آنالیز بیوانفورماتیک کدونی ژنوم کلروپلاست گونه‌های دیپلوئید پنبه و مقایسه با ژنوم دو گونه تتراپلوئید

فرشید طلعت[1*]، سمانه حسنی‌نژاد[2] و مهدی بدری انرجان[3]

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

[2] **ایران، ارومیه، دانشگاه آزاد اسلامی واحد ارومیه، گروه زیست فناوری(بیوتکنولوژی)**

[3] **ایران، ارومیه، دانشگاه ارومیه، دانشکده کشاورزی، گروه اصلاح نباتات**

تاریخ دریافت: ۹۷/۲/۱۰ تاریخ پذیرش: ۹۷/۱۰/۲۵

چکیده

تجزیه کاربردی کدون اهمیت ویژه‌ای در بهینه‌سازی سامانه بیان ژن‌ها در تولید پروتئین دارد. جنس گوسیپیوم مهمترین گیاه تولیدکننده الیاف در دنیای جدید می‌باشد. در این تحقیق توالی کامل ژنوم کلروپلاست دو گونه *G.thurberi* و *G.arboreum* توسط نرم‌افزار کدون W مورد مطالعه و تجزیه قرار گرفت. تجزیه کدون‌های دارای ترادف (RSCU) برای ۵۷ ژن کدکننده پروتئین در ژنوم کلروپلاست این دو گونه بمنظور دستیابی به فاکتورهای دخیل در اریبی کدون صورت پذیرفت. تمامی کدون-های ترجیح داده شده دارای بازهای آلی آدنین و تیمین در انتهای کدون بودند که با عنایت به غنای ژنوم کلروپلاست در خصوص این دو باز پدیده‌ای طبیعی است. آنالیز تطبیقی و روش برآورد تعداد موثر کدون‌ها بصورت نمودار تعداد کدون بمنظور تجزیه کاربرد کدون‌های مترادف انجام گرفت. نمودار ENC Vs GC3، عمده ژن‌های آنالیز شده را معادل یا دقیقا در سمت چپ منحنی قابل انتظار GC3 گروهبندی کرد که بیانگر تاثیر محدودیت‌های ترکیب کدونی در تنظیم استفاده ار کدون‌هاست. بر اساس تجزیه تطبیقی، اریبی مشاهده شده در ژنوم کلروپلاست گونه‌های مورد مطالعه با طول ژن، اریبی جهش در ژن‌ها، سطح هیدروپاتی ژن‌های پروتئینی، عملکرد ژن و انتخاب یا بیان ژن بصورت ماهرانه‌ای در کاربرد کدون‌ها تاثیرگذار است. نتایج این پژوهش در فراهم آوردن زمینه برای مطالعات تکامل ملکولی قابل استفاده خواهد بود.

واژه های کلیدی: آنالیز تطبیقی، ژنوم کلروپلاست، نمودارNC

* نویسنده مسئول، تلفن: ۰۹۱۰۴۰۵۵۶۸۲، پست الکترونیکی: f.talat@areeo.ac.ir

مقدمه

در بررسی‌های تکامل ملکولی گیاهان، ژنوم کلروپلاست بدلیل اندازه کوچک، تعداد نسخه‌های بالا، ساختار حفاظت شده گسترده و عملکرد شناخته شده بسیاری از ژن‌های آن در سطح ملکولی برای آنالیزهای ترجیح کدونی بسیار مورد توجه قرار گرفته است (۹). سازماندهی ژنوم کلروپلاست کاملاً حفاظت‌شده می‌باشد و از یک کروموزوم حلقوی با ساختار چهارجزئی تشکیل شده و دارای دو ناحیه تکرار معکوس (۲۵ کیلوباز) می‌باشد که منطقه تک نسخه بزرگ (۸۰ کیلوباز) را از ناحیه تک نسخه کوچک (۲۰کیلوباز) جدا می‌کنند. اندازه این ژنوم حلقوی شکل از ۳۵ تا ۲۱۷ کیلوباز متغیر است اما این محدوده برای اکثر ژنوم‌های پلاستید متعلق به موجودات فتوسنتزکننده ۱۶۵-۱۱۵ کیلوباز می‌باشد (۲). در مقایسه با ژنوم‌های هسته‌ای و میتوکندری، ژنوم پلاستید در میان گونه‌ها کاملا حفاظت

شده می‌باشد (۸) گرچه در مقایسه با توالی کامل کلروپلاست تفاوت‌هایی نیز گزارش شده است (۴ و ۱۳) که این موضوع بطور عمده به گسترش یا عدم گسترش تکاملی و یا از دست دادن ناحیه تکرار معکوس، بازآرایی ژنوم و تکرارهای پراکنده نسبت داده شده است (۴، ۱۱ و ۱۳). ژنوم کلروپلاست در گونه‌های مختلف دارای سازماندهی بسیار محافظت‌شده می‌باشد که اکثر ژن‌های آن در سطح ملکولی شناخته شده‌اند و به همین دلیل برای مطالعات و بررسی‌های تکامل ملکولی بسیار مناسب است. کلروپلاست یک محفظه ایده‌آل برای تجمع پروتئین‌های خاص و یا محصولات بیوسنتزی آنها است که در صورت انباشته شدن در سیتوپلاسم مضر خواهند بود و علاوه بر این خاموش کردن ژن در آن مشاهده نشده است (۱۲). به همین دلیل ژنوم کلروپلاست برای بررسی برخی از صفات مهم زراعی از قبیل مقاومت در برابر علف‌کش‌ها، مقاومت در برابر حشرات، مقاومت به بیماری‌ها و تحمل شوری مهندسی شده است. طلعت و وانگ (۲۰۱۵) با انجام آنالیزهای مقایسه‌ای روی ژنوم کلروپلاست گونه دیپلوئید *Gossypium thurberi* و دو گونه آلوتتراپلوئید به این نتایج دست یافتند که ژنوم کلروپلاست *G.thurberi* دارای ساختار چهار جزئی حفاظت شده بطول ۱۶۰۲۶۴ جفت‌باز بوده که نواحی LSC (Large Single Copy) و SSC (Small Single Copy) توسط دو ناحیه تکرار معکوس از هم جدا شده‌اند. این ژنوم حاوی ۱۱۳ ژن منفرد و ۲۰ ژن دو نسخه‌ای است. ژن‌های منفرد کد کننده ۷۹ پروتئین، ۴ ژن RNA ریبوزومی و ۳۰ ژن RNA ناقل هستند که در میان تمام ژن‌های پلاستیدی تنها ۱۸ ژن حاوی یک یا دو اینترون گزارش شدند و در مقایسه با DNA کلروپلاست دو گونه آلوتتراپلوئید، *rps18* تنها ژن مضاعف شده درگونه *G.thurberi* است (۱۲). بات و همکاران (۲۰۱۴) با انجام آنالیزهای ژنومی ترجیح کدونی و بررسی اثر فشار جهشی، انتخاب طبیعی و ویژگی‌های میزبان روی تکامل ویروس دریافتند که در کل گرایش ترجیح کدونی کمی وجود دارد و فاکتور اصلی اثرگذار فشار جهشی است که تحت تاثیر میزبان است. همچنین علاوه بر این انتخاب طبیعی، محیط و شرایط جغرافیایی نیز بر ترجیح کدونی اثر می‌گذارند (۱). این تحقیق با هدف مطالعه و مقایسه توالی کامل ژنوم کلروپلاست گونه‌های دیپلوئید *G.thurberi* و *G.arboreum* با گونه‌های تتراپلوئید *G.barbadense* و *G.hirsutum* گیاه پنبه، آنالیز ساختار ژنوم، محتوای ژن، سازماندهی توالی‌های تکراری، ترجیح کدونی و مقایسه بین ژنوم‌های موجود انجام گرفت.

