

# Large-Scale Development and Character Analysis of SSR Markers and Database Build in Brassicaceae

Shuai Yang<sup>1,2\*</sup>, Hui Li<sup>2</sup>, Xin Hou<sup>1</sup>, Li Zhang<sup>1\*</sup>

<sup>1</sup>College of Plant Protection, Shandong Agricultural University, Taian Shandong

<sup>2</sup>College of Life Sciences, Jinan University, Jinan Shandong

Email: \*longtoutengs@163.com, \*lilizhang324@163.com

Received: May 4<sup>th</sup>, 2017; accepted: May 21<sup>st</sup>, 2017; published: May 24<sup>th</sup>, 2017

## Abstract

Brassicaceae is an important family in the plant kingdom. The Simple Sequence Repeats (SSRs) play a vital role in the study of Brassicaceae. By using 13 known sequenced Brassicaceae species with bioinformatics and comparative genomics methods, a total of 1,786,619 SSR loci and 1,919,464 pair of primers have been developed. The results show that the SSRs are widely distributed in the Brassicaceae species' genomes, 1 - 3 bases duplication have a high ratio among these genomes and gene sequences, AT/TA repeats units have a high numbers in all of the 2 base duplication. In addition, 435,414 specific SSR primers could be used to analyze the correlation between the species of Brassicaceae. 11 pairs of universal primers' developed shows that there exist some consistent base fragments and could be amplified across different species. In this study, we constructed the world's first SSR molecular marker database platform (BSSRD, Brassicaceae Simple Sequence Repeats Database <http://biodb.sdau.edu.cn/BSSRD>) which will play an important role in the construction of genetic map, gene mapping and genetic breeding of Brassicaceae.

## Keywords

Brassicaceae, SSR, Specific Primers, Universal Primers, Database

# 大规模开发及特性分析十字花科SSR分子标记及其数据库的构建

杨 帅<sup>1,2\*</sup>, 李 慧<sup>2</sup>, 侯 欣<sup>1</sup>, 张 丽<sup>1\*</sup>

<sup>1</sup>山东农业大学植物保护学院, 山东 泰安

<sup>2</sup>济南大学生命科学院, 山东 济南

\*通讯作者。

Email: \*longtoutengs@163.com, \*lilizhang324@163.com

收稿日期: 2017年5月4日; 录用日期: 2017年5月21日; 发布日期: 2017年5月24日

## 摘要

十字花科在植物界中是一种极具价值的科, 简单重复序列(Simple Sequence Repeats, SSR)在十字花科的研究中发挥着及其重要的作用。本研究利用已知13个十字花科物种的基因组, 借助生物信息学和比较基因组学的方法, 获得了1,786,619个SSR位点及1,919,464对SSR引物。结果显示SSR位点广泛分布于十字花科物种的基因组中, 其中1~3单元的重复在基因组和基因序列中占有率较高, 二单元重复中的AT/TA重复单元的数目占总数目的大部分。利用435,414对物种特异性SSR, 可以进行十字花科物种关联分析。11对通用引物的开发, 说明了十字花科物种内存在一致性区段, 可以进行引物跨物种扩增。本研究构建了世界首个十字花科SSR分子标记数据库平台(BSSRD, Brassicaceae Simple Sequence Repeats Database <http://biodb.sdau.edu.cn/BSSRD>), 该平台将会在以后十字花科植物的遗传图谱的构建, 基因定位和遗传育种中发挥重要的作用。

## 关键词

十字花科, SSR, 特异引物, 通用引物, 数据库

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在植物界中, 十字花科是一种非常重要的科, 其拥有 375 个属, 3200 个种, 广泛分布于世界各地, 其中在中国就有 95 个属, 约 411 余种[1]。十字花科植物的种类繁多, 既有富含食用价值的蔬菜, 如白菜、甘蓝等植物, 又有极具观赏价值的紫罗兰等植物, 此外还有极高经济价值的油菜, 亚麻荠等油料作物, 是所有开花植物中最有经济价值的科之一[2]。更重要的是, 由于十字花科植物拟南芥的生长周期短, 结实多, 广泛生长等优点, 成为了植物界内研究最早的模式物种之一, 在遗传育种等分子生物学研究中扮演着非常重要的角色[3]。近年来, 在基因水平上通过一些分子标记来提高十字花科对人的营养利用价值, 经济价值和科研价值的研究成为了一个热点[4] [5]。

