

فصلنامه (علمی)

مقدمه

ادامه حیات موجودات عالی در گرو گیاهان عالی به علت نهاد پدیده فتوسنتز برای تامین اکسیژن و بیوسنتز برای تامین غذا است. کلروپلاست‌ها به مثابه کانون‌های اصلی فتوسنتز، با تامین اکسیژن و غذا، متضمن ادامه حیات در روی کره زمین می‌باشند. لذا بکارگیری روش‌های مهندسی ژنتیک و بیوتکنولوژی در جهت بهبود کم و کیف محصولات نباتی و یا کاهش هزینه‌های تولید، مستلزم شناخت اجزاء، عملکردها و فرآیندها می‌باشد، که با مطالعه ژن‌ها و کنش‌های آن میسر خواهد شد. لذا برای تامین نیازهای غذایی جمعیت فزاینده نوع بشر، کاربرد مهندسی در ساختار و عملکرد ژن‌ها امری اجتناب ناپذیر می‌باشد. از این روی مطالعه بیوانفورماتیکی و شناخت پلاست‌ها به عنوان اندامک‌های درون سلولی با محتوای ماده ژنتیکی و کنترل کلیه مکانیزم‌های آنزیمی مسیر بیوسنتز کربوهیدرات‌ها و فتوسنتسم‌های I و II مطالعه آنها را بیش از پیش ضروری می‌نماید. در این میان علم بیوانفورماتیک به عنوان دانشی بین رشته‌ای در علوم رایانه‌ای، آمار و احتمالات، زیست‌شناسی سلولی مولکولی، ژنتیک و سایر علوم مرتبط، با فعالیت‌هایی نظیر تعیین توالی ژنوم موجودات گوناگون و کشف ژن‌ها، تنظیم ساختار پروتئین‌ها، پیش‌بینی بیان ژن، مدل‌سازی فرگشت و بازسازی فیلوژنیک از دست آوردهای شگرف علوم و فنون در قرن گذشته و حاضر بحساب می‌آید. از سال ۱۹۶۰ به بعد که نظریه Endosymbiotic به طور رسمی پذیرفته شد، تحقیقات زیادی روی عملکرد و منشاء پلاست یاخته‌های یوکاریوتی انجام شده است. طبق این نظریه، پلاست‌ها و میتوکندری‌ها به عنوان واحدهای پروکاریوتی، به صورت همزیست در یاخته‌های یوکاریوتی تکامل یافته‌اند. به بیان صریح، ژنوم پلاستی‌ریسوب فرگشتی یک نوع ژنوم سیانوباکتریایی می‌باشد [۱]. از جمله شواهد این پدیده، علاوه بر دلایل کرومولوژیک، دلایل بیوشیمیایی زیادی مبنی بر مانستگی ساختار پلاست‌ها با یاخته‌های پروکاریوتی وجود دارد، نظیر، شکل حلقوی DNA، عدم وجود ساختار نوکلئوزومی، تشابه ریبوزوم‌ها و همچنین کیفیت نسخه‌برداری، پاسخ به بلوک‌های سنتزی، تقسیم دوتایی و غیره اشاره نمود [۲]. از خصوصیات ژنوم کلروپلاست گیاهان عالی می‌توان به مواردی نظیر، کم بودن تعداد ژن‌ها، حفاظت

شدید ماده ژنتیکی و ساختار حلقوی آن اشاره کرد، ژنوم کلروپلاست در چهار ناحیه، ناحیه کپی یگانه بلند به نام (LSC) با 80Kb و ناحیه کپی یگانه کوتاه به نام (SSC) با 20Kb و دو ناحیه تکرار معکوس (IR)، هر کدام با 25Kb قابل تشخیص است که، نواحی کپی یگانه توسط دو ناحیه تکرار معکوس از هم جدا شده‌اند. تعیین توالی ژنوم کلروپلاست، از بخش‌های مهم مطالعات پلاست‌ها می‌باشد، از سویی شناخت عملکرد و رابطه متقابل کلروپلاست با قسمت‌های دیگر سلول در مطالعات مربوط به کلروپلاست‌ها نهفته است [۳]. سبزیجات چلیپائی (Cruciferae)، یا سبزیجات صلیبی یا گل‌میان به عنوان گیاهان خوراکی، در علوم تغذیه، گروه بزرگی از گیاهان متنوع با طعم‌های قوی را تشکیل می‌دهند. سبزیجاتی مانند کلم بروکلی، کلم بروکسل، کلم پیچ، گل کلم و کیل (kale) از این خانواده هستند. گیاهان تیره شب‌بویان (Cruciferae) علفی با برگ‌هایی غالباً متناوب، ساده یا منقسم و بدون گوشواره هستند. گیاهان متعلق به خانواده شب‌بو مصارف فراوانی دارند و عمدتاً در صنایع غذایی مصرف زیادی داشته و به عنوان علفه، گیاهان زینتی، گیاهان دارویی و همچنین به عنوان گیاهان شهدآور، مورد استفاده قرار می‌گیرند. مصرف تعداد زیادی از گیاهان این تیره به علت وجود گلوکزید اسانس‌های گوگرددار در آنها است بر اثر وجود همین مواد در تغذیه نیز بکار می‌روند و یا به عنوان دارو جهت سم‌زدایی مواد سرطان‌زا مانند آفلاتوکسین مصرف می‌شوند. بعضی از این گیاهان دانه‌های روغن‌دار دارند و از آنها روغن‌های مختلف برای مصارف تغذیه درمانی و یا صنعتی تهیه می‌گردد. در پژوهش حاضر به شناخت، ساختار و عملکرد ژن‌های موجود در DNA کلروپلاست و چگونگی اشتقاق گونه‌های پلی‌پلوئیدی از گونه‌های دیپلوئید و همچنین انطباق و افتراق توالی ژنومی در خانواده Brassicaceae خواهیم پرداخت. در این راستا ژنوم کامل DNA کلروپلاست دو نوع دیپلوئید خانواده شب‌بو *B. oleracea* (C)، *B. rapa* (A) و $n=10$ با دو نوع آلوتتراپلوئید *B. napus* (AC)، *B. juncea* (AB) و $n=19$ و $n=18$ مطالعه و ارزیابی قرار گرفت و از باب تعیین موقعیت هر کدام از ژن‌ها، تعداد نوکلئوتیدها، نواحی بین‌ژنی، اینترون‌ها، نواحی تکرار معکوس و ژن‌های کد کننده پروتئین و همچنین

هم پوشانی ژن‌ها، طول نواحی رمز کننده (CDS، tRNA و rRNA)، نواحی غیر رمز کننده (اینترن و IGS) به دست آمد.

از نرم افزار REPuter قالب FASTA توالی ژنومی هر کدام از گونه‌ها را دریافت کرده و بعد از تحلیل توالی های تکراری نتایج را در قالب فایل های TEXT که شامل طول تکرار قسمت اول، نقطه آغازین قسمت اول، جهت تطابق، طول تکرار قسمت ثانوی، نقطه آغازین قسمت ثانوی، فاصله توالی تکراری و مقدار محاسباتی (e-value) هر کدام از توالی های تکراری اعم از توالی های مستقیم، معکوس، متمم و پالیندرومی را مشخص کردیم

از موضوعات مهم در مطالعه مقایسه‌ای ژنوم اندامک‌های سلولی در گونه‌های یک خانواده و بررسی فیلوژنیک آنها مطالعه محتوی و درصد نوکلئوتیدهای مختلف (A، T، C و G) می‌باشد. از طرفی مطالعه درصد محتوی AT و GC تمام قسمت‌های ژنوم اعم از پروتئین‌ها، نواحی اینترون و نواحی بین ژنی و غیره از اهمیت زیادی برخوردار است. در مطالعه حاضر، برای بررسی محتوای GC و AT ژنوم و قسمت‌های مختلف آن از نرم‌افزار DNA Assembler Baser v5.15.0 استفاده گردید.

برنامه Counter Percent Base DNA با دریافت فایل قالب FASTA هر کدام از گونه‌ها و یا هر فایل ژنومی، مقدار درصد هر کدام از بازهای تشکیل دهنده توالی ژنوم را به همراه مقدار درصد GC و AT هر کدام از گونه‌ها را محاسبه و ارائه می‌دهد. علاوه بر تعیین محتوی منونوکلئوتیدهای نواحی مختلف ژنوم هر چهار گونه با استفاده از نرم افزار DNAMAN، جداسازی و تهیه بانک اطلاعاتی قسمت‌های مختلف ژنوم هر چهار گونه اعم از CDS، EXON، INTRON، GENES، IR، SSC و LSC و SOURCE در برنامه Excel تهیه و ماکروها و برنامه‌های لازم جهت شناسائی و شمارش و تلفیق منونوکلئوتیدها در Excel تهیه و اجرا شد، عمده دستورهای بکار رفته در تحلیل این قسمت از مطالعه بصورت زیر می‌باشند.

=LEN(cellNo)

=LEN(cellNo)-

LEN(SUBSTITUTE(cellNo,"char",""))

ژن‌های rRNA و tRNA، ژن‌های تک و یا مضاعف در IR تحلیل و مقایسه خواهند شد.