مواد و روشها

توالی کامل ژنوم کلروپلاست گونه‌های *G.thurberi*، *G.arboreum*، *G.barbadense* و *G.hirsutum* با فرمت FASTA بترتیب با شماره دسترسی و طول (NC_015204.1 و 160,264)، (HQ_325740.1 و 160,230)، (NC_008641.1 و 160.317) و (NC_007944.1 و 160,301) از سایت NCBI دانلود شدند. برای شناسایی موقعیت هر یک از ژن‌ها در هر چهار ژنوم کلروپلاست مورد بررسی با توجه به اطلاعات موجود در بانک ژن مربوط به هر ژنوم در سایت NCBI مونتاژی از توالی هر ژنوم به همراه ژن‌هایی که در روی توالی مشخص شده‌اند توسط word (2010) بدست آمد. اینترون-های مربوط به ژن‌های حاوی اینترون نیز با رنگ مجزا مشخص شدند که با این کار هم موقعیت ژن‌ها، نواحی IGS (Intragenic Spacer) و همچنین ژن‌هایی که دارای اشتراک هستند نیز مشخص شدند. نقشه ساختار ژنوم و توزیع ژن با استفاده از نرم‌افزار OGDRAW (http://ogdraw.mpimpgolm.mpg.de) (V1.1): و با دادن اطلاعات مربوط به شماره دسترسی هر ژنوم رسم شد. برای تشخیص توالی‌های تکراری و موقعیت آنها از نسخه آنلاین نرم‌افزار REPuter (http://bibiserv.techfak.uni-bielfeld.de/reputer) استفاده شد. RSCU (Relative Synonymous

Codon Usage) های کدون‌های مختلف در هر نمونه ژنی توسط CodonW در نرم افزار Mobyle در آدرس (http://mobyle.pasteur.fr/cgibin/portal.py) محاسبه گردید. برای بدست آوردن درصد مربوط به نوکلئوتیدهای A، T، C، G و همچنین AT وGC در ژنوم‌ها جهت تشکیل جدول محتوای GC از نرم افزار Visual Bioinformatics (V 2.1.0) استفاده شد. برای رسم نمودارهای مربوط به بحث کاربرد کدون و آنالیزهای تطابقی بترتیب از نرم-افزارهای SPSS (V.22) و Minitab (V.16) استفاده شد.

نتایج

محتوی ژنی: ژن‌های کدکننده در ژنوم کلروپلاست *G.thurberi*، *G.arboreum*، *G.barbadense* و *G.hirsutum* در جدول ۱ آمده است.

جدول ۱- ژن‌های کد شونده توسط ژنوم کلروپلاست چهار گونه

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

§ نمایانگر ژن‌های موجود در ناحیه IR، * نمایانگر ژن‌های با ۱ اینترون و ** نمایانگر ژن‌های با یک اینترون می‌باشد.

نتایج نشان میدهند که *G.thurberi* حاوی ۱۳۴ ژن، *G.arboreum* ۱۲۸ ژن، *G.barbadense* و *G.hirsutum* دارای ۱۲۹ ژن بوده که شامل ۷۹ ژن کدکننده پروتئین، ۴ ژن RNA ریبوزومی و ۳۰ ژن RNA ناقل می‌باشد که بعبارت دیگر میتوان گفت در کل ۱۱۳ تک ژن و ۱۸ ژن دو نسخه‌ای واقع در ناحیه تکرار معکوس تقریبا مشمول هر چهار ژنوم مورد بررسی میباشند. در هر یک از ژنوم‌های کلروپلاست *G.thurberi*، *G.arboreum*، *G.barbadense* و *G.hirsutum* پنج ژن با عملکرد ناشناخته بنام *ycf* شناسایی و بعنوان ژن‌های ضروری گیاهان در نظر گرفته - شده که در میان گونه‌ها کاملاً محافظت شده‌اند. با توجه به نتایج بدست آمده در مطالعات ژنوم کلروپلاست، دو ژن

حائز اهمیت می‌باشند یکی از آنها ژن *rps12* میباشد و به ۲ قطعه که یکی با اگزون انتهای ´۵ واقع در LSC و دیگری با اگزون انتهای ´۳ واقع در IR (Inverted Repeat) جدا می‌گردد. ژن دیگر *matk*، 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.

محتوی *GC*: محتوی *GC* از ویژگی‌های مهم ژنوم کلروپلاست می‌باشد درصد *GC* در ژنوم *G.thurberi*، *G.arboreum*، *G.barbadense* و *G.hirsutum* شبیه هم بوده و برابر۳۷/۲ درصد می‌باشد. مناطق کدکننده و غیر کدکننده در هر چهار ژنوم دارای محتوی *GC* کمی بوده و بترتیب ۴۰/۴ و ۳۳/۱ درصد گزارش شد. تفاوت در محتوی *GC* برای چهار ناحیه ژنوم کلروپلاست چهار گونه تحت بررسی در جدول ۲ آورده شده است و نتایج نشان داد که ناحیه *IR* از لحاظ *GC* غنی‌ترین بود و میزان آن برای هر چهار گونه تقریباً ۴۳ درصد گزارش شد درحالیکه *GC* در*SSC* و*LSC* برای گونه‌های تحت بررسی *G.thurberi*، *G.arboreum*، *G.barbadense* و *G.hirsutum* بترتیب برابر ۳۱ و۳۵ درصد بود. ژن‌های *rRNA* دارای بیشترین محتوای *GC* حدود ۵۵/۵ درصد و توالی‌های کد کننده پروتئینی کمترین *GC* حدود ۳۷/۳ درصد بودند و در منطقه غیرکدکننده محتوی *GC* در *IGS* و اینترون بترتیب برای هر گونه برابر با ۳۵/۵۸ و ۳۶/۷۵ درصد بود.

