

雷蒙德氏棉和亚洲棉 SSR 重复序列类型和丰度差异比较

董 薇, 杜雄明*, 赖童飞

(中国农业科学院棉花研究所, 农业部棉花遗传改良重点开放实验室, 河南安阳 455000)

摘要:利用生物信息学相关方法对 NCBI 数据库中 39277 条亚洲棉 (*Gossypium arboreum* L.) 及 63588 条雷蒙德氏棉 (*Gossypium raimondii* L.) 的 EST 序列共计 64.08 Mb (约相当于基因组的 6.02%) 分别进行拼接, 比较分析了两棉种简单重复序列的重复类型及丰度差异。结果表明: 拼接后的一致性序列中, 亚洲棉比雷蒙德氏棉存在较高比例的单条序列簇; 亚洲棉一至六碱基 6 种重复类型的 SSR 间隔距离分别比雷蒙德氏棉的大; 亚洲棉中二到六碱基等 5 种重复类型的比例高于雷蒙德氏棉, 但单碱基重复类型比例低于雷蒙德氏棉; 两棉种的单碱基、四碱基和六碱基重复类型的优势重复基序和丰度也不同。因此, 认为利用四、五、六碱基三种重复类型设计的 SSR 引物可以有效地鉴定亚洲棉和雷蒙德氏棉的基因组差异。对四倍体棉种的 A、D 亚组供体棉种中 SSR 的差异分析, 为 SSR 标记开发、棉种起源和进化的深入研究提供更多有用的信息。

关键词:雷蒙德氏棉; 亚洲棉; EST 拼接; SSR 丰度

中图分类号: S562.024 文献标识码: A

文章编号: 7807-1002-(2008)06-0418-07

Comparison of SSR Types and Abundance Differences between *G. raimondii* and *G. arboreum*

DONG Wei, DU Xiong-ming*, LAI Tong-fei

(Cotton Research Institute, CAAS; Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, Anyang, Henan 455000, China)

Abstract: The 39277 *G. arboreum* and 63588 *G. raimondii* ESTs from NCBI databases which were in a total sizes of 64.08 Mb (approximately 6.02% of total cotton genome) were respectively assembled by bioinformatical methods, and the differences of Simple Sequence Repeat (SSR) types and abundance between these two species were analyzed. The main results were shown as follow: The singleton ratios of *G. arboreum* were higher than those of *G. raimondii* in the assembled tentative consensus sequences; The interval distances of six SSR types which were from mononucleotide to hexa-nucleotide repeats in *G. arboreum* were farther than those of the six SSR types in *G. raimondii*, respectively; The percentages of the five repeat types which were from di- to hexa-nucleotides were higher in *G. arboreum* than that of the five repeat types in *G. raimondii*. However, the percentages of mononucleotide repeat type were lower in *G. arboreum*. Moreover, the advanced SSR motifs and abundance of mono-, tetra-, and hexa-nucleotides repeat types were also different between two species. Therefore, it is suggested that the primers designed according to the tetra- or more nucleotides repeat types can effectively identify the genome differences between *G. arboreum* and *G. raimondii*. The differential analysis of SSR types and abundance in the donor species for A and D sub-genome of the tetraploid species provides more useful information for SSR marker design and profound research of the origin and evolution of *Gossypium* species.

Key words: *Gossypium raimondii*; *Gossypium arboreum*; EST assembly; SSR abundance

收稿日期: 2007-02-07 作者简介: 董薇 (1981-), 女, 硕士; sophia-110@sohu.com; * 通讯作者, dxm630723@163.com
基金项目: 国家 973 计划项目 (2004CB117301), “十五”国家重点科技攻关计划项目 (2004BA525B05)