简单重复序列(simple sequence repeats, SSR), 又称为微卫星 DNA 序列, 是广泛存在于整个基因组中的一类分子标记[6]。SSR 一般定义为 1~6 个碱基串联重复组成几十 bp 的序列, 其两端的序列是相对保守的单拷贝序列。SSR 具有的广泛分布, 共显性高, 重复性好, 多态性高, 易于检测等生物学优点使得其成为一种常用的分子标记, 并广泛应用于植物遗传多样性分析、遗传图谱构建、基因定位和分子标记辅助育种等领域[7] [8] [9]。但是据我们查询所知, 现在对于十字花科植物 SSR 的研究仅局限于单个物种, 并且仅利用表达序列标签(Express Sequence Tags, EST)序列检测的 SSR 并不完整[10] [11]。

随着现在第二代测序技术的发展成熟以及测序价格的降低, 到目前为止, 十字花科植物中阿拉伯岩芥菜[12], 玉山筷子芥[13], 拟南芥[14], 油菜[15], 甘蓝[16], 芜菁[17], 亚麻荠[18], 芥菜[12], 亚拉巴马水芹[12], 盐芥近缘种[19] [20], 水蒜芥[12], 小盐芥[21]等 13 个物种已经完成了全基因组测序并进行

了基因的预测, 这给本研究提供了丰富的数据基础。现在分子生物学技术以及生物信息学技术的发展, 也给本研究提供了后续个性化分析技术和平台。本研究将基于十字花科物种的基因组信息, 进行 SSR 标记标记的大规模开发和相关数据库的构建。

## 2. 材料与方法

### 2.1. 实验材料

从十字花科基因组数据库 BRAD (<http://brassicadb.org/brad/index.php>) [22] 以及美国生物信息中心 NCBI (<https://www.ncbi.nlm.nih.gov/>) 等公共数据库中下载到十字花科物种的全基因组序列以及其基因序列, 统计各个物种的详细信息(见表 1)。

### 2.2. 方法技术

#### 2.2.1. SSR 位点扫描

利用 Micro Satellite identification tool-MISA [23] 软件包中的 Perl 脚本进行批量扫描十字花科物种的 SSR 位点, 本研究设定是否为 SSR 的标准为: 1) 一单元至少十次重复, 二单元至少六次重复, 三单元至少五次重复, 四单元至少五次重复, 五单元至少五次重复, 六单元至少五次重复。2) 两个 SSR 之间距离小于 100 bp 时组合为一个复合 SSR。

#### 2.2.2. SSR 位点统计

利用 Perl 脚本对扫描得到的 SSR 做了以下分类统计: 1) 各个基因组和基因序列 SSR 位点数目; 2) 各个物种的 1~6 重复单元数目; 3) 各个物种不同的重复单元前两位的重复片段数目。

#### 2.2.3. 序列截取

计算每个 SSR 位点在基因组序列上的物理位置, 利用 SSR 位置信息从每个十字花科物种的基因组上截取 SSR 起始终止位点前后各 60 bp 的序列, 并将截取的序列中 N 值小于 10% 的序列筛选出来进行下一步引物的设计。

**Table 1.** The detail information of Brassicaceae species

**表 1.** 十字花科物种详细信息表

物种名	拉丁名	版本号	基因组总长(Mb)	基因总长(Mb)	发表文献	发布日期
阿拉伯岩芥菜	<i>Aethionema arabicum</i>	V1.0	199.4	43.5	<i>Nature Genetics</i> [12]	2013.06
玉山筷子芥	<i>Arabidopsis lyrata</i>	V1.0	206.7	39	<i>Nature Genetics</i> [13]	2011.04
拟南芥	<i>Arabidopsis thaliana</i>	TAIR10	119.7	32.1	<i>Nature</i> [14]	2000.12
油菜	<i>Brassica napus</i>	V5.0	850.3	100	<i>Science</i> [15]	2014.08
甘蓝	<i>Brassica oleracea</i>	V1.1	385	37.4	<i>Nature Communications</i> [16]	2014.05
芜菁	<i>Brassica rapa</i>	V1.5	284.9	49.4	<i>Nature Genetics</i> [17]	2011.08
亚麻芥	<i>Camelina sativa</i>	V2.0	641.5	113	<i>Nature Communications</i> [18]	2014.04
芥菜	<i>Capsella rubella</i>	V1.0	134.8	35.3	<i>Nature Genetics</i> [12]	2013.06
亚拉巴马水芹	<i>Leavenworthia alabamica</i>	V1.0	174.2	45.4	<i>Nature Genetics</i> [12]	2013.06
盐芥近缘种 1	<i>Schrenkiella parvula</i>	V7.0	137.3	34.8	<i>Mitochondrial DNA</i> [17]	2015.08
水蒜芥	<i>Sisymbrium irio</i>	V1.0	259.5	55	<i>Nature Genetics</i> [12]	2013.06
盐芥近缘种 2	<i>Thellungiella halophila</i>	V1.0	243.1	35.7	<i>Plant Physiology</i> [20]	2010.11
小盐芥	<i>Thellungiella salsuginea</i>	V2.0	233.7	34.1	<i>PNAS</i> [21]	2012.07

