

BlatSNP 使用说明

BlatSNP 是一个基于 Blat 设计的在线 SNP 搜索平台，这个平台中包括了 60 个物种的全基因组和 CDS 序列，可以为用户提供相应的序列片段的 SNP 扫描分析。

1 输入框中的序列必须以 fasta 格式输入，长度尽量不要超过 10,000bp。序列格式如下：

```
>AT1G01010
ATGGAGGATCAAGTTGGGTTTGGGTTCCGTCGGAACGACGAGGAGCTCGT
TGGTCACTATCTCCGTAACAAAATCGAAGGAAACACTAGCCGCGACGTTG
AAGTAGCCATCAGCGAGGTCAACATCTGTAGCTACGATCCTTGGAAC TTG
CGCTTCCAGTCAAAGTACAAAATCGAGAGATGCTATGTGGTACTTCTTCTC
TCGTAGAGAAAACAACAAAAGGGAATCGACAGAGCAGGACAACGGTTTCTG
GTAAAATGGAAGCTTACCGGAGAAATCTGTTGAGGTCAAGGACCAGTGGGGA
TTTTGTAGTGAGGGCTTTCGTGGTAAGATTGGTCATAAAAAGGGTTTTGGT
GTTCTCGATGGAAGATAACCTGACAAAACCAAATCTGATTGGGTTATCC
ACGAGTTCCTACTACGACCTCTTACCAGAACATCAGAGGACATATGTCATC
```

2 上传：可以将序列以 fasta 文件的形式上传，上传文件尽量不要超过 100K，数据越大将需要越长的计算时间。

3 物种选择：共有 60 个物种，涵盖了大部分目前已经测序的物种。名称以拉丁名显示。

4 类型选择：选择所要比对的库的类型，包括 cds 和 genome

5 最小联配长度：选择可以搜索 SNP 的最小的联配片段,即如下图的片段长度：默认值为 20

gcagaggttggtatcctagaagagcctaacgaggtaggc

gcagaggttggtgtcctcgaagagcacaacgaggtaggc

6 Gap 是不是 SNP? 默认值为是

gcagaggttggtatcctagaagagcctaacgaggtaggc

gcagaggttggtgtcctc-aagagcacaacgaggtaggc

7 最大连续 SNP 个数：默认是为 3

8 输出格式：

0 ->编号，物种，库片段名，库 SNP 位置，预测片段名称，预测片段 SNP 位点

```

~
10 id, species, database name, database SNP position, query name, query SNP position
11
12 =====
13
14 1 arabidopsis scaffold_1 536489 AT1G01010 10
15 2 arabidopsis scaffold_1 536499 AT1G01010 20
16 3 arabidopsis scaffold_1 536503 AT1G01010 24
17 4 arabidopsis scaffold_1 536509 AT1G01010 30
18 5 arabidopsis scaffold_1 536596 AT1G01010 102
19 6 arabidopsis scaffold_1 536616 AT1G01010 122
20 7 arabidopsis scaffold_1 536622 AT1G01010 128
21 8 arabidopsis scaffold_1 536628 AT1G01010 134
22 9 arabidopsis scaffold_1 536661 AT1G01010 167
23 10 arabidopsis scaffold_1 536675 AT1G01010 181
24 11 arabidopsis scaffold_1 536679 AT1G01010 185
25 12 arabidopsis scaffold_1 536680 AT1G01010 186
26 13 arabidopsis scaffold_1 536834 AT1G01010 211
27 14 arabidopsis scaffold_1 536854 AT1G01010 231
28 15 arabidopsis scaffold_1 536865 AT1G01010 242
29 16 arabidopsis scaffold_1 536872 AT1G01010 249

