

Introduction of BlatSNP

BlatSNP is a web tool to detect SNPs basen on the Blat, this platform includes genome and CDS sequence of 60 species, and can provide SNP mining for the related segments.

1. Input area, the sequence input shold be in the format of **FASTA**, and the sequence may not longer than 10,000bp.

```
>AT1G01010
ATGGAGGATCAAGTTGGGTTTGGGTTCCGTCCGAACGACGAGGAGCTCGT
TGGTCACTATCTCCGTAACAAAAATCGAAGGAAACACTAGCCGCGACGTTG
AAGTAGCCATCAGCGAGGTCAACATCTGTAGCTACGATCCTTGGAACCTG
CGCTTCCAGTCAAAGTACAAAATCGAGAGATGCTATGTGGTACTTCTTCTC
TCGTAGAGAAAAACAACAAAAGGGAATCGACAGAGCAGGACAACGGTTTCTG
GTAAATGGAAGCTTACCGGAGAAATCTGTTGAGGTCAAGGACCAGTGGGGA
TTTTGTAGTGAGGGCTTTCGTGGTAAGATTGGTCATAAAAAGGGTTTTGGT
GTTCTCGATGGAAGATAACCCTGACAAAACCAAATCTGATTGGGTTATCC
ACGAGTTCCTACTACGACCTCTTACCAGAACATCAGAGGACATATGTCATC
```

2. Upload: you also can upload your file in the format of fasta and with a name such as upload.fasta. The file should not larger than 100KB, and it will take a few minutes.

3. Species: 60 species in all, including almost all sequenced species. All species are listed as Latin name.

4. Type: choose CDS or genome as database

5. The min alignment length: choose the min length of your alignment segments, default is 20.

gcagaggttggtatcctagaagagcctaacgaggttaggc

gcagaggttggtgtcctcgaagagcacaacgaggttaggc

6. Is gap SNP? Default is yes

gcagaggttggtatcctagaagagcctaacgaggttaggc

gcagaggttggtgtcctc-aagagcacaacgaggttaggc

7. The max continue SNPs : how many continue SNPs is still effectively? The default is 3.

8. Output style

0 -> id, species, database name, database SNP position, query name, query SNP position

```

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 -> id, species, database name, database SNP position, query name, query SNP position, type

```

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 -> id, species, database name, database SNP position, query name, query SNP position, mutation type, type

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

```
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  
ttaaccaaccagaatgatcgaaccaatgatggagatgaacaacaagacccaatgatcatgttcttgagatagttccactgtgcattc  
ttaaccaacaagaatgatccaactaatgacggagatgaacaacaagacccaatgatcatgttcttgagatagttccactgtgcattc  
2  
arabidopsis scaffold_1 536499 AT1G01010 20 g->c transversion  
ttaaccaaccagaatgatcgaaccaatgatggagatgaacaacaagacccaatgatcatgttcttgagatagttccactgtgcattc  
ttaaccaacaagaatgatccaactaatgacggagatgaacaacaagacccaatgatcatgttcttgagatagttccactgtgcattc  
3  
arabidopsis scaffold_1 536503 AT1G01010 24 c->t transversion  
ttaaccaaccagaatgatcgaaccaatgatggagatgaacaacaagacccaatgatcatgttcttgagatagttccactgtgcattc  
ttaaccaacaagaatgatccaactaatgacggagatgaacaacaagacccaatgatcatgttcttgagatagttccactgtgcattc  
4  
arabidopsis scaffold_1 536509 AT1G01010 30 t->c transversion  
ttaaccaaccagaatgatcgaaccaatgatggagatgaacaacaagacccaatgatcatgttcttgagatagttccactgtgcattc  
ttaaccaacaagaatgatccaactaatgacggagatgaacaacaagacccaatgatcatgttcttgagatagttccactgtgcattc  
5  
arabidopsis scaffold_1 536596 AT1G01010 102 g->a transversion  
tccaaaacaatccgctctgcttccaccgtgttggatagagttcttgatcagcttccagcatttccactcgtattcgtttctgagataatcac  
tccaaaacaatccgctctgcttccaccgtgttggatagagttcttgatcagcttccagcatttccactcgtattcgtttctggaactatcac
```