微卫星 (Microsatellite) 或简单重复序列 (Simple Sequence Repeat, SSR) 是由 1~6 个核苷酸为重复单位组成的重复序列, 由于重复次数的不同及重复程度的不完全造成了每个位点的多态性。这种序列绝大部分分布于基因的内含区和内含子等非编码区, 同时少量存在于基因的外显子和调控区^[1], 并且不同物种的基因组对简单重复序列的偏好性也不同^[2]。近来 EST 大规模测序的数据为我们从中筛选 SSR 提供了平台, 在许多作物中已有相关报道^[3-6]。国内外研究学者也开展了棉花基因组 SSR 分布的研究, 在 34819 条亚洲棉 ESTs 序列中包含 SSRs 的 EST 频率是 2.73%; 在陆地棉的 2 个棉纤维和胚珠 cDNA 文库中发现 13505 条 ESTs 序列, 包含 SSRs 的分布频率为 7.15%^[7]; 在 63485 条雷蒙德氏棉 ESTs 序列中, 包含 SSR 的 EST 分布频率是 4.45%^[8]。以上研究结果仅针对某一个棉种的 EST 数据库进行 SSR 丰度分析, 主要目的是开发更多的 SSR 标记以建立更饱和棉花遗传图谱。对 SSR 丰度的研究可能会由于搜索软件和鉴定标准不同而产生差别, 难以比较不同种间的 SSR 分布差异。本研究采用统一的标准对四倍体棉种的 A、D 亚组供体棉种亚洲棉和雷蒙德氏棉进行 EST 拼接, 并利用 Microsatellite Identification tool 软件进行 SSR 位点搜索, 较为系统地比较和分析了雷蒙德氏棉和亚洲棉的 SSR 重复序列类型和丰度差异, 为 SSR 标记开发、棉种起源和进化的深入分析提供更多有用的信息。

1 材料和方法

1.1 EST 序列

2006 年 12 月在 NCBI 的 dbEST 数据库中 (<http://www.ncbi.nlm.nih.gov/dbEST/>) 分别以 *Gossypium raimondii*、*Gossypium arboreum* 为关键词进行搜索, 得到两个雷蒙德氏棉的 cDNA 文库: GR_Ea 和 GR_Eb, 分别是来自棉花第一片真叶和开花前 3 d 到开花后 3 d 的胚珠 EST 序列, 共计 63588 条。亚洲棉的 cDNA 文库 GA_Ea, 包括亚洲棉开花后 7~10 d 的胚珠 EST 序列 39277 条, 以 GenBank 格式保存。

1.2 简单重复序列的鉴定方法

应用 Perl 语言所编写的程序鉴定简单重复

序列。在 www.activestate.com 下载 Perl 脚本安装程序 Activeperl-5.6.1, 于本实验室 IBM 服务器进行运算。通过 Perl 脚本 EST-trimmer.pl 除去 EST 中的 5' Poly T、3' Poly A 及一些低质量片段 (<100 bp), 得到的高质量 EST 序列, 以 FASTA 格式保存。利用 Seqverter 将处理后序列拆分为单条序列, 通过 DNASTAR 软件中的 seqman 程序进行 EST 序列的聚类 and 拼接。在拼接前, 进一步除掉载体序列, DNASTAR 的拼接参数采用默认值。使用另一个 Perl 脚本程序 MISA 在拼接后的一致性序列 (Tentative consensus sequence, TCs) 中进行 SSR 查找。SSR 的查找标准为: 单碱基重复 ≥ 25 bp, 2、3、4、5、6 碱基重复 ≥ 20 bp, 即在 (N)n、(NN)n、(NNN)n、(NNNN)n、(NNNNN)n、(NNNNNN)n 中, n 值分别为 ≥ 25 、10、7、5、4、4, 查找标准参照 Svetlana^[9] 等人的报道。根据计算结果, 分析两棉种 TCs 中的 SSR 频率和分布特点。由于 DNA 双链为互补链, 所以 (A)n 相应于 (T)n、(AC)n 相应于 (CA)n, 其它简单重复序列也有类似情况^[10]。

2 结果与分析

2.1 EST 序列的拼接

在 39277 条亚洲棉及 63588 条雷蒙德氏棉的原始 EST 序列中, 分别产生了 39174 条及 63217 条高质量序列 (表 1), 拼接前长度分别为 23 Mb (占基因组的 1.35%) 及 41.08 Mb (占基因组的 4.67%), 拼接后形成的一致性序列 (TCs) 总长减为 10 Mb 及 16 Mb。亚洲棉的 14130 条 TCs 中包括 6153 条 contigs 和 7977 条 singletons, 分别占 43.55% 和 56.45%; 17912 条雷蒙德氏棉 TCs 中则为 8460 条 contigs 和 9452 条 singletons, 分别占 47.23% 和 52.77%。可见, 雷蒙德氏棉比亚洲棉产生了较高比例的高质量序列, 但序列拼接后, 亚洲棉比雷蒙德氏棉有较多比例的单条序列簇 (singleton)。其中 contig 是指由两条及两条以上的 ESTs 拼接成的重叠群, 单条序列产生的簇被称为 singleton。