#### 2.2.4. 引物设计

将上述截取的序列利用 Emboss (<http://emboss.open-bio.org/>) 软件包中的 e-primer3 (<http://emboss.bioinformatics.nl/cgi-bin/emboss/eprimer3>) 进行引物的设计, 设置参数为: MinTM = 55, MaxTM = 80, MinGC = 40, MaxTM = 60, Product Size = 100~300, numberturn = 5。引物设计完成后将引物提取出并将重复的引物去除。

#### 2.2.5. 引物验证

利用 Electronic PCR-e-PCR (<https://www.ncbi.nlm.nih.gov/tools/epcr/>) [24] 将各个物种所有的引物在其各自的基因组上进行电子 PCR 模拟扩增, 将能够在其物种上完整扩增的引物提取出来作为十字花科 SSR 的储备引物。

#### 2.2.6. 特异性扩增引物

利用 Shell 语言提取在各个基因组中能够进行特异性扩增的引物, 对所有特异性扩增引物进行数目和引物在基因组中的分布统计。

#### 2.2.7. 通用扩增引物

将所有已验证过的 SSR 扩增的引物与所有的物种的基因组进行电子模拟 PCR 扩增, 通过 Perl 以及 Shell 语言将在所有物种中均能够扩增的引物提取出来作为十字花科物种的通用 SSR 引物并提取出。

#### 2.2.8. BSSRD 数据库的构建

使用 LAMP (Linux, Apache, MySQL, PHP/Perl) 框架进行数据库的构建。首先通过 Dreamweaver 2014CC 软件构建用户使用的前台界面, 将所有的 SSR 位点和引物信息导入 MySQL 后台数据库中, 通过编写 PHP 和 Perl 脚本实现前台界面和后台数据的交互构建出十字花科 SSR 数据库 BSSRD (<http://biodb.sdau.edu.cn/BSSRD>)。SSR 开发以及引物设计技术路线见图 1。

### 3. 结果与分析

#### 3.1. SSR 位点数目及特征

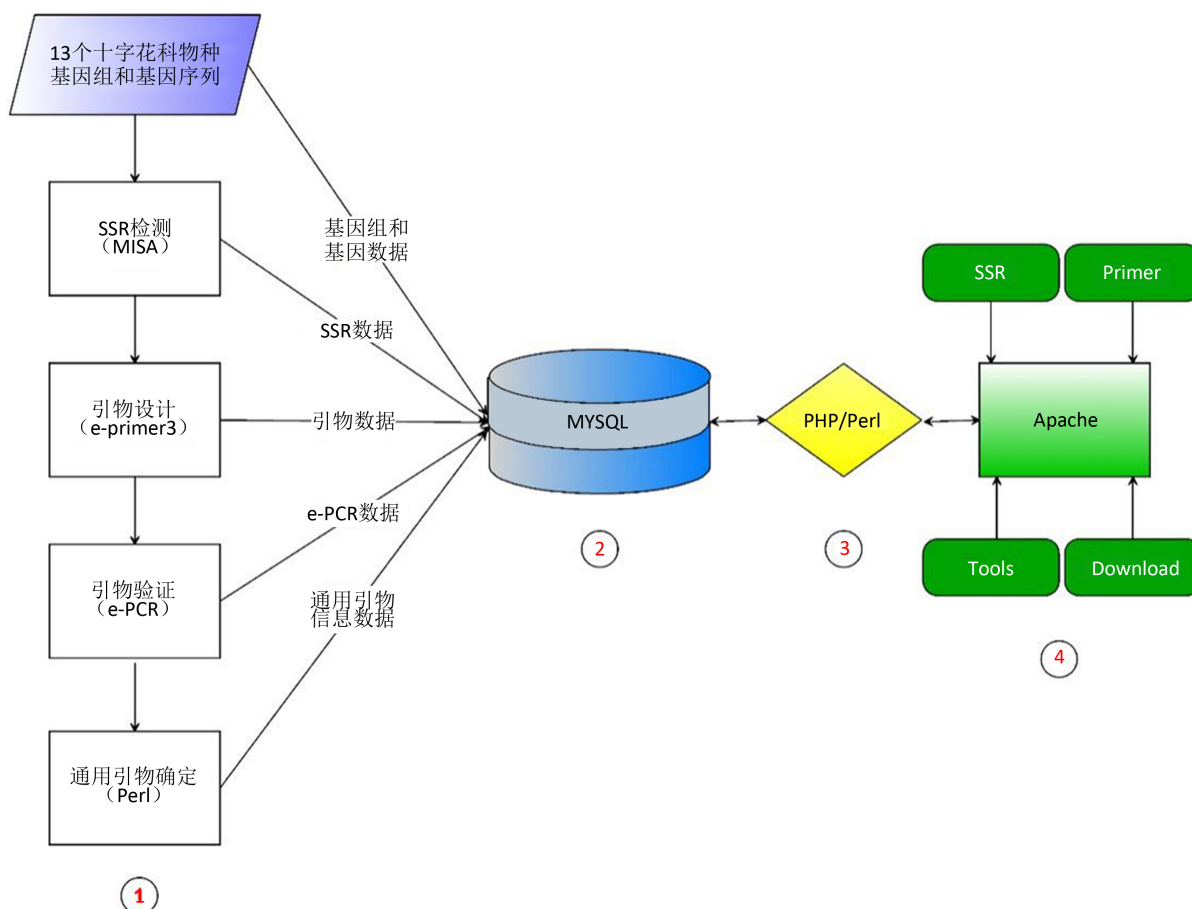
分别对十字花科物种进行了全基因组序列和基因序列的 SSR 扫描, 在全基因组中共发现了 1,786,619 个 SSR 位点, 在基因序列中共发现了 63,915 个 SSR 位点, 占总基因组序列 SSR 的 3.58%。在十字花科物种的全基因组序列中, 平均每 100,000,000 个碱基中都会存在 324~596.8 个 SSR。拥有两个较大基因组的物种 *B. napus* 和 *C. sativa* 其 Genome-SSR 和 Gene-SSR 也居于前两位。各物种 Gene-SSR 占 Genome-SSR 的比值在 (*B. oleracea*) 1.86%~(*A. lyrata*) 6.14% 之间趋于相对的平稳 (见表 2)。

