cDNA配列やアミノ酸配列があれば、Blast検索で有用な情報が得られるが、遺伝子があるかも分からない長いDNA配列を検索しても有用な情報が得られない。

そこで、長いDNA配列の中にどこに遺伝子が
隠されているかをまず探し出す。

DNA配列の中から遺伝子を予測する

遺伝子予想ソフト

MEGANTE : https://megante.dna.affrc.go.jp/

Softberry/FGENESH: http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind

GENSCAN : http://genes.mit.edu/GENSCAN.html

-2（11/15）演習

第９回目-1（11/22）KEGGを用いた解析

連鎖解析

出席と宿題の提出　提出方法

リンク集

データベース検索：

1. PubMed: 論文検索
2. Entrez：総合データベース
3. Google：何でも検索
4. Google Scholar：科学に関する検索
5. 特許データベース

ホモロジー検索：

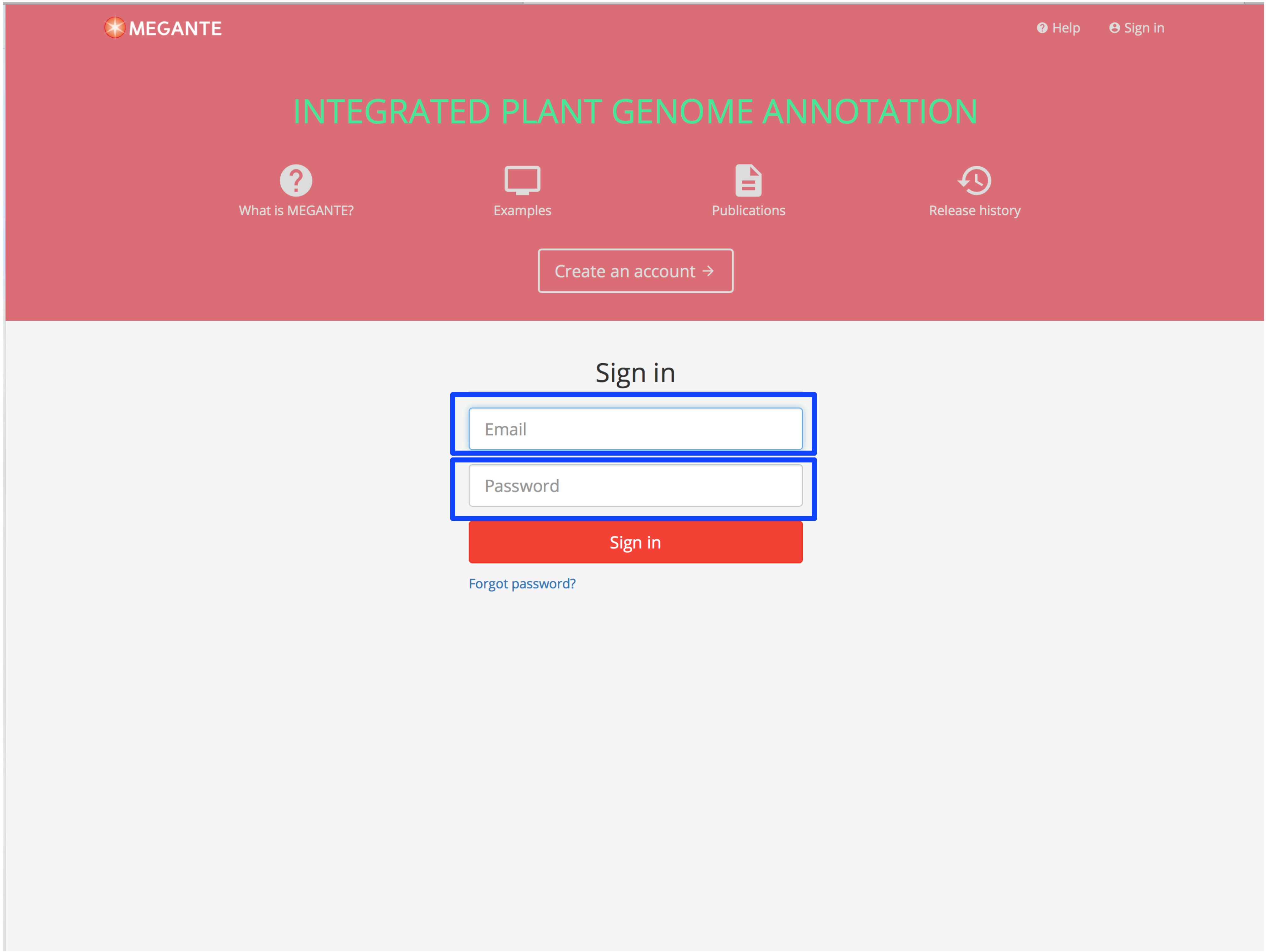
1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. MEGANTE ：遺伝子の予測 ◀
8. Softberry/FGENESH ：遺伝子の予測
9. GeneMark ：遺伝子の予測
10. GENESCAN ：遺伝子の予測
11. clustalW： アライメント
12. InterProScan： ドメイン解析
13. PSORT： 細胞内局在性予測
14. GEO：遺伝子発現解析
15. Gene Ontology：用語を用いた検索
16. SSPN：スプライシングの予測
17. ORF Finder：ORFの予測
18. TFSEARCH ：転写因子結合部位の予測
19. MOTIF ：タンパク質モチーフ検索
20. pI/Mw： 等電点、分子量の予測
21. RAP-DB：イネのゲノムデータベース

