人参EST资源的SSR信息分析

杨成君，王军*
东北林业大学林学院，哈尔滨 150040

提要：从7 055条人参EST序列中搜索出791个SSR，其出现频率为11.21%，平均长度为21.37 bp，平均分布频率为1/5.7 kb。二核苷酸重复是主要的重复类型，占全部EST-SSR的56.89%，其次是三核苷酸重复的占全部SSR的21.11%。AT、GAA是二核苷酸和三核苷酸中出现次数最多的重复基元类型，分别占28.89%和10.18%。

关键词：人参；表达序列标签；简单重复序列

Analysis of SSR Information of EST Resource in *Panax ginseng* C. A. Meyer

YANG Cheng-Jun, WANG Jun*
College of Forestry, Northeast Forestry University, Harbin 150040, China

Abstract: Seven hundred and ninety one microsatellites (SSRs) were isolated from 7 055 *Panax ginseng* expressed sequence tags (ESTs). The occurrence frequency of SSR was 11.21%, the average length of *Panax ginseng* searched was 21.37 bp and the average distance of distribution was 1/5.7 bp. The dinucleotide repeat was the dominant repeat type in *Panax ginseng* EST-SSR, accounting for 56.89% of the total EST-SSR, and then trinucleotide repeat accounts for 21.11% of the total EST-SSR. AT and GAA were the most frequent motifs, accounting for 28.89% and 10.18% in dinucleotide and trinucleotide repeats, respectively.

Key words: *Panax ginseng*; expressed sequence tag (EST); simple sequence repeat (SSR)

作为高多态性遗传标记的简单重复序列(simple sequence repeats，SSRs)已广泛用于遗传连锁图谱构建，种质资源遗传多样性分析，品种鉴定和基因诊断等研究领域。简单重复序列是由1~6个核苷酸为重复单元的串联重复序列，广泛分布于真核生物基因组的编码区和非编码区(Gupta等1996)。SSR标记具有多态性高、共显性、重复性好、数量丰富、技术简单和特异性强等特点(Powell等1996)。源自传统基因组的SSR开发需要构建基因组文库、探针杂交和克隆测序等繁杂的操作程序(Röder等1998)，不仅费时费力，而且开发成本高。EST-SSR可以直接从数据库存在的EST序列中开发，相对简便、经济，而且由于这些EST序列来自于基因表达区域，非常保守，采用SSR两端的保守序列设计引物，可以缩短开发周期和节省开发成本。近年来大量快速增长的EST数据已成为SSR的重要来源，各种植物中约有1%~5%的EST含有可用于建立标记的SSR(Kantety等2002)。迄今，尚无采用现有人参EST资源建立EST-SSR的报道。本文对现有人参EST资源中的SSR进行了分析，从而为建立其EST-SSR标记及其在人参遗传育种中的应用建立了基础。

材料与方法

五加科人参属的人参(*Panax ginseng* C. A. Meyer)的EST共计7 055条，其中6 614条来自NCBI (美国国家生物技术信息中心)的dbEST数据库(http://www.ncbi.nlm.nih.gov/dbEST/index.html)；另外441条从本校林学院人参叶片cDNA文库测序获得。

人参EST中SSR筛选的登陆网站 为(http://www.gramene.org/db/searches/ssrtool)，用SSRIT (simple sequence repeat identification tool)软件在线搜索EST-SSR。搜索SSR的长度标准为：二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸的最少重复次数7、5、4、4、3次以上。对二、三、四、五和六核苷酸重复基元分为完全重复、不完全重复，并对搜索的SSR的频率和

收稿 2007-10-29 修定 2008-01-18
资助 国家林业局野生动植物保护项目(010-413255)。
* 通讯作者(E-mail：junwang1966@yahoo.com.cn；Tel：0451-82191829)。

长度进行了统计和分析。

实验结果

1 人参EST-SSR的发生频率

分析获得的7 055条人参EST结果显示，其EST的总长度为4 507.55 kb，平均长度为639.92 bp。对7 055条人参EST进行SSR搜索的结果表明，含有SSR的EST共计661条，占全部EST总数的9.37%，含有791个SSR。在661条含有SSR的EST中，只含有1个SSR的有603条，含有2个SSR的有28条，含有3个SSR的14条，含有4个SSR的有9条，含有5个SSR的有2条，含有6、7、8、11、12个SSR的EST分别都是1条。