不同 SSR 重复单元具有较大的偏向性, 在全基因组序列和基因序列中的 SSR 重复单元也存在差异。在全基因组序列中一单元重复的 SSR 数目占了总数目的一半以上为 68.4%, 其次是二单元重复和三单元重复的 23.32% 和 7.59% (见图 2(a)), 然而在基因序列中的三单元重复的 SSR 数目高达 79.47%, 其次为一单元重复和二单元重复的 11.12% 和 8.47% (见图 2(b))。

对不同十字花科物种基因组序列的不同 SSR 重复片段分布频率做了一个统计, 将每个物种每个重复单元数目前两位的重复片段提取出来。各个物种各个重复单元类型的前两位 SSR 片段基本上是相同的, 其中二碱基重复的 AT/TA, 三碱基重复的 GAA 及 TCT 和四碱基类型的 AAAT/TAAA 在每个物种中数目都是非常高。在三碱基重复单元中大部分物种数目前两位为 GAA 和 TCT, 然而 *S. parvula* 物种为 GAA 和 TTC (见表 3)。

#### 3.2. SSR 引物设计及验证

利用十字花科物种的全基因组序列以及基因序列扫描到的 SSR 进行批量的设计了引物并进行了电子



**Figure 1.** The route of SSR development and primer design. 1) SSR development, primer design and confirm of universal primers; 2) Put all the data to MySQL; 3) The connection by PHP and Perl scripts; 4) Foreground display by Apache

**图 1.** SSR 开发以及引物设计技术路线图。1) SSR 开发，引物设计以及特意引物和通用引物的确定；2) 将所有数据导入 MySQL 数据库中；3) PHP 和 Perl 语言连接前台和后台；4) Apache 前台页面展示

**Table 2.** The statistics of Genome-SSR and Gene-SSR in Brassicaceae

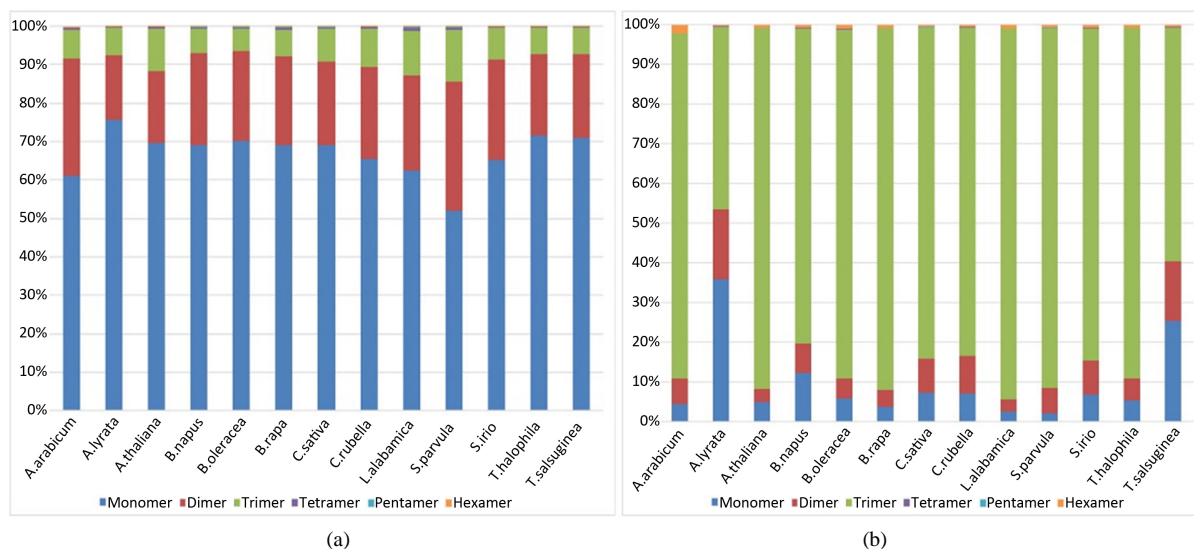
**表 2.** 十字花科物种 Genome-SSR 和 Gene-SSR 数目表

