

## آنالیز بیوانفورماتیک ژنوم کلروپلاست پنبهٔ دیپلوئید (*Gossypium herbaceum*) و مقایسه آن با دو گونهٔ آلوتتراپلوئید

هاجر نصراللهی<sup>۱</sup>، فرشید طلعت<sup>۲\*</sup>، ایرج برنوسی<sup>۳</sup>، مهدی بدری انرجان<sup>۴</sup>

۱ و ۴. کارشناس ارشد اصلاح نباتات، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران

۲. 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

۳. دانشیار گروه اصلاح نباتات و بیوتکنولوژی گیاهی، دانشکده کشاورزی دانشگاه ارومیه، ارومیه، ایران

(تاریخ دریافت: ۱۳۹۶/۱۵ - تاریخ تصویب: ۱۳۹۶/۱۲/۲۱)

## Comparative bioinformatics analyses of the chloroplast genome of a diploid Cotton (*G. herbaceum*) with two allotetraploid species

Hajar Nasrollahi<sup>1</sup>, Farshid Talat<sup>2\*</sup>, Iraj Bernousi<sup>3</sup>, Mehdi Badri Anarjan<sup>4</sup>

1, 4. Former M.Sc. Student of Plant Breeding, Urmia University, Urmia, Iran

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

3. Associate Professor, Department of Plant Breeding and Plant Biotechnology, Faculty of Agriculture, Urmia University, Urmia, Iran  
(Received: Oct. 7, 2017 - Accepted: Mar. 12, 2018)

### Abstract

Diploid genome contains 8 genomes designated as A, B, C, D, E, F, G and K which have been identified in the genus *Gossypium*. The genome of A is limited only in two species as *G. herbaceum* and *G. arboreum*, and it is transferred from *G. herbaceum* to other species. The chloroplast genome (CP) of *G. herbaceum* has 160,140 bp lengths with protected quadripartite structure. Single copy regions of chloroplast genome are separated by two inverted repeat regions with a large copy region with 88,709 bp also the single copy region and each small inverted repeat regions have 20,221 and 25,605bp. The plastid's genome has 113 single copy genes and 19 duplicated copy genes. Single copy genes are encoding of 79 genes for protein production, four ribosomal RNA genes and 31 transfer RNA genes. Result showed that among all plastid genes only 18 genes appeared to have 1-2 intron/s and when compared with chloroplast genome of two allotetraploid species. *Ycf15* gene as the only duplicated gene, *rpl22* was in *G. herbaceum* and in the two species has studied *G. herbaceum*, *G. barbadense*. But *ycf15* gene in *G. barbadense* and both *ycf15* and *rpl22* genes were lost in *G. barbadense* and *G. hirsutum*. Though the high level of SSR protection in the chloroplast genome. SSRs are useful for genetic variation analysis since they have high efficiency against genomic SSRs.

**Keywords:** *Gossypium herbaceum*, Chloroplast Genome, Sequencing.

### چکیده

هشت ژنوم دیپلوئید شامل ژنوم‌های A, B, C, D, E, F, G و K در جنس *Gossypium* شناسایی شده‌اند. ژنوم A فقط به دو گونه *herbaceum* و *arboreum* پنبه‌های دنیای قدیم محدود شده است و این ژنوم از *G. herbaceum* به گونه‌های دیگر انتقال پیدا کرده است. به منظور توالی‌یابی ژنوم کلروپلاست *G. herbaceum* (A1) و مقایسه آن با دو گونه *G. hirsutum* (AD1) و *G. barbadense* (AD2) توالی‌های ژنوم سه گونه از سایت مرکز ملی اطلاعات بیوتکنولوژی برداشت شده است. موقعیت هر کدام از ژن‌ها، تعداد نوکلئوتیدها، نواحی بین ژنی و اینترون‌ها مشخص گردید. ژنوم کلروپلاست *G. herbaceum* دارای ۱۶۰۱۴۰ جفت‌باز بوده و ساختمان چهاربندی حفاظت شده دارد. نواحی تک نسخه‌ای ژنوم کلروپلاست، توسط دو ناحیه تکرار معکوس از هم جدا شده که ناحیه تک نسخه‌ای بزرگ دارای ۸۸۷۰۹ جفت‌باز، ناحیه تک نسخه‌ای کوچک ۲۰۲۲۱ جفت‌باز بوده و هر ناحیه تکرار معکوس ۲۵۶۰۵ جفت‌باز دارد. ژنوم پلاستییدی ۱۱۴ ژن تک نسخه‌ای و ۱۹ ژن دو نسخه‌ای دارد که ژن‌های تک نسخه‌ای شامل ۷۹ ژن رمزکننده پروتئین، ۴ ژن rRNA ریبوزومی و ۳۱ ژن tRNA است. نتایج نشان داد میان ژن‌های پلاستییدی فقط ۱۸ ژن دارای ۱-۲ اینترون هستند که در مقایسه با ژنوم کلروپلاست دو گونه آلوتتراپلوئیدی ژن *ycf15* تنها ژن دو نسخه‌ای موجود در *G. herbaceum* بود. در *G. herbaceum* و *G. barbadense* ژن *rpl22* وجود داشت ولی ژن *ycf15* در *G. barbadense*، ژن‌های *rpl22* و *ycf15* در *G. hirsutum* از دست رفته‌اند. با وجود میزان بالای حفاظت SSRها در ژنوم کلروپلاست این SSRها به دلیل بازده بالا در مقابل SSR ژنومی برای آنالیز تنوع ژنتیکی مفید هستند.