مواد و روش‌ها

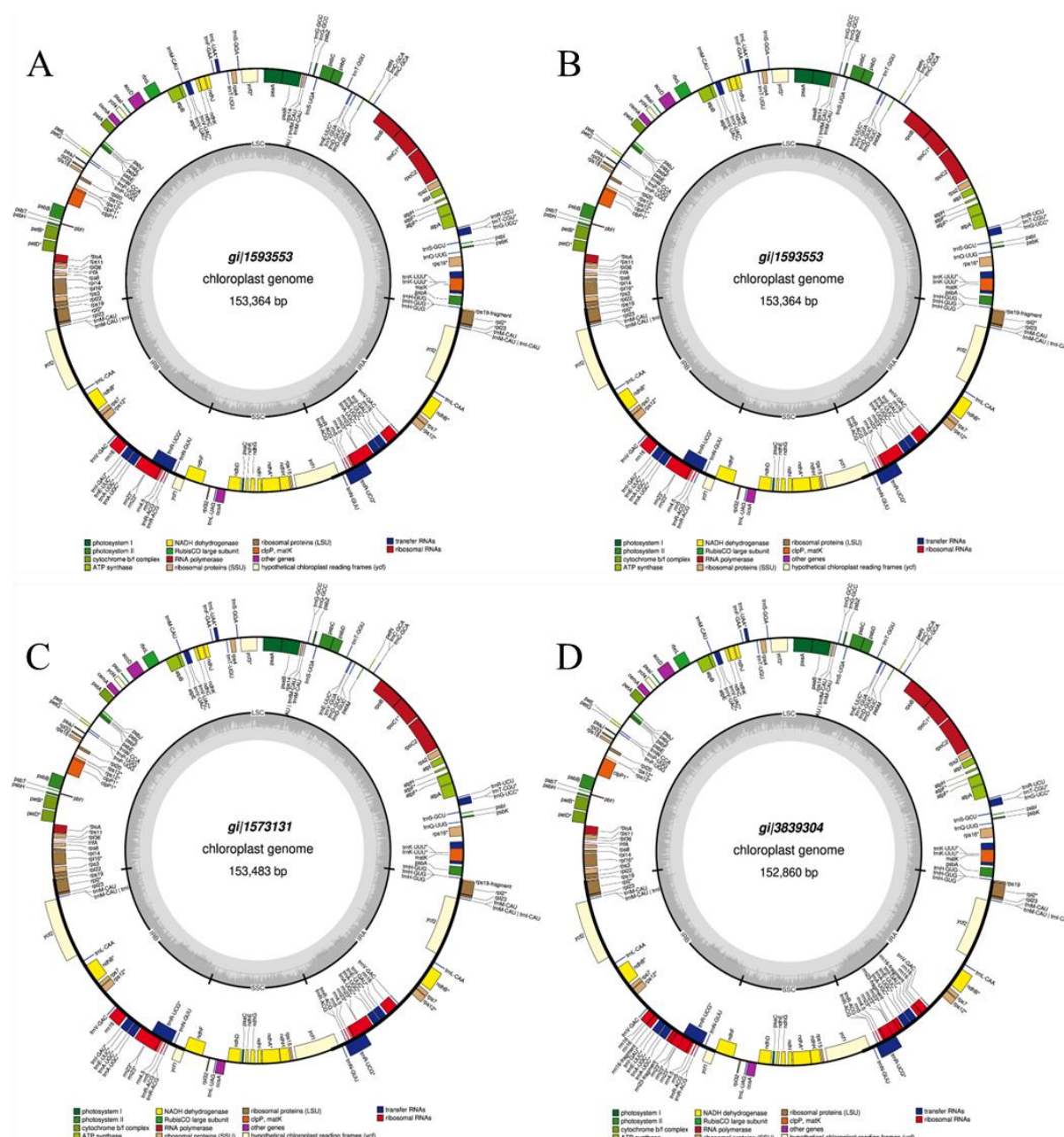
دانلود توالی کامل ژنوم

در تحقیق حاضر، توالی کامل ژنوم کلروپلاست گونه‌های مورد بحث شامل دو گونه دیپلوئید *oleracea B.* (NC_041167.1)، *rapa B.* (NC_040849.1؛ n=9) C، *napus B.* (NC_016734.1؛ n=19) AC، *junceae* (NC_028272.1؛ n=18) AB، از پایگاه NCBI به فرمت FastA و Genbank بارگیری گردیدند.

آنالیزهای انجام شده

ابتدا، نقشه ساختار ژنوم‌ها و توزیع ژن‌ها در توالی ژنوم‌ها در هر ۴ گونه مشخص گردید. برای این امر، از نرم‌افزار OGDRAW استفاده کردیم که با دریافت شماره دسترسی ژنوم‌ها ژنوم اندامک کلروپلاست هر کدام از گونه‌ها، قالب گرافیکی و نقشه ژنوم مربوطه را به صورت حلقوی (شکل واقعی کروموزوم) ترسیم و ژن‌های ژنوم را در محیط حلقه به ترتیب واقعی خود مکان‌یابی و نام گذاری انجام شد. خروجی‌های گرافیکی برنامه شامل فتوسیسستم‌های I و II، کمپلکس‌های b/f سیتوکروم، ATP سنتتاز، ترانسکریپشن‌های پلی‌سیسترونیک، اینترون‌ها، دهیدروژناز NADH، زیرواحدهای بزرگ روبیسکو، RNA پلیمرازها، پروتئین‌های ریبوزمی SSU و LSU، چهارچوب‌های تقریر تصویری کلروپلاست‌ها *ycf*، RNAs ناقل، RNAs ریبوزمی، مبادی تکثیر، (Open ORFs) Reading (Frames)، *matK* و *P clp* و سایر ژن‌ها می‌باشد.

در ادامه برای حاشیه نویسی (Annotation) از نرم‌افزار GeSeq استفاده شد. در این بخش از آنالیز، توالی ژنوم گیاهان در قالب FASTA بارگیری شده در نرم‌افزار جهت بررسی اسم ژن‌های ژنوم گونه‌ها، ژن‌های مربوط به گروه‌های مختلف ژنی، شناسایی موقعیت تک تک ژن‌ها، محل و تعداد دقیق نوکلئوتیدهای هر ژن در توالی، آگزونها و اینترون‌ها و سایر نواحی ژنوم‌ها بارگزاری شد. در ادامه این آنالیز، نواحی بین ژنی،



شکل ۱- نقشه حاشیه نویسی شده ژنوم کلروپلاست چهار گونه مورد مطالعه با استفاده از برنامه GeSeq: A: *B. oleracea*, B: *B. juncea*, C: *B. rapa*, D: *B. napus*.

B. juncea

نتایج و بحث

نتایج حاصل از اجرای برنامه OGDraw و حاشیه نویسی (Annotation) با استفاده از نرم افزار GeSeq برای نمایش نقشه کلی ژنوم گونه‌های مورد مطالعه در اشکال ۱ تا ۴ ارائه شده است. ژنوم کلروپلاست هر چهار ژنوم مورد بررسی حلقوی بوده و گونه *B. oleracea* با یک ترکیب نوکلئوتیدی به طول ۱۵۳۳۶۴ جفت باز می‌باشد. ژنوم کلروپلاست *B. rapa* متشکل از

۱۵۳۴۸۳ جفت باز، *B. napus* متشکل از ۱۵۲۸۶۰ جفت باز و *B. juncea* نیز متشکل از ۱۵۳۴۸۳ جفت باز می‌باشند. دو ناحیه کپی یگانه بزرگ و ناحیه کپی یگانه کوچک توسط دو ناحیه تکرار معکوس IRa و IRb از هم جدا شده‌اند. ترکیب نوکلئوتیدی ژنوم کلروپلاست چهار گونه مورد مطالعه بر اساس داده‌های برنامه OGDraw در جدول ۱ ارائه شده است. داده‌های جدول ۱ حاکی از این است، طول مربوط به ناحیه LSC در ژنوم

تکرار معکوس، همچنین نواحی کپی یگانه و کل ژنوم قابل مشاهده می‌باشند.

با بکارگیری برنامه DNAMAN که مشتمل بر نمایش خطی کل ژنوم گونه‌های مورد مطالعه به صورت دای نوکلئوتید (بازهای جفت شده) که موقعیت قسمت‌های مختلف ژنوم اعم از قسمت‌های کد کننده CDS توالی‌ها، ژن‌ها Genes، نواحی میان ژنی اینترون‌ها Introns و نواحی اگزونی Exons بررسی و با استفاده از برنامه DNAMAN برای نمایش Dotplot ژنوم هر کدام از گونه‌های مورد مطالعه که بصورت Self_alignment تهیه شده است، بکار گرفته شد، این روش مانسته آنالیز Self_Blast در ماژول Blast، سایت NCBI می‌باشد. بررسی نتایج تکنیک Self_Blast، حاکی از این است که ژنوم کلروپلاست گونه‌های مورد مطالعه از چهار بخش کاملاً مشخص تشکیل شده‌اند، نتایج حاصل از این تکنیک در شکل ۲ ارائه شده است.

کلروپلاست هر چهار گونه یکسان و برابر ۳۲۷۵۹ جفت باز می‌باشد. ولی ناحیه SSC در دو گونه *rapa B.* و *juncea B.* ۱۷۷۷۵ جفت باز برابر ولی در مورد دو گونه دیگر متفاوت می‌باشند. در ضمن کوتاه‌ترین SSC با ۱۷۷۶۰ جفت باز مربوط به گونه *napus B.* می‌باشد.