パスワード：大文字、小文字、数字が必要

これまでに、沢山の研究であきらかになった情報が集積されていて、
正確な遺伝子の存在場所や機能が簡単に推測できるようになっている。

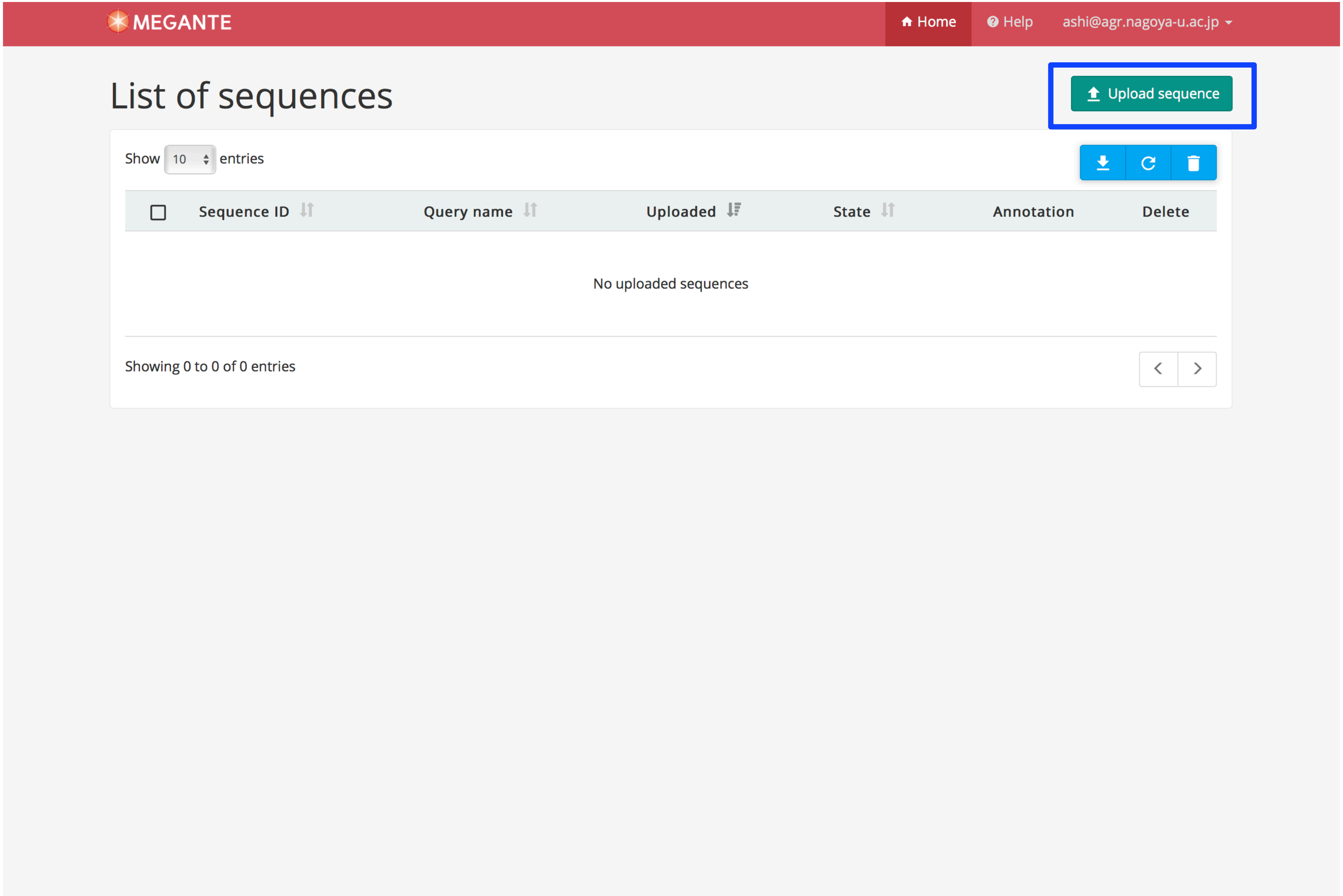

MEGANTE

Help

Sign in

INTEGRATED PLANT GENOME ANNOTATION

Sign in

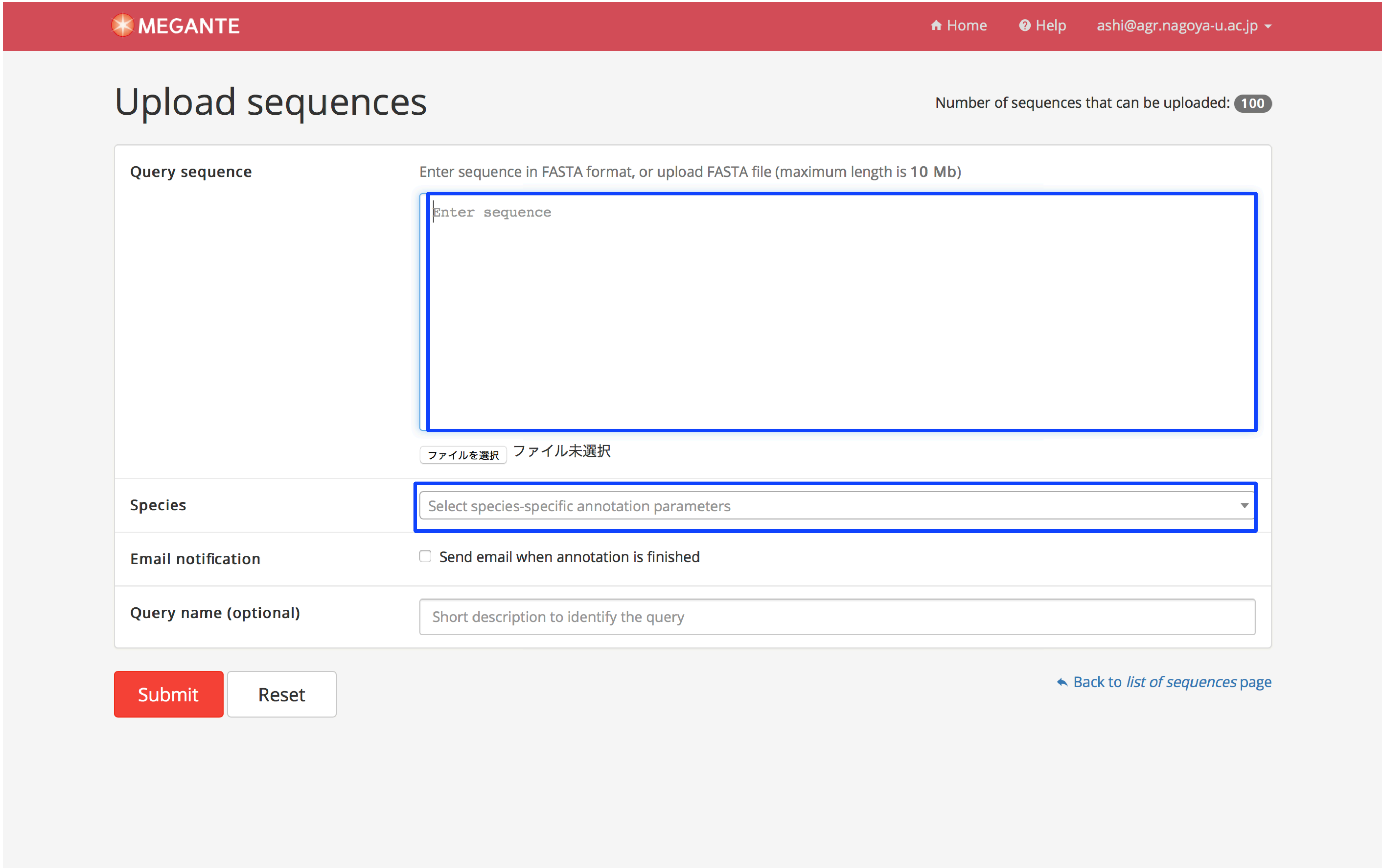


List of sequences

☐	Sequence ID	Query name	Uploaded	State	Annotation	Delete

No uploaded sequences

Showing 0 to 0 of 0 entries



Upload sequences

Number of sequences that can be uploaded: 100

Query sequence

Enter sequence in FASTA format, or upload FASTA file (maximum length is **10 Mb**)

Enter sequence

ファイルを選択 ファイル未選択

Species

Select species-specific annotation parameters

Email notification

☐ Send email when annotation is finished

Query name (optional)

Short description to identify the query

Back to *list of sequences* page

1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. Softberry/FGENESH ：遺伝子の予測
8. GeneMark ：遺伝子の予測
9. GENESCAN ：遺伝子の予測
10. clustalW： アライメント
11. InterProScan： ドメイン解析
12. PSORT： 細胞内局在性予測
13. GEO：遺伝子発現解析
14. Gene Ontology：用語を用いた検索
15. SSPN：スプライシングの予測
16. ORF Finder：ORFの予測
17. TFSEARCH ：転写因子結合部位の予測
18. MOTIF ：タンパク質モチーフ検索
19. pI/Mw： 等電点、分子量の予測
20. RAP-DB：イネのゲノムデータベース