人参EST-SSR的重复类型比较丰富，一至六核苷酸都能观测到，其他类型还有七核苷酸、八核苷酸，但是这两种核苷酸的基元种数和SSR数都极少。从占全部SSR比例和出现频率来看(表1)，二核苷酸所占全部SSR比例较大，占56.89%，其次是三核苷酸，占21.11%，其他类型所占比例较小；SSR出现频率为11.21%，其中二核苷酸出现的频率最高(为6.38%)，其次是三核苷酸(为2.37%)，六核苷酸为1.29%，其他类型均不足1%。可见，在人参EST-SSR中二核苷酸重复

表1 人参EST中SSR的出现频率

Table 1 Occurrence frequency of SSRs in ESTs of *P. ginseng*

重复类型	SSR数	占全部SSR比例/%	出现频率/%	平均分布频率/kb^{-1}
二核苷酸	450	56.89	6.38	10.02
三核苷酸	167	21.11	2.37	26.99
四核苷酸	49	6.19	0.69	91.99
五核苷酸	24	3.03	0.34	187.81
六核苷酸	91	11.50	1.29	49.53
其他	10	1.47	0.14	450.76
总计	791	100.00	11.21	5.70

表2 人参EST-SSR的重复基元

Table 2 Repeat motifs in EST-SSRs of *P. ginseng*

重复类型	基元种类	重复基元
二核苷酸	11	AT、TA、CT、TC、GA、AG、TG、CA、AC、GT、GC
三核苷酸	40	GAA、ATC、AGA、CAG、AGC、TAT、ATT、AAG、AAT、TGC、CTT、ATA、CCT、GAT、GCT、GTA、CAC、CTC、GCA、TCT、CCA、GAG、GGA、TAA、TCC、ATG、CTA、ACC、AGG、CCG、CGA、CGC、CGG、GAC、GCC、GGC、GGT、GTG、TCA、TGA
四核苷酸	27	TGCA、AGCT、CTTT、AAAG、ACTC、TCTT、TTTA、AATA、AGTG、TATG、 TCAA、AATT、ACAG、AGAA、AGAC、ATAC、CTCC、GAAG、GTAC、TAAA、TAAT、TACA、TACG、TCAC、TCTA、TGTA、TTAT
五核苷酸	19	ACCAA、AGAAA、AAATA、AACAA、AATCA、AGTTT、CAATT、CACTG、CTGTC、CTTGT、GAGGA、GCACA、GCACG、TCCAA、TCTGC、TGCCC、TGGCT、TGTTG、TTTCT
六核苷酸	63	AAAAAG、AAAAAT、AAACAG、AAAGAA、AAGCAG、AATTTA、AGAGAA、AGATTC、AGCCTG、AGGAAG、AGGGGT、ATAAAA、ATCATA、ATGGGA、ATTTTT、CAAAAT、CACCTC、CAGTTG、CATCTC、CATGAG、CCAGAA、AAAAAG、AAAAAT、AAACAG、AAAGAA、AAGCAG、AATTTA、AGAGAA、AGATTC、AGCCTG、AGGAAG、AGGGGT、ATAAAA、ATCATA、ATGGGA、ATTTTT、CAAAAT、CACCTC、CAGTTG、CATCTC、 CATGAG、CCAGAA
其他	9	CTGGCTA、GAGTGAG、AATTTAA、AGTTTTG、ATTATTA、CCCTCCT、GAGGGAA、TCTCAAC、ATATAGCG

占主导地位。从平均分布频率(EST长度与EST-SSR数目的比值)来看，人参EST中平均每5.7 kb就出现1个SSR。其中二核苷酸每10.02 kb就出现一次，出现频率最高，其次是三核苷酸、六核苷酸、四核苷酸、五核苷酸、其他类型。这表明人参EST中的SSR数量非常丰富。

2 人参EST-SSR的特点

在人参的EST-SSR中，共观察到169种重复基元。二、三、四、五、六核苷酸酸重复基元分别有11、40、27、19、63和其他类型9种(表2)。二核苷酸和三核苷酸重复基元约占重复基元的32.5% (表3)，其中二核苷酸重复基元AT、TA、CT、TC、GA、AG出现的次数较多，分别占二核苷酸重复的28.89%、24.00%、17.78%、12.22%、8.89%、5.56%，TG、CA、AC、GT、GC出现次数较少，分别占二核苷酸的比例不足1%，其中GC只出现1次；在三核苷酸重复基元中，GAA、ATC、AGA、CAG、AGC、TAT出现较多，分别占三核苷酸重复的10.18%、7.19%、6.59%、6.59%、5.39%、5.39%，其他三核苷酸重复基元都不足5%。四、五、六核苷酸重复类型约占总重复基元的67.5%，但是其重复基元各自出现频率都很低，而且所占比例相近。