نتایج خروجی برنامه GeSeq نشان داد ناحیه ژن‌های کدکننده در گونه *napus B.* با ۱۳۳۴۰۰ جفت باز در ناحیه ژن‌های کدکننده و اختصاص ۸۷/۲۷ درصد از کل ژنوم به ناحیه ژن‌های کدکننده در مقایسه با *oleracea B.* با ۸۰/۲۵ درصد و *rapa B.* با ۸۰/۱۸ درصد *juncea B.* با ۸۰/۱۸ درصد ناحیه کدکننده بیشتری را شامل می‌شود (شکل ۱). درصد مربوط به ژن‌های کدکننده پروتئین و rRNA و tRNA در مورد هر یک از چهار ژنوم مورد بررسی در جدول ۲ ارائه شده است. مقادیر و درصد‌های مختص هر کدام از نواحی ژنوم اعم از نواحی ژن‌های کدکننده، اگزون‌ها Exons، اینترون‌ها Introns-inter-genic Regions، نواحی بین ژنی Regions Intra-genic، نواحی

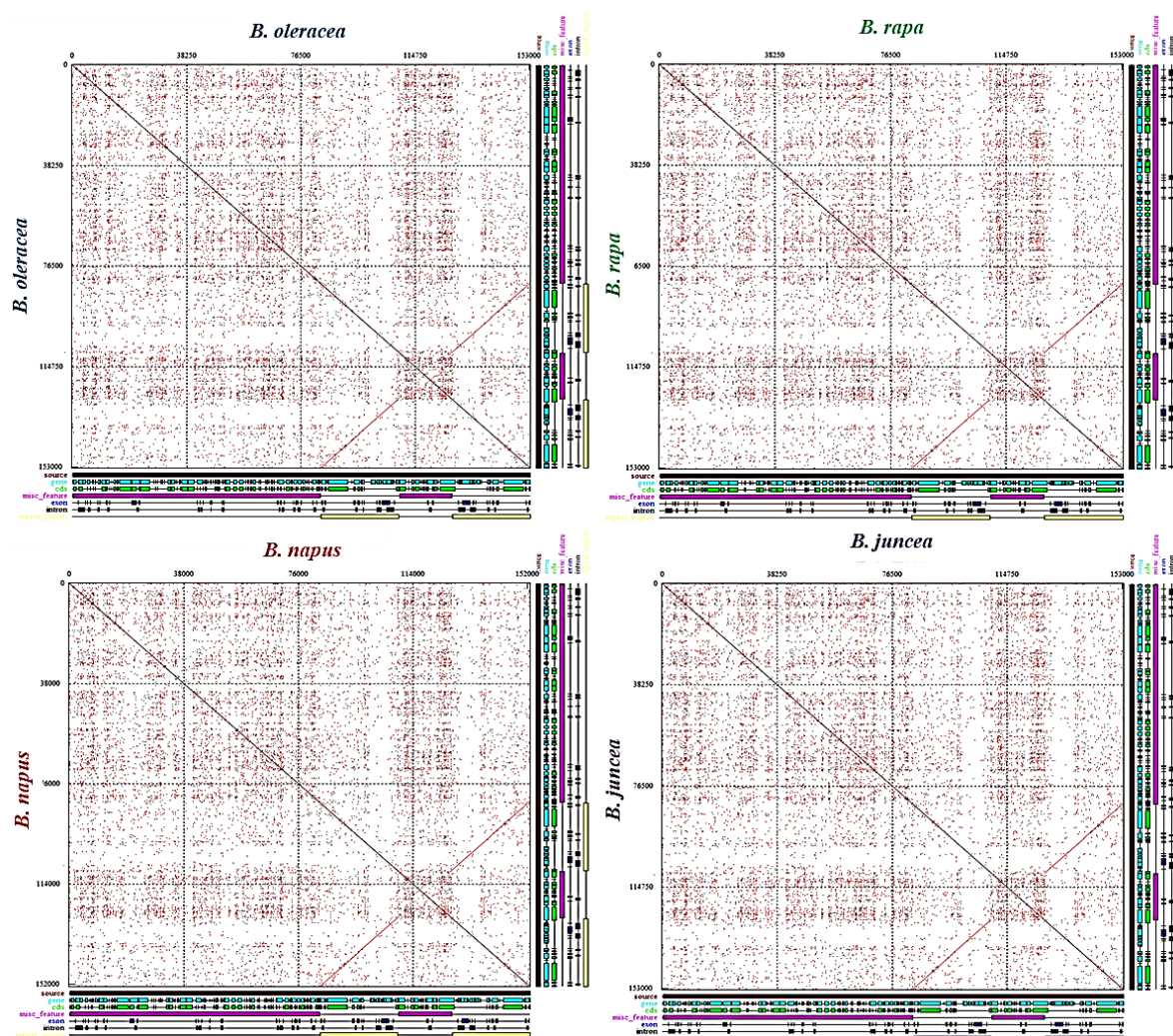
جدول ۱- ترکیب نوکلئوتیدی ژنوم کلروپلاست چهار گونه مورد مطالعه خانواده Brassicaceae

Nc distribution length in Brassicaceae family CP.Genome Global								
No.	Brassica family	Features						Source
		Genes	Intron	LSC	SSC	IRa	IRb	Complete Genome
1	<i>B. oleracea</i>	123080	30178	32759	17834	26197	26197	153364
2	<i>B. rapa</i>	123061	30197	32759	17775	26213	26213	153483
3	<i>B. napus</i>	133400	29866	32759	17760	26035	26035	152860
4	<i>B. juncea</i>	123059	30195	32759	17775	26211	26211	153483

جدول ۲- توزیع نوکلئوتیدی نواحی مختلف ژنومی در توالی ژنوم‌های خانواده Brassicaceae

CP.Genome Normalized Nc distribution length in Brassicas family															
No.	Brassica family	Coding Region					Intermittent region	Non Coding Region			Misc Features				Source
		Protein	tRNA	rRNA	Total	CDS	Exon	IGS+Junk	Intron	Total	LSC	SSC	IRa	IRb	Complete Genome
1	<i>B. oleracea</i>	60.63	13.72	5.90	80.25	52.53	12.20	68.05	19.68	87.73	21.36	11.63	17.08	17.08	100.00
2	<i>B. rapa</i>	60.56	13.72	5.90	80.18	52.42	12.19	67.99	19.67	87.66	21.34	11.58	17.08	17.08	100.00
3	<i>B. napus</i>	60.63	13.57	13.07	87.27	52.15	12.16	75.11	19.54	94.65	21.43	11.62	17.03	17.03	100.00
4	<i>B. juncea</i>	60.56	13.72	5.90	80.18	52.42	12.19	67.98	19.67	87.66	21.34	11.58	17.08	17.08	100.00

CP.Genome Nc distribution length in Brassicas family															
No.	Brassica family	Coding Region					Intermittent region	Non Coding Region			Misc Features				Source
		Protein	tRNA	rRNA	Total	CDS	Exon	IGS+Junk	Intron	Total	LSC	SSC	IRa	IRb	Complete Genome
1	<i>B. oleracea</i>	92984	21046	9050	123080	80559	18715	104365	30178	134543	32759	17834	26197	26197	153364
2	<i>B. rapa</i>	92947	21064	9050	123061	80463	18715	104346	30197	134543	32759	17775	26213	26213	153483
3	<i>B. napus</i>	92681	20745	19974	133400	79715	18584	114816	29866	144682	32759	17760	26035	26035	152860
4	<i>B. juncea</i>	92947	21062	9050	123059	80463	18715	104344	30195	134539	32759	17775	26211	26211	153483



شکل ۲- خروجی Dotplot ژنوم چهار گونه مورد مطالعه به صورت Self_aligment

است. جداول مذکور حاوی اطلاعاتی از قبیل، طول توالی تکرار، مبدا شروع قسمت اول تکرار، نوع توالی تکرار، نقطه شروع قسمت دوم تکرار و مقدار محاسباتی e می‌باشند. توالی‌های تکرار ژنوم کلروپلاست *B. oleracea* شامل یک توالی تکرار مکمل (C)، ۱۰ توالی تکراری مستقیم (F)، ۲۵ توالی تکرار پالیندرومیک (P)، و ۱۴ توالی تکرار معکوس (R)، می‌باشد (جدول S۳). توالی‌های تکرار ژنوم‌های کلروپلاست گونه‌های *B. rapa* و *B. juncea* مشابه بوده که هر کدام شامل یک توالی تکرار مکمل (C)، ۱۲ توالی تکراری مستقیم (F)، ۲۳ توالی تکرار پالیندرومیک (P)، و ۱۲ توالی تکرار معکوس (R)، می‌باشند (جدول S۴ و S۶). گونه *B. napus* شامل یک توالی تکرار مکمل (C)، ۸ توالی تکراری مستقیم (F)، ۲۶ توالی تکرار پالیندرومیک (P)، و ۱۵ توالی تکرار معکوس (R)، می‌باشد

ساختار ژنوم کلروپلاست، مشابه پروکاریوت‌ها بوده و داشتن توالی‌های تکراری در مقیاس بزرگ، غیرمعمول به نظر می‌رسد. در مطالعه حاضر، بعد از حاشیه نویسی و مطالعه موقعیت و مشخصات نواحی کدکننده و هر کدام از ژن‌ها، شناسایی چگونگی توزیع توالی‌های تکراری گونه‌های *B. oleracea*، *B. rapa*، *B. juncea* و *B. napus* از نرم افزار REPuter استفاده شد. در کل ۴ نوع تکرار شامل توالی‌های تکرار مستقیم^۱، معکوس^۲، مکمل^۳ و پالیندرومیک^۴ شناسایی شدند. ۵۰ توالی تکراری که بیش از ۲۰ جفت باز طول دارند برای ژنوم هر چهار گونه مورد بررسی در جداول S۳ تا S۶ ارائه شده

¹ forward or direct match

² reverse match

³ complement match

⁴ palindromic or inverted match

(جدول S5). نتایج نشان دادند که کوتاه‌ترین توالی تکرار در دو گونه *B. rapa* و *B. juncea* با ۲۰ جفت باز و بلندترین آن با ۴۲ جفت باز می‌باشد، در حالی که در *B. oleracea* کوتاه‌ترین توالی تکرار مشتمل بر ۲۰ و بلندترین آن ۴۷ جفت باز و در مورد گونه *B. napus* کوتاه‌ترین توالی تکرار شامل ۲۰ و بلندترین آن ۴۰ جفت باز می‌باشد.