جدول۲- محتوی GC ژنوم کلروپلاست چهار گونه

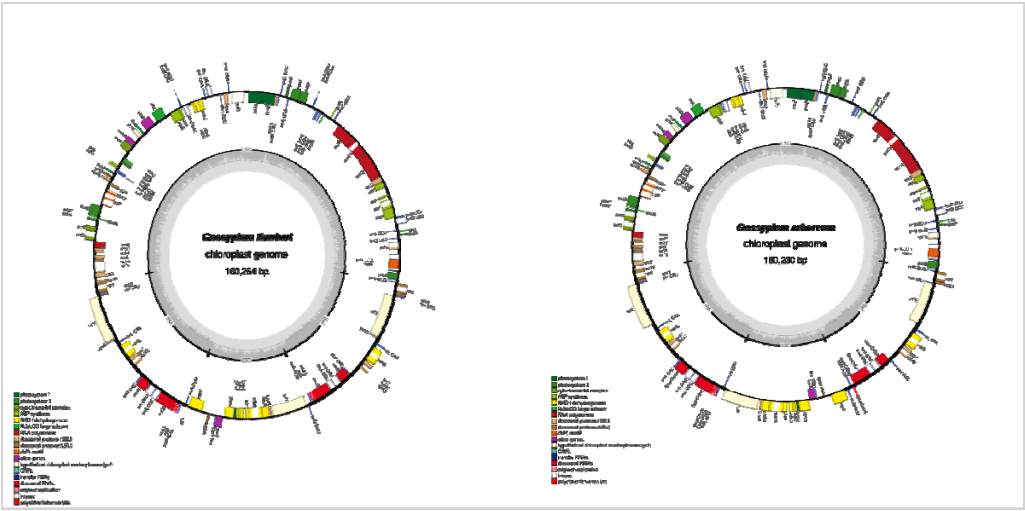
G.thurberi											
-	Coding	Region	-	-	Non	Coding	Region	-	-	-	-
-	Protein	tRNA	rRNA	Total	IGS	Intron	Total	Complete Genome	LSC	SSC	IR
Length	۷۹۷۴۰	۲۷۷۵	۸۹۷۰	۹۱۴۸۵	۴۹۹۱۵	۲۰۴۳۶	۷۰۳۵۱	۱۶۰۲۶۴	۸۸۷۳۷	۲۰۲۷۱	۲۵۶۲۸
Proportion	۴۹/۷۶	۱/۷۳	۵/۶۰	۵۷/۰۸	۳۱/۱۵	۱۲/۷۵	۴۳/۹۰	۱۰۰/۰۰	۵۵/۳۷	۱۲/۶۵	۱۵/۹۹
T%	۳۱/۰۵	۲۳/۱۴	۲۲/۳۲	۳۰/۳۴	۳۴/۳۲	۳۲/۲۸	۳۳/۷۳	۳۱/۸۳	۳۳/۱۵	۳۴/۴۳	۲۸/۲۵
A%	۲۹/۸۵	۲۴/۵۸	۲۲/۱۷	۲۹/۳۱	۳۴/۱۰	۳۰/۹۷	۳۳/۱۹	۳۰/۹۵	۳۱/۶۷	۳۳/۹۲	۲۸/۵۴
C%	۱۹/۲۸	۲۶/۱۳	۷۹۲۷	۲۰/۵۶	۱۵/۹۹	۱۹/۱۲	۱۹/۹۰	۱۸/۹۹	۱۸/۱۱	۱۶/۵۴	۲۰/۶۶
G%	۱۸/۴۲	۲۰/۱۶	۲۷/۷۱	۱۹/۷۶	۱۵/۵۸	۱۷/۶۳	۱۶/۱۸	۱۸/۲۳	۱۷/۰۹	۱۵/۰۹	۲۲/۲۹
A+T%	۶۰/۹۰	۴۷/۷۱	۴۴/۴۹	۵۹/۶۵	۶۸/۴۲	۶۳/۲۵	۶۶/۹۲	۶۲/۷۸	۶۴/۸۱	۶۸/۳۸	۵۷/۰۵
C+G%	۳۷/۶۹	۵۲/۲۹	۵۵/۵۱	۴۰/۳۵	۳۱/۵۸	۳۶/۷۵	۳۳/۰۸	۳۷/۲۲	۳۵/۱۹	۳۱/۶۲	۴۲/۹۵
G.arboreum											
-	Coding	Region	-	-	Non	Coding	Region	-	-	-	-
-	Protein	tRNA	rRNA	Total	IGS	Intron	Total	Complete Genome	LSC	SSC	IR
Length	۷۹۲۵۳	۲۷۶۹	۸۳۴۹	۹۰۳۷۱	۴۹۳۳۳	۲۰۵۲۶	۶۹۸۵۹	۱۶۰۲۳۰	۸۸۷۲۱	۲۰۲۸۷	۲۵۶۱۱
Proportion	۴۹/۴۶	۱/۵۳	۵/۲۱	۵۶/۴۰	۳۰/۸۰	۱۲/۸۰	۴۳/۶۰	۱۰۰/۰۰	۵۵/۳۷	۱۲/۶۶	۱۵/۹۸
T%	۳۱/۴۰	۲۲/۸۰	۲۲/۳۰	۳۰/۲۰	۳۴/۴۰	۳۲/۰۰	۳۳/۷۰	۳۱/۸۰	۳۳/۲۰	۳۳/۹۰	۲۸/۵۰
A%	۳۰/۳۰	۲۴/۰۰	۲۲/۳۰	۲۹/۴۰	۳۴/۰۰	۳۱/۱۰	۳۳/۲۰	۳۱/۰۰	۳۱/۶۰	۳۴/۴۰	۲۸/۵۰
C%	۱۹/۳۰	۲۷/۲۰	۲۷/۷۰	۲۰/۴۰	۱۶/۰۰	۱۹/۰۰	۱۶/۸۰	۱۸/۸۰	۱۸/۱۰	۱۵/۱۰	۲۱/۵۰
G%	۱۹/۰۰	۲۶/۰۰	۲۷/۷۰	۲۰/۱۰	۱۵/۶۰	۱۷/۹۰	۱۶/۳۰	۱۷/۴۰	۱۷/۱۰	۱۶/۶۰	۲۱/۵۰
A+T%	۶۱/۷۰	۴۶/۷۰	۴۴/۵۰	۵۹/۶۰	۶۸/۴۰	۶۳/۱۰	۶۶/۹۰	۶۲/۸۰	۶۴/۸۰	۶۸/۲۰	۵۷/۰۰
C+G%	۳۸/۳۰	۵۳/۲۰	۵۵/۵۰	۴۰/۴۰	۳۱/۶۰	۳۶/۹۰	۳۳/۱۰	۳۷/۳۰	۳۵/۲۰	۳۱/۷۰	۴۳/۰۰
G.barbadense											
-	Coding	Region	-	-	Non	Coding	Region	-	-	-	-
-	Protein	tRNA	rRNA	Total	IGS	Intron	Total	Complete Genome	LSC	SSC	IR