```

1 ->编号, 物种, 库片段名, 库 SNP 位置, 预测片段名称, 预测片段 SNP 位点, SNP 类型

```
id, species, database name, database SNP position, query name, query SNP position, type
```

```

=====
1 arabidopsis scaffold_1 536489 AT1G01010 10 replacement
2 arabidopsis scaffold_1 536499 AT1G01010 20 transversion
3 arabidopsis scaffold_1 536503 AT1G01010 24 transversion
4 arabidopsis scaffold_1 536509 AT1G01010 30 transversion
5 arabidopsis scaffold_1 536596 AT1G01010 102 transversion
6 arabidopsis scaffold_1 536616 AT1G01010 122 replacement
7 arabidopsis scaffold_1 536622 AT1G01010 128 transversion
8 arabidopsis scaffold_1 536628 AT1G01010 134 transversion
9 arabidopsis scaffold_1 536661 AT1G01010 167 transversion
10 arabidopsis scaffold_1 536675 AT1G01010 181 replacement
11 arabidopsis scaffold_1 536679 AT1G01010 185 transversion
12 arabidopsis scaffold_1 536680 AT1G01010 186 transversion
13 arabidopsis scaffold_1 536834 AT1G01010 211 transversion
14 arabidopsis scaffold_1 536854 AT1G01010 231 transversion
15 arabidopsis scaffold_1 536865 AT1G01010 242 transversion
16 arabidopsis scaffold_1 536872 AT1G01010 249 transversion
17 arabidopsis scaffold_1 536874 AT1G01010 251 replacement
18 arabidopsis scaffold_1 536877 AT1G01010 254 transversion

```

2 ->编号, 物种, 库片段名, 库 SNP 位置, 预测片段名称, 预测片段 SNP 位点, SNP 变异方式, SNP 类型

```
id, species, database name, database SNP position, query name, query SNP position, mutation type, type
```

```

=====
1 arabidopsis scaffold_1 536489 AT1G01010 10 c->a replacement
2 arabidopsis scaffold_1 536499 AT1G01010 20 g->c transversion
3 arabidopsis scaffold_1 536503 AT1G01010 24 c->t transversion
4 arabidopsis scaffold_1 536509 AT1G01010 30 t->c transversion
5 arabidopsis scaffold_1 536596 AT1G01010 102 g->a transversion
6 arabidopsis scaffold_1 536616 AT1G01010 122 t->g replacement
7 arabidopsis scaffold_1 536622 AT1G01010 128 a->g transversion
8 arabidopsis scaffold_1 536628 AT1G01010 134 t->a transversion
9 arabidopsis scaffold_1 536661 AT1G01010 167 t->c transversion
10 arabidopsis scaffold_1 536675 AT1G01010 181 a->c replacement
11 arabidopsis scaffold_1 536679 AT1G01010 185 t->c transversion
12 arabidopsis scaffold_1 536680 AT1G01010 186 a->t transversion
13 arabidopsis scaffold_1 536834 AT1G01010 211 g->c transversion
14 arabidopsis scaffold_1 536854 AT1G01010 231 c->t transversion
15 arabidopsis scaffold_1 536865 AT1G01010 242 g->a transversion
16 arabidopsis scaffold_1 536872 AT1G01010 249 t->c transversion
17 arabidopsis scaffold_1 536874 AT1G01010 251 c->a replacement
18 arabidopsis scaffold_1 536877 AT1G01010 254 g->a transversion
19 arabidopsis scaffold_1 536887 AT1G01010 264 g->a transversion
20 arabidopsis scaffold_1 536889 AT1G01010 266 t->g replacement
21 arabidopsis scaffold_1 536904 AT1G01010 281 g->c transversion

```

3 ->编号, 物种, 库片段名, 库 SNP 位置, 预测片段名称, 预测片段 SNP 位点, SNP 变异方式, SNP 类型, 联配的库序列, 联配的预测序列

id, species, database name, database SNP position, query name, query SNP position, mutation type, type, database sequence, query sequence

#####

```
1
arabidopsis scaffold_1 536489 AT1G01010 10 c->a replacement
ttaaccaccagaatgatcaaccaatgatggagatgaacaacaagaccaatgatcatgtctttgagatgtccactgtgcatc
ttaaccacaagaatgatccaactaatgacggagatgaacaacaagaccaatgatcatgtctttgagatgtccactgtgcatc

2
arabidopsis scaffold_1 536499 AT1G01010 20 g->c transversion
ttaaccaccagaatgatcaaccaatgatggagatgaacaacaagaccaatgatcatgtctttgagatgtccactgtgcatc
ttaaccacaagaatgatccaactaatgacggagatgaacaacaagaccaatgatcatgtctttgagatgtccactgtgcatc

3
arabidopsis scaffold_1 536503 AT1G01010 24 c->t transversion
ttaaccaccagaatgatcaaccaatgatggagatgaacaacaagaccaatgatcatgtctttgagatgtccactgtgcatc
ttaaccacaagaatgatccaactaatgacggagatgaacaacaagaccaatgatcatgtctttgagatgtccactgtgcatc

4
arabidopsis scaffold_1 536509 AT1G01010 30 t->c transversion
ttaaccaccagaatgatcaaccaatgatggagatgaacaacaagaccaatgatcatgtctttgagatgtccactgtgcatc
ttaaccacaagaatgatccaactaatgacggagatgaacaacaagaccaatgatcatgtctttgagatgtccactgtgcatc

5
arabidopsis scaffold_1 536596 AT1G01010 102 g->a transversion
tccaaacatccagctctgctccactgtgttgatgaggtttcttgatcagctcttcagccatttccactgattcgtttctgagaatatac
tccaaacatccagctctgctccactgtgttgatgaggtttcttgatcagctcttcagccatttccactgattcgtttctgagaatatac
```