解析用配列

1. ある遺伝子AのgenomeDNA配列

2. ある遺伝子AのcDNA配列

3. ある遺伝子XのgenomeDNA配列

4. ある遺伝子XのcDNA配列

5. ある遺伝子YのgenomicDNA

6. ある遺伝子YのcDNA

7. Z1遺伝子のgenomicDNA

8. Z1遺伝子のCDS

9. Z1遺伝子の予想アミノ酸配列

10. Z2遺伝子のgenomicDNA

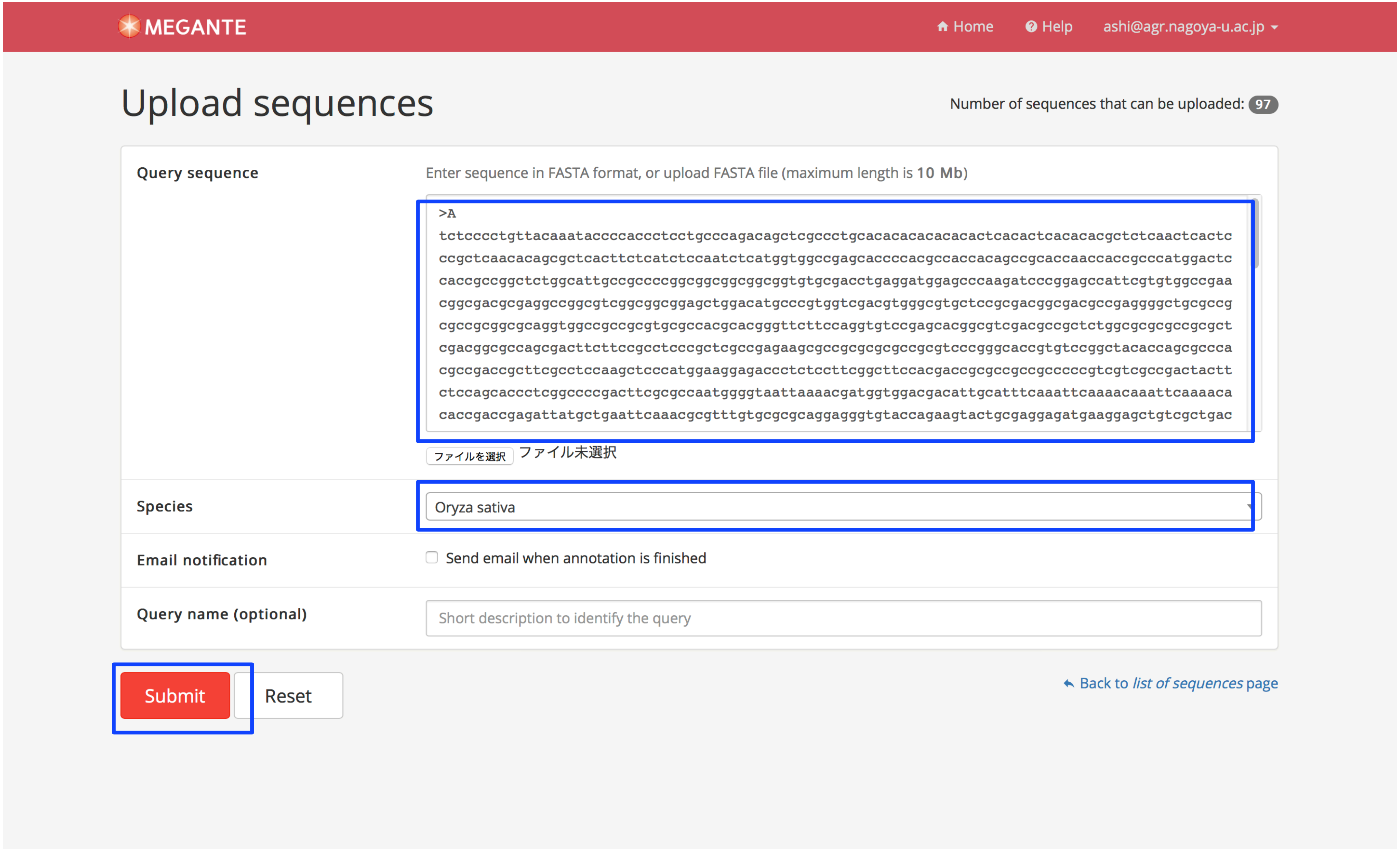
11. Z2遺伝子のCDS

12. Z2遺伝子のアミノ酸配列

13. 野生型SD1-genomicDNA配列

14. 野生型SD1-cDNA配列

15. カルロース-cDNA



Upload sequences

Number of sequences that can be uploaded: 97

Query sequence

Enter sequence in FASTA format, or upload FASTA file (maximum length is **10 Mb**)

```
>A
tctcccctgttacaaataccccaccctcctgcccagacagctcgccctgcacacacacacactcacactcacacacgctctcaactcactc
ccgctcaacacagcgctcacttctcatctccaatctcatggtggccgagcaccccacgccaccacagccgcaccaaccaccgcccatggactc
caccgccggctctggcattgccgccccggcggcggcggtgtgcgacctgaggatggagcccaagatcccggagccattcgtgtggccgaa
cggcgacgcgaggccggcgtcggcggcggagctggacatgcccgtggtcgacgtgggcgtgctccgcgacggcgacgccgaggggctgcgccg
cgccgcggcgcaggtggccgccgcgtgcgccacgcacgggttcttccaggtgtccgagcacggcgtcgacgccgctctggcgcgcgccgcgct
cgacggcgccagcgacttcttccgcctcccgctcgccgagaagcgccgcgcgcgccgcgtcccgggcaccgtgtccggctacaccagcgccca
cgccgaccgcttcgcctccaagctcccatggaaggagaccctctccttcggcttccacgaccgcgccgccgcccccgtcgtcgccgactactt
ctccagcaccctcggccccgacttcgcgccaatggggtaattaaaacgatggtggacgacattgcatttcaaattcaaaacaaattcaaaaca
caccgaccgagattatgctgaattcaaacgcgtttgtgcgcgcaggagggtgtaccagaagtactgcgaggagatgaaggagctgtcgctgac
```

ファイル未選択

Species

Oryza sativa

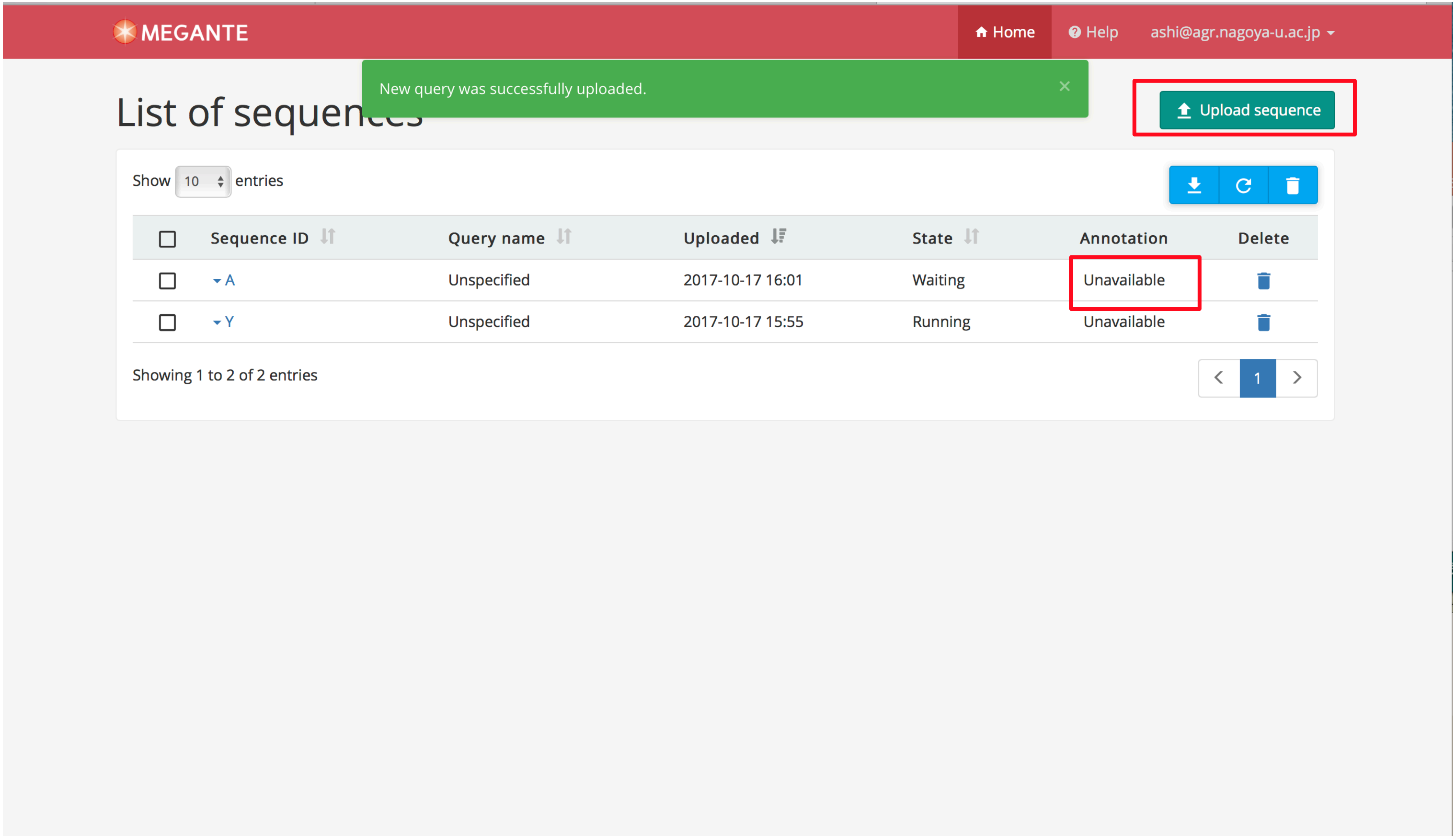
Email notification

☐ Send email when annotation is finished

Query name (optional)