Length	۷۸۶۷۵	۲۷۹۱	۹۰۵۰	۹۰۵۱۶	۴۸۵۵۶	۲۱۲۴۵	۶۹۸۰۱	۱۶۰۳۱۷	۸۸۸۹۷	۲۰۰۳۶	۲۵۶۹۲
Proportion	۴۹/۰۷	۱/۷۴	۵/۶۴	۵۶/۴۶	۳۰/۲۸	۱۳/۲۷	۴۳/۵۳	۱۰۰/۰۰	۵۵/۴۵	۱۲/۴۹	۱۶/۰۳
T%	۳۱/۵۰	۲۳/۳۰	۲۲/۳۰	۳۰/۴۰	۳۴/۳۰	۳۲/۰۰	۳۳/۷۰	۳۱/۸۰	۳۳/۱۰	۳۴/۴۰	۲۸/۵۰
A%	۳۰/۲۰	۲۳/۹۰	۲۲/۳۰	۲۹/۳۰	۳۴/۱۰	۳۱/۲۰	۳۳/۲۰	۳۳/۹۰	۳۱/۷۰	۳۳/۸۰	۲۸/۵۰
C%	۱۹/۶۰	۲۶/۷۰	۲۷/۷۰	۲۰/۶۰	۱۶/۶۰	۱۹/۰۰	۱۶/۸۰	۱۹/۰۰	۱۸/۱۰	۱۶/۶۰	۲۱/۵۰
G%	۱۸/۷۰	۲۶/۱۰	۲۷/۷۰	۱۹/۷۰	۱۹/۶۰	۱۷/۸۰	۱۶/۲۰	۱۸/۲۰	۱۷/۱۰	۱۵/۳۰	۲۱/۵۰
A+T%	۶۱/۷۰	۴۷/۱۰	۴۴/۶۰	۵۹/۷۰	۶۸/۷۰	۶۳/۲۰	۶۶/۹۰	۶۲/۳۰	۶۴/۸۰	۶۸/۱۰	۵۶/۹۰
C+G%	۳۸/۳۰	۵۲/۸۰	۵۵/۴۰	۴۰/۳۰	۳۱/۶۰	۳۶/۸۰	۳۳/۱۰	۳۷/۲۰	۳۵/۲۰	۳۱/۹۰	۴۳/۱۰