ژنوم کلروپلاست، یک ژنوم محافظت شده است و بنابراین به نظر می‌رسد فاقد توالی‌های تکراری ساده باشد، در حالیکه، حضور SSR در ژنوم کلروپلاست بسیاری از گونه‌ها گزارش شده است. Kim و همکاران (۲۰۰۶) [۴] گزارش کردند که SSR در ژنوم کلروپلاست‌ها 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 [۵]. به لحاظ قابلیت نشانگری توالی‌های تکراری ساده SSRs علاوه بر مطالعه تکرارهای مرسوم، مستقیم، معکوس، مکمل و پالیندرومی توالی‌های تکراری ساده^۱ (SSR) نیز وجود مطالعه شدند. توالی‌های تکراری ساده هر چهار ژنوم کلروپلاست مورد بررسی، شناسایی و در جداول S8-S11 ارائه شد. نتایج نشان داد که ژنوم کلروپلاست دو گونه *B. juncea* و *B. rapa* هر دو دارای ۱۷ توالی تکراری ساده دو نوکلئوتیدی با طول توالی بین ۳۲ تا ۵۴ بوده، و گونه *B. oleracea* دارای ۱۸ توالی تکراری ساده دو نوکلئوتیدی با طول توالی بین ۳۲ تا ۴۶ و گونه *B. napus* با ۱۸ توالی تکراری ساده دو نوکلئوتیدی با طول توالی بین ۳۲ تا ۶۷ می‌باشد. ژنوم گونه‌های *B. rapa*، *B. juncea* و *B. napus* هرکدام دارای ۵۸ توالی تکرار ساده سه نوکلئوتیدی و *B. oleracea* دارای ۵۹ توالی تکرار سه نوکلئوتیدی می‌باشد. توالی‌های تکرار ساده با دو نوکلئوتید، سه نوکلئوتید، چهار نوکلئوتید و بیشتر و همچنین موتیف‌های ناقص هر چهار گونه مورد مطالعه در جدول ۷ ارائه شده است.

محتوای ژنی

ژن‌های کدکننده در ژنوم کلروپلاست گونه‌های مورد مطالعه در جدول ۱۲ آورده شده است. نتایج نشان می‌دهند که در هر چهار

ژنوم کلروپلاست یاد شده، کل ژنوم حاوی ۱۱۷ ژن بوده که شامل ۸۰ ژن کد کننده پروتئین، ۴ ژن RNA ریپوزومی و ۳۳ ژن RNA ناقل می‌باشد که به عبارت دیگر می‌توان گفت در کل، ۷۸ ژن تک نسخه‌ای و ۳۹ ژن دو و یا چند نسخه‌ای^۲ واقع در ناحیه تکرار معکوس مشمول هر چهار ژنوم مورد بررسی می‌باشند. همه ژن‌های مورد بررسی بر اساس عملکرد ژن می‌توانند به عنوان ژن‌های سیستم ژنتیک عملکردی، سیستم فتوسنتزی و تعدادی نیز با عملکرد ناشناخته طبقه‌بندی شوند.

در هر یک از ژنوم کلروپلاست گونه‌های مورد مطالعه تعداد ۴ ژن با عملکرد ناشناخته بنام *ycf* شناسایی، و به عنوان ژن‌های ضروری گیاهان در نظر گرفته شده، که در میان گونه‌ها به شدت محافظت شده‌اند [۶]. با توجه به نتایج بدست آمده، در ژنوم کلروپلاست، دو ژن حائز اهمیت می‌باشند. یکی از آنها ژن *rps12* است و به ۲ قطعه که یکی با اگزون انتهایی ۵ واقع در LSC و دیگری با اگزون انتهایی ۳ واقع در IR (توسط یک اینترون) جدا می‌گردد، که تنها ژن شکافته شده در تمام گونه‌ها می‌باشد [۷]. *matk*، ژن دیگری است که ۱/۵ کیلوباز طول داشته و در ناحیه‌ی اینترون *trnK/UUU* شناسایی شده است، که تنها ژن واقع در اینترون بوده و پروتئین *maturase K* را کد می‌کند. گزارش شده است که این ژن (*matk*) دارای هر دو قطعه متغیر و محافظت شده می‌باشد و این ویژگی موجب شده که در مطالعات فیلوژنتیکی بطور مکرر مورد استفاده قرار گیرد [۸-۱۱].

ژنوم کلروپلاست یک ژنوم محافظت شده است، ولی تحقیقات نشان داده‌اند که تعدادی از ژن‌ها در طی تکامل، حذف یا منتقل شده‌اند. ژن *infA*، فاکتور^۱ آغازین ترجمه را کد می‌کند [۱۱، ۱۲]. این ژن احتمالاً، دفعات زیادی به ژنوم هسته‌ای منتقل گردیده است، که این امر باعث شده تا *infA* به عنوان شناخته شده‌ترین ژن متحرک کلروپلاست در گیاهان گزارش شود [۱۱]. ژن *infA* در کلروپلاست *Cassava* حذف گردیده است [۱۳]. نتایج نشان داد که *infA* در هر چهار ژنوم کلروپلاست گونه‌های مورد مطالعه حضور دارد. ژن *infA* در ژنوم کلروپلاست قهوه به صورت دست نخورده یافت شده است و در بعضی از گونه‌ها نیز به عنوان ژن کاذب شناخته شده است [۱۴]. نتایج حاصل نشان داد که ژن *trnP-GGG* در هر چهار ژنوم مورد بررسی حذف

² Duplicated

¹ simple sequence repeats

کلروپلاست هر چهار گونه‌ی مورد بررسی شناسایی شد و همچنین حضور این ژن در پنبه-های دیپلوئید در ناحیه‌ی IRb گزارش شده است [۱۶]. علاوه بر این، در سه لگوم *Glycine*، *Lotus* و *Medicago* نیز حذف گردیده است.

گردیده است. طبق مطالعات، این ژن در نهاندانگان نیز وجود ندارد ولی در *Cryptomeria japonica*، شناسایی شده است [۱۵]. از این رو چنین به نظر می‌رسد که این ژن شاید قبل از شکل‌گیری نهاندانگان حذف گردیده است. ژن *rpl22*، که پروتئین‌های بزرگ واحد ۲۲ ریبوزومی را کد می‌کند، در ژنوم

جدول ۷- تعداد و نوع توالی‌های تکراری ساده SSRs گونه‌های مورد مطالعه خانواده Brassicaceae

Simple Sequence Repeats (SSRs) Brassicaceae Family					
No.	Brassicaceae Family	2Nt SSRs	3Nt SSRs	4-more Nt SSRs	Imperfect Motifes
1	<i>B. oleracea</i>	18	43	310	930
		32-46	32-59	32-150	32-172
2	<i>B. rapa</i>	17	58	312	933
		32-54	32-63	32-171	32-193
3	<i>B. napus</i>	18	58	317	928
		32-67	32-96	32-195	32-242
4	<i>B. juncea</i>	17	58	312	932
		32-54	32-63	32-171	32-193

جدول ۱۲- ژن‌های کدکننده ژنوم کلروپلاست *B. juncea*، *B. napus*، *B. rapa*، *B. oleracea*

Title	Group	Gene name
Protein genes	Subunit of Acetyl-CoA-carboxylase	<i>AccD</i>
	Small subunit of rubisco	<i>rbcL</i>
	Subunit 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>petN</i> , <i>petA</i> , <i>petL</i> , <i>petG</i> , <i>petB*</i> , <i>petD*</i>
	Subunits of photosystem I AND II	<i>psbA</i> , <i>psbK</i> , <i>psbI</i> , <i>psbM</i> , <i>psbD</i> , <i>psbC</i> , <i>psaB</i> , <i>psaA</i> , <i>psaI</i> , <i>psbJ</i> , <i>psbL</i> , <i>psbF</i> , <i>psbE</i> , <i>psaJ</i> , <i>psbT</i> , <i>psbN</i> , <i>psbH</i> , <i>psaC</i> , <i>psbZ</i> , <i>psbB</i>
	DNA dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC2</i> , <i>rpoC1*</i>
	Large subunit of ribosome	<i>rpl2*</i> §, <i>rpl14</i> , <i>rpl16*</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> §, <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i> ,
	Small subunit of ribosome	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> §, <i>rps8</i> , <i>rps11</i> , <i>rps12*</i> §, <i>rps14</i> , <i>rps15</i> , <i>rps16*</i> , <i>rps18</i> , <i>rps19</i>
	Others	<i>matK</i> , <i>cemA</i> , <i>clpP**</i> , <i>infA</i> , <i>ccsA</i>
	Function unknown	<i>ycf3**</i> , <i>ycf4</i> , <i>ycf1</i> §, <i>ycf2</i> §
RNA genes	Ribosomal RNA genes	<i>rrn16</i> §, <i>rrn23</i> §, <i>rrn4.5</i> §, <i>rrn5</i> §
	Transfer RNA genes	<i>trnY-GUA</i> , <i>trnW-CCA</i> , <i>trnV-UAC</i> , <i>trnV-GAC</i> , <i>trnT-UGU</i> , <i>trnT-GGU</i> , <i>trnT-CGU</i> , <i>trnS-UGA</i> , <i>trnS-GGA</i> , <i>trnS-GCU</i> , <i>trnR-UCU</i> , <i>trnR-UCG</i> , <i>trnR-ACG</i> , <i>trnQ-UUG</i> , <i>trnP-UGG</i> , <i>trnN-GUU</i> , <i>trnM-CAU</i> , <i>trnf trnM-CAU trnI</i> , <i>trnM-CAU</i> , <i>trnL-UAG</i> , <i>trnL-UAA</i> , <i>trnL-CAA</i> , <i>trnK-UUU</i> , <i>trnI-GAU</i> - <i>trnE</i> , <i>trnI-GAU</i> , <i>trnH-GUG</i> , <i>trnG-UCC</i> , <i>trnG-GCC</i> , <i>trnF-GAA</i> , <i>trnE-UUC</i> , <i>trnD-GUC</i> , <i>trnC-GCA</i> , <i>trnA-UGC</i>

§، * و ** به ترتیب نشان‌دهنده‌ی موقعیت ژن در ناحیه‌ی IR، ژن‌های حاوی یک اینترون و ژن‌های حاوی دو اینترون می‌باشد.