所有 contig 序列中, contig 大小 (拼接为 contig 的 EST 序列数量) 在亚洲棉中最大为 244, 最小为 2, 而在雷蒙德氏棉中为 2~219 (图 1, 图 2)。

图中反应了拼接为 contig 的 EST 组成大小分布情况,可以看出由 2 个 EST 序列拼接的 contig 个数都是最多,分别为 2803 和 4076 个,并且随着 contig 大小的增加,contig 个数有递减的趋势,拼

接为 contig 的 EST 数量在 20 以内时,这两个棉种的 contig 个数呈直线下降。亚洲棉及雷蒙德氏棉中最大的 contig 长度分别为 5.08 kb 和 4.24 kb。

表 1 亚洲棉和雷蒙德氏棉间的 EST 序列拼接结果

Table 1 Assembly results of the EST sequences between *G. arboreum* and *G. raimondii*

棉种	基因组大小 /Mb	EST 数量/个		全部序列长度/Mb		一致性序列数 /个	重叠群数 /个	单条序列数 /个
		原始序列	高质量序列	拼接前序列长度	一致性序列长度			
亚洲棉	1700	39277	39174	23	10	14130	6153	7977
雷蒙德氏棉	880	63588	63217	41.08	16	17912	8460	9452

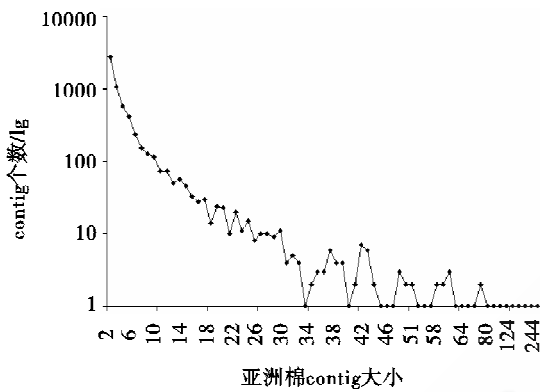


图 1 亚洲棉 contig 组成大小分布图

Fig. 1 Distribution of in contig size in *G. arboreum*

2.2 EST-SSRs 的丰度分析

利用 MISA 鉴定了不同重复类型的 SSR 在 TCs 中的分布(表 2)。在亚洲棉的 14130 个 TCs 中,获得 405 个 SSR,占 TCs 序列的 2.87%,其中,包含 SSR 的 TCs 序列占 2.52%(356 条),含有一个以上 SSR 的 TCs 序列占 0.2%(28 条),出现的复合型 SSR 占 0.19%(27 个);在 17912 条雷蒙德氏棉 TCs 中,获得 1082 个 SSR,占 TCs 序列的 6.04%,其中,包含 SSR 的 TCs 序列占 5.26%(942 条),含有一个以上 SSR 的 TCs 序列占 0.63%(112 条),出现的复合型 SSR 占 1.05%(188 个)。显然,雷蒙德氏棉较之亚洲棉,有较大比例的简单重复序列。

两棉种的 SSR 优势重复类型分析:雷蒙德氏棉拼接序列全长为 16 MB,鉴定的 1082 个简单重复序列总长度是 58.54 kb,占总序列长度的 0.36%,平均相隔 14.95 kb 就出现一个 SSR。但亚洲棉 SSR 平均间隔距离较大,每隔 25.54 kb 才出现一个 SSR。此外,亚洲棉中的 6 种重复类型的 SSR 间隔距离都分别比雷蒙德氏棉的大(表 3)。各种重复类型在两棉种中的分布较一致,单