**واژه‌های کلیدی:** *G. herbaceum*، ژنوم کلروپلاست، توالی‌یابی.

### مقدمه

پنبه (*Gossypium*) متعلق به خانواده پنیرکیان (*Malvaceae*) می‌باشد. *G. herbaceum* معمولاً به‌عنوان Kapa یا پنبه (Cotton) شناخته می‌شود (Sharma et al., 2005). دو درصد از محصول تجاری پنبه متعلق به گونه *herbaceum* است (Blenda et al., 2006 & Nguyen et al., 2004). *G. herbaceum* به‌دلیل دارا بودن فیبر کشت شده و در تهیه پوشاک قابل استفاده است. گونه‌های جنس *Gossypium* منابع وسیعی از تنوع برای اصلاح پنبه‌های زراعی هستند ولی فقط گونه‌های *arboreum*, *barbadense* و *hirsutum* در عمل قابل کشت هستند (Wendel et al., 2009). کروموزوم‌های پلاستییدی به طور نسبی تعداد ژن کم، سطح حفاظت بالایی داشته و اطلاعات با ارزشی را برای برآورد ارتباط فیلوژنتیکی بین گیاهان و جلبک‌های سبز دارند (Turmel et al., 2009; Chan et al., 2011; Greiner et al., 2011). ژنوم کلروپلاست یک سازمان‌بندی بالایی حفاظتی با تک کروموزوم حلقوی ترکیب‌شده و ساختار چهاربعدی دارد که ساختمان چهاربعدی شامل دو ناحیه تکراری (IR) که بوسیله دو ناحیه LSC و SSC جدا شده است (Palmer, 1991; Raubeson & Jansen, 2005; Saski et al., 2005). انتشار توالی ژنوم کلروپلاست (*AD*<sub>1</sub>) *G. hirsutum* و (*AD*<sub>2</sub>) *barbadense* مدرک معتبری برای بیان موقعیت مهم پنبه در تیره پنیرکیان است (Lee et al., 2006). کل ساختار کلروپلاست حاوی دو نسخه از نواحی تکراری بوده که طول آنها از ۵ تا ۷۶ جفت‌باز متغیر است و این نسخه‌ها بوسیله ناحیه تک نسخه‌ای بزرگ (LSC) و ناحیه تک نسخه‌ای کوچک (SSC) جدا شده است (Segiura, 1989). آنالیز توالی‌ها نشان دادند که ساختار ژنوم کلروپلاست و محتویات ژنی بین گیاهان به‌شدت حفاظت شده است (Daniell et al., 2006; Lesberg & Duvall, 2009).

(2009). گسترش روزافزون حجم داده‌های ژنومی، نیاز به ذخیره آنها، بازیابی، 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 (Talat et al., 2015). (Talat & Wang, 2015) با بررسی ژنوم کلروپلاست *Gossypium thurberi* اعلام کردند از بین ژن‌های پلاستییدی فقط ۱۸ ژن یک تا دو اینترون داشتند و در مقایسه با دو گونه آلوتتراپلوئید زراعی تنها ۱۸ ژن تکراری در *G. thurberi* بود، با وجود سطح زیاد حفاظتی SSRهای ژنوم کلروپلاست، SSRها به‌علت بازده بالایشان در مقابله با SSRهای ژنومی برای تجزیه تنوع ژنتیکی مفید هستند همچنین محتوای کم GC یک مشخصه مهم ژنوم پلاستییدی است. ژنوم کلروپلاست جامع-ترین ژنوم مورد مطالعه گیاه است که به‌دلیل ساختار ملکولی منحصربه‌فرد و وراثت تک والدی ویژگی‌های زیادی برای تجزیه و تحلیل‌های تکاملی دارد. اهداف موردنظر در این پژوهش شناسایی نواحی اینترونی در ژن‌های ژنوم کلروپلاست *G. herbaceum*، استفاده از SSRها برای آنالیز تنوع ژنتیکی در پنبه، شناسایی پلی‌مورفیسم در بین گونه‌های مختلف پنبه، مطالعه ساختار ژنوم، محتوای ژن، سازمان‌دهی توالی‌های تکراری و Codon Usage است. کاربرد توالی‌یابی ژنوم کلروپلاست *G. herbaceum* برای توضیح تعامل ژن با محصولاتش، یافتن روابط فیلوژنتیک میان ژن‌ها، توالی‌های پروتئینی، به خط کردن توالی‌ها در بانک‌های اطلاعاتی برای یافتن شباهت‌ها، تفاوت‌ها و یافتن ژن‌ها در میان توالی‌های ژنومی است. تاکنون مطالعه جامعی در خصوص ژنوم‌های کلروپلاستی توالی‌یابی شده و شباهت‌های ساختاری بین گونه‌های مختلف پنبه صورت نگرفته است که در این مطالعه مقایسه توالی کامل ژنوم کلروپلاست *G. herbaceum* با دو گونه *G. barbadense hirsutum* و آنالیز بیوانفورماتیکی

*herbaceum* ۸۷۰۹ جفت باز و طول ناحیه SSC ۲۰۲۲۱ جفت باز بوده، همچنین طول هر کدام از نواحی IR<sub>A</sub> و IR<sub>B</sub> ۲۵۶۰۵ جفت باز است (شکل ۱). در *G. herbaceum* طول ناحیه رمزکننده ژن ۹۳۱۵۲ جفت باز است که شامل ۵۸/۱۷ درصد از کل ژنوم پلاستییدی می-باشد، طول ناحیه رمزکننده ژن در *G. hirsutum* ۵۶/۲۲ درصد و در *G. barbadense* ۵۶/۸۳ درصد است. طول ناحیه غیر رمزکننده ژن در گونه دیپلوئید ۴۵/۲۵ درصد که تقریباً شبیه طول ناحیه رمزکننده ژن در *G. hirsutum* با ۴۴/۷۱ درصد و در *G. barbadense* با ۴۴/۵۳ درصد است. در *G. herbaceum* اینترون ۱۳/۷۱ درصد و نواحی بین ژنی (IGS) ۳۱/۵۴ درصد از ژنوم پلاستییدی است که مشابه *G. barbadense* با ۳۱/۳۳ درصد ناحیه بین ژنی، ۱۳/۱۲ درصد اینترون و *G. hirsutum* با ۳۱/۵ درصد ناحیه بین ژنی و ۱۳/۱۴ درصد اینترون است. ژنوم پلاستییدی در *G. herbaceum* برای پروتئین ها ۵۰/۸۱ درصد، ژن های tRNA با ۱/۷۰ درصد و tRNA با ۵/۶۵ درصد را شامل شده که شبیه *G. barbadense* به ترتیب با درصد های ۴۹/۵۸، ۱/۶، ۵/۶۴ و *G. hirsutum* با درصد های ۴۸/۸۲، ۱/۷۵ و ۵/۶۴ است (جدول ۱). نتایج این تحقیق مشابه ژنوم کلروپلاست *G. turberi* بوده که توسط Talat et al. (2015) گزارش شده است و بر اساس گزارشات آنها ناحیه رمزکننده شامل ۵۷/۰۸ درصد ژنوم پلاستییدی بود. نتایج به دست آمده توسط Wu et al. (2009) برای ناحیه رمزکننده در ژنوم گیاه *Bambusa oldhamii* ۵۳/۴ درصد گزارش شده است در حالی که کمتر از ۶۰ درصد برای *Glycine max* توسط Saski et al. (2005) گزارش گردیده است. Daniell et al. (2008) نتایجی مشابه این تحقیق با درصد های پروتئین (۴۹/۸۲ درصد)، tRNA (۱/۷ درصد) و rRNA (۵/۶ درصد) در *Manihot esculenta* گزارش کرده اند.