کاربرد کدون

تعداد آمینه اسیدهای تشکیل دهنده پروتئین‌های موجودات زنده جمعا ۲۰ عدد می‌باشد، این ۲۰ عدد اسید آمینه با ۶۴ کدون شناسائی و انتقال و فیوژن می‌شوند. فراوانی کاربرد کدون‌های مترادف در میان گونه‌های مختلف، اندامک و حتی ژن‌ها متفاوت است. کاربرد کدون‌ها در ناحیه ژن‌های کدکننده ژنوم کلروپلاست هر چهار گونه ژنوم مورد مطالعه بررسی شد و در جداول S1۶-S1۳ که شامل کدون‌های استفاده شده و آمینواسیدهای مربوطه و تعداد کدون‌های استفاده شده با RSCU مربوطه آورده شده است. AUG (کدون آغازی) و UGG به ترتیب، متیونین و تریپتوفان را کد می‌کنند و مقدار RSCU آنها طبق نتایج بدست آمده، یک می‌باشد. مقدار RSCU سه کدون پایانی UGA، UAA و UAG به ترتیب برای *B. juncea*، *B. napus*، *B. rapa* و *B. oleracea* برابر (۰/۹۳)، (۰/۷۵، ۱/۳۲)، (۰/۷۵، ۱/۱۱)، (۰/۷۵، ۱/۱۴)، (۱/۰۲، ۱/۱۴)، (۰/۴۴، ۱/۰۳) و (۰/۷۶، ۱/۲۱) گزارش شد. این نتایج حاکی از این است که هر چهار ژنوم، کدون UAA را بیشتر از ۲ کدون دیگر به عنوان کدون پایانی ترجیح می‌دهند. مقدار RSCU اغلب کدون‌ها بیشتر از یک گزارش شده است، که این نشان می‌دهد، این کدون‌ها با تناوب بیشتری مورد استفاده قرار گرفته‌اند. همی نتایج گزارش شده تا به اینجا، در مورد *Alsoephila* [۱۷] و *Panax schinseng* Nees و [۱۸] نیز مشابه گزارش شده است.

نظر به اهمیت ترکیب بازی ژنوم و اهمیت آن در بررسی‌های فیلوژنتیک و ارتباط آن با سنج‌های مختلف کاربرد کدون، جداول S1۷-S2۴ سنج‌های مختلف کاربرد کدون در ناحیه ژن‌های کد کننده هر یک از ژنوم‌های مورد بررسی را برای ژن‌هایی که تعداد کدون آنها بیشتر از ۱۰۰ عدد می‌باشد و با استفاده از برنامه CodonW بدست آمده است، ارائه شده است. با استفاده از این جداول، نمودار توزیع Nc، که شاخص‌های Enc و GC3s در آن دخیل هستند، برای هر چهار ژنوم کلروپلاست مورد مطالعه تهیه و در شکل ۳ آورده شده است. با توجه به اینکه نتایج مربوط به این آنالیز برای هر چهار ژنوم تقریباً یکسان بدست آمد، لذا ژنوم کلروپلاست گونه *B. oleracea* را مرجع قرار داده و توضیحات بر اساس آن صورت گرفت. مقادیر ENC از ۳۵/۸۷ تا ۵۹/۷۹ متغیر است، که دارای میانگین ۴۹/۹۵ و انحراف استاندارد ۵/۴۹ می‌باشد. ناهمگنی کاربرد کدون بیشتر به دلیل تنوع مقادیر GC3 است که از ۰/۱۵۹ تا

۰/۵۷۵ متغیر بوده و دارای میانگین ۰/۳۹۱ و انحراف استاندارد ۰/۱۰۰۱ می‌باشد. Wright پیشنهاد کرده که نمودار مقادیر NC در مقابل GC3s می‌تواند برای کشف تنوع کاربرد کدون در میان ژن‌ها موثر باشد [۱۹]. او چنین بیان کرده است که مقایسه توزیع واقعی ژن‌ها با توزیع قابل انتظار، نشان می‌دهد که انحراف کاربرد کدون تحت تاثیر چه عواملی بجز محدودیت‌های ترکیبی می‌تواند باشد. اگر انحراف کاربرد کدون بطور کامل تحت تاثیر GC3 بوده باشد، مقادیر NC باید روی منحنی قابل انتظار بین GC3s و NC باشد. شکل ۳ نشان می‌دهد که تعداد کمی از نقاط، روی منحنی قرار گرفته اند و متمایل 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 مقادیر NC کم، در زیر منحنی قابل انتظار قرار گرفته‌اند. این نتایج نشان می‌دهند که اکثر ژن‌های انتخابی در *B. oleracea* کاربرد کدونی مستقل از ترکیب نوکلئوتیدی مشتق از فشار جهشی و متأثر از عوامل دیگر دارند، که مستقل از محدودیت‌های ترکیبی می‌باشند.

تاثیر کمتر فشار جهشی و دخیل بودن عوامل دیگر در سوگیری کاربرد کدون عملاً آنالیزهای تطبیقی با ترکیب‌های بازی را سوال بر انگیز می‌نمایاند. لذا برای ژن‌های کد کننده انتخاب شده برای هر چهار گونه مورد مطالعه به لحاظ قرار گرفتن داده‌ها در ناحیه Lowbias که از آنالیزهای آماری و آزمون مربع کای حاصل شد آنالیزهای تطبیقی را میسر نشان نمی‌دهد. مطالعات قبلی، عمدتاً روی انحراف کدونی در ژنوم هسته‌ای متمرکز بوده‌اند و چنین بیان شده است که انحراف کدونی، منعکس کننده تعادل بین انحرافات جهشی و انتخاب طبیعی برای بهینه‌سازی ترجمه می‌باشد [۲۰]. گزارش شده است که کدون‌های بهینه در میکروارگانیزم‌های به سرعت در حال رشد مثل اشرشیاکلی یا ساکارومایسز سرویسیا [۲۱] منعکس کننده ترکیب مخزن ژنومی tRNA می‌باشند. چنین تصور می‌شود که کدون‌های بهینه برای کمک به دستیابی به سرعت‌های بالا در ترجمه و دقت بالاتر است. تصور بر این است که انحراف کدونی با طول ژن ارتباط دارد، و چنین استدلال می‌شود که این نتیجه انتخاب است و برای به حداکثر رساندن سرعت ترجمه، به حداقل رساندن پیامدهای تصحیح کردن^۱، یا به حداکثر رساندن دقت و صحت ترجمه می‌باشد.

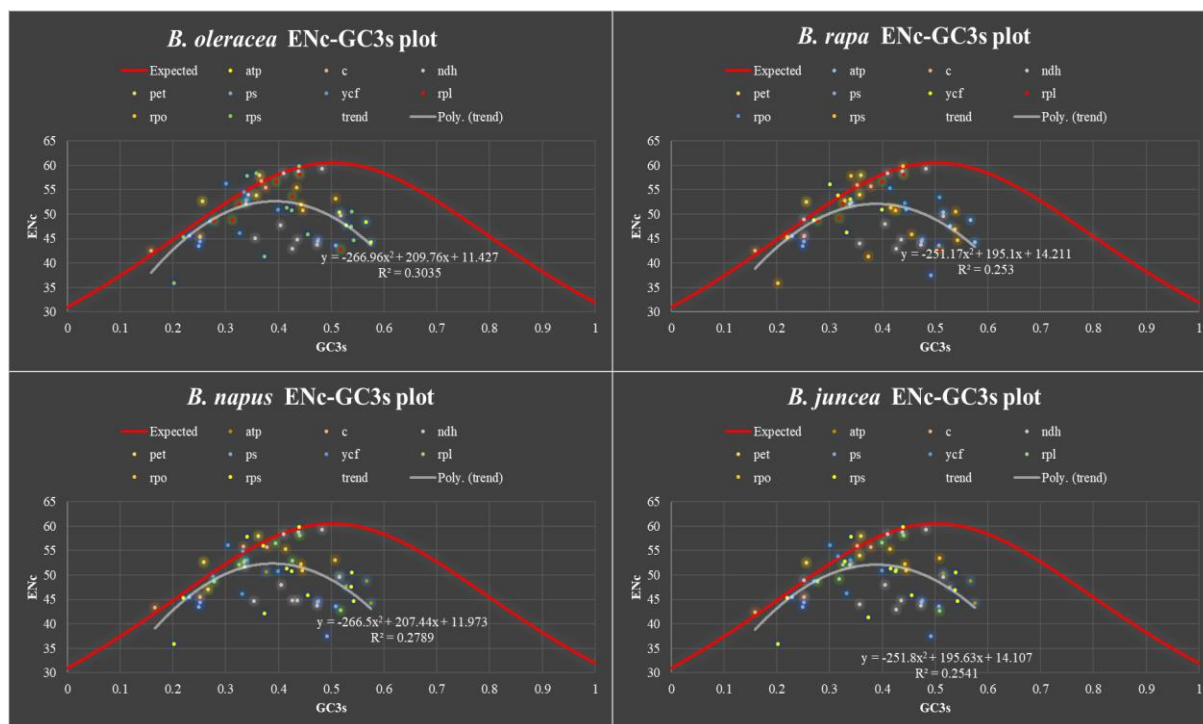
¹ proofreading

تحقیقات بیشتر در مورد تجزیه و تحلیل مقایسه‌ای انحراف کدونی و عوامل دخیل در شکل دادن به الگوهای کاربرد کدون در میان ژن‌های میتوکندریایی، کلروپلاست و هسته در آینده، ممکن است بتواند به روشن شدن رابطه بین ژنوم هسته، کلروپلاست و میتوکندری در مطالعات بیوانفورماتیکی کلروپلاست‌ها 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.