Item	A. <i>arabicum</i>	A. <i>lyrata</i>	A. <i>thaliana</i>	B. <i>napus</i>	B. <i>oleracea</i>	B. <i>rapa</i>	C. <i>sativa</i>	C. <i>rubella</i>	L. <i>alabamica</i>	S. <i>parvula</i>	S. <i>irio</i>	T. <i>halophila</i>	T. <i>salsuginea</i>	Total
Genome-SSRs	94,773	103,106	50,092	362,529	174,199	145,843	367,769	80,450	61,345	44,482	85,325	110,234	106,472	1,786,619
Gene-SSRs	3,042	6,335	2,647	9,743	3,233	4,568	13,008	3,549	3,167	2,724	3,856	3,308	4,735	63,915
Ratio	3.21%	6.14%	5.28%	2.69%	1.86%	3.13%	3.54%	4.41%	5.16%	6.12%	4.52%	3.00%	4.45%	3.58%
No. of SSRs (Mb)	475.3	498.8	418.5	426.4	452.5	511.9	573.3	596.8	352.2	324.0	328.8	453.5	455.6	-

模拟 PCR 扩增验证，最终，在全基因组序列中共有 1,919,464 对引物被确定，同时在基因序列中有 54,619 对引物被确定。在不同的物种的全基因组序列中引物的密度从(*L. alabamica*) 112.94 对/Mb 到(*T. halophila*) 1177.68 对/Mb 不等，在基因序列中为(*A. arabicum*) 42.65 对/Mb 到(*C. sativa*) 139.69 对/Mb 不等(见表 4)。

### 3.3. 特异性引物和通用引物

在全基因组序列和基因序列中分别找到 435,414 和 30,141 对特异性扩增引物，同时 11 对能在十字花科物种中扩增的 SSR 引物被确定(Universal-primer-01~ Universal-primer-01)，其中有三条来自于 *C. rubella*，



**Figure 2.** The distribution of different repeat unit in Genome-SSR and Gene-SSR

**图 2.** Genome-SSR 和 Gene-SSR 不同单元重复数目分布图

**Table 3.** The statistics of different SSR top two repeat units

**表 3.** 不同 SSR 重复单元数目前两位的重复片段统计

Species	Dimer		Trimer		Tetramer		Pentamer		Hexamer	
<i>A.arabicum</i>	AT(11,437)	TA(8993)	GAA(464)	TCT(386)	AAAT(51)	TTTA(44)	AAAAT(14)	TTTTA(8)	TAAAAT(7)	ACATAC(4)
<i>A.lyrata</i>	AT(5993)	TA(4539)	GAA(730)	TCT(672)	AAAT(31)	AAAG(19)	GAAGA(3)	TTCGG(3)	TTCTTG(3)	AGAAGG(2)
<i>A.thaliana</i>	AT(2641)	TA(2054)	GAA(514)	TCT(475)	TTTG(13)	AAAG(8)	AAACA(3)	AGAGA(3)	CGACCA(3)	CAAAC(2)
<i>B.napus</i>	AT(28,081)	TA(19,905)	GAA(1435)	TCT(1422)	AAAT(148)	TTTA(96)	TTCGG(15)	TTTTG(15)	TGGGCT(19)	CCCAAG(17)
<i>B.oleracea</i>	AT(13,474)	TA(9200)	TCT(642)	GAA(607)	AAAT(77)	TATT(44)	TATTT(34)	AAAAT(34)	CCCAAG(7)	CTCTCC(5)
<i>B.rapa</i>	AT(10,743)	TA(8039)	GAA(678)	TCT(610)	AAAT(184)	TATT(161)	TTTTA(10)	TTCGG(6)	CCCAAG(7)	TTTTGT(4)
<i>C.sativa</i>	AT(25,786)	TA(17,597)	GAA(2528)	TCT(2368)	AAAT(87)	TTTC(63)	TTTTG(80)	ACAAA(61)	GAGGTA(20)	GGGTTA(9)
<i>C.rubella</i>	AT(5423)	TA(4218)	GAA(696)	TCT(646)	AAAT(55)	ATTT(54)	TTTTG(7)	AATAA(5)	CAAAA(2)	AAAAAC(1)
<i>L.alabamica</i>	AT(6477)	TA(5003)	GAA(510)	TCT(438)	TTTA(55)	AAAT(53)	AAAAT(13)	TTTTA(10)	TATTTA(3)	TTTTAA(3)
<i>S.parvula</i>	AT(4463)	TA(3132)	GAA(610)	TTC(537)	TTTA(20)	AAAT(18)	AAAAG(5)	AAATA(3)	TCACTC(4)	TTTTGT(4)
<i>S.irio</i>	AT(6953)	TA(4700)	GAA(578)	TCT(576)	AAAT(29)	ATTT(17)	TTTTA(6)	AAAAT(5)	CATAAA(2)	GGTGCT(2)
<i>T.halophila</i>	AT(7128)	TA(5283)	GAA(698)	TCT(614)	AAAT(26)	AAAG(21)	AAAAT(7)	TTTTA(4)	AAAAAT(2)	CCCGAG(2)
<i>T.salsuginea</i>	AT(6960)	TA(5015)	GAA(685)	TCT(590)	TTTA(20)	AAAT(19)	AAAAT(7)	GAAAA(4)	AAAAAC(2)	AGAGAA(2)