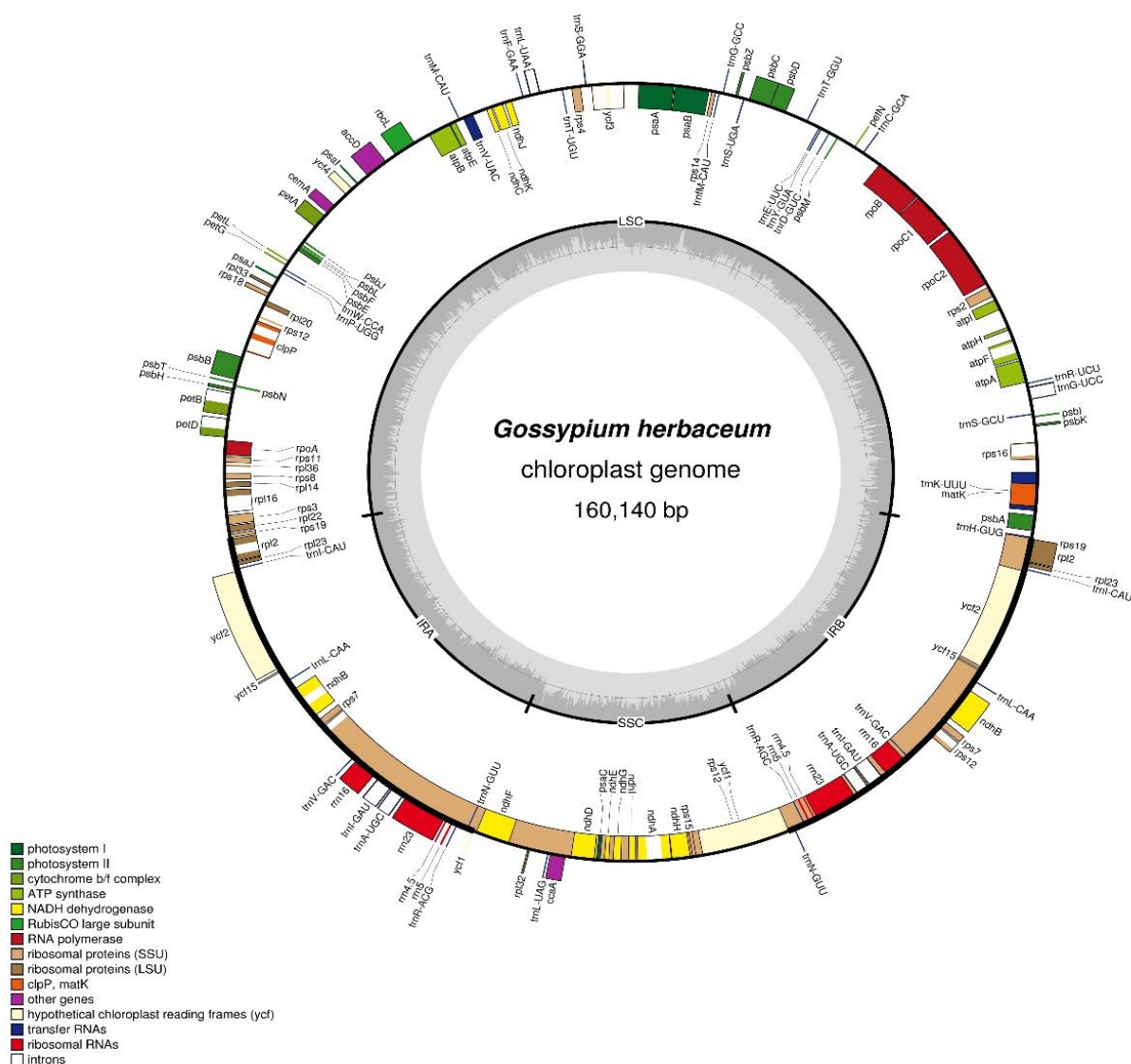
توالی کامل ژنوم *G. herbaceum* برای اولین بار انجام شده است.

## مواد و روش ها

توالی ژنوم کلروپلاست گونه دیپلوئید *G. herbaceum* با دو گونه آلوتتراپلوئید *G. hirsutum* و *G. barbadense* مورد مطالعه در این تحقیق ابتدا از پایگاه اطلاعاتی NCBI (www.ncbi.nlm.nih.gov/nucore/3180844) (64 دانلود و توالی های در دسترس با استفاده از نرم افزار DNAMAN نسخه ۹ گردآوری شده است. در ادامه Annotation ژن با استفاده از نرم افزار آنالین DOGMA (http://dogma.cccb.utexas.edu/Wyman,2004/)، ساختمان نقشه ژنومی و توزیع ژن با نرم افزار آنالین OGDRAW (http://ogdraw.mpimpgolmmpg.de/) انجام شده که برای شروع و اجرا، فایل Bank Gene یا شماره دسترسی وارد شده است. در آخر RSCU کدون های مختلف در هر نمونه ژنی از CodonW در Mobylye به آدرس (http://mobile.pasteur.fr/cgibin/portal) محاسبه و برای شناسایی توالی تکراری و موقعیت توالی ها از نسخه آنالین REPuter (http://bibiserv.techfak.uni-bielefeld.de/reputer) استفاده شده است.

## نتایج و بحث

ژنوم کلروپلاست گونه دیپلوئید *G. herbaceum* یک DNA حلقوی با ۱۶۰۱۴۰ جفت باز بوده که کوتاه تر از ژنوم کلروپلاست گونه های آلوتتراپلوئید *G. hirsutum* با طول ۱۶۰۲۵۶ جفت باز و ژنوم کلروپلاست *G. barbadense* با طول ۱۶۰۳۱۷ جفت باز است، همچنین در ساختار ژنوم گونه دیپلوئید دو ناحیه تک نسخه ای بزرگ (LSC) و تک نسخه ای کوچک (SSC) وجود دارد که بوسیله دوناچیه تکرار معکوس (IR<sub>B</sub>) و IR<sub>A</sub> از هم جدا شده است. طول ناحیه LSC در *G.*



شکل ۱. نقشه ژنوم کلروپلاست گونه دیپلوئید *G. herbaceum*

جدول ۱. محتوای GC در گونه دیپلوئید *G. herbaceum*

GC content		Length(bp)	Proportion%	T%	A%	C%	G%	A+T%	C+G%
Coding gene	Protein	81373	50.81	31.4	30.4	19.5	18.7	61.8	38.2
	trna	2729	1.70	23.1	24.0	26.4	26.4	47.1	52.9
	rrna	9050	5.65	22.3	22.3	27.7	27.7	44.5	55.5
	Total	93152	58.17	30.2	29.4	20.5	19.8	59.7	40.3
	IGS	50507	31.54	34.4	33.6	16.2	15.7	68.0	32.0
Noncoding gene	Intron	21953	13.71	32.6	30.4	19.4	17.6	63.0	37.0
	Total	72460	45.25	33.9	32.8	17.1	16.2	66.7	33.3
	Complete genome	160140	100	31.8	30.9	19.0	18.3	62.7	37.3
	LSC	88709	55.39	33.2	31.6	18.2	17.1	64.7	35.3
	SSC	20221	12.63	34.4	33.9	16.6	15.1	68.3	31.7
	IRA	25605	15.99	28.5	28.5	20.7	22.3	57.0	43.0
IRB						22.3	20.7		

ژن matK و ژن‌های هم‌پوشان دوبار آنالیز شده‌اند. IGS معرف نواحی بین ژنی است.

