

东北师范大学



三七基因 SSR 位点分析

组 长：张美齐

组 员：初立爽、董行刚、曲春杰、

任梦晨、王 卫、于真悦、

赵 雪

指导教师：王花英

2017年10月14日

三七基因 SSR 位点分析

摘要： 随着高通量测序技术的发展，使得基因组测序成为挖掘功能基因、筛选分子标记、阐明代谢途径等的有力工具。本文以已有三七基因序列为基础，利用 Perl 语言程序在选出的 10 条最长序列中搜索 SSR 位点，采用生物信息学方法分析数据，得知三七转录组 SSR 不但出现频率高，而且类型丰富，为人参起源与驯化的研究提供了数据基础。

关键字： SSR 位点 、 筛选 、 三七 SSR 位点应用

1 前言

人参属(Panax)是五加科的一个小属，包含约 11 个种(Wen, 2001)，其药用价值较高，全属植物均可药用。最知名的有人参 (Panax ginseng)、三七 (Panaxnotoginseng) 以及西洋参(Panaxquinquefolius)，并作为人工栽培药材，在全世界广泛应用^[1]。三七别称田七、金不换，为五加科人参属植物，《本草纲目拾遗》中记载：“人参补气第一，三七补血第一，味同而功亦等，故称人参三七，为中药之最珍贵者”，可见三七是一种重要而名贵的药材^[2]，扬名中外的中成药“云南白药”和“片仔癀”即以三七为主要原料制成。三七味甘、微苦，性温，归肝、肾经，具有止血、散瘀、消肿、止痛等功效。三七的分布范围狭窄，仅分布于 23° 30' N 附近的中、高海拔地区^[3]，中国的三七主产于云南西南的山区。三七的栽培条件甚为苛刻，性喜温凉，忌阳光直射，需透光率为 15% 的人工阴棚才可正常生长，宜在海拔为 1 500 m 左右的地区种植^[4]。

简单重复序列 (simple sequence repeat, SSR) 又称微卫星 DNA，SSR 标记是较为理想的遗传标记，具有数量丰富、成本低、技术简单、分布广泛、多态性强和共显性遗传易于分析等特点^[5]。DNA 分子标记是鉴定种质资源遗传多样性的重要手段，也是一种重要的辅助育种方法^[6]。目前已利用 SSR 标记构建了许多物种的染色体遗传图谱，并被广泛应用于基因定位及亲缘关系分析、品种鉴定动植物育种、基因标记与做图、种质资源保存、评价和利用等领域^[7, 8]。本研究统计了三七一段基因组序列中的 A、T、C、G、N 含量，分析了三七转录组中的 SSR 位点 (2~3bp 重复)，并设计简单重复序列 (SSR) 引物，了解人参起源，为人参驯化提供帮助。

2 材料和方法

2.1 三七基因组概况

三七基因组数据由指导老师提供，来源于目前已经测定出三七基因组的文献

[9]。利用 perl 语言知识设计脚本,用 seqkit 测定出已有的三七基因组中的 A(腺嘌呤)、T(胸腺嘧啶)、G(鸟嘌呤)、C(胞嘧啶)、N(氮)的含量,并测定三七基因组是否拼接到染色体水平,利用脚本统计每一条染色体的 A(腺嘌呤)、T(胸腺嘧啶)、G(鸟嘌呤)、C(胞嘧啶)的含量。

2.2 SSR 位点筛选

为了检测三七中的 SSR 位点,对组装得到的 Unigene 序列进行 SSR 分析。利用 misa.pl 搜索和定位 SSR,搜索的标准为重复单元长度 2~3 bp,二核苷酸至少重复 6 次,三核苷酸至少重复 5 次,同时筛选被间隔小于或等于 100 bp 碱基间隔的复合型 SSR。单核苷酸重复容易发生错配而测序失败,所以没有选择。

2.3 SSR 引物设计及筛选

调用 Primer 5.0 程序对含有 SSR 位点的两端序列设计引物,引物设计标准:退火温度(T_m)在 55~64°C,上、下游引物的 T_m 值相差不大于 5°C^[10];引物长度在 2-3bp;GC 量在 40%~60%;尽量避免出现引物二级结构如发卡结构、二聚体、错配和引物二聚体。将设计出的引物通过以下方式筛选:(1)引物不能存在 SSR;(2)将获得的引物比对到 Unigene 序列,引物的 5' 端允许有 3 个碱基的错配,3' 端允许有 1 个碱基的错配;(3)去掉比对到不同 Unigene 上的引物,筛选唯一匹配的引物;(4)使用 SSRfinder 校验 SSR,使用产物序列来寻找 SSR,检验结果是否与 misa 结果相同,并筛选出相同的 SSR 产物^[11]。