*S. irio* 和 *T. halophila* 分别占用两条，其余来自于 *B. oleracea*, *A. arabicum*, *S. parvula*, *T. salsuginea* 分别占一条(见表 5)。

### 3.4. BSSRD 数据库的构建和应用

基于实验中的所有的数据包括物种的序列信息, SSR 位点信息, e-PCR 信息以及各类引物信息, 构建了十字花科 SSR 数据库(BSSRD, Brassicaceae Simple Sequence Repeats Database <http://biodb.sdau.edu.cn/BSSRD>), 该数据库包括 SSR, Primer, Tools 三个部分, 用户可以在 SSR 和 Primer 界面中利用 SSR ID 和 Primer ID 等个性化搜索有目的的进行 SSR 及其引物的检索, 同时在 Tools 界面能够使用一些简单的工具, 包括界面化的

**Table 4.** Statistics of the SSR primers in Brassicaceae  
**表 4.** 十字花科物种中 SSR 引物的统计

Species	No. of Genome Primes	Genome Primer (Mb)	No. of Genome Primes	Gene Primer (Mb)
<i>A. arabicum</i>	43,217	216.74	1915	42.65
<i>A. lyrata</i>	92,470	447.36	3131	80.49
<i>A. thaliana</i>	25,620	214.04	1899	56.86
<i>B. napus</i>	586,594	689.87	9908	97.91
<i>B. oleracea</i>	210,585	546.97	2584	70.22
<i>B. rapa</i>	71,954	252.56	3659	76.07
<i>C. sativa</i>	403,231	628.58	16,162	139.69
<i>C. rubella</i>	33,039	245.10	2404	67.34
<i>L. alabamica</i>	21,417	122.94	2304	49.13
<i>S. parvula</i>	19,585	142.64	2018	55.75
<i>S. irio</i>	37,828	145.77	3072	54.18
<i>T. halophila</i>	286,293	1177.68	2650	73.41
<i>T. salsuginea</i>	87,631	374.97	2913	82.06
Average	147651.08	400.40	4201.46	72.75

**Table 5.** The detail information of 11 pair of universal primers  
**表 5.** 11 对通用引物信息表

编号	引物名称	前引物	后引物
Universal-primer-01	Bol-C08-17368332-17368462	GACACAGCTGCTTTCAAAAA	ATTCAATGTCTCCCCACCTG
Universal-primer-02	Cru-scaffold_1-17431124-17431282	GAGTGGCAGTGAGCTGTCAA	AGGGACAAACAGCGTCAGAG
Universal-primer-03	Aar-AA_scaffold6649-155781-155912	TGACATGGAAGATTCGGTTC	ACTCGTCAAACATTGCTTCTTG
Universal-primer-04	Spa-Sp6-18047624-18047764	AGGGCACGTGCATTACTAGC	AGATGGGCACATGAAGAACT
Universal-primer-05	Tha-scaffold_22-1282063-1282197	CATTCTTCTCTTTGGCCTTT	GCACCTGTGTCCACTTCGTA
Universal-primer-06	Tsa-ch1-01-1287940-1288076	CCTTGTACAAACGCTCCTTCA	TCCTGTCTTTGATTAGGACAG
Universal-primer-07	Sir-SI_scaffold2431-6850-6980	CGAGCGCCTAAAGTTAAAAA	ATGACAGAGCTGCGGTGAC
Universal-primer-08	Cru-scaffold_1-429966-430108	TATCATTCATGGGCCCTAC	TGGTGGCTTGAGATTCTTGA
Universal-primer-09	Cru-scaffold_6-26067-26247	GGAAGATGGGCACATGAAGA	TTGAGCAAGATGGAGAAGCA
Universal-primer-10	Sir-SI_scaffold2301-134273-134433	GAGTGGCAGTGAGCTGTCAA	AGGGACAAACAGCGTCAGAG
Universal-primer-11	Tsa-ch4-82-2961373-2961507	CATTCTTCTCTTTGGCCTTT	GCACCTGTGTCCACTTCGTA