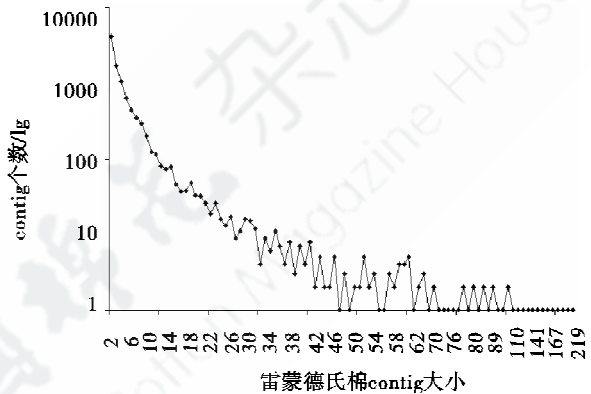


图 2 雷蒙德氏棉 contig 组成大小分布图

Fig. 2 Distribution of in contig size in *G. raimondii*

碱基重复类型数目最多,亚洲棉和雷蒙德氏棉中分别占重复序列总数的 36.36% 和 50.37%。数目依次递减的重复类型均为三、二、六、五和四碱基重复类型,可以看出,二、三碱基重复类型在重复序列总数占有极大的比例,仅次于单碱基重复类型,其它三种碱基重复类型不到总数的 25%(表 3)。最多的重复基序在二、三、五碱基重复类型中相同,分别为 AT/TA、AAG/CTT、AAAAG/CTTTT。单碱基重复类型在雷蒙德氏棉中为 A/T 最多,而亚洲棉最丰富的单碱基重复类型是 G/C。四碱基重复类型中最多的重复基序在亚洲棉中为 AAAT/ATTT,在雷蒙德氏棉中则为 AAAG/CTTT 和 ACAT/ATGT。六碱基重复类型中分布最广的重复基序也不同,分别为 AGGCTC/AGTCCG、AAAAAT/ATTTTT。总之,亚洲棉的 6 种重复类型的 SSR 间隔距离都分别比雷蒙德氏棉的大,二到六碱基等 5 种重复类型的比例高于雷蒙德氏棉,但单碱基重复类型比例低于雷蒙德氏棉,两棉种的单、四、六碱基重复类型的优势重复基序也不同。

表 2 SSRs 在一致性序列中的数量分布

Table 2 Quantity distribution of SSRs in tentative consensus sequences

棉种	TCs 数 /个	TCs 总长度 /Mb	搜索的 SSR		包含 SSR 的 TCs		包含 1 个以上 SSR 的 TCs		复合型 SSRs	
			总数/个	比例/%	个数/个	比例/%	个数/个	比例/%	个数/个	比例/%
亚洲棉	14130	10	405	2.87	356	2.52	28	0.20	27	0.19
雷蒙德氏棉	17912	16	1082	6.04	942	5.26	112	0.63	188	1.05

表 3 亚洲棉及雷蒙德氏棉中的 SSR 重复类型分布特点

Table 3 SSR Repeat type distribution between *G. arboreum* and *G. raimondii*

重复类型	SSR 个数/个		百分比频率/%		间隔距离/ (kb · SSR ⁻¹)		优势重复基序	
	亚洲棉	雷蒙德氏棉	亚洲棉	雷蒙德氏棉	亚洲棉	雷蒙德氏棉	亚洲棉	雷蒙德氏棉
单碱基	143	545	36.36	50.37	69.90	29.68	G/C	A/T
二碱基	76	164	18.43	15.16	137.94	98.62	AT/TA	AT/TA
三碱基	90	199	21.62	18.39	117.56	81.27	AAG/CTT	AAG/CTT
四碱基	19	39	4.67	3.60	544.50	414.71	AAAT/ATTT	AAAG/CTTT ACAT/ATGT
五碱基	38	61	9.09	5.64	279.61	265.14	AAAAAG/CTTTT	AAAAAG/CTTTT
六碱基	39	74	9.83	6.84	279.61	218.56	AGGCTC/AGTCCG	AAAAAT/ATTTTT
总数	405	1082	100	100	25.54	14.95		