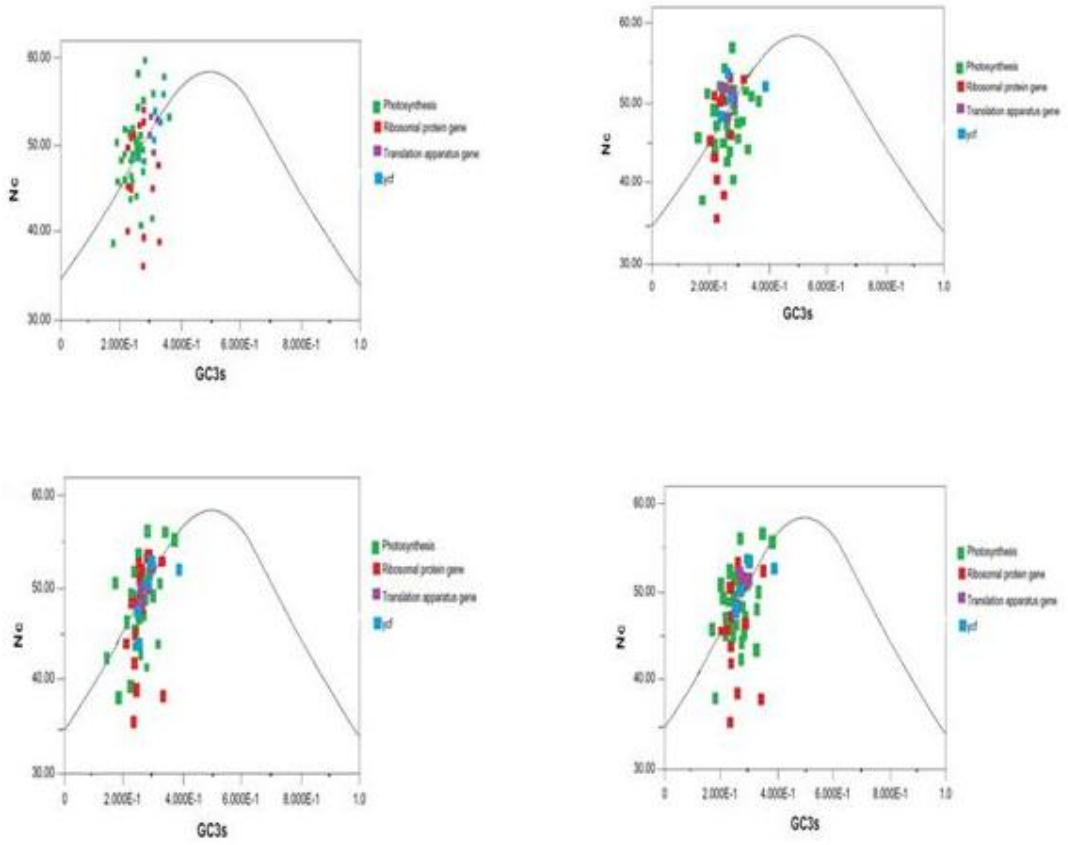
G.hirsutum

-		Coding	Region	-	Non	Coding	Region	-	-	-	-
-	Protein	tRNA	rRNA	Total	IGS	Intron	Total	Complete Genome	LSC	SSC	IR
Length	۷۸۵۳۱	۲۸۰۱	۹۰۴۸	۹۰۳۸۰	۴۸۷۹۸	۲۱۱۲۳	۶۹۹۲۱	۱۶۰۳۰۱	۸۸۸۶۲	۲۰۵۰۹	۲۵۴۶۴
Proportion	۴۸/۹۸	۱/۷۴	۵/۶۴	۵۶/۳۸	۳۰/۴۴	۱۳/۲۰	۴۳/۶۱	۱۰۰/۰۰	۵۵/۴۳	۱۲/۷۹	۱۵/۸۸
T%	۳۱/۴۰	۲۲/۸۰	۲۲/۳۰	۳۰/۲۰	۳۴/۰۰	۳۱/۹۰	۳۳/۷۰	۳۱/۷۰	۳۳/۲۰	۳۳/۹۰	۲۸/۵۰
A%	۳۰/۴۰	۲۳/۹۰	۲۲/۳۰	۲۹/۴۰	۳۳/۹۰	۳۱/۳۰	۳۳/۲۰	۳۱/۰۰	۳۱/۶۰	۳۴/۴۰	۲۸/۵۰
C%	۱۹/۳۰	۲۷/۱۰	۲۷/۷۰	۲۰/۳۰	۱۶/۰۰	۱۸/۹۰	۱۶/۹۰	۱۸/۸۰	۱۸/۱۰	۱۵/۱۰	۲۱/۵۰
G%	۱۹/۰۰	۲۶/۲۰	۲۷/۷۰	۲۰/۱۰	۱۵/۶۰	۱۷/۹۰	۱۶/۳۰	۱۸/۴۰	۱۷/۱۰	۱۶/۵۰	۲۱/۵۰
A+T%	۶۱/۸۰	۴۶/۷۰	۴۴/۵۰	۵۶/۹۰	۶۸/۷۰	۶۳/۲۰	۶۶/۸۰	۶۲/۵۰	۶۴/۸۰	۶۸/۳۰	۵۷/۰۰
C+G%	۳۸/۲۰	۵۳/۳۰	۵۵/۵۰	۴۰/۴۰	۳۱/۶۰	۳۶/۸۰	۳۳/۲۰	۳۷/۲۰	۳۵/۲۰	۳۱/۷۰	۴۳/۵۰

نمودار NC: نمودار NC (Number of Codons) برای هر چهار ژنوم مورد مطالعه مطابق با شکل ۲ رسم شد و نتایج مربوط به آن برای تمام گونه‌ها تقریباً یکسان بدست آمد به همین جهت تنها یک گونه *G.thurberi* را مرجع قرار داده و تجزیه و تحلیل بر اساس آن انجام گرفت. شکل ۱ نشان می‌دهد که تعداد قابل توجهی از نقاط روی منحنی قرار گرفته‌اند و به سمت ناحیه‌ای که از نظر GC فقیر است می‌باشند که این از ترکیب نوکلئوتیدی شدید منشأ می‌گیرد اما موضوع جالب توجه این است که اکثر نقاط با مقادیرNC کم در زیر منحنی قابل انتظار با فاصله زیاد قرار گرفته‌اند و این نتایج نشان می‌دهند که برخی ژن‌ها در *G.thurberi* کاربرد کدونی مستقل از ترکیب نوکلئوتیدی مشتق از فشار جهشی و متأثر از عوامل دیگری که مستقل از محدودیت-های ترکیبی می‌باشند.



شکل ۱- سازماندهی ژنی و ساختار ژنوم دو گونه دیپلوئید



شکل ۲- نمودار NC چهار گونه

یادداشت: از بالا و از چپ به راست بترتیب: *G.thurberi*، *G.arboreum*، *G.barbadense* و *G.hirsutum*

آنالیزهای تطبیقی: آنالیزهای تطبیقی برای دستیابی به فاکتورهای مسئول توزیع ژن‌ها در نمودار آنالیز تطبیقی (COA=(Correspondence Analysis) Plot) بوسیله تعیین همبستگی بین چهار محور (axis1–axis4) با شاخص‌های ترجیح کدونی برای ۵۷ ژن کدکننده با طول بیش از ۱۰۰ کدون در گونه‌های مورد مطالعه انجام شد. نتایج در جدول ۳ و شکل ۳ نشان داده شده است..