Short description to identify the query

Back to *list of sequences* page



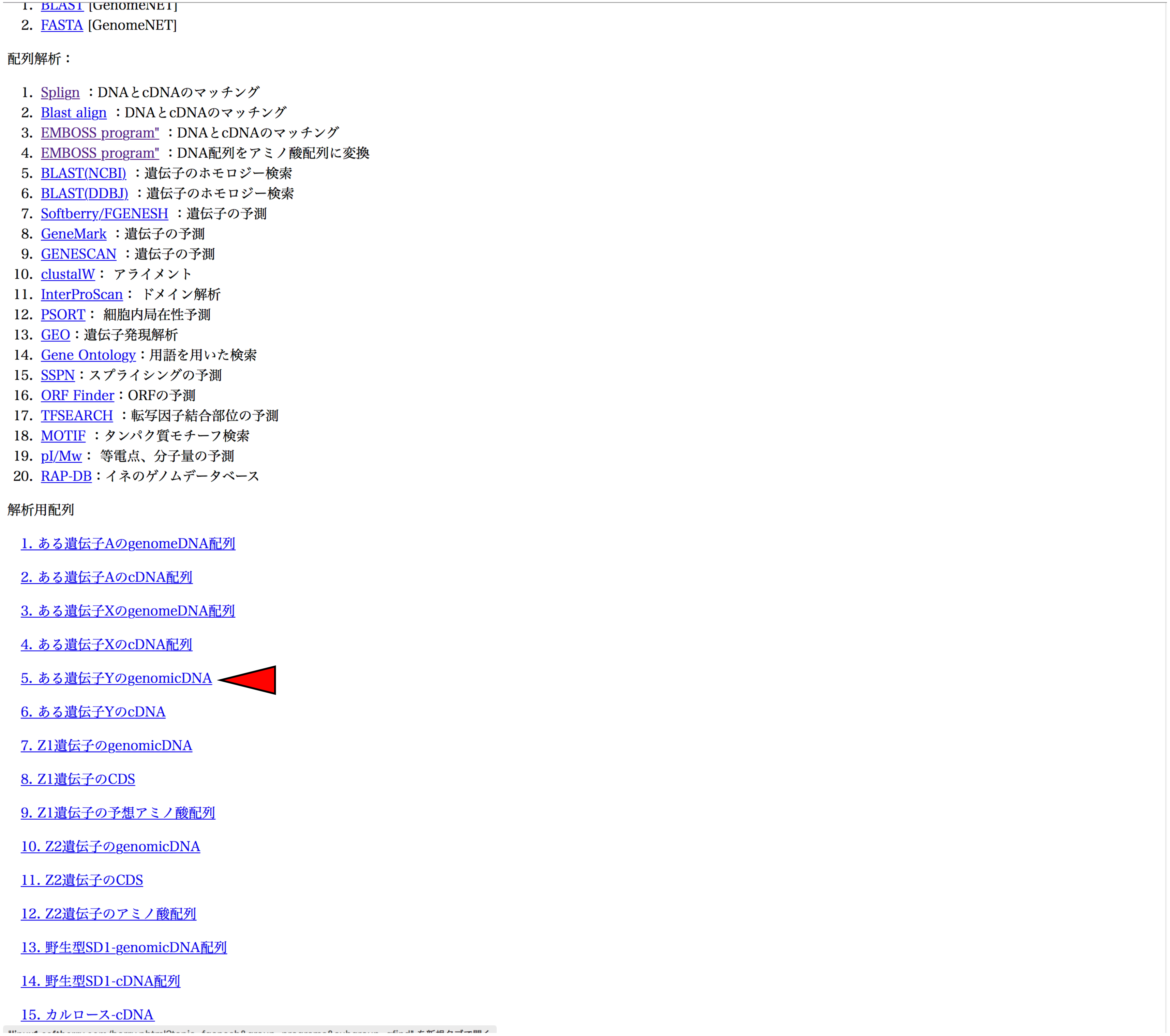
New query was successfully uploaded.

List of sequences

Show 10 entries

☐	Sequence ID	Query name	Uploaded	State	Annotation	Delete
☐	A	Unspecified	2017-10-17 16:01	Waiting	Unavailable	🗑
☐	Y	Unspecified	2017-10-17 15:55	Running	Unavailable	🗑

Showing 1 to 2 of 2 entries



1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. Softberry/FGENESH ：遺伝子の予測
8. GeneMark ：遺伝子の予測
9. GENESCAN ：遺伝子の予測
10. clustalW： アライメント
11. InterProScan： ドメイン解析
12. PSORT： 細胞内局在性予測
13. GEO：遺伝子発現解析
14. Gene Ontology：用語を用いた検索
15. SSPN：スプライシングの予測
16. ORF Finder：ORFの予測
17. TFSEARCH ：転写因子結合部位の予測
18. MOTIF ：タンパク質モチーフ検索
19. pI/Mw： 等電点、分子量の予測
20. RAP-DB：イネのゲノムデータベース