ژنوم‌های بازآرایی نشده هم وجود دارد که این تکرارها بیش‌تر موارد در نواحی بین ژنی و اینترون‌ها شایع‌تر هستند.

جدول ۲. توالی تکراری ژنوم کلروپلاستگونه دیپلوئید

<i>G. herbaceum</i>			
Number	Size	Location	Match Direction
1	66	psaB, psaA	F
2	41	ycf3, ndhA	F
3	38	IGS, ndhA	F
4	36	ycf3, IGS	F
5	31	IGS	F
6	30	IGS	F
7	29	IGS	F
8	29	IGS	F
9	26	ycf2	F
10	26	ycf2	F
11	23	IGS	F
12	23	IGS	F
13	23	IGS	F
14	22	petG, IGS	F
15	22	trnG-UCC, IGS	F
16	22	clpP, IGS	F
17	21	trnS-GCU, trnS-UGA	F
18	21	IGS	F
19	20	IGS	F
20	20	trnV-UAC, trnA-UGC	F
21	48	IGS	P
22	38	ndhA, IGS	P
23	36	ycf3, IGS	P
24	29	IGS	P
25	29	IGS	P
26	29	IGS, trnS-GGA	P
27	26	psaA	P
28	26	ycf2	P
29	26	ycf2	P
30	24	IGS	P
31	23	trnG-UCC, clpP	P
32	23	IGS	P
33	23	IGS	P
34	22	trnG-UCC, IGS	P
35	22	AtpF	P
36	22	clpP, IGS	P
37	21	trnS-UGA, trnS-GGA	P
38	20	IGS	P
39	20	trnF-GAA	P
40	20	IGS	P
41	20	trnV-UAC, trnA-UGC	P
42	31	IGS	R
43	26	IGS	R
44	25	ycf1	R
45	23	IGS	R
46	23	IGS	R
47	21	IGS	R

در توالی تکراری ژنوم کلروپلاست سه گونه *G. barbadense* و *G. hirsutum* و *G. herbaceum* سه نوع توالی تکراری پیدا شده است: جورشدگی مستقیم، جورشدگی معکوس و جورشدگی پالیندرومی. تعداد کل تکرارها برای هر کدام از سه گونه یکسان بوده و برابر ۴۹ تکرار است که در توالی تکراری *G. herbaceum* ۲۰ تکرار مستقیم، ۸ تکرار معکوس و ۲۱ تکرار پالیندرومی وجود دارد که اکثریت تکرارها در نواحی بین‌ژنی قرار گرفته و تعداد خیلی کمی از تکرارها در نواحی *ycf2* و *trnG* قرار دارند همچنین طول اکثر تکرارها ۲۰-۴۰ جفت‌باز بوده و بزرگترین تکرار ۶۶ جفت‌باز دارد که در نواحی *psaA* و *psaB* قرار گرفته است. در توالی تکراری *G. hirsutum* ۲۰ تکرار مستقیم، ۹ تکرار معکوس و ۲۰ تکرار پالیندرومی وجود دارد که اکثریت تکرارها در نواحی بین‌ژنی قرار داشته و تعداد کمی در نواحی اینترون و *ycf2* قرار دارند همچنین طول اکثر تکرارها ۲۰-۴۰ جفت‌باز بوده و بزرگترین تکرار ۷۲ جفت‌باز دارد که در ناحیه *psaA* و *psaB* قرار گرفته است. در توالی تکراری *G. barbadense* ۲۳ تکرار مستقیم، ۹ تکرار معکوس و ۱۷ تکرار پالیندرومی وجود دارد که اکثریت تکرارها در نواحی بین‌ژنی قرار داشته و تعداد خیلی کمی از تکرارها در نواحی اینترون و *ycf2* قرار گرفته‌اند و بزرگترین تکرار ۷۲ جفت‌باز طول دارد که در نواحی *psaA* و *psaB* قرار گرفته است (جدول ۲).  
Talat *et al.* (2015) در *G. turberi* گزارش کردند که تعداد کل تکرارها ۶۶ بوده و در آن چهار نوع تکرار وجود دارد یعنی علاوه بر سه تکرار نامبرده شده در *G. herbaceum* تکرار جورشدگی کامل نیز وجود دارد و بزرگترین تکرار در *G. turberi* ۷۲ جفت‌باز طول دارد. Saski *et al.* (2005) گزارش دادند که در ژنوم کلروپلاست غیر از ناحیه تکرار معکوس (IR) توالی‌های تکرار دیگری نیز وجود دارد. بر اساس گزارشات Jansen *et al.* (2006) در ژنوم کلروپلاست *Vitis venifera* تکرارهای پراکنده در