وسعت IR و اینترون‌ها

محدوده IR در بین گونه‌ها معمولاً متفاوت است و گستردگی و ادغام IR از لحاظ سائز ژنوم، مهم تلقی می‌شود. گستردگی IR، غالباً سائز بزرگ ژنوم و شبه ژن‌های موجود در مناطق اتصال IR، LSC یا SSC را تحت تاثیر قرار می‌دهد. در هر چهار ژنوم مورد مطالعه، تعداد ۱۰ ژن حاوی اینترون (۲ تایی) می‌باشند، که مشابه *Panax schinseng* بوده [۱۸]، ولی اینترون در ژن *clpP* و *rpoC1* در *B. oldhamii* و *D. latiflorus* یافت نشده

است [۲۲]. تعداد و محل اینترون در کلروپلاست، به نظر محافظت شده می‌آید. مقایسه اینترون‌ها طبق جدول ۲۵ حاکی از این است که هر چهار گونه دارای ۲۹ ژن دارای ۱ یا ۲ اینترون می‌باشند که ۱۰ عدد از آنها، ژن‌های tRNA، و ۱۹ عدد دیگر ژن‌های کد کننده پروتئین می‌باشند. طول‌ترین اینترون در ژنوم *B. rapa* با ۲۵۷۳ جفت باز، که در مورد *B. oleracea* و *B. juncea* برابر ۲۵۷۲ جفت باز و در مورد *B. napus* برابر ۲۴۵۸ جفت باز می‌باشد. که جایگاه آنها در مورد هر چهار ژنوم در ژن *trnR-UCG* قرار دارند، کوتاه‌ترین اینترون با ۱۶ جفت باز در *trnI-GAU* جای دارد، طول و موقعیت اینترون‌های مربوط به کلیه ژن‌های دو گونه *B. rapa* و *B. juncea* محافظت شده می‌باشد. در حالیکه ژن‌های حاوی اینترون در گونه *B. oleracea* از لحاظ طول اینترون، تفاوت‌های جزئی و در مورد گونه *B. napus* تفاوت‌ها چشمگیر نشان می‌دهد. بطوریکه گونه *B. napus* فقط در ژن‌های حاوی اینترون زیر (*petD*, *ndhB*), *TrnI*-, *TrnA-UGC*, *rnn23*, *rps12*, *rpoC1*, *rpl2*, *GAU*, *TrnL-UAA* و *TrnV-UAC*) با سه گونه دیگر مشابه هستند.



شکل ۳- نمودار مقادیر NC3s در برابر GC3s برای ژنوم کلروپلاست چهار گونه مورد مطالعه

جدول ۲۵ - مقایسه طول توالی‌های اینترونی در گونه‌های مورد مطالعه خانواده Brassicaceae

Intron Analysis				<i>B. oleracea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>
No	Feature	Gene		Length	Length	Length	Length
1	intron	<i>atpF</i>	C	721	715	716	715
2	intron	<i>clpP1</i>	C	936	933	571	933
3	intron	<i>clpP1</i>	C	570	571	935	571
4	intron	<i>ndhA</i>	C	1098	1091	1101	1091
5	intron	<i>ndhB</i>	C	679	679	679	679
6	intron	<i>ndhB</i>		679	679	679	679
7	intron	<i>petB</i>		781	781	780	781
8	intron	<i>petD</i>		733	733	733	733
9	intron	<i>rpl16</i>	C	1064	1068	1076	1068
10	intron	<i>rpl2</i>	C	687	687	687	687
11	intron	<i>rpl2</i>		687	687	687	687
12	intron	<i>rpoC1</i>	C	778	785	785	785
13	intron	<i>rps12</i>	C	537	537	537	537
14	intron	<i>rps12</i>		537	537	537	537
15	intron	<i>rps16</i>	C	859	866	859	866
16	intron	<i>rrn23</i>		199	199	199	199
17	intron	<i>rrn23</i>	C	199	199	199	199
18	intron	<i>trnA-UGC</i>		800	800	40	800
19	intron	<i>trnA-UGC</i>		40	40	800	40
20	intron	<i>trnA-UGC</i>		800	800	800	800

ادامه جدول ۲۵ - مقایسه طول توالی‌های اینترونی در گونه‌های مورد مطالعه خانواده Brassicaceae

Intron Analysis				<i>B. oleracea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>
No	Feature	Gene		Length	Length	Length	Length
21	intron	<i>trnA-UGC</i>	C	800	800	800	800
22	intron	<i>trnA-UGC</i>	C	800	800	800	800
23	intron	<i>trnA-UGC</i>	C	40	40	40	40
24	intron	<i>trnE-UUC</i>		809	809	781	809
25	intron	<i>trnE-UUC</i>	C	809	809	781	809
26	intron	<i>trnG-UCC</i>		717	717	718	717
27	intron	<i>trnI-GAU</i>		809	809	16	809
28	intron	<i>trnI-GAU</i>		16	16	781	16
29	intron	<i>trnI-GAU</i>	C	809	809	781	809
30	intron	<i>trnI-GAU</i>	C	16	16	16	16
31	intron	<i>trnK-UUU</i>	C	2557	2561	2559	2561
32	intron	<i>trnL-UAA</i>		313	313	313	313
33	intron	<i>trnL-UAA</i>		313	313	313	313
34	intron	<i>trnR-UCG</i>	C	2572	2573	2458	2572
35	intron	<i>trnR-UCG</i>		2572	2573	2458	2572
36	intron	<i>trnT-CGU</i>		712	712	713	712
37	intron	<i>trnV-UAC</i>	C	601	609	600	609
38	intron	<i>trnV-UAC</i>	C	17	17	17	17
39	intron	<i>ycf3</i>	C	730	730	790	730
40	intron	<i>ycf3</i>	C	782	784	731	784

Nicotiana (درصد، ۳۷/۸۶) *Solanum lycopersicum*

tabacum (درصد، ۳۷/۸۵) *Atropa belladonna* (۳۷/۵۶)

(درصد) و *Glycin max* (درصد، ۳۴) نیز گزارش شده است.

مناطق کد کننده و غیر کد کننده در هر چهار ژنوم، دارای محتوای

محتوای GC

محتوای GC کل ژنوم کلروپلاست، گونه‌های *B. oleracea*، *B.*

rapa، *B. juncea* و *B. napus* شبیه هم بود و در حدود ۳۶

درصد می‌باشد، که عدد مشابه آن برای گیاهان دیگر نظیر

GC کم و در حدود ۴۰ درصد می‌باشند. تفاوت در محتوای GC، برای ۴ ناحیه ژنوم کلروپلاست هر چهار ژنوم مورد بررسی واقع گردید که در جداول S2۹-S2۶ منعکس شده است. نتایج نشان داد که ناحیه IR از لحاظ GC غنی‌ترین می‌باشد، و این مطلب با یافته‌های Kim و همکاران (۲۰۰۹) [۷] مطابقت دارد. چنین به نظر می‌رسد که ژن‌های ریپوزومی (*rna4.5*, *rna5*, *rna16* و *rna23*) و نواحی کدکننده [۷، ۱۳، ۱۷، ۲۳] مسئول محتوای غنای GC در IR باشند. محتوای GC در ناحیه IR در *B. oleracea* ۴۲/۳۵ درصد بود، در حالی که محتوای GC در SSC و LSC به ترتیب ۲۹/۰۹ و ۳۳/۸۱ درصد گزارش شد. توزیع محتوای GC در هر منطقه از ژنوم هر چهار گونه مورد بررسی، با گونه‌های دیگر مشابه بود [۴، ۷، ۲۳]. با توجه به مطالعات Gao و همکارانش (۲۰۰۹) [۱۷]، محتوای GC در طول ژنوم کلروپلاست *Alsophila*، دارای توزیع نابرابر می‌باشد. محققان با مطالعاتی که بر روی *Alsophila spinulosa* انجام دادند، گزارش کردند که محتوای GC در ژنوم کلروپلاست، در بین گروه‌های ژنی، که از لحاظ عملکرد متفاوت می‌باشند، یکسان نیست؛ $rRNA (55.18\%) > tRNA$ ؛ $genetic (43.85\%) > photosynthetic (54.55\%)$ ؛ $NADH (39.54\%) > system (40.80\%)$ [۱۷]. در مورد، *B. oleracea*، *B. rapa*، *B. napus* و *B. juncea* نیز نتایج مشابه اخذ شد. در ناحیه کد کننده *B. oleracea*، ژن‌های rRNA دارای بیشترین محتوای GC، ۵۵/۳۶ درصد و ژن‌های کد کننده پروتئین دارای کمترین محتوای GC، به مقدار ۳۶/۵۹ درصد بود. در منطقه غیرکدکننده، محتوای GC در IGS+Junk، برابر ۳۸/۴۲ درصد و در اینترون‌ها برابر ۴۰/۳۹ درصد گزارش شد (جدول ۳۰). محتوای GC، از ویژگی‌های مهم ژنوم

کلروپلاست به شمار می‌رود، که با تعداد سایت‌های اتصال microRNA [۲۴]، موقعیت عناصر عملکردی [۲۵]، سرعت نوترکیبی و توزیع ژن [۲۶]، ویرایش RNA اندامکی [۲۷] و تنظیم بیان ژن [۲۸] در ارتباط است. محتوای ژنی در طول 5'UTR، مناطق توالی کد کننده و 3'UTR متفاوت است [۲۵]. محتوای ژنی همچنین بر روی RNAi تاثیر دارد، زیرا محتوای GC، با دسترسی به سایت هدف RNAi شدیدا در ارتباط است و با فعالیت RNAi ارتباط منفی دارد [۲۹]. محتوای GC پایین از مشخصات قابل توجه ژنوم پلاستید می‌باشد، که احتمال می‌رود پس از درون همزیستی (endosymbiosis) توسط همانندسازی DNA و اصلاح آن، تشکیل شده باشد [۳۰]. در ویروس‌ها، محتوای GC به ساختار ژن‌ها وابسته نمی‌باشد. ولی با موقعیت آن در ارتباط است [۳۱].