解析用配列

1. ある遺伝子AのgenomeDNA配列

2. ある遺伝子AのcDNA配列

3. ある遺伝子XのgenomeDNA配列

4. ある遺伝子XのcDNA配列

5. ある遺伝子YのgenomicDNA ◀

6. ある遺伝子YのcDNA

7. Z1遺伝子のgenomicDNA

8. Z1遺伝子のCDS

9. Z1遺伝子の予想アミノ酸配列

10. Z2遺伝子のgenomicDNA

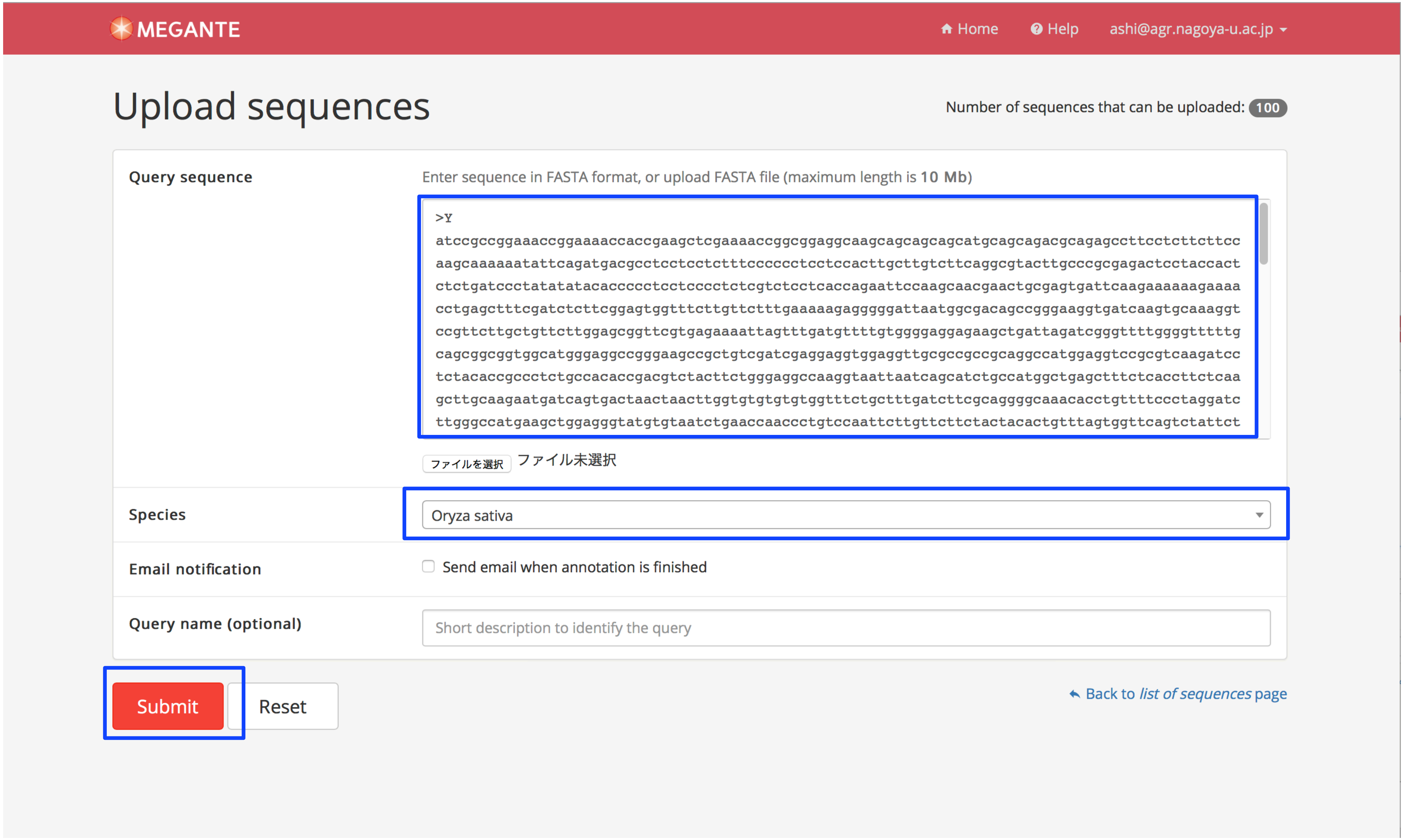
11. Z2遺伝子のCDS

12. Z2遺伝子のアミノ酸配列

13. 野生型SD1-genomicDNA配列

14. 野生型SD1-cDNA配列

15. カルロース-cDNA



Upload sequences

Number of sequences that can be uploaded: 100

Query sequence

Enter sequence in FASTA format, or upload FASTA file (maximum length is **10 Mb**)

```
>Y
atccgccggaaaccggaaaaccaccgaagctcgaaaaccggcggaggcaagcagcagcagcatgcagcagacgcagagccttcctcttcttcc
aagcaaaaaatattcagatgacgcctcctcctctttccccccctcctccacttgcttgtcttcaggcgtacttgcccgcgagactcctaccact
ctctgatccctatatatacacccccctcctcccctctcgtctcctcaccagaattccaagcaacgaactgcgagtgattcaagaaaaaagaaaa
cctgagctttcgatctcttcggagtggtttcttgttctttgaaaaagaggggattaatggcgacagccgggaaggtgatcaagtgcaaaggt
ccgttcttgctgttcttggagcggttcgtgagaaaattagtttgatgttttgtggggaggagaagctgattagatcgggttttggggtttttg
cagcggcggtggcatgggaggccgggaagccgctgtcgatcgaggaggtggaggttgcgccgccgcaggccatggaggtccgcgtcaagatcc
tctacaccgccctctgccacaccgacgtctacttctgggaggccaaggtaattaatcagcatctgccatggctgagctttctcaccttctcaa
gcttgcaagaatgatcagtgactaactaacttggtgtgtgtgtggtttctgctttgatcttcgcaggggcaaacacctgttttccctaggatc
ttgggccatgaagctggagggtatgtgtaatctgaaccaaccctgtccaattcttgttcttctactacactgtttagtggttcagtctattct
```

Species

Oryza sativa

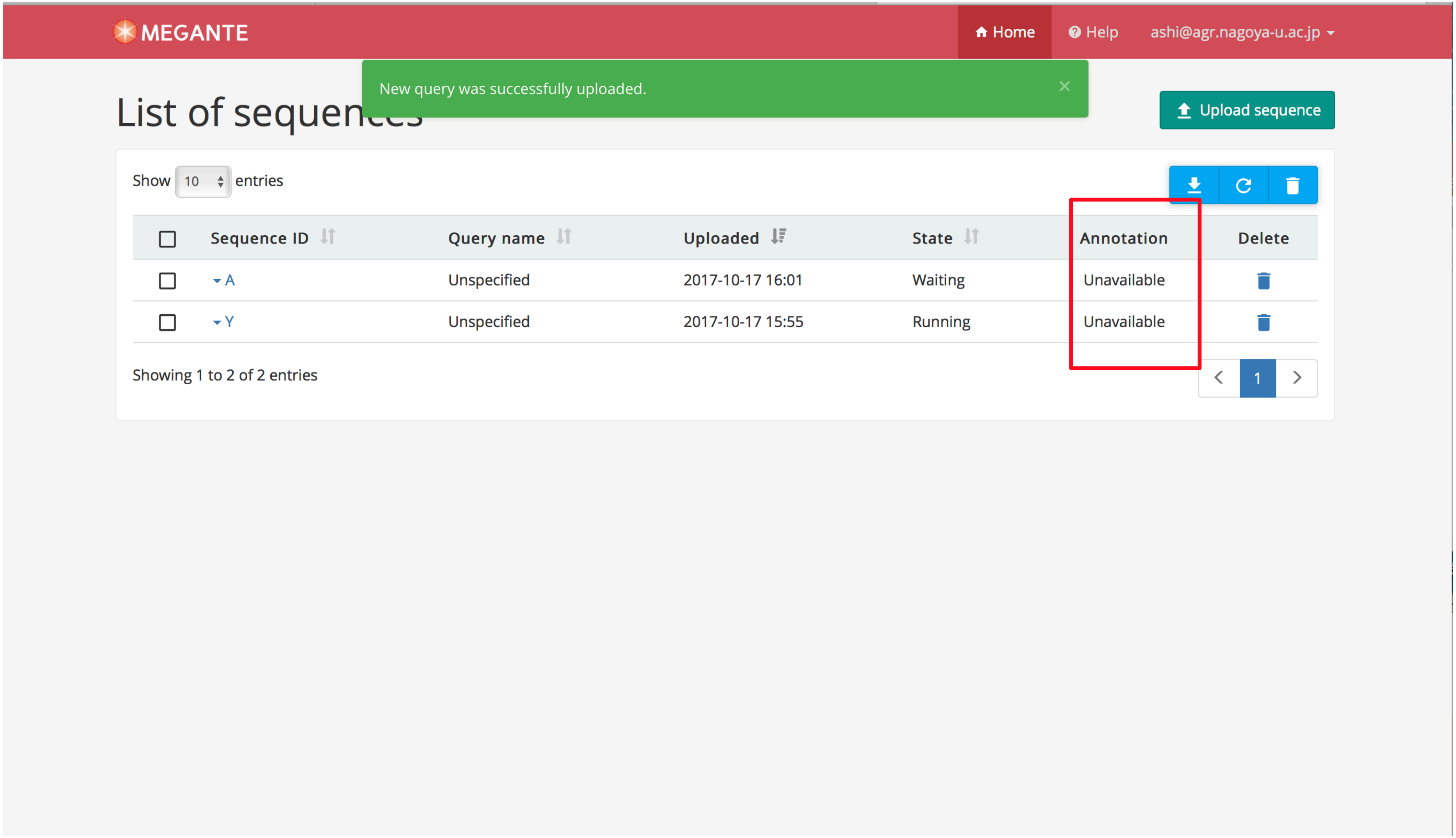
Email notification

☐ Send email when annotation is finished

Query name (optional)