	CATT	1	13
Pentanucleotide	TTAGA	1	14
Total		5	16

در *G. herbaceum* ۷۹ ژن رمزکننده پروتئین، ۴ ژن rRNA و ۳۱ ژن tRNA که در مجموع ۱۱۴ ژن تک نسخه‌ای را شامل می‌شوند، همچنین دارای ۱۹ ژن دو نسخه‌ای دارد در ناحیه تکرار معکوس قرار دارند. تعداد ژن‌های دو نسخه‌ای در *G. hirsutum* و *G. barbadense* ۱۷ عدد است (جدول ۴). در *G. herbaceum* ۵ ژن (ycf1, ycf2, ycf3, ycf4, ycf5), برای *G. hirsutum* (ycf1, ycf2, ycf3, ycf4) و *G. barbadense* (ycf1, ycf2, ycf3, ycf4) هر کدام چهار ژن با عملکرد ناشناخته به دست آمده است. دو ژن rpl22 و ycf15 در *G. herbaceum* وجود داشته اما در *G. hirsutum* این دو ژن از دست رفته و در *G. barbadense* تنها ژن ycf15 از دست رفته است. ژن‌های ycf در گیاهان ضروری هستند و به شدت در بین گونه‌ها حفاظت شده می‌باشد (Steane, 2005). ژن trnP-GGG در *Cryptomeria japonica* توسط Hairo et al. (2008) گزارش شده است در حالی که این ژن در *G. herbaceum* و *G. hirsutum* از دست رفته و در هر سه گونه دو ژن rps12 وجود دارد که دارای یک اینترون هستند و rps12 به وسیله این اینترون به دو قطعه تقسیم شده است که یک اگزون (5'-end) در ناحیه LSC و دیگری (3'-end) در ناحیه IR<sub>A</sub> قرار گرفته است و در دو گونه پنبه آلوتراپلوئید در سایت مرکز ملی اطلاعات بیوتکنولوژی ژن rps12 به عنوان سیستم پیرایش ترانس معرفی شده است. ژن matK ۱۵۱۵ جفت‌باز طول دارد و ژن matK در اینترون trnK-UUU پیدا شده و تنها ژنی است که درون یک اینترون قرار گرفته و رمزکننده maturaseK است، همچنین قطعات این ژن هم حفاظت شده و هم متغیر است (Wilson, 2004).

تجزیه کدون یوسج برای هر سه گونه نشان داد که ATG و TGG به ترتیب کدون‌های آمینواسید

48	21	IGS	R
49	20	ndhC	R

IGS معرف نواحی بین‌ژنی، F معرف جورشدگی مستقیم، R معرف جورشدگی معکوس، P معرف جورشدگی پالیندرومی.

توالی‌های تکراری ساده یا SSRها به تعداد کم در توالی‌های ژنوم کلروپلاست سه گونه وجود دارد. در گونه دیپلوئید *G. herbaceum* پنج توالی تکراری ساده که بیش از ده جفت‌باز طول دارند به دست آمده است که توالی‌های تکراری ساده به ترتیب دی‌نوکلئوتیدی، تری‌نوکلئوتیدی، تترا نوکلئوتیدی و پنتا نوکلئوتیدی هستند. در *G. hirsutum* هجده توالی تکراری ساده به دست آمده که بیش از ده جفت‌باز طول دارند و طولانی‌ترین SSR سی جفت‌باز طول داشته که پنتا نوکلئوتیدی است و بیش‌ترین توالی تکراری ساده دی‌نوکلئوتیدی می‌باشد که بیست و سه جفت‌باز طول دارد. در *G. barbadense* شش توالی تکراری ساده به دست آمده که دارای طول بیش از ده جفت‌باز هستند که طولانی‌ترین توالی تکراری ساده در آن دی‌نوکلئوتیدی بوده و دارای طول با بیست و سه جفت‌باز است (جدول ۳). Raubeson et al. (2007) توالی تکراری ساده مربوط به گونه نیلوفر زرد جنس *Ranunculus* و ۲۴ ژنوم کلروپلاست دیگر را گزارش کردند که در آن تعداد تکرار مونونوکلئوتیدی  $8 \leq$  نوکلئوتید، تکرار دی‌نوکلئوتیدی  $8 \leq$  نوکلئوتید و تکرار تری‌نوکلئوتیدی  $9 \leq$  نوکلئوتید است که در مجموع ۸، ۸ و ۹ توالی تکراری ساده و در تکرارهای با طول بیش‌تر، توالی‌های تکراری ساده طول بیش‌تری داشتند (مونونوکلئوتیدی  $10 \leq$ ، دی‌نوکلئوتیدی  $10 \leq$  و تری‌نوکلئوتیدی  $12 \leq$ ). توالی‌های تکراری ساده کلروپلاست به دلیل کارایی بالایشان در مقابل توالی‌های تکراری ساده ژنومی در تجزیه تنوع ژنتیکی مفید هستند (Rebeiro et al., 2002).

جدول ۳. توالی تکراری ساده در گونه دیپلوئید *G. herbaceum*

Repeat	Repeat sequence	Number	Max (bp)
Dinucleotide	CT	1	16
Trinucleotide	TTA	1	13
Tetranucleotide	AAAT	1	14

هر کدام دارای یک یا دو اینترون هستند که مانند *G. hirsutum* و *G. herbaceum* است. مقایسه اینترون‌ها در جدول ۵ بین *G. herbaceum*، *G. hirsutum* و *G. herbaceum* آمده است و نشان داد ۱۸ ژن در ژنوم کلروپلاست به دست آمده که یک یا دو اینترون دارند، ۶ ژن دارای اینترون در ژن‌های رمزکننده tRNA پیدا شده است و بقیه اینترون‌ها در ژن‌های رمزکننده پروتئین‌ها به دست آمده است. طولانی‌ترین اینترون trnK-UUU با ۲۴۹۲ جفت‌باز طول در *G. herbaceum* بوده و تنها اینترون در پنبه است که درون ژن matK قرار گرفته است و کوتاه‌ترین اینترون rps12 با ۵۳۶ جفت‌باز طول است که در ناحیه تکرار معکوس به دست آمده است، همچنین دو ژن ycf3 و clpP در ناحیه تک نسخه‌ای بزرگ به دست آمده و هر کدام با دو اینترون تقسیم شده است که نتایج به دست آمده مشابه *G. turberi* است (Talat & Wang, 2015).

متیونین به عنوان کدون شروع و تریپتوفان دارای تشابه نسبی کدونی برابر یک هستند که تشابه نسبی کدونی کدون‌های پایان TGA، TAA و TAG در گونه *G. herbaceum* به ترتیب ۱/۲۲، ۱/۰۱ و ۰/۷۷، در گونه *G. hirsutum* به ترتیب ۱/۲۰، ۰/۹۴ و ۰/۸۵ و در گونه *G. barbadense* ۱/۲۵، ۰/۹۶ و ۰/۷۹ است. براساس تشابه نسبی کدون یوسیج به دست آمده برای سه گونه، کدون TAA به عنوان کدون پایان ترجیح داده شد، همچنین ارزش تشابه نسبی کدون یوسیج بیش‌تر از یک نشان‌دهنده فراوانی کدون‌های مهم‌تر است و براساس تشابه نسبی کدون یوسیج کدون‌های A و T در جایگاه سوم ترجیح داده شده‌اند (جدول ۵). Choudhury et al. (2017) اعلام کردند تجزیه و تحلیل الگوی کدون یوسیج برای فهم سازمان‌دهی ژنتیکی و ملکولی یک ژن مهم است همچنین کدون یوسیج به بیان ژن‌های هترولوگ، طراحی پرایمر و ژن مصنوعی کمک کرده است. در *G. herbaceum* ۱۸ ژن به دست آمده که