3 实验结果

3.1 ATGC 含量

整个基因组的碱基中,含有 26.99%的 A,27.01%的 G,14.43%的 T,14.43%的 C。从基因组中选出最长的 10 条序列,其 ATCG 分布如下。

表 1 各序列碱基占比表

编号	名称	碱基数	各碱基所占比例%			
			A	G	T	C
1	Scaffold1	1196860	29.47	29.71	16.16	16.68
2	Scaffold2	1134356	27.34	27.65	15.25	15.66
3	Scaffold19	1092634	28.8	28.93	15.69	15.9
4	Scaffold3	1021613	28.89	29.14	16.33	15.05
5	Scaffold13	955161	29.59	29.06	15.9	15.73
6	Scaffold222	947491	28.63	28.06	15.7	15.15
7	Scaffold39	912679	26.63	26.78	14.55	14.38
8	Scaffold207	906998	27.84	27.86	14.38	15.08

9	Scaffold4	895980	28.31	28.11	14.98	15.51
10	Scaffold5	882481	29.56	29.3	15.34	16.23

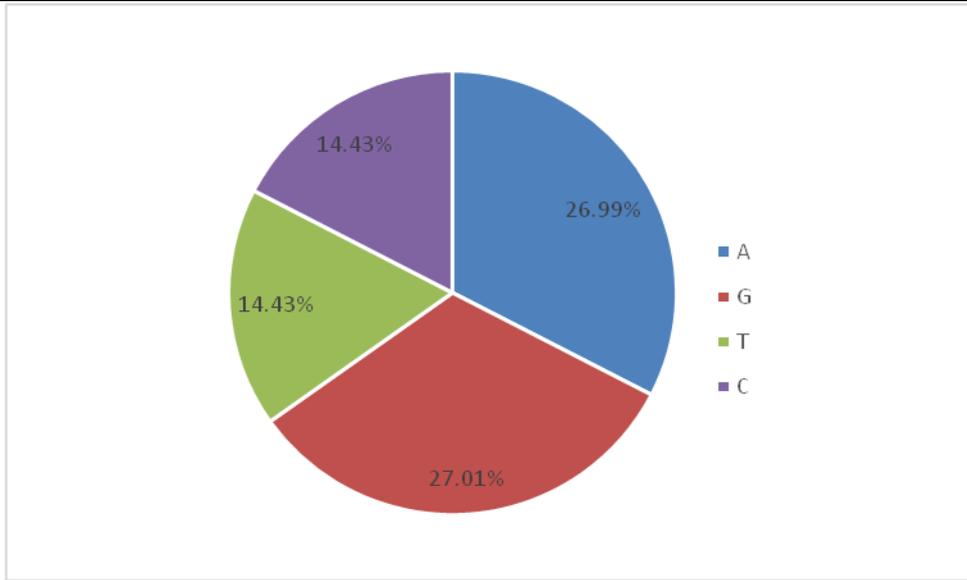


图 1 基因组中各碱基所占比例

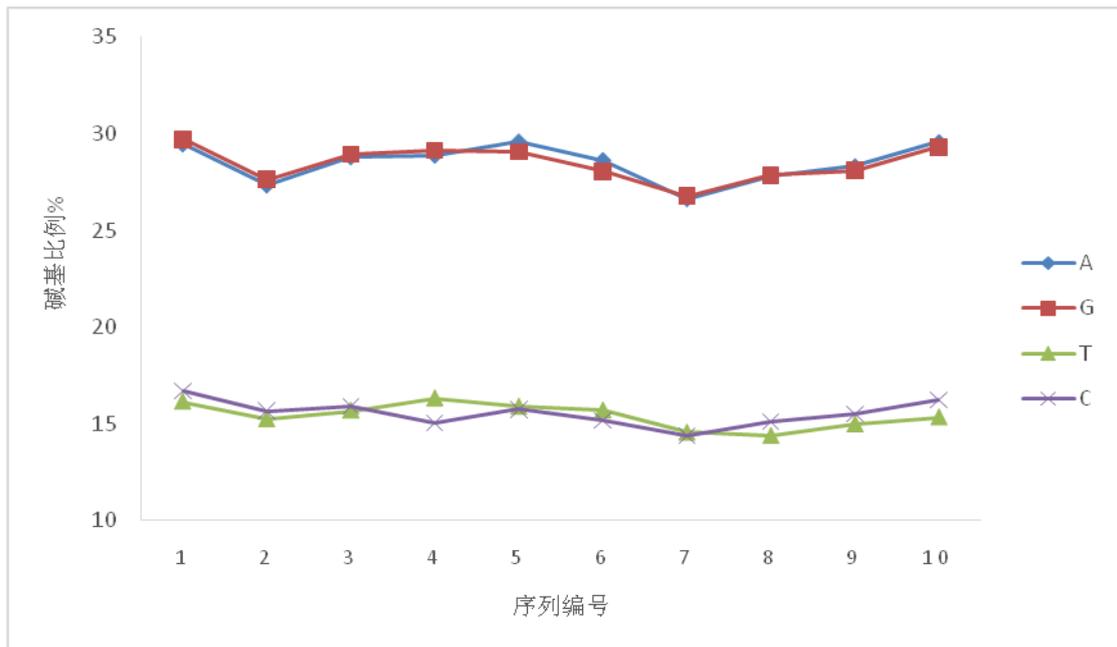


图 2 所选序列中各碱基所占比例

3.2 SSR 位点筛选

本研究从上述所选 10 序列中分布筛选出 1770 个 SSR 位点，每条序列中均含有丰富的 SSR 位点（见表 1）。高于玉米（1.5%）、水稻（4.7%）、大麦（3.4%）、高粱（3.6%）和小麦（3.2%）等作物^[12]，这种差异可能与 SSR 搜索标准、数据库大小和物种等有关。其中出现频率较高的是二核苷酸的重复基元 AT/AT 和 AG/CT，GC 重复基元非常少，仅有 1 个 GC 重复基元。而大豆、番茄、棉花、