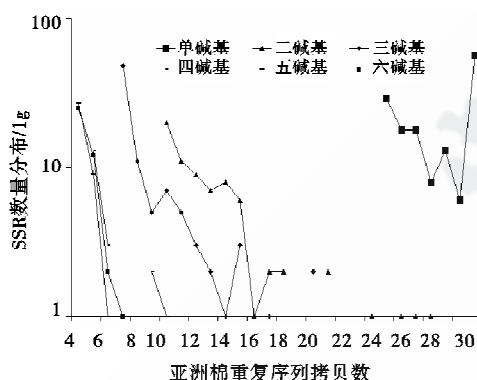


图 3 亚洲棉 6 种重复类型的拷贝数

Fig. 3 Copy number of six repeat types in *G. arboreum*

不同重复类型的重复拷贝数分布:从整体上看,随着重复序列拷贝数的增加,SSR 频率有快速递减的趋势。两棉种中单碱基重复拷贝数为 25 时,SSR 频率最高,拷贝数大于 30 的单核苷酸重复类型分布广泛,雷蒙德氏棉中最高的单碱基重复拷贝数为 126,亚洲棉中最高为 80。雷蒙德氏棉中二碱基重复拷贝数主要在 10~21 之间,最高的二碱基重复拷贝数为 26,而亚洲棉二碱基重复拷贝数主要分布于 10~18 之间,并在拷贝数为 22~26 时有零星分布。三碱基重复拷贝数分布略有差异,亚洲棉中拷贝数变化范围较窄,为 7~17,而雷蒙德氏棉中变幅为 7~21,随着重复拷贝数的增加,其 SSR 数量分布呈下降、上升再下降的趋势。两棉种的四、五、六碱基重复拷贝数较少,基本都在 4~6 之间,较为稳定。从上可见,雷

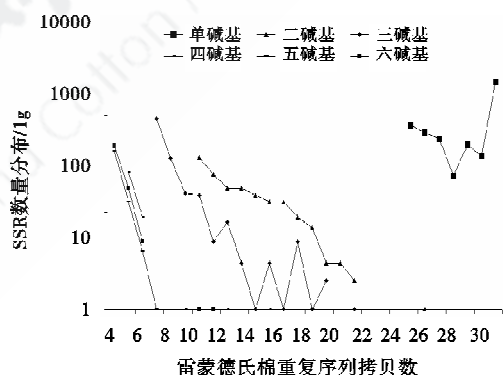


图 4 雷蒙德氏棉 6 种重复类型的拷贝数

Fig. 4 Copy number of six repeat types in *G. raimondii*

蒙德氏棉和亚洲棉两棉种的 6 种 SSR 重复类型的重复拷贝数变化趋势都较为一致(图 3,4)。

不同重复类型 SSR 的重复基序和丰度分布:在雷蒙德氏棉的 1082 个 SSR 中,有 107 种不同的重复基序,亚洲棉的 405 个 SSR 中,包括 72 个不同的重复基序。在雷蒙德氏棉中,84.03% 的单碱基重复基序为 A/T,亚洲棉中的 G/C 重复占单碱基重复的 54.73%。雷蒙德氏棉和亚洲棉的二碱基重复基序的丰度趋势基本一致,按百分比频率分布依次为 AT/TA>AG/CT>AC/GT,所不同的是雷蒙德氏棉中的 AT/TA 较为丰富,而 AG/CT 基序则是亚洲棉较丰富(图 5)。三碱基重复基序在两棉种的分布较为一致,依次为 AAG/CTT>AAT/ATT>AGT/ATC>ACC/GGT>AGC/CGT,只是 ACT/ATG 基序在雷蒙