جدول ۴. محتویات ژن در گونه دیپلوئید *G. herbaceum*

Gene Coded	Group	Gene Name	
	Subunit of Acetyl-CoA-carboxylase	accD	
	Large subunit of rubisco	rbcL	
	subunit of NADH-dehydrogenase	ndhA* ndhB*§ ndhC ndhD ndhE ndhF ndhG ndhH ndhI ndhJ ndhK	
	Subunit of ATP synthase	atpA atpB atpE atpF* atpH atpI	
	Subunit of cytochrome b/f complex	petA petB* petD* petG petL petN ccsA	
Protein Gene	Subunit 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	
	ribosomal RNA gene	rrn16§ rrn23§ rrn4.5§ rrn5§	
RNA Gene	transfer RNA gene	trnA-UGC*§ trnC-GCA trnD-GUC trnE-UUC	
		trnF-GAA trnM-CAU trnG-UCC* trnG-GCC	
		trnH-GUG trnI-CAU§ trnI-GAU*§ trnK-UUU*	
		trnL-CAA§ trnL-UAA* trnL-UAG trnM-CAU	
		trnN-GUU§ trnP-UGG trnQ-UUG trnR-ACG§	
		trnR-AGC§ trnR-UCU trnS-GCU trnS-GGA trnS-UGA	
		trnT-GGU trnT-UGU trnV-GAC§ trnV-UAC*	
		trnW-CCA trnY-GUA	

\* دارای یک اینترون؛ \*\* دارای دو اینترون؛ § در ناحیه IR قرار گرفته‌اند.

جدول ۵. نتایج کدون یوسج در گونه دیپلوئید *G. herbacum*

AA	Codon	Number	RSCU	AA	Codon	Number
Phe	<b>TTT</b>	2417	1.22	Ser	<b>TCT</b>	1163
	TTC	1561	0.78		<b>TCC</b>	888
	<b>TTA</b>	1143	1.25		<b>TCA</b>	922
Leu	<b>TTG</b>	1105	1.24	Pro	TCG	653
	<b>CTT</b>	1059	1.19		<b>CCT</b>	680
	CTC	693	0.78		CCC	640
	CTA	814	0.92		<b>CCA</b>	789
	CTG	523	0.59		CCG	452
Ile	<b>ATT</b>	1805	1.18	Thr	<b>ACT</b>	707
	ATC	1184	0.78		<b>ACC</b>	592
	<b>ATA</b>	1594	1.04		<b>ACA</b>	684
Met	ATG	876	1	Ala	ACG	373
	<b>GTT</b>	809	1.38		<b>GCT</b>	498
Val	GTC	442	0.75	Cys	GCC	405
	GTA	692	1.18		<b>GCA</b>	441
	GTG	402	0.69		GCG	280
Tyr	<b>TAT</b>	1524	1.37	<b>TGT</b>	661	
	TAC	694	0.63	TGC	465	
Ter	<b>TAA</b>	1265	1.22	Ter	<b>TGA</b>	1044
	TAG	791	0.77	Trp	TGG	690
His	<b>CAT</b>	943	1.46	Arg	CGT	404
	CAC	351	0.54		CGC	263
Gln	<b>CAA</b>	1046	1.41	Ser	<b>CGA</b>	612
	CAG	441	0.59		CGG	380
Asn	<b>AAT</b>	1881	1.44	Arg	AGT	700
	AAC	735	0.56		AGC	524
Lys	<b>AAA</b>	2220	1.37	Gly	<b>AGA</b>	1097
	AAG	1019	0.63		<b>AGG</b>	653
Asp	<b>GAT</b>	1031	1.43	Gly	GGU	579
	GAC	412	0.57		GGC	364
Glu	GAA	1338	1.4	GGA	817	
	GAG	576	0.6	GGG	574	

جدول ۶. مقایسه اینترون‌ها در *G. herbaceum*، *G. hirsutum* و *G. barbadense*

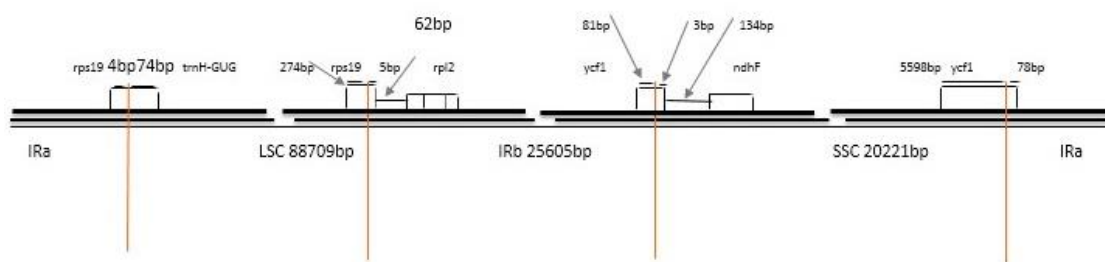
Intron	<i>G. herbaceum</i>	<i>G. hirsutum</i>	<i>G. barbadense</i>	Sequence Identity
trnK-UUU	2492	2542	2535	99/33
rps16	873	869	870	99
trnG-UCC	778	762	763	99
atpF	791	804	805	99/33
rpoC1	751	753	753	99
ycf3-1	780	789	789	99/67
ycf3-2	761	777	777	99/33
trnL-UAA	589	573	582	99
trnV-UAC	603	607	609	99/67
rps12	536	536	536	99/33
clpP-1	679	683	679	99/33
clpP-2	911	891	890	99
petB	758	818	821	99/33
petD	758	757	754	99/33
rpl16	1114	1140	1135	99/33
rpl2	693	693	688	99/33
ndhB	689	683	683	99/33
trnI-GAU	955	954	959	99/67
trnI-UGC	797	797	795	99/33
ndhA	1068	1076	1076	99