Short description to identify the query

Back to *list of sequences* page



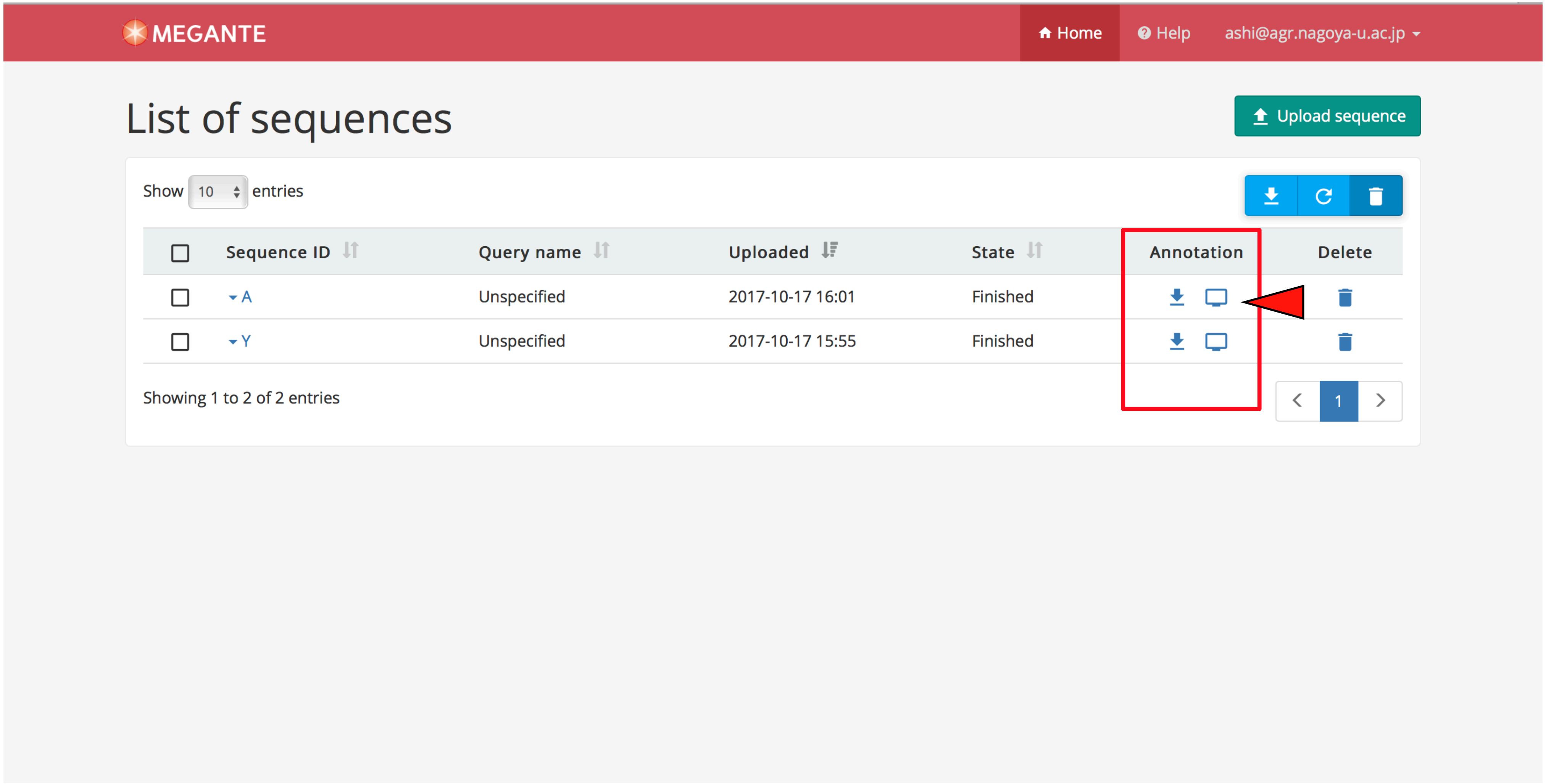
New query was successfully uploaded.

List of sequences

Show 10 entries

☐	Sequence ID	Query name	Uploaded	State	Annotation	Delete
☐	A	Unspecified	2017-10-17 16:01	Waiting	Unavailable	🗑
☐	Y	Unspecified	2017-10-17 15:55	Running	Unavailable	🗑

Showing 1 to 2 of 2 entries

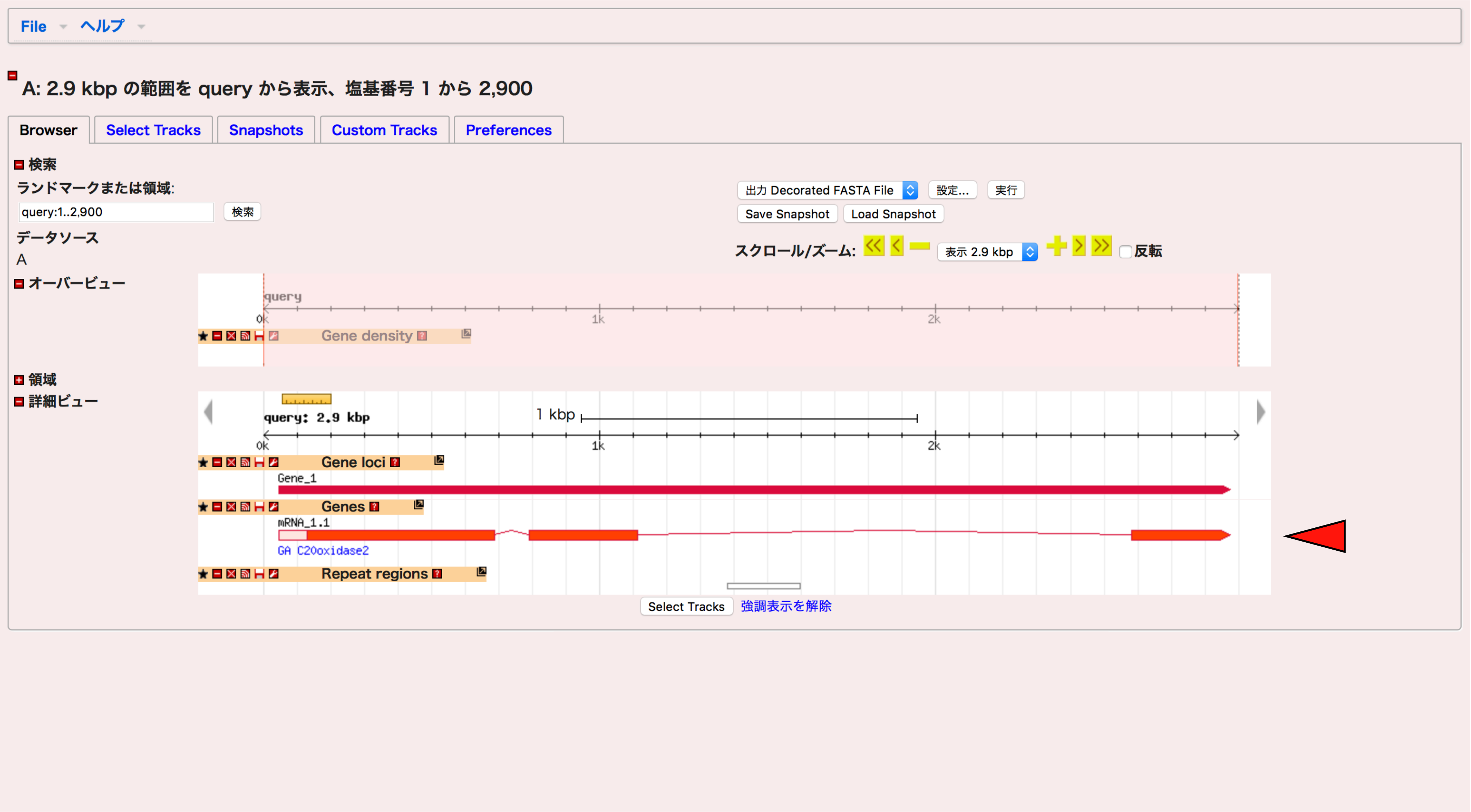


List of sequences

Show 10 entries

☐	Sequence ID	Query name	Uploaded	State	Annotation	Delete
☐	A	Unspecified	2017-10-17 16:01	Finished	⬇ 🖥	🗑
☐	Y	Unspecified	2017-10-17 15:55	Finished	⬇ 🖥	🗑