从二至六核苷酸重复基元的重复特性来看，人参EST-SSR分完全重复和不完全重复两类，分别为599和182个(表4)。总体上二至六核苷酸的完全重复多于不完全重复。人参EST中二至六核苷酸SSR平均长度为21.37 bp，但不同的SSR长度有很大差异，最短为14 bp，最长为482 bp。

讨 论

表达序列标签在国际公共数据库中呈指数增长，为遗传分析提供了丰富资源。本文对7 055条人参EST进行了搜索，发现791个SSR，其出现频率为11.21%，平均分布频率1/5.7 kb。这个比率低于茶树(1/2.61 kb) (金基强等2006)、油菜(1/4.34 kb) (李小白等2007)、水稻(1/3.4 kb) (Cardle等2000)，高于大麦(1/6.3 kb) (Thiel等2003)、大豆(1/7.4 kb) (Cardle等2000)、拟南芥(1/14.9 kb) (Morgante等2000)等植物。表明人参EST-SSR比较丰富。EST-SSR的频率一般是通过

表3 人参EST中二核苷酸和三核苷酸重复基元

Table 3 Dinucleotide and trinucleotide repeat motifs in ESTs of *P. ginseng*

重复类型	重复基元	数量	发生频率/%	所占比例/%
二核苷酸	AT	130	1.84	28.89
	TA	108	1.53	24.00
	CT	80	1.13	17.78
	TC	55	0.78	12.22
	GA	40	0.57	8.89
	AG	25	0.35	5.56
	TG	4	0.06	0.89
	CA	3	0.04	0.67
	AC	2	0.03	0.44
	GT	2	0.03	0.44
	GC	1	0.01	0.22
三核苷酸	GAA	17	0.24	10.18
	ATC	12	0.17	7.19
	AGA	11	0.16	6.59
	CAG	11	0.16	6.59
	AGC	9	0.13	5.39
	TAT	9	0.13	5.39
	ATT	7	0.10	4.19
	AAG	6	0.09	3.59
	AAT	6	0.09	3.59
	TGC	6	0.09	3.59
	CTT	5	0.07	2.99
	ATA	4	0.06	2.40
	CCT	4	0.06	2.40
	GAT	4	0.06	2.40
	GCT	4	0.06	2.40
	GTA	4	0.06	2.40
	CAC	2	0.03	1.20
	CTC	3	0.04	1.80
	GCA	3	0.04	1.80
	TCT	3	0.04	1.80
	CCA	2	0.03	1.20
	GAG	2	0.03	1.20
	GGA	2	0.03	1.20
	TAA	2	0.03	1.20
	TCC	2	0.03	1.20
	ATG	2	0.03	1.20
	CTA	2	0.03	1.20
	ACC	1	0.01	0.60
	AGG	1	0.01	0.60
	CCG	1	0.01	0.60
	CGA	1	0.01	0.60
	CGC	1	0.01	0.60
	CGG	1	0.01	0.60
	GAC	1	0.01	0.60
	GCC	1	0.01	0.60
	GGC	1	0.01	0.60
	GGT	1	0.01	0.60
	GTG	1	0.01	0.60
	TCA	1	0.01	0.60
	TGA	1	0.01	0.60

表4 人参EST中SSR的重复长度和性质

Table 4 The repeat length and character of SSRs in *P. ginseng* ESTs

重复类型	长度		重复性质	
	变化范围/bp	平均长度/bp	完全重复	不完全重复
二核苷酸	14~482	25.41	331	119
三核苷酸	15~84	19.27	117	50
四核苷酸	16~36	20.08	43	6
五核苷酸	20~25	21.04	22	2
六核苷酸	18~36	21.03	86	5
总计	14~482	21.37	599	182

搜索数据库中的EST序列而估算出的，由于搜索SSR重复类型和长度等标准的不同以及分析数据的多少不同，可能是造成统计结果的不一致的原因，这种差异也可能是物种间真实的SSR信息差异。