نتیجه‌گیری کلی

نتایج حاصل از کاربرد نرم‌افزارهای بیوانفورماتیکی و بررسی‌ها و تحلیل اطلاعات حاصل از اجرای نرم‌افزارهای مختلف در آنالیز مقایسه‌ای ژنوم کامل کلروپلاست، چهار گونه مورد مطالعه خانواده Brassicaceae حاکی از این است که در بین گونه‌های انتخاب شده، گونه *B. napus* با ۱۵۲۸۶۰ جفت باز دارای کوتاه‌ترین طول نوکلئوتیدی بوده و دو گونه *B. rapa* و *B. juncea* هرکدام با ۱۵۳۴۸۳ جفت باز دارای طول مساوی بوده و طول توالی نوکلئوتیدی در گونه *B. oleracea* برابر ۱۵۳۳۶۴ جفت باز می‌باشد. مانند ژنوم کلروپلاست اغلب گیاهان فتوسنتز کننده، هر چهار ژنوم کلروپلاست مورد مطالعه، از یک ساختار چهار جزئی، که شامل دو ناحیه تکرار معکوس IR، یک ناحیه کپی یگانه بزرگ LSC و یک ناحیه کپی یگانه

جدول ۳۰- طول نوکلئوتیدی نواحی مختلف ژنوم کلروپلاست در گونه‌های مورد مطالعه خانواده Brassicaceae

CP.Genome Nc distribution length in Brassicas family														
No.	Brassica family	Coding Region					Intermittent region	Non Coding Region			Misc Features			Source
		Protein	tRNA	rRNA	Total	CDS	Exon	IGS+Junk	Intron	Total	LSC	SSC	IR	Complete Genome
1	<i>B. oleracea</i>	92984	21046	9050	123080	80559	18715	104365	30178	134543	32759	17834	52394	153364
2	<i>B. rapa</i>	92947	21064	9050	123061	80463	18715	104346	30197	134543	32759	17775	52426	153483
3	<i>B. napus</i>	92681	20745	19974	133400	79715	18584	114816	29866	144682	32759	17760	52070	152860
4	<i>B. juncea</i>	92947	21062	9050	123059	80463	18715	104344	30195	134539	32759	17775	52422	153483

کوچک SSC برخوردار می‌باشند. در بررسی نواحی کد کننده هر چهار ژنوم، مشخص شد که گونه *B. napus* با ۷۹۷۱۵ جفت باز در ناحیه CDS دارای کوتاه‌ترین طول توالی در این ناحیه و دو گونه *B. juncea* و *B. rapa* هر کدام با ۸۰۴۶۳ جفت باز در ناحیه CDS دارای طول بازی مساوی بوده و در مورد *B. oleracea* طول این ناحیه برابر ۸۰۵۵۹ جفت باز می‌باشد. نواحی غیر کد کننده اینترونی در گونه *B. rapa* با بیشترین طول (۳۰۱۹۷ جفت باز) که فقط دو جفت باز با *B. juncea* (۳۰۱۹۵ جفت باز) اختلاف دارد. طول این ناحیه در *B. oleracea* برابر ۳۰۱۷۸ جفت باز بوده و گونه *B. napus* با ۲۹۸۶۶۰ جفت باز دارای کوتاه‌ترین طول در ناحیه غیر کد کننده اینترونی می‌باشد. طول توالی ناحیه کپی یگانه بلند LSC با ۳۲۷۵۹ جفت باز در مورد هر چهار گونه مساوی بوده ولی طول ناحیه کپی یگانه کوتاه در *B. rapa* و *B. juncea* مساوی و هر کدام برابر ۱۷۷۷۵ جفت باز و در مورد گونه *B. napus* برابر ۱۷۷۶۰ جفت باز که کوتاه‌ترین طول در ناحیه کپی یگانه کوتاه می‌باشد. طول توالی در نواحی تکرار معکوس IRa و IRb در چهار گونه متفاوت بود، بطوریکه طول این ناحیه در گونه *Napus* با مجموع ۵۲۰۷۰ جفت باز کوتاه‌ترین و در گونه *B. rapa* با طول ۵۲۴۲۶ جفت باز، بلندترین طول ناحیه تکرار معکوس در بین اعضای مورد مطالعه را به خود اختصاص داده بود و این در حالی است که با گونه *B. juncea* با ۵۲۴۲۲ جفت باز، فقط در ۴ جفت باز تفاوت داشت، طول ناحیه تکرار معکوس در گونه *B. oleracea* برابر ۵۲۳۹۴ جفت باز است. در مورد هر چهار گونه، تعداد ۵۰ توالی تکراری که بیش از ۲۰ جفت باز طول دارند شناسایی شد. توالی‌های تکرار ژنوم کلروپلاست *B. oleracea* شامل یک توالی تکرار مکمل (C)، ۱۰ توالی تکراری مستقیم (F)، ۲۵ توالی تکرار پالیندرومیک (P) و ۱۴ توالی تکرار معکوس (R)، می‌باشد. توالی‌های تکرار ژنوم‌های کلروپلاست گونه‌های *B. rapa* و *B. juncea* مشابه بوده که هر کدام شامل یک توالی تکرار مکمل (C)، ۱۲ توالی تکراری مستقیم (F)، ۲۳ توالی تکرار پالیندرومیک (P) و ۱۲ توالی تکرار معکوس (R)، می‌باشند. گونه *B. napus* شامل یک توالی تکرار مکمل (C)، ۸ توالی تکراری مستقیم (F)، ۲۶ توالی تکرار پالیندرومیک (P) و ۱۵ توالی تکرار معکوس (R)، می‌باشد. مشخص شد که

کوتاه‌ترین توالی تکرار در دو گونه *B. rapa* و *B. juncea* ۲۰ جفت باز و بلندترین آن ۴۲ جفت باز می‌باشد، در حالی که در *B. oleracea* کوتاه‌ترین توالی تکرار مشتمل بر ۲۰ و بلندترین آن ۴۷ جفت باز و در مورد گونه *B. napus* کوتاه‌ترین توالی تکرار شامل ۲۰ و بلندترین آن ۴۰ جفت باز می‌باشد. بررسی نتایج مربوط به توالی‌های تکرار حاکی از این است که اغلب تکرارها در ژن ycf2 و فواصل بین ژنی (IGS) قرار دارند، علاوه بر ۴ نوع تکرار گزارش شده، تعدادی توالی تکراری ساده (SSR) نیز شناسایی شدند که تکرارهای دو نوکلئوتیدی و سه نوکلئوتیدی و موتیف‌های نادقیق نیز شناسایی شدند. مطالعه ژن‌های کد کننده در ژنوم کلروپلاست گونه‌های مورد مطالعه حاکی از این است که در هر کدام از ژنوم‌ها، کل توالی، حاوی ۱۱۷ ژن بوده که شامل ۸۰ ژن کد کننده پروتئین، ۴ ژن rRNA ریبوزومی و ۳۳ ژن rRNA ناقل می‌باشد. در ضمن مشخص شد که در کل، ۷۸ ژن تک نسخه‌ای و ۳۹ ژن دو و یا چند نسخه‌ای^۱ واقع در ناحیه تکرار معکوس وجود دارد. همه ژن‌های مورد بررسی بر اساس عملکرد ژن می‌توانند به عنوان ژن‌های سیستم ژنتیک عملکردی، سیستم فتوسنتزی، و تعدادی نیز با عملکرد ناشناخته طبقه‌بندی شوند. در هر یک از ژنوم کلروپلاست گونه‌های مورد مطالعه تعداد ۴ ژن با عملکرد ناشناخته بنام ycf شناسایی شد. در ژنوم کلروپلاست، دو ژن حائز اهمیت می‌باشند. یکی از آنها ژن rps12 است و به ۲ 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* با ۱/۵ کیلوباز طول که در ناحیه اینترون *trnK/UUU* واقع شده است، و تنها ژن واقع شده در اینترون می‌باشد. به نظر می‌رسد که در هر چهار ژنوم مورد مطالعه، تعداد و محل اینترون در کلروپلاست‌ها، محافظت شده باشند. مقایسه اینترون‌ها نشان داد که در هر چهار گونه ۲۹ ژن با ۱ یا ۲ اینترون وجود دارد که ۱۰ عدد از آنها ژن‌های tRNA و ۱۹ عدد دیگر ژن‌های کد کننده پروتئین می‌باشند. طویل‌ترین اینترون در ژنوم *B. rapa* با ۲۵۷۳ جفت باز که در مورد *B. oleracea* و *B. juncea* برابر ۲۵۷۲ جفت باز و در مورد *B. napus* برابر ۲۴۵۸ جفت باز می‌باشد که جایگاه آنها در مورد هر چهار ژنوم در ژن *trnR-UCG* قرار

¹ Duplicated

مقاله حاضر، حاوی فایل تکمیلی می‌باشد. این فایل شامل برخی جداول که با علامت S مشخص شده است.