Showing 1 to 2 of 2 entries

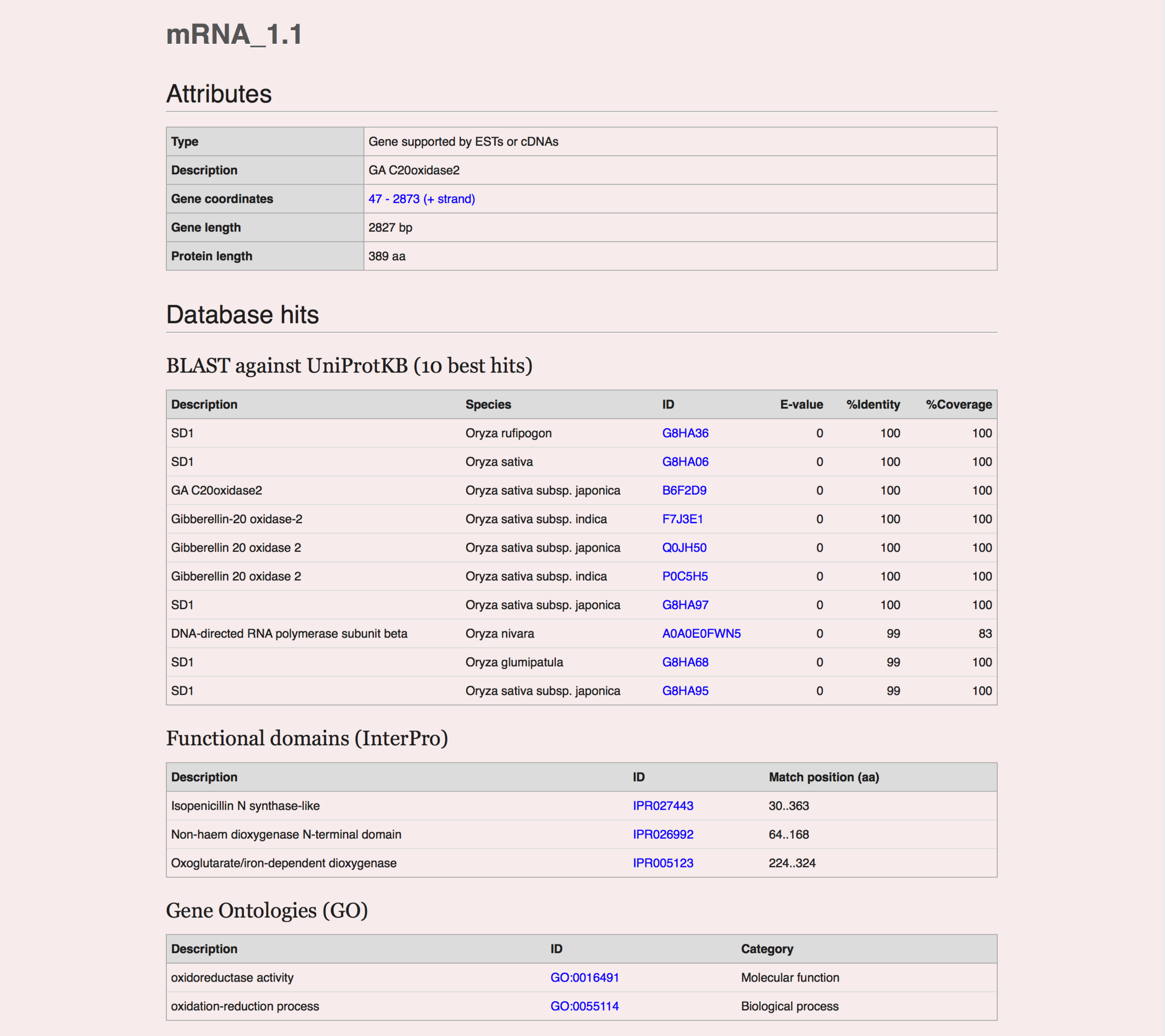


A: 2.9 kbp の範囲を query から表示、塩基番号 1 から 2,900

ランドマークまたは領域: query:1..2,900

データソース: A

スクロール/ズーム: 表示 2.9 kbp



mRNA_1.1

Attributes

Type	Gene supported by ESTs or cDNAs
Description	GA C20oxidase2
Gene coordinates	47 - 2873 (+ strand)
Gene length	2827 bp
Protein length	389 aa

Database hits

BLAST against UniProtKB (10 best hits)

Description	Species	ID	E-value	%Identity	%Coverage
SD1	Oryza rufipogon	G8HA36	0	100	100
SD1	Oryza sativa	G8HA06	0	100	100
GA C20oxidase2	Oryza sativa subsp. japonica	B6F2D9	0	100	100
Gibberellin-20 oxidase-2	Oryza sativa subsp. indica	F7J3E1	0	100	100
Gibberellin 20 oxidase 2	Oryza sativa subsp. japonica	Q0JH50	0	100	100
Gibberellin 20 oxidase 2	Oryza sativa subsp. indica	P0C5H5	0	100	100
SD1	Oryza sativa subsp. japonica	G8HA97	0	100	100
DNA-directed RNA polymerase subunit beta	Oryza nivara	A0A0E0FWN5	0	99	83
SD1	Oryza glumipatula	G8HA68	0	99	100
SD1	Oryza sativa subsp. japonica	G8HA95	0	99	100

Functional domains (InterPro)

Description	ID	Match position (aa)
Isopenicillin N synthase-like	IPR027443	30..363
Non-haem dioxygenase N-terminal domain	IPR026992	64..168
Oxoglutarate/iron-dependent dioxygenase	IPR005123	224..324

Gene Ontologies (GO)

Description	ID	Category
oxidoreductase activity	GO:0016491	Molecular function
oxidation-reduction process	GO:0055114	Biological process

Type	Start	End	Length
5'UTR	47	130	84 bp
CDS	131	687	557 bp
CDS	790	1111	322 bp
CDS	2583	2873	291 bp

```
>mRNA_1.1 position=query:47..2873 (+ strand)
CTGCACACAC ACACACACTC ACACTCACAC ACGCTCTCAA CTCACTCCCG CTCAACACAG CGCTCACTTC TCATCTCCAA
TCTCATGGTG GCCGAGCACC CCACGCCACC ACAGCCGCAC CAACCACCGC CCATGGACTC CACCGCCGGC TCTGGCATTG
CCGCCCCGGC GGCGGCGGCG GTGTGCGACC TGAGGATGGA GCCCAAGATC CCGGAGCCAT TCGTGTGGCC GAACGGCGAC
GCGAGGCCGG CGTCGGCGGC GGAGCTGGAC ATGCCCGTGG TCGACGTGGG CGTGCTCCGC GACGGCGACG CCGAGGGGCT
GCGCCGCGCC GCGGCGCAGG TGGCCGCCGC GTGCGCCACG CACGGGTTCT TCCAGGTGTC CGAGCACGGC GTCGACGCCG
CTCTGGCGCG CGCCGCGCTC GACGGCGCCA GCGACTTCTT CCGCCTCCCG CTCGCCGAGA AGCGCCGCGC GCGCCGCGTC
CCGGGCACCG TGTCCGGCTA CACCAGCGCC CACGCCGACC GCTTCGCCTC CAAGCTCCCA TGGAAGGAGA CCCTCTCCTT
CGGCTTCCAC GACCGCGCCG CCGCCCCCGT CGTCGCCGAC TACTTCTCCA GCACCCTCGG CCCCGACTTC GCGCCAATGG
GGTAATTAAA ACGATGGTGG ACGACATTGC ATTTCAAATT CAAAACAAAT TCAAAACACA CCGACCGAGA TTATGCTGAA
TTCAAACGCG TTTGTGCGCG CAGGAGGGTG TACCAGAAGT ACTGCGAGGA GATGAAGGAG CTGTCGCTGA CGATCATGGA
ACTCCTGGAG CTGAGCCTGG GCGTGGAGCG AGGCTACTAC AGGGAGTTCT TCGCGGACAG CAGCTCAATC ATGCGGTGCA
ACTACTACCC GCCATGCCCG GAGCCGGAGC GGACGCTCGG CACGGGCCCG CACTGCGACC CCACCGCCCT CACCATCCTC
CTCCAGGACG ACGTCGGCGG CCTCGAGGTC CTCGTCGACG GCGAATGGCG CCCCGTCAGC CCCGTCCCCG GCGCCATGGT
CATCAACATC GGCGACACCT TCATGGTAAA CCATCTCCTA TTCTCCTCTC CTCTGTTCTC CTCTGCTTCG AAGCAACAGA
ACAAGTAATT CAAGCTTTTT TTTCTCTCTC GCGCGAAATT GACGAGAAAA ATAAGATCGT GGTAGGGGCG GGGCTTTCAG
CTGAAAGCGG GAAGAAACCG ACCTGACGTG ATTTCTCTGT TCCAATCACA AACAATGGAA TGCCCCACTC CTCCATGTGT
TATGATTTAT CTCACATCTT ATAGTTAATA GGAGTAAGTA ACAAGCTACT TTtttcatat tatagttcgt ttgatttttt
ttttttaaag tttttttagt tttatccaaa tttattgaaa aacttagcaa cgtttataat accaaattag tctcatttag
tttaatattg tatatatttt gataatatat ttatgttata ttaaaaatat tactatattt ttctataaac attattaaaa
gccatttata atataaaatg gaaggagtaA TTAATATGGA TCTCCCCCGA CATGAGAATA TTTTCCGATG GTGTGACGAC
GCCATGTAAG CTTCGGTGGG CCTGGACGGC CAGAGGTGCC AACAGCCACG TCCAACAACC CCTGGGTCCC CCCCTAACAC
TCCAAACAGT AGTGAGTAGT GTCTCGTCGC GTTTTAGTAT TTGATGACAA ACAAAGTGTG AGTTGAGTTA GCCACCACCA
ACTTGCACAC GAGCACATAC ATTTGTGTCC ATTCTCGCCA GTCACTTCCA TCTCTAGTCC TAACTCCTAT CTAGCGATGT
AAGCGGATAA TTTCATCATC CGTATATAAA CCTGTTTGTT ATAGTTAATT TCCTATATAA TACTATAACA GTATACATTT
TAAAAGAAAA CAAAATTAGG ATAAACAGGC CCTGCTCCTA TCCATCCATG GCACTTGGAA GGACCAGACT CGGTCATGCC
ATGCCAAGCC AAGATATGGG TTATGGAAGA GTAGAGAAGA GGAGAGATGA GAGATAAGCA TGCGTTCTCC TCCTCGTTGG
ATGTGTATTT TGGAGGGATT TGTGTAGTAG TAGCAGCGGC GCCGCGGGGA CGGATGCGGA TGGTGGCGCT TTCGGTGGCG
TTTTCCCGGG GGGGTTTTGG TTTGGCGCTT GGGGGGGATG GCATGGCGCG GCGTGCGGCT GCACGCCACA CACACGCGCG
CGCACGCACG TACGTCGTCG TCGCCGCGGG CGGACGGTAG CTTAGGGTGG TGTGTTCCGC GCGCGGGCGC GGATTGTTCC
ATGCCGATCG ATTTGGCGCC ACCCTCGCCG CGGCTCTTGT CGCGTCGTGC GCCTCTCTCG CGCGGTTTGT CCTTGTCGCG
TTGCTCAGCC GGCGACGGGG GCACGGACAT TGGCGATGTA GCCCTGCACG TGTCGGCCTC TCCGTTGATG AATGATGATG
TATGTATGTA TTTTTTTTTG TCTGAAGGAA TTTGTGGGGA ATTGTTGTGT GTGCAGGCGC TGTCGAACGG GAGGTATAAG
AGCTGCCTGC ACAGGGCGGT GGTGAACCAG CGGCGGGAGC GGCGGTCGCT GGCGTTCTTC CTGTGCCCGC GGGAGGACAG
GGTGGTGCGG CCGCCGCCGA GCGCCGCCAC GCCGCAGCAC TACCCGGACT TCACCTGGGC CGACCTCATG CGCTTCACGC
AGCGCCACTA CCGCGCCGAC ACCCGCACGC TCGACGCCTT CACGCGCTGG CTCGCGCCGC CGGCCGCCGA CGCCGCCGCG
ACGGCGCAGG TCGAGGCGGC CAGCTGA
```