جدول ۳- همبستگی شاخص‌های ترجیح کدونی با محورهای توزیع ژنی در چهار ژنوم

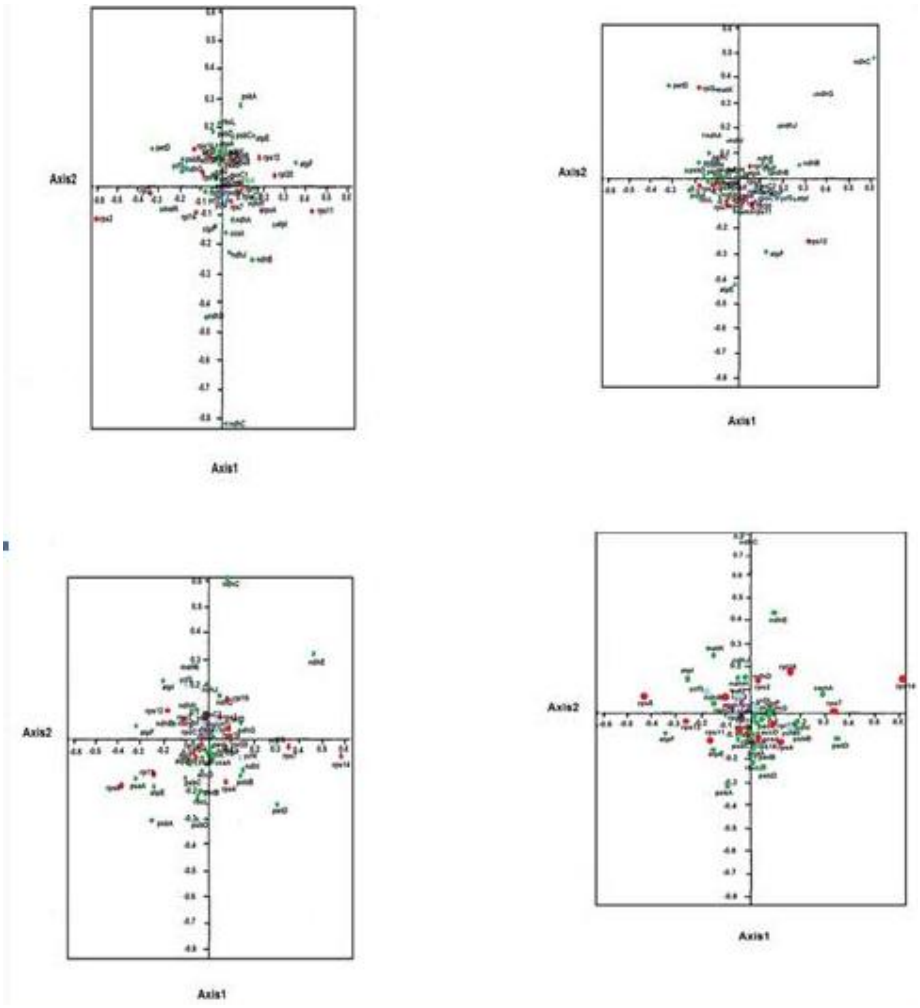
G.thurberi											
-	ENc	CAI	Fop	$T3_s$	$C3_s$	$A3_s$	$G3_s$	$GC3_s$	GC	L_sym	Gravy
Axis1	−۰/۰۴۹	−۰/۰۷۰	۰/۱۱۲	−۰/۱۳۱	۰/۲۰۳	−۰/۰۵۰	−۰/۰۸۷	۰/۱۰۵	۰/۱۰۸	۰/۰۹۲	۰/۰۱۲
Axis2	−۰/۲۹۴*	۰/۲۲۳	۰/۳۸۰**	−۰/۱۶۹	۰/۳۷۴**	−۰/۰۸۵	−۰/۱۲۴	۰/۲۵۲	۰/۲۹۴*	−۰/۰۷۰	−۰/۲۶۱
Axis3	−۰/۲۶۰	۰/۲۸۹*	۰/۱۳۲	۰/۱۷۷	۰/۰۰۷	−۰/۱۳۷	−۰/۳۴۴**	−۰/۲۰۱	۰/۰۹۹	−۰/۰۹۵	۰/۲۱۴
Axis4	−۰/۰۰۱	۰/۴۷۴**	۰/۲۹۹*	۰/۱۶۴	۰/۲۲۷	−۰/۲۲۸	−۰/۱۹۸	۰/۰۲۰	۰/۱۸۰	۰/۰۲۰	۰/۳۴۸**
G.arboreum											
-	ENc	CAI	Fop	T3s	C3s	A3s	G3s	GC3s	GC	L_sym	Gravy
Axis1	۰/۲۰۹	−۰/۳۵۸**	۰/۳۷۲**	−۰/۰۱۰	−۰/۲۵۲	۰/۱۱۲	۰/۲۶۸*	−۰/۰۵۹	−۰/۲۶۴*	۰/۰۶۰	۰/۱۴۱
Axis2	۰/۱۸۹	۰/۱۶۸	۰/۱۴۳	۰/۳۷۰**	−۰/۳۱۴*	۰/۱۳۹	−۰/۲۴۸	−۰/۵۵۹*	−۰/۲۸۰*	−۰/۱۴۳	−۰/۳۰۰
Axis3	۰/۰۳۳	۰/۵۲۶*	۰/۴۳۶*	۰/۱۴۹	۰/۳۴۱*	−۰/۰۵۷	−۰/۳۹۲*	−۰/۰۲۱	۰/۲۴۵	۰/۰۱۳	۰/۳۶۸**
Axis4	۰/۵۵۲*	۰/۲۱۷	۰/۳۳۵	۰/۰۰۱	۰/۱۱۶	−۰/۰۵۷	−۰/۲۹۰	−۰/۰۸۵	۰/۰۸۹	−۰/۱۶۱	۰/۰۳۹

G.barbadense

-	ENc	CAI	Fop	$T3_s$	$C3_s$	$A3_s$	$G3_s$	$GC3_s$	GC	L_sym	Gravy
Axis1	−۰/۲۴۱	−۰/۰۷۸	−۰/۲۵۲	۰/۰۳۶	−۰/۴۰۸*	۰/۲۲۰	۰/۰۳۶	−۰/۲۷۷*	−۰/۲۱۳	−۰/۱۰۶	۰/۰۲۳
Axis2	۰/۰۸۵	−۰/۴۵۷*	−۰/۵۱۹*	۰/۰۴۳	−۰/۴۲۶**	۰/۲۳۲	۰/۲۰۳	−۰/۲۳۷	−۰/۳۹۹**	۰/۰۱۴	۰/۱۰۹
Axis3	۰/۴۶۳**	−۰/۲۵۳	−۰/۱۰۹	−۰/۲۱۴	۰/۱۰۰	−۰/۰۳۰	۰/۴۷۵**	۰/۴۰۶**	۰/۰۴۶	۰/۲۲۰	−۰/۱۹۴
Axis4	۰/۰۷۸	۰/۴۴۲**	۰/۳۰۴	۰/۱۵۹	۰/۰۹۲	−۰/۰۶۷	۰/۲۷۷*	−۰/۱۳۸	۰/۱۲۳	−۰/۰۲۱	۰/۳۶۴*

G.hirsutum

-	ENc	CAI	Fop	$T3_s$	$C3_s$	$A3_s$	$G3_s$	$GC3_s$	GC	L_sym	Gravy
Axis1	−۰/۱۸۵	۰/۰۹۷	−۰/۰۳۳	۰/۰۰۹	−۰/۲۴۰	۰/۱۸۰	−۰/۰۳۰	−۰/۱۸۸	−۰/۰۵۲	−۰/۰۹۳	−۰/۰۵۹
Axis2	۰/۱۳۱	−۰/۲۱۲*	−۰/۴۵۰**	۰/۰۸۳	−۰/۴۸۵**	۰/۲۲۶	۰/۲۰۶	−۰/۲۷۵*	−۰/۳۷۹**	۰/۰۰۵	۰/۱۸۶
Axis3	−۰/۴۴۲**	۰/۴۳۰**	۰/۲۲۹	۰/۳۳۵*	−۰/۰۳۲	−۰/۰۸۴	−۰/۵۲۹**	−۰/۴۲۲**	−۰/۰۳۰	−۰/۲۱۴	۰/۳۲۵*
Axis4	−۰/۱۲۱	−۰/۴۳۱*	−۰/۳۷۷*	−۰/۱۰۵	−۰/۳۰۵*	۰/۱۹۰	۰/۲۵۶	−۰/۰۴۹	−۰/۲۴۳	−۰/۰۱۹	−۰/۲۸۸