منابع

- [1] Gray, M W. Organellar evolution michael w. Gray and david f. Spencer. Evolution of microbial life, 1996; 54: 109.
- [2] Majd A, Shariatzadeh M A. Cell and molecular biology, Ayge Publications, 2013; 18th edition.
- [3] Talat F, Wang K. Chloroplast genome study, new tool in plant biotechnology; Gossypium Spp. As a model crop. J. Curr. Res. Sci. 2014; 2: 838.
- [4] Kim JS, Jung JD, Lee JA, Park HW, Oh KH, Jeong WJ, Choi DW, Liu JR Cho KY. Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome. Plant cell reports, 2006; 25: 334-340.
- [5] Ribeiro MM, LeProvost G, Gerber S, Vendramin GG, Anzidei M, Decroocq S, Marpeau A, Mariette S, Plomion C. Origin identification of maritime pine stands in France using chloroplast simple-sequence repeats. Annals of Forest Science. 2002; 59(1): 53-62.
- [6] Yang X, Wang Y, Gong W, Li Y. Comparative Analysis of the Codon Usage Pattern in the Chloroplast Genomes of Gnetales Species. International Journal of Molecular Sciences. 2024; 25(19): 10622. <https://doi.org/10.3390/ijms251910622>
- [7] Kim E H, Li X P, Razeghifard R, Anderson J M, Niyogi K K, Pogson B J, Chow W S. The multiple roles of light-harvesting chlorophyll a/b-protein complexes define structure and optimize function of Arabidopsis chloroplasts: a study using two chlorophyll b-less mutants. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 2009; 1787(8): 973-984.
- [8] Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssallem IS, Yu J. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PloS one. 2010; 5(9): e12762.
- [9] Wilson CA. Phylogeny of Iris based on chloroplast *matK* gene and *trnK* intron sequence data. Molecular Phylogenetics and Evolution. 2004; 33(2): 402-12.
- [10] Ohsako T, Ohnishi O. Nucleotide sequence variation of the chloroplast *trnK/matK* region in two wild Fagopyrum (polygenaceae) species, *F. leptopodum* and *F. statice*. Genes & genetic systems. 2001; 76(1): 39-46.
- [11] Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ. Many

دارند، کوتاه‌ترین اینترون با ۱۶ جفت باز در *trnI-GAU* جای دارد، طول و موقعیت اینترون‌های مربوط به کلیه ژن‌های دو گونه *B. juncea* و *B. rapa* محافظت شده می‌باشد. در حالیکه ژن‌های حاوی اینترون در گونه *B. oleracea* از لحاظ طول اینترون، تفاوت‌های جزئی و در مورد گونه *B. napus* تفاوت‌ها چشمگیر نشان می‌دهد. بطوریکه گونه *B. napus* فقط در ژن‌های حاوی اینترون زیر (*rpoC1*, *rpl2*, *petD*, *ndhB*)، *TrnL*، *TrnI-GAU*، *TrnA-UGC*، *rnn23*، *rps12* و *UAA* و *TrnV-UAC*) با سه گونه دیگر مشابه هستند. محتوای GC کل ژنوم کلروپلاست، گونه‌های *B. rapa*، *B. oleracea*، *B. juncea* و *B. napus* شبیه هم و در حدود ۳۶ درصد بود، مناطق کدکننده و غیر کدکننده در هر چهار ژنوم، دارای محتوای GC کم و در حدود ۴۰ درصد می‌باشند. تفاوت در محتوای GC، ۴ ناحیه ژنوم کلروپلاست برای هر چهار ژنوم مورد بررسی واقع گردید و مشخص شد که ناحیه IR از لحاظ GC غنی‌ترین ناحیه می‌باشد، محتوای GC در ناحیه IR در *B. oleracea* ۴۲/۳۵ درصد بود، در حالی که محتوای GC در SSC و LSC به ترتیب ۲۹/۰۹ و ۳۳/۸۱ درصد می‌باشد. توزیع محتوای GC در هر منطقه از ژنوم هر چهار گونه مورد بررسی، با سایر گونه‌های مطالعه شده مشابه بود. در این تحقیق، به شواهدی رسیدیم که نشان می‌دهد گونه‌های *B. rapa* و *B. juncea* از نظر ژنتیکی قرابت خیلی نزدیکی به هم دارند که داده‌های فنوتیپی این دو گیاه نیز این امر را تأیید می‌کند، گونه *B. napus* از نظر قرابت ژنتیکی فاصله نسبتاً زیادی با سه گونه دیگر دارد. انتخاب ژن‌های بیشتر در آنالیزهای کاربرد کدونی شاید بتواند به عمق شکاف موجود در قرابت ژنتیکی گونه *B. juncea* با سایر گونه‌های خانواده بیافزاید و اطلاعات دقیقی در مورد انشقاق و تبار این گونه بدهد و حقایق زیادی را در مورد تکامل اعضای این خانواده روشن بنماید.

سپاسگزاری

نگارندگان مقاله از مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان آذربایجان غربی به دلیل حمایت از انجام پژوهش حاضر سپاسگزاری می‌کنند.

فایل تکمیلی

- [12] parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *The Plant Cell*. 2001; 13(3): 645-58.
- [13] Sands JF, Cummings HS, Sacerdot C, Dondon L, Grunberg-Manago M, Hershey JW. Cloning and mapping of *infA*, the gene for protein synthesis initiation factor IFI. *Nucleic acids research*, 1987; 15(13): 5157-5168.
- [14] Daniell H, Wurdack KG, 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. *Theor Appl Genet*. 2008; 116(5): 723-737.
- [15] Samson N, Bausher MG, Lee SB, Jansen RK, Daniell H. The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plant Biotechnology Journal*. 2007; 5(2): 339-53.
- [16] Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC plant biology*, 2008; 8: 1-20.
- [17] Talat F, Gheibi M. Bioinformatics analysis of the chloroplast genome of *G.stocksii* and *G.longicalyx* diploid species and comparing them with tetraploid species. *Developmental Biology*, 2023; 1(15): 37-48. <http://doi.org/10.30495/jdb.2023.1959300.1305>.
- [18] Gao L, Yi X, Yang YX, Su YJ, Wang T. Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evol Biol*, 2009; 9:130.
- [19] Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA research*. 2004; 11(4): 247-61.
- [20] Zavala B, Dineen L, Fisher KJ, Opulente DA, Harrison MC, Wolters JF, Shen XX, Zhou X, Groenewald M, Hittinger CT, Rokas A. Genomic factors shaping codon usage across the *Saccharomycotina* subphylum. *G3: Genes, Genomes, Genetics*. 2024; jkae 207. <https://doi.org/10.1093/g3journal/jkae207>
- [21] Yu Z. Codon usage bias analysis of 22 *Anemone*. *Highlights in Science, Engineering and Technology*. 2024; 102: 328-37. <https://doi.org/10.54097/04rwse67>
- [22] Coghlan A, Wolfe KH. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast*, 2000; 16(12): 1131-1145.
- [23] Wu FH, Kan DP, Lee SB, Daniell H, Lee YW, Lin CC, Lin NS, Lin CS. Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. *Tree Physiology*. 2009; 29(6): 847-56.
- [24] Mardanov ES, Zamchuk LA, Skulachev MV, Ravin NV. The 5' untranslated region of the maize alcohol dehydrogenase gene contains an internal ribosome entry site. *Gene*. 2008; 420(1):11-6.
- [25] Davis N, Biddlecom N. On the relationship between GC content and the number of predicted microRNA binding sites by MicroInspector. *Comput Biol Chem*. 2008; 32(3): 222-6.
- [26] Zhang B, Pan X, Anderson TA. MicroRNA: a new player in stem cells. *Journal of cellular physiology*. 2006; 209(2): 266-9.
- [27] Wang D, Yang B. Analysis of codon usage bias of thioredoxin in *apicomplexan* protozoa. *Parasites & Vectors*. 2023; 16(1): 431. <https://doi.org/10.1186/s13071-023-06002-w>
- [28] Smith G. (2009). *Democratic Innovations: Designing Institutions for Citizen Participation*. Cambridge University.
- [29] Paterson A, Lan TH, Wendel J, Reischmann K, Chang C, Lin YR, Liu SC, Burow M, Kowalski S, Katsar C and DelMonte T., 1996. Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence.
- [30] Chen X, Zhao Y, Xu S, Zhou Y, Zhang L, Qu B, Xu Y. Analysis of Codon Usage Bias in the Plastid Genome of *Diplandrorchis sinica* (Orchidaceae). *Current Issues in Molecular Biology*. 2024; 46(9): 9807-20. <https://doi.org/10.3390/cimb46090582>
- [31] Fang J, Hu Y, Hu Z. Comparative analysis of codon usage patterns in 16 chloroplast genomes of suborder Halimedineae. *BMC genomics*. 2024; 25(1): 945. <https://doi.org/10.1186/s12864-024-10825-x>
- Khrustalev V V, Barkovsky E V. Mutational pressure is a cause of inter-and intragenomic differences in GC-content of simplex and varicello viruses. *Computational Biology and Chemistry*, 2009; 33(4): 295-302.