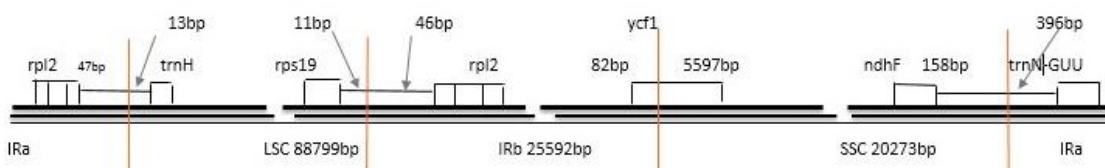
آمده است. Ycf1 در SSC/IR<sub>b</sub> قرار گرفته است بنابراین ycf1 در IR<sub>b</sub> در حاشیه IR<sub>b</sub> و SSC قرار گرفته است. در *Spinacia* ycf1 طولانی‌ترین کپی با ۱۴۴۵ جفت‌باز طول را دارد در حالی که در *G. barbadense* کوتاه‌ترین کپی با ۷۹ جفت‌باز طول به‌دست آمده است. در *G. hirsutum* ycf1 در IR<sub>b</sub>/SSC قرار گرفته است و ۸۲ جفت‌باز طول دارد. اندازه ناحیه IR در سه گونه مورد مطالعه در یک محدوده سایز ژنوم قرار دارند. حاشیه ناحیه تکرار معکوس معمولاً بین گونه‌ها متفاوت است و گستردگی و فشردگی ناحیه تکرار معکوس به اندازه سایز ژنوم اهمیت دارد. ژن‌های کاذب معمولاً در نقاط تکرار معکوس و بین LSC/SSC قرار داشته و اندازه ناحیه تکرار معکوس در *Vitis vinifera* ۲۶۳۵۸ جفت‌باز است که در محدوده سایز ژنوم‌های توالی‌یابی شده دولپه‌ای‌ها قرار دارد (Jansen et al., 2006).

اندازه ناحیه تکرار معکوس در سه گونه (*hirsutum*, *barbadense*, *herbaceum*) به‌ترتیب برابر ۲۵،۵۶۰۵، ۲۵،۵۹۱ و ۲۵،۵۹۲ جفت‌باز به‌دست آمده است، اندازه ناحیه IR<sub>A</sub> و IR<sub>B</sub> در هر ژنوم با هم برابر و همه ژن‌های تکراری در ناحیه تکرار معکوس قرار گرفته است. در *G. herbaceum* IR<sub>b</sub> /LSC، درون ژن rps19 یافت شده و قسمتی از ژن rps19 در بانک ژن به‌عنوان ژن کاذب معرفی شده است، چهار جفت‌باز طول دارد و در IRA /LSC قرار گرفته است. در *G. herbaceum* با ۵ جفت‌باز طول کوتاه‌ترین و در *Spinacia oleracea* با ۱۴۴ جفت‌باز طول بلندترین کپی به‌دست آمده است (شکل ۲). با توجه به شکل‌های ۳ و ۴ در گونه‌های *Solanum*، *Spinacia* و *Arabidopsis* دو ژن ycf1 و ndhF با هم هم‌پوشانی دارند که طولانی‌ترین هم‌پوشانی در *Arabidopsis* با ۳۷ جفت‌باز طول و کوتاه‌ترین هم‌پوشانی در *Solanum* با ۱۷ جفت‌باز طول به‌دست

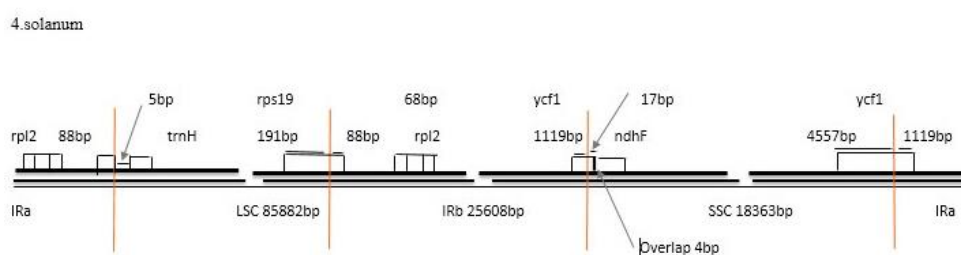
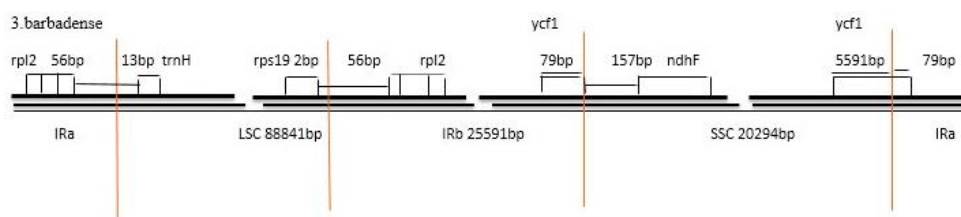
## 1. herbaceum



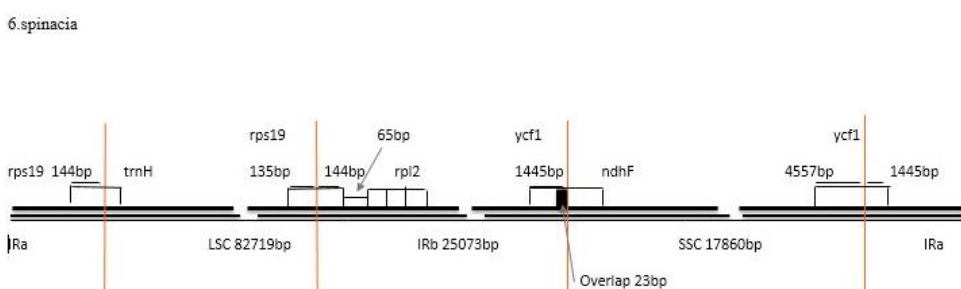
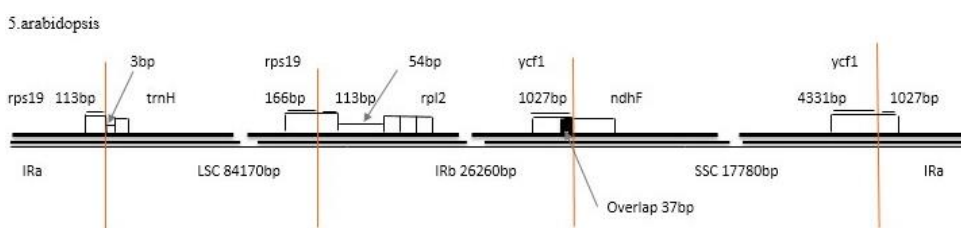
## 2. hirsutum



شکل ۲۰ مقایسه بین ناحیه تکرار معکوس و تک نسخه‌ای بزرگ و کوچک در *G. hirsutum* و *G. herbaceum*



شکل ۳. مقایسه بین ناحیه تکرار معکوس و تک نسخه‌ای بزرگ و کوچک در *G. barbadens* و *Solanum*



شکل ۴. مقایسه بین ناحیه تکرار معکوس و تک نسخه‌ای بزرگ و کوچک در *Spinacia* و *Arabidopsis*

## REFERENCES

Bausher M, Singh ND, Lee SB, Jansen RK, Daneill H (2006) The complete chloroplast genome sequence of *Citrus sinensis* (L.) obsek var *Ridye pineapple* organization and phylogenetic relationships to the angiosperms. J. BMC. Plant. Bio. doi: 10.1186/1471-2229-6-21.