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

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 -> 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

```
-----
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
ttaaccaacagaatgatcgaaccaatgatggagatgaacaacaagacaccaatgatcatgttttgagatgttccactgtgcattc
ttaaccaacaagaatgatccactaatgatggagatgaacaacaagacaccaatgatcatgttttgagatgttccactgtgcattc

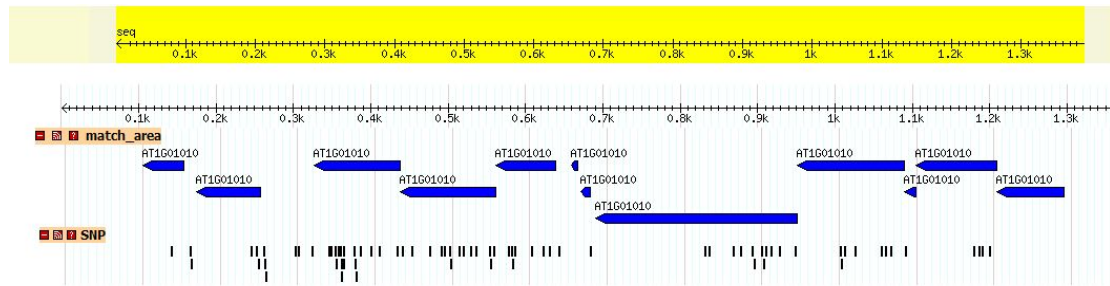
2
Arabidopsis_lyrata scaffold_1 536480 536567 536499 AT1G01010 1290 1203 1184 g->c transversion
ttaaccaacagaatgatcgaaccaatgatggagatgaacaacaagacaccaatgatcatgttttgagatgttccactgtgcattc
ttaaccaacaagaatgatccactaatgatggagatgaacaacaagacaccaatgatcatgttttgagatgttccactgtgcattc

3
Arabidopsis_lyrata scaffold_1 536480 536567 536503 AT1G01010 1290 1203 1180 c->t transversion
ttaaccaacagaatgatcgaaccaatgatggagatgaacaacaagacaccaatgatcatgttttgagatgttccactgtgcattc
ttaaccaacaagaatgatccactaatgatggagatgaacaacaagacaccaatgatcatgttttgagatgttccactgtgcattc

4
Arabidopsis_lyrata scaffold_1 536480 536567 536509 AT1G01010 1290 1203 1174 t->c transversion
ttaaccaacagaatgatcgaaccaatgatggagatgaacaacaagacaccaatgatcatgttttgagatgttccactgtgcattc
ttaaccaacaagaatgatccactaatgatggagatgaacaacaagacaccaatgatcatgttttgagatgttccactgtgcattc

5
Arabidopsis_lyrata scaffold_1 536583 536686 536596 AT1G01010 1202 1099 1086 g->a transversion
tccaaaacatccagctcgttccacgtgttggatagagtttcttgatcagcttccagcatttccactcgtatcgttttctgagaatatacac
tccaaaacatccagctcgttccacgtgttggatagagtttcttgatcagcttccagcatttccactcgtatcgttttctggaactatacac
```

9. All output data will be provided in a tar.gz file. The file include two files, the first one is the sequence you input and another is the information of SNPs. There will also be a website link a dynamic map. The map including two parts, the blue blocks are the match area, the red blocks are SNPs. There is also a statics file containing a statics of all the SNPs.



 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. In the page of Perl script, there are three simple Perl script for users to pretreat the data, or used in other area.

Note: if you have any suggestion, please mail to LongYang(lyang@sdau.edu.cn). You also can give us you data through any ways and we will mine it as soon as possible. All sequence data in this platform were download from the net, please cite the sequence you used. The sequence information will be listed in the result.

Thank you

Tobacco Lab, SDAU

2014-5-14