杨树、拟南芥、大麦等作物中 SSR 的主要类型是三核苷酸重复^[13-14]。GC 重复基元在多数植物中很难见到，出现频率都很低^[15-17]。

表 2 各长度 SSR 位点所占比例

基序长度	数量	比例
1	417	23.56%
2	1104	62.37%
3	222	12.54%
4	13	0.73%
5	5	0.28%
6	9	0.51%

3.3 SSR 位点引物设计

通过 primer3 我们对其中的 4 条序列设计了引物，结果如图:

ID	FORWARD PRIMER1 (5'-3')	REVERSE PRIMER1 (5'-3')
scaffold1	CAAAATTTTCACATATCACCTCA	CATTCGGCTGAATTTGTTTTT
scaffold2	AAAAGCCCCCGTAAGAAAAA	CCTCAAATCTCAAGCCATCA
scaffold2	TTGTCCTTCGTTGCAGATTG	GTCGTTGCCTGAAAAATCGT
scaffold2	ACGATTTTTCAGGCAACGAC	GAATGGACCTATTTTGGGGG

4 讨论

从这 10 条选出的最长序列可以看出，三七转录组 SSR 不但出现频率高，而且类型丰富；从多态性潜能的角度考虑，搜索到的这些 SSR 也具有较高的可用性。

SSR 引物在不同物种中具有一定的通用性，其通用程度取决与 SSR 侧翼序列的保守程度及 SSR 进化的稳定性。SSR 引物的通用性对于增加基因组信息较少的物种的标记数目，提高标记的利用价值具有显著的作用。

物种的 SSR 分子标记不仅同科不同属间具有转移扩增能力，而且在遗传背景差异大的植物中也同样有转移扩增的能力。随着分子标记技术的发展，SSR 标记技术被广泛地应用于物种遗传多样性及种质鉴定研究中。而遗传多样性是生物多样性的的重要组成部分，也是物种对环境变化适应能力的重要体现，研究遗传多

样性可以揭示物种或种群的进化历史（起源的时间、地点和方式），预测种源的适应性及估算基因资源的分布，推动保护生物学研究，对种质资源的收集、保存、评价和利用具有十分重要的意义^[18]。

