biotechnology*. 2015 Sep; 13(3): 47.
- [5] Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine*. 2001; 40(04): 346-58.
- [6] Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17

Bioinformatics analysis of the chloroplast genomes of *G.stocksii* and *G.longicalyx* diploid species and comparing them with tetraploid species

Talat F.^{1*}, Gheibi M.²

¹ Assistant Professor, Seed and Plant Improvement Research Department, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia, Iran

² Master Graduate of Biology (Biotechnology), Islamic Azad University of Urmia, Iran

* (Corresponding author): f.talat@areeo.ac.ir

DOI:10.30495/JDB.2023.1959300.1305

<https://dorl.net/dor/20.1001.1.2008692.1401.15.1.4.8>

Received: June 2022

Accepted: December 2022

Abstract

Cotton (*Gossypium* spp.) included 5 tetraploid ($2n=4x=52$) and more than 45 diploid ($2n=2x=26$) species. Tetraploid species of cotton are formed through hybridization between an A- genome and a D- genome. Genome mapping analyses showed that there is high level of similarity between diploid A- and D- genomes with tetraploid At- and Dt- genomes. *G. stocksii* and *G.longicalyx* belong to D genome, which is an important wild bio-source for the cotton breeding and genetic research. Present study was conducted to study and compare the complete chloroplast sequence of two diploid species of cotton (*G.stocksii* and *G.longicalyx*) compared with two tetraploid species (*G.hirsutum* and *G.barbadense*), analyses of their genome structure, gene content and organization, repeated sequences and codon usage has been carried out. *G.longicalyx* chloroplast (cp) genome is 160,248 bp in length which longer than *G.stocksii* and shorter than two tetraploid species. *G.stocksii* chloroplast (cp) genome is 159,039 bp in length which shorter than *G.longicalyx* and shorter than two tetraploid species. In both species single copy regions of cp genome is separated by the two inverted repeats. The plastidic genomes of *G.stocksii* and *G.longicalyx* comprised of 79 protein coding genes, 4 ribosomal RNA genes and 30 transfer RNA genes, 113 single genes in total and 20 duplicated genes in inverted repeats. In both diploid species 5 *yef* genes with unknown function have been identified. These genes considered essential for plants and are highly conserved in plant taxa.

Keywords: Diploid, Tetraploid, Chloroplast, Genome, Sequencing, Bioinformatics.