人参EST-SSR的重复类型以二核苷酸重复为主，占全部SSR的56.89%，其次是三核苷酸，占全部SSR的21.11%。李永强等(2004)认为大多数植物的EST-SSR主要是二、三核苷酸重复类型。但是不同的重复基元类型又各有特点。例如，大麦(Thiel等2003)、甘蔗(Cordeiro等2001)、水稻、大豆和小麦(Cardle等2000)、棉花(李华盛等2005)、苜蓿(Eujayl等2004)、牛尾草(Saha等2004)均以三核苷酸重复为主，而茶树(金基强等2006)、猕猴桃(Fraser等2004)、杏树、桃树(Jung等2005)等则以二核苷酸重复为主，黄瓜(Kong等2007)、油菜(李小白等2007)、白菜(忻雅等2006)、谷子(Jia等2007)的二、三核苷酸重复类型所占比例相差不多。人参EST-SSR的二核苷酸重复中出现次数最多的是AT，其他重复基元出现次数较多的依次是TA、CT、TC、GA、AG，在茶树(金基强等2006)、小麦(陈军方等2005)、油菜(李小白等2007)二核苷酸重复都是AG/CT重复出现次数最多。黄瓜(Kong等2007)二核苷酸重复GA/CT，TA/TA出现次数最多。人参EST-SSR的二核苷酸重复中GC出现次数最少，与其他报道的物种类似(Kantety等2002；陈军方等2005)。在人参EST-SSR三核苷酸重复基元中GAA出现次数最多，黄瓜(Kong等2007)中也是GAA出现次数最多，而小麦(陈军方等2005)、水稻、高粱和玉米(Kantety等2002)中AAC是主要的三核苷酸重复类型。在其他物种中AAG、AAT也是常见的三核苷酸重复基元(Gupta等1996)。

一般来说，SSR标记是来自基因组文库的筛选，这个过程很复杂，得到结果的费用昂贵。目前，数据库中EST数量不断增加，已成为SSR标记建立的一种新的途径和丰富来源。并且在许多物种的SSR已成功地从EST中发展出来。EST-SSR属于或者接近功能基因，为基因定位提供了一种有效方法。EST-SSR标记作为一种新型的SSR标记，已经应用到许多物种的遗传作图、遗传多样性分析、物种分类和系统发育研究及其比较基因组学等诸多方面的研究。人参EST-SSR标记的开发尚未见报道，本文结果表明人参EST中含有丰富的SSR，值得重视。

参考文献

陈军方, 任正隆, 高丽锋, 贾继增(2005). 从小麦EST序列中开发新的SSR引物. 作物学报, 31: 154~158

金基强, 李素芳, 龚晓春, 卢美贞, 姚艳玲, 忻雅, 崔海瑞(2006). 茶树EST资源中SSR的信息分析. 科技通报, 22 (4): 471~476

李华盛, 范术丽, 沈法富(2005). 从棉花ESTs数据库中筛选微卫星标记的初步研究. 棉花学报, 17 (4): 211~216

李小白, 张明龙, 崔海瑞(2007). 油菜EST资源的SSR信息分析. 中国油料作物学报, 29 (1): 20~25

李永强, 李宏伟, 高丽锋, 何蓓如(2004). 基于表达序列标签的微卫星标记(EST-SSRs)研究进展. 植物遗传资源学报, 5 (1): 91~95

忻雅, 崔海瑞, 卢美贞, 姚艳玲, 金基强, 林容杓, 崔水莲(2006). 白菜EST-SSR信息分析与标记的建立. 园艺学报, 33 (3): 549~554

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics, 156: 847~854

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001). Microsatellite markers from sugarcane (*Saccharum*

spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci, 160: 1115~1123

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MA (2004). *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet, 108: 414~422

Fraser LG, Harvey CF, Crowhurst RN, de Silva HN (2004). EST-derived microsatellites from *Actinidia* species and their potential for mapping. Theor Appl Genet, 108: 1010~1016

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996). Microsatellites in plants: a new class of molecular markers. Curr Sci, 70: 45~54

Jia XP, Shi YS, Song YC, Wang GY, Wang TY, Li Y (2007). Development of EST-SSR in foxtail millet (*Setaria italica*). Genet Resour Crop Evol, 54: 233~236

Jung S, Abbott A, Jesudurai C, Tomkins J, Main D (2005). Frequency, type, distribution and annotation of simple sequence repeats in *Rosaceae* ESTs. Funct Integr Genomics, 5: 136~143

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol, 48: 501~510

Kong Q, Xiang C, Yu Z, Zhang C, Liu F, Peng C, Peng X (2007). Mining and charactering microsatellites in *Cucumis melo* expressed sequence tags from sequence database. Mol Ecol Notes, 7: 281~283

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang LJ, May GD (2004). Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet, 109: 783~791

Morgante M, Hanafey M, Powell W (2000). Microsatellites are prefertially associated with nonrepetitive DNA in plant genomes. Nat Genet, 30: 194~200

Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends Plant Sci, 1: 215~222

Röder MS, Korzun V, Wendehake K, Plashke J, Tixier MH, Leroy P, Ganal MW (1998). A microsatellite map of wheat. Genetics, 149: 2007~2023

Thiel T, Michalek W, Varshney RK, Ganer A (2003). Exploiting EST database for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet, 106: 411~422