而且 SSR 标记位点能被整合到现有的植物遗传图谱中构建分子标记遗传连锁图。分子标记遗传连锁图表示各标记所对应的 DNA 片段在染色体上的相对位置，是分子标记运用于作物遗传育种的基础。分子标记连锁图是进行基因定位、基因克隆、实现辅助选择育种的技术平台，在遗传学、功能基因组学以及遗传育种等领域已显示出了十分重要的作用。基于 SSR 标记技术的连锁图谱的构建，为数量性状基因定位克隆提供了物质基础。

测序技术的进步及各主要作物的遗传图谱的成功构建，为基因的精细定位和图位克隆奠定了坚实的基础。所以本研究的结果为进一步开发新的三七功能基因 SSR 标记奠定了基础，这种标记的建立对于加速三七功能基因资源的开发利用、丰富其分子标记类型、遗传资源评价、绘制遗传图谱、实现特定性状的辅助选择和进行比较基因组学研究都具有重要的意义。

参考文献:

- [1] 潘跃芝, 张亦弛, 龚洵, 李富生. 4 种人参属植物基因组大小的测定[J]. 植物分类与资源学报, 2014, 36(02):233-236.
- [2] 黄福量. 三七的正伪品鉴别 [J]. 医药前沿, 2012, 2(17):321.
- [3] 李晓琳, 邵爱娟, 陈敏等. 三七种子研究进展 [J]. 种子, 2011, 30(7):63-65.
- [4] 王振峰, 高云涛, 张文斌等. 不同生长年限三七中总皂苷含量的变化特征 [J]. 安徽农业科学, 2012, 40(15):8458-8459, 8463.
- [5] Liu T, Zhu S, Fu L, et al. Development and characterization of 1827 expressed Sequence tag-derived simple sequence repeat markers for ramie (*Boehmeria nivea* L. Gaud) [J]. PLoS One, 2013, 8(4): e60346.
- [6] 黄映萍. DNA 分子标记研究进展[J]. 中山大学研究生学刊: 自然科学医学版, 2010, 31(2): 27-36.
- [7] 闫秋良. 基于生物信息学方法从牛和绵羊表达序列标签中筛选 SSR 标记的初步研究[D]. 西安:西北农林科技大学, 2007.
- [8] 朴红梅, 王玉民, 刘宪虎, 王景余, 金成海. 简单重复序列的研究与应用[J]. 吉林农业科学, 2004, (06):11-15.
- [9] Dan Zhang, Wei Li, En-hua Xia, Qun-jie Zhang, Yuan Liu, Yun Zhang, Yan Tong, Yuan Zhao, Yong-chao Niu, Jia-huan Xu, Li-zhi Gao. The Medicinal Herb *Panax notoginseng* Genome

Provides Insights into Ginsenoside Biosynthesis and Genome Evolution[J]. *Molecular Plant*,2017,10(6):.

[10] 李翠婷,张广辉,马春花,孟珍贵,陈军文,杨生超. 野三七转录组中 SSR 位点信息分析及其多态性研究[J]. *中草药*, 2014, 45(10):1468-1472.

[11] 王东,曹玲亚,高建平. 党参转录组中 SSR 位点信息分析[J]. *中草药*, 2014, 45(16):2390-2394.

[12] Kantety R V, Rota M L, Matthews D E, et al. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat [J]. *Plant Mol Biol*, 2002, 48: 501-510.

[13] Cardle L, Ramsay L, Mibourne D, et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. *Genetics*, 2000, 156: 847-854.

[14] Varshney R K, Graner A, Sorrells M. Genic mirosatellite markers in plants: features and applications [J]. *Trends Biotechnol*, 2005, 23(1): 48-55.

[15] Lidan S, Weiru Y, Qixiang Z, et al. Genome-wide characterization and linkage mapping of simple sequence repeats in Mei (*Prunus mume* Sieb. et Zucc.) [J]. *PLoS One*, 2013, 8(3): e59562.

[16] Cardle L, Ramsay L, Milbourne D, et al. Computational and experimental characterization of rhyssically clutered simple sequence repeats in plants [J]. *Genetics*, 2000, 156: 847-854.

[17] Marshall H, Wayne W X, Brian W K, et al. Identification of differential gene expression in *Brassica rapa* Nectaries through expressed sequence tag analysis [J]. *PLoS One*, 2010, 5(1): e8782.

[18] 张炜. 利用小麦 SSR 研究其近缘物种的遗传多样性及其微卫星序列的进化[J]. 四川农业大学, 硕士学位论文, 2008.