شکل ۳- نمودار COA برای چهار گونه

یادداشت: از بالا و از چپ به راست بترتیب: *G.thurberi*، *G.arboreum*، *G.barbadense* و *G.hirsutum*

طبق گزارشات نتایج مربوط به این آنالیز، در هر چهار ژنوم مورد مطالعه بسیار مشابه هم بود و لذا برای ممانعت از تکرار مطالب توضیحات مربوط به این آنالیز بر مبنای اعداد جدول و نمودار مربوط به ژنوم کلروپلاست *G.thurberi* صورت گرفت. محور اول ۱۰/۳ و سه محور بعدی بترتیب ۷/۹۳، ۹ و ۷/۱۶ درصد از تغییرات ترجیح کدونی را بیان می‌کند، همچنین نتایج نشان داد که محور دوم با C3 (r=0.374, p<0.01)، GC (r=0.294, p<0.05) و Fop (r=0.380, p<0.01) همبستگی مثبت نشان داد، محور سوم با G3 (r=-0.344, p<0.01) همبستگی مثبت و با CAI (r=0.289, p<0.05) همبستگی منفی داشت همچنین محور چهارم با CAI (r=0.474, p<0.01) و Fop (r=0.299, p<0.05) همبستگی مثبت نشان داد که وجود ارتباط معنی‌دار بین هر کدام از محورها و بازهای نوکلئوتیدی احتمال اثر ترکیب نوکلئوتیدی در شکل‌گیری ترجیح کدونی را نشان می‌دهد. با توجه به این در هر چهار گونه تعداد دو محور ارتباط معنی‌دار با CAI (Frequency of Optimal Codon Usage) و Fop (Codon Adaptation Index) داشت پیش‌بینی شد که سطوح بیان ژن بطور اندک الگوهای ترجیح کدونی را تحت تأثیر قرار دهد (انتخاب ضعیف) و نیز بدلیل معنی‌داری Gravy تنها در یک محور در گونه‌های *G.barbadense* و*G.hirsutum* نیروی انتخاب اثرگذار بر ترجیح کدونی ضعیف می‌باشد. طبق نتایج بدست آمده ترکیب نوکلئوتیدی (مشتق از فشار جهشی) فاکتور مهم اثرگذار بر تغییرات ترجیح کدونی همتراز در ژنوم کلروپلاست این گونه‌ها گزارش شد.

بحث و نتیجه گیری

نمودار NC که مقادیر NC عمود بر مقادیر GC3s می‌باشد و محورهای y و x را تشکیل می‌دهند که برای نشان دادن تغییر ترجیح کدونی هم‌تراز (SCU=Synonymous Codon Usage) بین شماری از ژن‌ها در بین گونه‌ها یا درون یک گونه استفاده می‌شود (۱۰). نمودار مقادیرNC عمود بر GC3s می‌تواند برای کشف تنوع کاربرد کدون در میان ژن-ها موثر باشد و مقایسه توزیع واقعی ژن‌ها با توزیع قابل انتظار نشان می‌دهد که گرایش ترجیح کدونی تحت تاثیر چه عواملی به جز محدودیت‌های ترکیبی می‌تواند باشد و گرایش ترجیح کدونی بطور کامل تحت‌تاثیرGC3 بوده باشد و مقادیرNC باید روی منحنی قابل انتظار بین GC3s وNC باشد (۳). طبق آنالیزهای ترجیح کدونی برای چهار گونه موجود کدون پایان TAA با RSCU=1 بعنوان کدون پایان ترجیح داده شده مشابه با *Alsophila* (۵) و *Panax schinseng Nees* (۷) گزارش شد. مقدار RSCU اغلب کدون‌ها بیشتر از یک گزارش شده است که این نشان می-دهد این کدون‌ها با تناوب بیشتری مورد استفاده قرار گرفته‌اند. اغلب کدون‌ها از A یا T در جایگاه سوم خود (بدون تشابه در جایگاه اول و دوم) استفاده می‌کنند. بررسی ساختار کدون‌ها نشان می‌دهد محتوای A+T در جایگاه سوم ۷۱/۴ درصد می‌باشد که بیشتر از مقدار آن در ناحیه کدکننده پروتئین و کل ژنوم در هر چهار ژنوم کلروپلاست می‌باشد و این نشان داد که چهار گونه مورد بررسی کدون‌هایی را که در جایگاه سومشان A یا T دارند بیشتر ترجیح می‌دهند. در بررسی نتایج آنالیزهای ترجیح کدونی عامل اثرگذار بر ترجیح کدونی در چهار گونه موجود ترکیب نوکلئوتیدی بهمراه انتخاب ضعیف گزارش شد که مشابه با یافته‌های (۱۰) بود. برخلاف نتیجه بدست آمده در پژوهش خو و همکاران (۲۰۱۱) در بررسی عامل اثرگذار بر ترجیح کدونی در ژنوم کلروپلاست *Oncicidium gower* بعلت معنی‌دار بودن شاخص آمینواسیدی Gravy دریافتند که ترجیح کدونی متأثر از سطح هیدروپسی هر پروتئین می‌باشد (۱۴). مطالعات قبلی روی ترجیح کدونی عمدتا روی گرایش کدونی در ژنوم هسته‌ای متمرکز بوده‌اند و چنین بیان شده که گرایش کدونی منعکس کننده تعادل بین گرایشات جهشی و انتخاب طبیعی برای بهینه‌سازی ترجمه می‌باشد (۶). تا به امروز مطالعات جامعی بر روی ژنوم کلروپلاست‌های ثبت