SSR 的检测和 e-primer3 的在线使用(见图 3)。

#### 4. 讨论

由 SSR 检测结果可以看出, SSR 广泛分布于各个十字花科物种基因组序列中, 相比较其他分子标记来说, SSR 是一种研究基因组的理想的分子标记[25]。从每个物种的 Gene-SSR 和 Genome-SSR 的比值来看, 我们发现十字花科的物种拥有相对比较接近的 SSR 比例, 这也从侧面揭示了十字花科物种 SSR 分布的广泛性和统一性。

**BSSRD** Welcome to the Brassicaceae Simple Sequence Repeats Database

Home  
Methodology  
SSR  
Primer  
Tools  
Download  
About

Version 1.0.0

LINKS : BRAD NCBI TAIR PlantGDB JGI EMBL DDBJ

Copyright©2017 : Shandong Agricultural University PostCode: 271018

**Figure 3.** The construction and use of BSSRD

**图 3.** BSSRD 的构建以及使用

通过对 SSR 各个碱基重复的比例以及 SSR 片段统计可以看出，一单元，二单元和三单元重复占了基因组和基因序列的大部分，并且 AT/TA 以及 GAA/TCT 所占比例非常高，在以后的十字花科遗传图谱的构建过程中，我们可以利用相对保守稳定的二单元和三单元重复[26]。

在各个十字花科物种中特异性扩增的引物对以后十字花科植物的基因的定位，遗传育种以及遗传图谱的构建能得到广泛的参考和应用[8]。同时 11 对通用引物被发掘出来也为以后十字花科同源物种和同源基因的鉴定提供了数据基础。

最后，基于本研究的所有数据，我们构建了十字花科 SSR 数据库 BSSRD 储存我们所得到的所有 SSR 位点以及设计的各种引物信息来用作非商业性质的科学研究。在该数据库中在线的 SSR 扫描工具以及引物设计的交互式界面也给科研工作者提供了很大的便利。当有新的基因组以及基因数据产生时我们会及时的去更新该数据库。我们相信这个数据库在以后的关于十字花科的遗传图谱的构建，基因定位，基因组组装和遗传育种等研究中将发挥重要的作用。

## 5. 结论

利用现有的十字花科物种的全基因组序列和基因序列，批量扫描了其中的 SSR 位点，进行了各项统计分析，对这些位点进行了引物设计以及特异性引物和通用引物的查找，最后我们构建了 BSSRD 十字花