Blenda A, Scheffler J, Scheffler B, Palmer M, Lacap JM, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yellambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D (2006) A cotton microsatellite database resource for *Gossypium* genomics. BMC. Gen. 7:132-142.

- Chan CX, Grass J, Yoon HS, Bhattacharya D (2011) Plastid origin and evolution: new models provide insights into old problems. *Plant. Phys.* doi: 10.1104/pp.111.173500
- Chatterjee A, Pakrashi SC (2006) The treatise on Indian medicinal plants. National Institute of Science Communication and Information Resources, New Delhi.
- Choudhury MN, Uddin A, Chakraborty S (2017) Nucleotide composition and codon usage bias of SRY gene. *J. Andr.* doi: 10.1111/and.12787
- Daniell H, Lee SB, Grerich J, Saski C, Quesada-Vargas T, Guda C, Tomkins J, Jansen RK (2006) Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analysis with other Solanaceae genomes. *Theor. Appl. Genet.* 112:1503-1518.
- Daniell H, Wurdack KJ, Kanagaraj A, Lee SB, Saski Ch, Jansen RK (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. *Theor. Appl. Genet.* doi: 10.1007/s00122-007-0706-y
- Goremykin V, Hirsch-Ernest K, Wolf S, Hellwig F (2003) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol. Bio. Evo.* doi: 10.1093/mbev/msg159
- Gray JC, Hird SM, Dyer TA (1990) Nucleotide sequence of a wheat chloroplast gene encoding the proteolytic subunit of an ATP-dependent protease. *Plant. Mol. Bio.* 15:947-950.
- Greiner S, Rauwolf U, Meurer J, Herrmann RG (2011) The role of plastids in plant speciation. *Mol. Eco.* 20:671-691.
- Hirao T, Watanabe A, Kurita M, Kondo T, Takata K (2008) Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. Chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC. Plant. Biol.* doi: 10.1186/1471-2229-8-70
- Ibrahim RIH, Azuma JI, Sakamoto M (2006) Complete nucleotide sequence of the cotton (*Gossypium barbadense* L.) chloroplast genome with a comparative analysis of sequences among 9 dicot plants. *Gene. Genet. Syst.* 81:311-321.
- Lee SB, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ (2006) The complete chloroplast genome sequence of *Gossypium hirsutum* organization and phylogenetic relationship to other angiosperms. *BMC. Gen.* 7:61.
- Leebens-Mack J, Raubenson LA, Cui L, Kuehl J, Fourcade M, Chumley T, Boore JL, Jansen RK, Pamphiliis CW (2005) Identifying the basal angiosperms in chloroplast genome phylogenetic: sampling one's way out of the Felsenstein zone. *Mol. Bio. Evo.* 22:1948-1963.
- Leseberg CH, Duvall MR (2009) The complete chloroplast genome of *Coixlacryma jobi* and a comparative molecular evolutionary analysis of plastomes in cereals. *J. Mol. Evo.* 69:311-318.
- Lin CHP, Huang JP, Wu CHSH, Hsu CHY, Chaw SHM (2010) Comparative chloroplast genomics reveals the evolution of *Pinaceae* genera and *Subfamilies*. *Gen. Bio. Evo.* doi: 10.1093/gbe/evq036
- Nguyen TB, Giband M, Brottier P, Risterucci AM, Lacape JM (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. *Theor. Appl. Genet.* 109:167-175.
- Palmer JD (1991) Plastid chromosomes, structure and evolution in the molecular biology of plastids edited by

- (Bogorad L, Vasil K, San Diego). Academic Press, USA.
- Raubeson LA, Peery RH, Chumley TW, Dziubek Ch, Fourcade HM, Boore JL, Jansen RK (2007) Comparative chloroplast genome: analyses including new sequences from the angiosperm *Nuphar advena* and *Ranunculus macranthus*. BMC. Geno. doi: 10.1186/1471-2164-8-174
- Rebeiro MM, Mariette S, Vendramin GG, Szmidt AE, Plomin C, Kremer A (2002) Comparison of genetic diversity estimates within and among populations of *Maritime pine* using chloroplast simple-sequence repeat and amplified fragment length polymorphism data. Mol. Eco. doi: 10.1046/j.1365-294X.2002.01490.x
- Sharma PC, Yenle MB, Dennis TJ (2005) Data base of medicinal plants used in Ayurveda. Central council for research in Ayurveda and siddha, India.
- Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu YQ, Chase MW, Foris JS, Stefanovic S, Rice DW, Palmer JD, Solits PS (2004) Genome-scale data, angiosperm relationships, and 'ending incongruence'. Tren. Plant. Sci. 9:477-483.
- Soltis DE, Soltis PS (2004) Amborella not a "basal angiosperm"? No so fast. J. Botan. doi:10.3732/ajb.91.997
- Sugiura M, Hirose T, Sugita M (1998) Evolution and mechanism of translation in chloroplasts. Ann. Rev. Genet. 32:437-459.
- Talat F, Wang K (2014) Chloroplast genome study, new tool in plant biotechnology *Gossypium* spp. as a model crop. J. curr. Rese. Sci. 838-850.
- Talat F, Wang K (2015) Comparative bioinformatics analysis of the chloroplast genome of a wild diploid *Gossypium* and two cultivated allotetraploid species. J. Biotech. doi:10.15171/ijb.1231
- Turmel M, Gagnon MC, O'Kelly CJ, Otis C, Lemieux C (2009) The chloroplast genome of the green *Algae pyramimona*, *Monomostix* and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplast of *Euglenids*. Mol. Bio. Evo. 26: 631-648.
- Turmel M, Otis C, Lemieux C (2009) The chloroplast genome of the green *Algae pedinomonas* minor, *Parachlorella kessleri* and *Oocystis solitaria* reveal a shared ancestry between the *chlorellales*. Mol. Bio. Evo. 26: 2317-2331.
- Walker JF, Michael JZ, Nancy CE (2014) Correction to comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (*Medieae, asteraceae*). Amer. J. Botan. doi:10.3732/ajb.1400049
- Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM (2009) Evolution and natural history of the cotton genus. Crop. Model. 3:3-22.
- Wilson C A (2004) Phylogeny of *Iris* based on chloroplast matK gene and trnK intron sequence data. Mol. Phylo. Evo. doi: 10.1016/j.ympev.2004.06.013
- Wu FH, Kan DP, Lee SB, Daniell H, Lee YW, Lin CC, Lin NS, Lin CS (2009) Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. Tree. physiol. doi: 10.1093/treephys/tpp015.