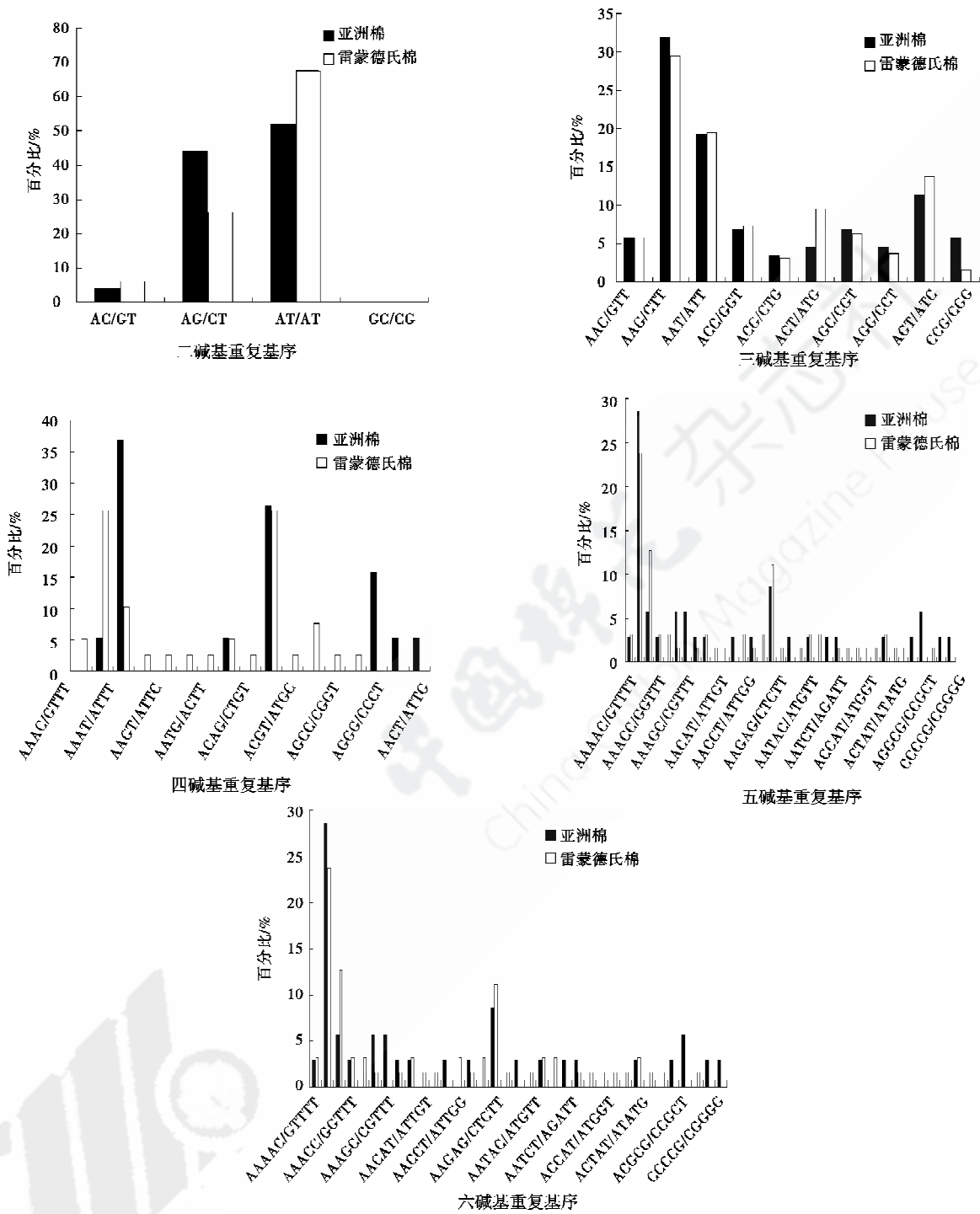


图5 基于重复基序(Motif)的简单重复序列分布

Fig. 5 Different motifs of SSRs distribution

德氏棉中比亚洲棉的丰度大。分布最少的三碱基重复基序在亚洲棉中为 ACG/CTG, 在雷蒙德氏棉中为 CCG/CGG。所有三碱基基序中 A/T 含量在两棉种均大于 G/C 含量, 分别为 55.68% 和 58.77%。四碱基重复类型的优势重复基序在两棉种中有所不同, 在亚洲棉中 AAAT/ATTT 基序最多, 占有四碱基重复类型的 36.84%, 其次

为 ACAT/ATGT, 占 26.32%; 雷蒙德氏棉中 AAAG/CTTT 和 ACAT/ATGT 中各占总数的 25.64%, 丰度最高。其它四碱基重复基序如 AAAC/GTTT、AAGG/CCCT、AAGT/ATTC、AATC/AGTT、AGAT/ATCT 等基序只存在于雷蒙德氏棉中, AGGG/CCCT、AACC/GGTT、AACT/ATTG 这三种基序只在亚洲棉中发现。