شده در NCBI انجام نگرفته و این مطالعه می‌تواند بعنوان یک پژوهش پیشرو در زمینه آنالیز ژنوم گونه‌های زراعی پنبه ثبت شود و مطمئنا نتایج حاصل از این پژوهش در ادامه مطالعات بنیادی در این زمینه مفید واقع خواهد شد. چنین تصور می‌شود که کدون‌های بهینه برای کمک به دستیابی به سرعت‌های بالا در ترجمه و دقت بالاتر است. در این تحقیق، به ارائه شواهدی پرداختیم که نشان می‌دهد، انحراف کدونی در ژنوم کلروپلاست*G.thurberi* ، *G.barbadense*، *G.arboreum* و *G.hirsutum* ارتباط نزدیکی ترکیب نوکلئوتیدی ناشی از فشار جهشی دارد. تحقیقات بیشتر در مورد تجزیه و تحلیل مقایسه‌ای گرایش کدونی و عوامل دخیل در شکل دادن به الگوهای ترجیح کدون در میان ژن‌های میتوکندریایی، کلروپلاست و هسته در آینده، ممکن است بتواند به روشن شدن رابطه بین ژنوم هسته، کلروپلاست و میتوکندری درجنس گوسیپیوم کمک کند، و این باعث فراهم آمدن یک تحقیق جامع و مفصل درباره فرضیه اندوسیمبیوتیک و یک چهارچوب از آن برای ساخت مدل‌های قویتر جهت بهبود درک ما از تکامل مولکولی و بویژه چگونگی تفسیر داده‌های مولکولی برای بازسازی الگوهای فیلوژنی می‌شود.

سپاسگزاری

نگارندگان مقاله از مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی آذربایجان غربی و دانشگاه آزاد اسلامی واحد ارومیه بدلیل حمایت از انجام پژوهش حاضر سپاسگزاری می‌کنند.

منابع

1- Butt, A. M., Nasrullah, I. and Tong. Y. (2014). Genome-Wide analysis of codon usage and influencing factors in (*Chikungunya viruses*). *PLOS ONE*. 0090905.

2- Daniell, H., Wurdack, K. J., Kanagaraj, A., Lee, S. B., Saski, C. and Jansen. R. K. (2008). 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. *Journal of Theoretical and Applied Genetics*. 116(5):723-737.

3- Duret, L. (2000). tRNA gene number and codon usage in the *C.elegans* genome are co-adapted for optimal translation of highly expressed genes. *Journal of Trends in Genetics*. 16:287–9.

4- Elizabeth, E. M., Lotzer, J. and Eberle, M. (1989). Codon usage in plant genes. *Journal of Nucleic Acids Research*. 17(2):477–98.

5- Gao, L., Yi, X., Yang, Y. X., Su, Y. J. and Wang, T. (2009). Complete chloroplast genome sequence of a tree fern (*Alsophila spinulosa*) insights into evolutionary changes in fern chloroplast genomes. *BMC Evolutionary Biology*. 10.1186/1471-2148-9-130.

6- Hershberg, R. and Petrov, D. A. (2008). Selection on Codon Bias. *Journal of Annual Review of Genetics*. 42.110807.091442.

7- Kim, K. J. and Lee, H. L. (2004). Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng Nees*) and comparative analysis of sequence evolution among 17 vascular plants. *Journal of DNA Research*. 11(4):247-261.

8- Marais, G., Domazet-Loso, T., Tautz, D. and Charlesworth, B. (2004). Correlated evolution of synonymous and non-synonymous sites in (*Drosophila*). *Journal of Molecular Evolution*. 59:771–779.

9- Martin, D. P., Posada, D., Crandall, K. A. and Williamson, C. (2005). A modified boot scan algorithm for automated identification of recombinant sequences and recombination breakpoints. *Journal of AIDS Research and Human Retroviruses*. 21(1):98–102.

10- Nair, R. R., Nandhini, M. B., Monalisha, E., Murugan, K., Sethuraman, T., Nagarajan, S., Rao, N. S. P. Ganesh, D. (2012). Synonymous codon usage in chloroplast genome of (*Coffea arabica*). *Journal of Bioinformation*. 10.6026/97320630081096.

11- Sharp, P. M. and Li, W. H. (1986). Codon usage in regulatory genes in (*Escherichia coli*) does not reflect selection for ‘rare’ codons. *Journal of Nucleic Acids Research*. 14(19):7737–49.

12- Talat, F. and Wang, K. (2015). Comparative Bioinformatics Analysis of the Chloroplast Genomes of a Wild Diploid Gossypium and two Cultivated Allotetraploid Species. *Iranian Journal of Biotechnology*. 10.15171/ijb.1231.

13- Wright, F. and Bibb, M. J. (1992). Codon usage in the G+C rich Streptomyces genome. *Journal of Gene.* 10.1016/0378-1119(92)90669-G.

14- Xu, C., Cai, X., Chen, Q., Zhou, H., Cai, Y. and Ben, A. (2011). Factors Affecting Synonymous Codon Usage Bias in Chloroplast Genome of (*Oncidium gower*). *Journal of Evolutionary Bioinformatics.* 7 271– 278.

Codon Usage Analysis of Chloroplast Genomes in *G.thurberi*, *G.arboreum* and Comparison with Two Tetraploid Species of Cotton

Talat F.[1], Hasani-Nejad S.[2] and Badri Anarjan M.[3]

[1] Dept. of Research Seed and Plant Improvement, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia, I.R. of Iran

[2] Dept. of Biotechnology, Islamic Azad University of Urmia, Urmia, I.R. of Iran

[3] Dept. of Plant Breeding, Urmia University, Urmia, I.R. of Iran

Abstract

Analysis of codon usage is very important to optimize the production of proteins in gene expression system. *Gossypium* spp. is the most important fiber crop in the modern world. In this research the complete nucleotide sequence of the chloroplast genomes of two wild cotton species was studied and analyzed using codon W software. Synonymous codon usage of 57 protein coding genes in chloroplast genome of *G.thurberi* and *G.arboreum* was analyzed to find out the possible factors contributing codon bias. All preferred synonymous codons were found to use A/T ending codons as chloroplast genomes are rich in AT. Correspondence analysis and method of effective number of codon as NC-plot were conducted to analyze synonymous codon usage. ENC Vs GC3 plot grouped majority of the analyzed genes on or just below the left side of the expected GC3 curve indicating the influence of base compositional constraints in regulating codon usage. According to the corresponding analysis, codon bias in the chloroplast genome of *G.thurberi* and *G.arboreum* are related to their gene length, mutation bias, gene hydropathic level of each protein, gene function and selection or gene expression only subtly affect codon usage. This study provided insights into the

Key words: Correspondence Analysis, Chloroplast Genome, NC Plot