4 ->编号, 物种, 库片段名, 库片段起点, 库片段终点, 库 SNP 位置, 预测片段名称, 预测片段起点, 预测片段终点, 预测 SNP 位点, SNP 变异方式, SNP 类型

id, species, database name, database start, database end, database SNP position, query name, query start, query end, query SNP position, mutation type, type

#####

```
1 arabidopsis scaffold_1 536480 536567 536489 AT1G01010 1 88 10 c->a replacement
2 arabidopsis scaffold_1 536480 536567 536499 AT1G01010 1 88 20 g->c transversion
3 arabidopsis scaffold_1 536480 536567 536503 AT1G01010 1 88 24 c->t transversion
4 arabidopsis scaffold_1 536480 536567 536509 AT1G01010 1 88 30 t->c transversion
5 arabidopsis scaffold_1 536583 536686 536596 AT1G01010 89 192 102 g->a transversion
6 arabidopsis scaffold_1 536583 536686 536616 AT1G01010 89 192 122 t->g replacement
7 arabidopsis scaffold_1 536583 536686 536622 AT1G01010 89 192 128 a->g transversion
8 arabidopsis scaffold_1 536583 536686 536628 AT1G01010 89 192 134 t->a transversion
9 arabidopsis scaffold_1 536583 536686 536661 AT1G01010 89 192 167 t->c transversion
10 arabidopsis scaffold_1 536583 536686 536675 AT1G01010 89 192 181 a->c replacement
11 arabidopsis scaffold_1 536583 536686 536679 AT1G01010 89 192 185 t->c transversion
12 arabidopsis scaffold_1 536583 536686 536680 AT1G01010 89 192 186 a->t transversion
13 arabidopsis scaffold_1 536831 536968 536834 AT1G01010 208 345 211 g->c transversion
14 arabidopsis scaffold_1 536831 536968 536854 AT1G01010 208 345 231 c->t transversion
15 arabidopsis scaffold_1 536831 536968 536865 AT1G01010 208 345 242 g->a transversion
16 arabidopsis scaffold_1 536831 536968 536872 AT1G01010 208 345 249 t->c transversion
17 arabidopsis scaffold_1 536831 536968 536874 AT1G01010 208 345 251 c->a replacement
18 arabidopsis scaffold_1 536831 536968 536877 AT1G01010 208 345 254 g->a transversion
19 arabidopsis scaffold_1 536831 536968 536887 AT1G01010 208 345 264 g->a transversion
20 arabidopsis scaffold_1 536831 536968 536889 AT1G01010 208 345 266 t->g replacement
21 arabidopsis scaffold_1 536831 536968 536904 AT1G01010 208 345 281 g->c transversion
```