五碱基重复基序和丰度在两棉种中存在很大差异,有 15 种五碱基重复基序只在雷蒙德氏棉中存在,而有 8 种五碱基重复基序只在亚洲棉中存在。另外,值得注意的是组成五碱基重复基序的 A/T 和 G/C 碱基含量在两棉种中有很大差别,亚洲棉中为 G/C 碱基较多,G/C 含量多达 70.60%,雷蒙德氏棉中基序组成碱基主要是 A/T,占组成碱基总数的 61.9%。然而,在两棉种中,最丰富的重复基序类型均为 AAAAG/CTTTT。同样,在六碱基重复类型和丰度中发现亚洲棉的组成基序的碱基 G/C 含量大于 A/T 含量,雷蒙德氏棉中 G/C 含量小于 A/T 含量。雷蒙德氏棉中分布最广六碱基重复基序为 AAAAAT/ATTTTT,亚洲棉中为 AGGCTC/AGTCCG。总体看来,两棉种在组成二、三碱基重复基序的种类上没有差别,只存在丰度上的变化,但单碱基,尤其是四碱基以上的重复基序种类和丰度在两个棉种间都有很大差异。根据这一特点,我们可以利用在两棉种中丰度差异很大的四核苷酸以上的 SSRs 来区别和鉴定这两个棉种的差异。

3 讨论

从 EST 序列数据库中开发的 SSR 标记,已经广泛应用于遗传图谱构建、基因定位、比较基因组学研究及新基因筛选和发掘等方面^[11]。本研究通过对雷蒙德氏棉和亚洲棉的 EST 序列进行拼接,得到 14130 条亚洲棉 TCs,鉴定 405 个 SSR,占 2.87%;17912 条雷蒙德氏棉 TCs 中,获得 1082 个 SSR,占 6.04%。与前人结果的差异主要归因于不同作物的 EST 测序计划完成进度不同,EST 序列来源的组织不同或选择发育时期不同。此外,采用的软件及搜索标准不同也是导致 SSR 分布差异的原因。

简单重复序列的不稳定性随着长度的增加而增加,长的简单重复序列更容易对基因组的组织、遗传变异、蛋白进化等产生影响^[12]。本研究基于前人的总结,将搜索标准定为单碱基重复的长度 ≥ 25 bp,其它重复类型长度 ≥ 20 bp,分别在亚洲棉、雷蒙德氏棉中查找 SSRs,以保证在两棉种中能找到高度变异的简单重复序列。在两棉种的重复类型比较中,雷蒙德氏棉的单碱基重复分布大于亚洲棉,亚洲棉的二到六碱基 5 种重复类型分布比雷蒙德氏棉的多。三碱基重复类型比二碱基重复类型的分布广,分布最少的重复类型为四碱基重复类型。与 Meng^[13] 等的研究结果一致,最

丰富的单碱基重复基序在雷蒙德氏棉中为 A/T,而亚洲棉中为 G/C。雷蒙德氏棉中最普遍的二、三、五、六碱基重复基序分别 (AT/TA)_n、(AAG/CTT)_n、(AAAAG/CTTTT)_n、(AAAAAT/ATTTTT)_n,结果与 Wang^[8] 等的结果一致,但四碱基重复基序不同。在水稻、拟南芥、小麦、玉米、高粱的 EST-SSR 中,出现频率最高的二碱基重复基序是 (AG/TC)_n,最多的三碱基重复基序在水稻、拟南芥、高粱、大麦中是 (CCG)_n,而小麦和玉米中分别为 (AAG)_n 和 (AAC)_n^[11]。为了避免高丰度表达基因的 ESTs 被重复鉴定,我们将数据库中的 EST 序列进行聚类以除去冗余部分,并在拼接前,除去 EST 中低质量序列、真核生物 mRNA 中的 polyA 尾,得到 5'-UTR 和 3'-UTR 末端 50 bp 碱基长度内不包含 (T)₅ 或 (A)₅ 的 ESTs,同时去除载体序列。在水稻公共 EST 数据库中,富含 GC 的 3 碱基简单重复序列最多^[11],而在本研究中两个棉种的三碱基重复基序 AT 含量较高,为 55.68% 和 58.77%。组成五、六碱基重复基序的碱基含量在两棉种中分布不同。在亚洲棉中组成基序的 GC 碱基含量大于 AT 含量,雷蒙德氏棉中 AT 含量高于 GC 含量。由于 EST 序列中含有非翻译区,有很多 (A/T)_n 分布在 EST 中,导致所查询的单碱基重复序列比例较高。此外,EST 序列在棉花中非常有限其中部分序列只进行单次测序,我们未能预测 EST 开放读码框并精确确定 SSR 在 5'UTR 和 3'UTR 中的位置。