科数据库来储存这些数据。

## 基金项目

本研究得到了山东省自然科学基金(ZR2015CL006)支持。

## 参考文献 (References)

- [1] 中国科学院中国植物志委员会, 周太炎, 郭荣麟, 蓝永珍. 中国植物志: 十字花科[J]. 1987.
- [2] 陈秋芳, 贾宏汝. 十字花科植物的研究价值及开发利用前景[J]. 安徽农业科学, 2007, 35(34): 11183-11185.
- [3] 曹仪植. 拟南芥[M]. 北京: 高等教育出版社, 2004.
- [4] 蔡光勤, 杨庆勇, 杨倩, 赵振兴, 陈浩, 吴健, 范楚川, 周永明. 拟南芥与芸薹属比较基因组学分析鉴定甘蓝型油菜千粒重 QTL 的候选基因[J]. 中国作物学会油料作物专业委员会第七次会员代表大会暨学术年会, 2013.
- [5] 李丽, 何伟明, 马连平, 刘庞源, 徐海明, 徐家柄, 郑晓鹰. 用 EST-SSR 分子标记技术构建大白菜核心种质及其指纹图谱库[J]. 基因组学与应用生物学, 2009, 28(1): 76-88.
- [6] Litt, M. and Luty, J.A. (1989) A Hypervariable Microsatellite Revealed by *in Vitro* Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. *American Journal of Human Genetics*, **44**, 397-401.
- [7] 程小毛, 黄晓霞. SSR 标记开发及其在植物中的应用[J]. 中国农学通报, 2011, 27(5): 304-307.
- [8] 罗冉, 吴委林, 张旸, 李玉花. SSR 分子标记在作物遗传育种中的应用[J]. 基因组学与应用生物学, 2010, 29(1): 137-143.
- [9] 唐荣华, 张君诚, 吴为人. SSR 分子标记的开发技术研究进展[J]. 西南农业学报, 2002, 15(4): 106-109.
- [10] 李小白, 张明龙, 崔海瑞. 油菜 EST-SSR 标记的建立[J]. *Journal of Molecular Cell Biology*, 2007, 40(2): 137-144.
- [11] 忻雅, 崔海瑞, 卢美贞, 姚艳玲, 金基强, 林容杓, 崔水莲. 白菜 EST-SSR 信息分析与标记的建立[J]. 园艺学报, 2006, 33(3): 549-554.
- [12] Haudry, A., Platts, A.E., Vello, E., Hoehn, D.R., Leclercq, M., Williamson, R.J., Forczek, E., Jolylopez, Z., Steffen, J.G. and Hazzouri, K.M. (2013) An Atlas of over 90,000 Conserved Noncoding Sequences Provides Insight into Crucifer Regulatory Regions. *Nature Genetics*, **45**, 891-898. <https://doi.org/10.1038/ng.2684>
- [13] Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J. and Gundlach, H. (2011) The *Arabidopsis lyrata* Genome Sequence and the Basis of Rapid Genome Size Change. *Nature Genetics*, **43**, 476. <https://doi.org/10.1038/ng.807>
- [14] Initiative, A.G. (2000) Analysis of the Genome Sequence of the Flowering Plant *Arabidopsis thaliana*. *Nature*, **408**, 796. <https://doi.org/10.1038/35048692>
- [15] Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C. and Samans, B. (2014) Early Allopolyploid Evolution in the Post-Neolithic *Brassica napus* Oilseed Genome. *Science*, **345**, 950-953. <https://doi.org/10.1126/science.1253435>
- [16] Yang, X. (2014) The *Brassica oleracea* Genome Reveals the Asymmetrical Evolution of Polyploid Genomes. *Nature Communications*, **5**, 3930.
- [17] Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I. and Cheng, F. (2011) The Genome of the Mesopolyploid Crop Species *Brassica rapa*. *Nature Genetics*, **43**, 1035-1039. <https://doi.org/10.1038/ng.919>
- [18] Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W.E., Tuteja, R., Spillane, C., Robinson, S.J., Links, M.G. and Clarke, C. (2014) The Emerging Biofuel Crop *Camelina sativa* Retains a Highly Undifferentiated Hexaploid Genome Structure. *Nature Communications*, **5**, 3706. <https://doi.org/10.1038/ncomms4706>
- [19] He, Q., Hao, G., Wang, X., Bi, H., Li, Y., Guo, X. and Ma, T. (2016) The Complete Chloroplast Genome of *Schrenkiella parvula* (Brassicaceae). *Mitochondrial DNA*, **27**, 1-2. <https://doi.org/10.1080/24701394.2016.1197219>
- [20] Oh, D.H., Dassanayake, M., Haas, J.S., Kropornika, A., Wright, C., D'Urzo, M.P., Hong, H., Ali, S., Hernandez, A. and Lambert, G.M. (2010) Genome Structures and Halophyte-Specific Gene Expression of the Extremophile *Thellungiella parvula* in Comparison with *Thellungiella salsuginea* (*Thellungiella halophila*) and *Arabidopsis*. *Plant Physiology*, **154**, 1040-1052. <https://doi.org/10.1104/pp.110.163923>
- [21] Wu, H.J. and Xie, Q. (2012) Insights into Salt Tolerance from the Genome of *Thellungiella salsuginea*. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 12219-12224. <https://doi.org/10.1073/pnas.1209954109>

- [22] Feng, C., Liu, S., Jian, W., Lu, F., Sun, S., Bo, L., Li, P., Wei, H. and Wang, X. (2011) BRAD, the Genetics and Genomics Database for Brassica Plants. *BMC Plant Biology*, **11**, 136. <https://doi.org/10.1186/1471-2229-11-136>
- [23] Martins, W., Sousa, D.D., Proite, K., Guimarães, P., Moretzsohn, M. and Bertoli D. (2006) New Softwares for Automated Microsatellite Marker Development. *Nucleic Acids Research*, **34**, e31. <https://doi.org/10.1093/nar/gnj030>
- [24] Schuler, G.D. (1997) Sequence Mapping by Electronic PCR. *Genome Research*, **7**, 541-550. <https://doi.org/10.1101/gr.7.5.541>
- [25] 刘何, 辛艳. 植物 SSR 分子标记技术的应用[J]. 天津农林科技, 2015(5): 34-37.
- [26] 任毅. 黄瓜高密度 SSR 遗传图谱构建及其应用[J]. 中国农业科学院, 2009.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [br@hanspub.org](mailto:br@hanspub.org)