5 ->编号, 物种, 库片段名, 库片段起点, 库片段终点, 库 SNP 位置, 预测片段名称, 预测片段起点, 预测片段终点, 预测 SNP 位点, SNP 变异方式, SNP 类型, 联配的库序列, 联配的预测序列。

```

-----
id, species, database name, database start, database end, database SNP position, query name, query start, query end, query SNP position, mutation type, type, database sequence, query sequence
=====
1
Arabidopsis_lyrata scaffold_1 536480 536567 536489 AT1G01010 1290 1203 1194 c->a replacement
ttaaccaaccagaatgatcgaaccaatgatgggatgaacaacaagacaccaatgatcatgttcttgagatgtccactgtgcattc
ttaaccaacaagaatgatcccaactaatgacgggatgaacaacaagacaccaatgatcatgttcttgagatgtccactgtgcattc

2
Arabidopsis_lyrata scaffold_1 536480 536567 536499 AT1G01010 1290 1203 1184 g->c transversion
ttaaccaaccagaatgatcgaaccaatgatgggatgaacaacaagacaccaatgatcatgttcttgagatgtccactgtgcattc
ttaaccaacaagaatgatcccaactaatgacgggatgaacaacaagacaccaatgatcatgttcttgagatgtccactgtgcattc

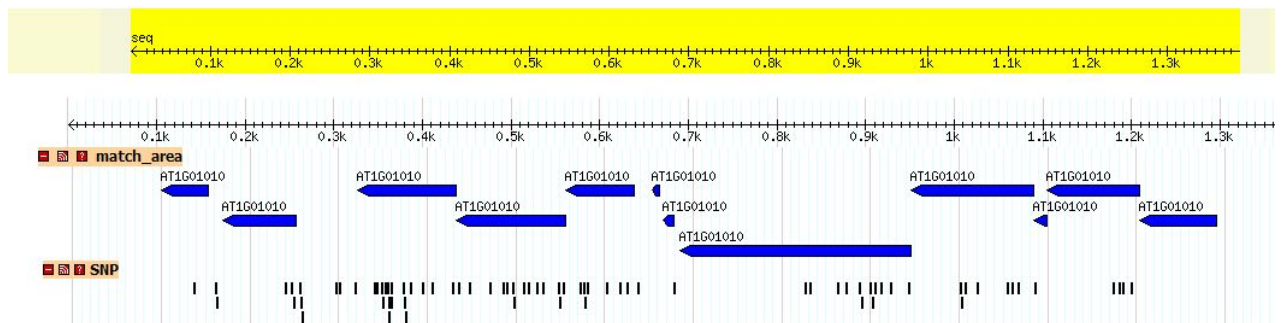
3
Arabidopsis_lyrata scaffold_1 536480 536567 536503 AT1G01010 1290 1203 1180 c->t transversion
ttaaccaaccagaatgatcgaaccaatgatgggatgaacaacaagacaccaatgatcatgttcttgagatgtccactgtgcattc
ttaaccaacaagaatgatcccaactaatgacgggatgaacaacaagacaccaatgatcatgttcttgagatgtccactgtgcattc

4
Arabidopsis_lyrata scaffold_1 536480 536567 536509 AT1G01010 1290 1203 1174 t->c transversion
ttaaccaaccagaatgatcgaaccaatgatgggatgaacaacaagacaccaatgatcatgttcttgagatgtccactgtgcattc
ttaaccaacaagaatgatcccaactaatgacgggatgaacaacaagacaccaatgatcatgttcttgagatgtccactgtgcattc

5
Arabidopsis_lyrata scaffold_1 536583 536686 536596 AT1G01010 1202 1099 1086 g->a transversion
tccaaaacatccggctcgttcacggtgtgttgatagaggtttcttgatgagcttcagcactttccactgattcgttttctgagatatacac
tccaaaacatccagctcgttcacggtgtgttgatagaggtttcttgatgagcttcagcactttccactgattcgttttctgagactatacac

```

9 所有的数据会以一个 tar.gz 格式的压缩包的形式提供下载，结果会包括输入序列以及 SNP 位点信息，同时还有一个 Gbrowse 动态展示。Gbrowse 动态展示中，数据分为两部分，上方的蓝色为联配到的位置，下方的红色色块为 SNP 位点存在情况。同时，还会提供一个总的统计 statics 文本，可以用记事本打开。



SNP number: 90

mutation type number

g_a 12
c_g 3
t_a 8
c_a 5
g_c 6
t_g 4
a_t 5
a_c 2
t_c 19
c_t 4
g_t 4
a_g 12

Replacement: 15

Transversion: 69

10. 本平台在 Perl script 界面提供了三个简单实用的 Perl 脚本，用于数据的预处理以及相关应用。

注：如果您有任何意见和建议，请致函杨龙（lyang@sdau.edu.cn）。您也可以将您所无法上传的大量数据通过邮箱获取他渠道发送给我们，我们会以最快的速度进行处理并给您反馈结果。同时，我们的数据库中所有的序列文件均来源于网络，请您在使用时根据需要进行相关测序文章的引用。感谢您的合作。

山东农业大学烟草实验室

2014-5-14 于泰安