在棉花二倍体 A、B、C、D、E、F 和 G 7 个染色体组中,D 染色体组的染色体相对大小和 DNA 相对含量都是最小的,因而普遍认为 D 染色体组是棉属进化中最原始的祖先。其它的染色体组可能是通过 DNA 重复顺序的增加而衍生形成的^[14]。亚洲棉和雷蒙德氏棉分别为 A₂ 和 D₅ 染色体组,基因组大小是 880 MB 和 1700 MB。前人研究表明,亚洲棉基因组中重复序列大于雷蒙德氏棉 D 基因组的重复序列^[15-16]。但是在本研究中亚洲棉的简单重复序列却远远少于雷蒙德氏棉,产生的原因其一是研究的亚洲棉 EST 数量少于雷蒙德氏棉,另一种推测为在编码序列中,简单重复序列的比例在亚洲棉中比雷蒙德氏棉中小。这些结果表明具有较多高度重复序列的亚洲棉,其简单重复序列反而低于高度重复序列较少的雷蒙德氏棉,可能是两个棉种进化差异造成的。

参考文献:

- [1] HANCOCK J M. The contribution of slippage-like processes to genome evolution[J]. *Mol Evol*, 1995, 41: 1038-1047.
- [2] BECKMANN J S, Weber J L. Survey of human and rat microsatellites[J]. *Genomics*, 1992, 12: 627-631.
- [3] MORGANTE M, Michael H, Wayne P, et al. Microsatellites are preferentially present with nonrepetitive DNA in plant genomes [J]. *Nat Genet*, 2002, 30: 194-200.
- [4] SAHA S, Karaca M, Jenkins J N, et al. Simple sequence repeats as useful resources to study transcribed genes of cotton [J]. *Euphytica*, 2003, 130: 355-364.
- [5] HAN Z G, Guo W Z, Song X L, et al. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton[J]. *Mol Gen Genome*, 2004, 272, 308-327.
- [6] 李华盛, 范术丽, 沈法富. 从棉花 ESTs 数据库中筛选微卫星标记的初步研究[J]. *棉花学报*, 2005, 17(4): 211-216.
- [7] HAN Z G, Wang C B, Song X L, et al. Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton[J]. *Theor Appl Genet*, 2006, 112: 430-439.
- [8] WANG C B, Guo W Z, Cai C P, et al. Characterization, development and exploitation of EST-derived microsatellites in *Gossypium raimondii* Ulbrich [J]. *Chinese Science Bulletin*, 2006, 51(5): 557-561.
- [9] TEMNYKH S, Declerck G, Lukashova A, et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential[J]. *Genome Res*, 2001, 11(8): 1441-1452.
- [10] LI C. Diversity of simple sequence repeats between *Arabidopsis thaliana* and rice[J]. *Acta Botanica Sinica*, 2004, 46(5): 603-609.
- [11] 李永强, 李宏伟. 基于表达序列标签的微卫星标记 (EST-SSRs) 研究进展[J]. *植物遗传资源学报*, 2004, 5(1): 91-9.
- [12] 李 灿. 水稻基因组第 4 号染色体简单重复的遗传分析及比较基因组学研究[D]. 中国科学院研究生院, 2004, 3-6.
- [13] MENG H J, Cao Q Q, Hu Z Y, et al. Analysis of SSR in Citrus sequences from EMBL database[J]. *Agricultural Sciences in China*, 2005, 4(7): 501-506.
- [14] 季道藩. 棉花百科知识[M]. 北京: 中国农业出版社, 2001: 4-5.
- [15] HANSON R E, Zhao X P, Paterson A H, et al. Fish indicates a dispersed repetitive element highly represented in the *Gossypium hirsutum* L. D-subgenome is absent from its putative ancestral diploid donor (*Gossypium raimondii* U.) [C]. // *Plant Genome IV Conference Abstract*. Town&Country Conference Center, San Diego, CA: 1995, January.
- [16] ZHAO X P, Si Y, Hanson R E, et al. Dispersed repetitive DNA has spread to new genome since polyploidy formation in cotton [J]. *Genome Research*, 1998, 8(5): 479-492. ●