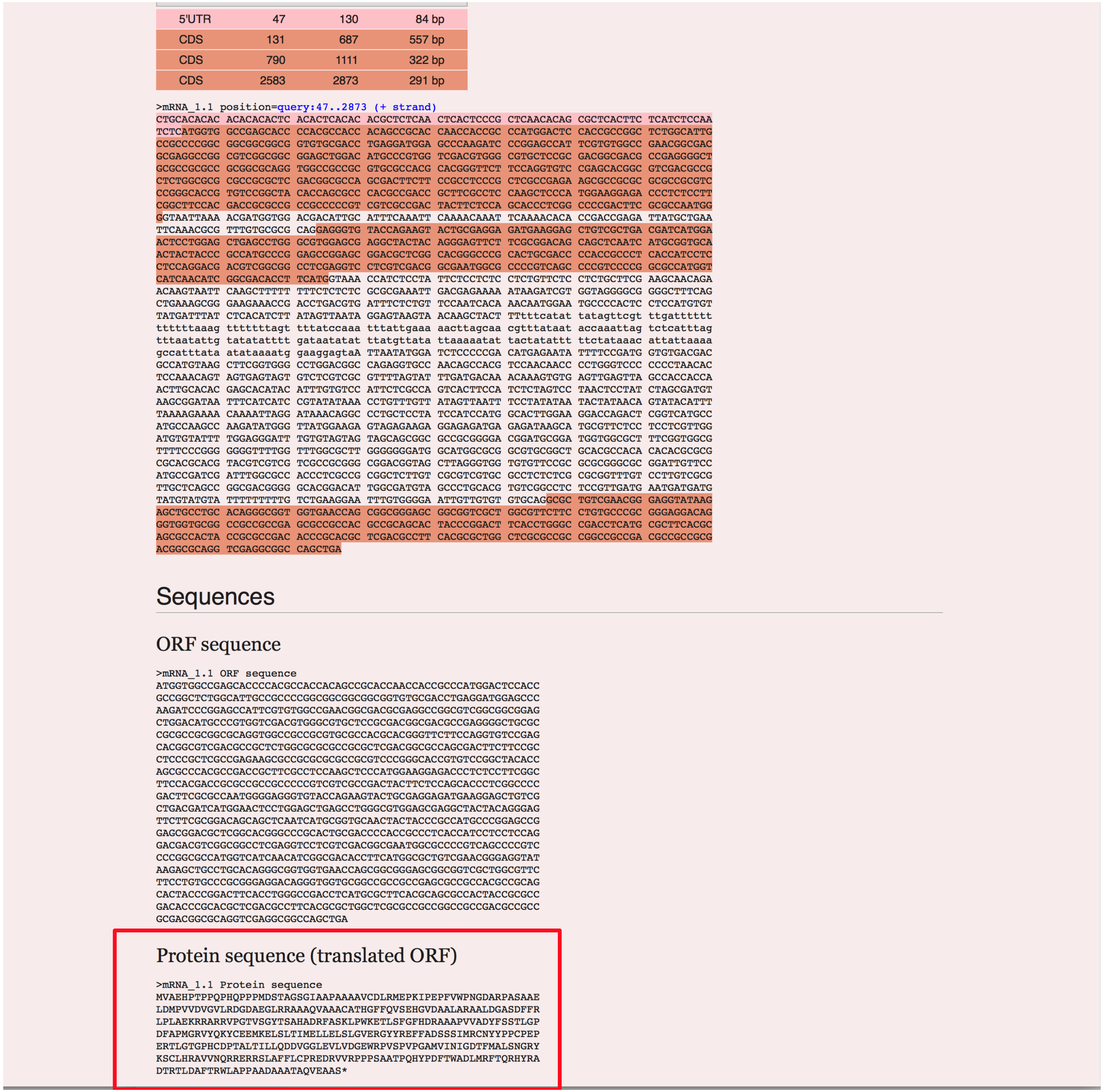
Sequences

ORF sequence

```
>mRNA_1.1 ORF sequence
ATGGTGGCCGAGCACCCCACGCCACCACAGCCGCACCAACCACCGCCCATGGACTCCACC
GCCGGCTCTGGCATTGCCGCCCCGGCGGCGGCGGCGGTGTGCGACCTGAGGATGGAGCCC
AAGATCCCGGAGCCATTCGTGTGGCCGAACGGCGACGCGAGGCCGGCGTCGGCGGCGGAG
CTGGACATGCCCGTGGTCGACGTGGGCGTGCTCCGCGACGGCGACGCCGAGGGGCTGCGC
CGCGCCGCGGCGCAGGTGGCCGCCGCGTGCGCCACGCACGGGTTCTTCCAGGTGTCCGAG
CACGGCGTCGACGCCGCTCTGGCGCGCGCCGCGCTCGACGGCGCCAGCGACTTCTTCCGC
CTCCCGCTCGCCGAGAAGCGCCGCGCGCGCCGCGTCCCGGGCACCGTGTCCGGCTACACC
AGCGCCCACGCCGACCGCTTCGCCTCCAAGCTCCCATGGAAGGAGACCCTCTCCTTCGGC
TTCCACGACCGCGCCGCCGCCCCCGTCGTCGCCGACTACTTCTCCAGCACCCTCGGCCCC
GACTTCGCGCCAATGGGGAGGGTGTACCAGAAGTACTGCGAGGAGATGAAGGAGCTGTCG
CTGACGATCATGGAACTCCTGGAGCTGAGCCTGGGCGTGGAGCGAGGCTACTACAGGGAG
TTCTTCGCGGACAGCAGCTCAATCATGCGGTGCAACTACTACCCGCCATGCCCGGAGCCG
GAGCGGACGCTCGGCACGGGCCCGCACTGCGACCCCACCGCCCTCACCATCCTCCTCCAG
GACGACGTCGGCGGCCTCGAGGTCCTCGTCGACGGCGAATGGCGCCCCGTCAGCCCCGTC
CCCGGCGCCATGGTCATCAACATCGGCGACACCTTCATGGCGCTGTCGAACGGGAGGTAT
AAGAGCTGCCTGCACAGGGCGGTGGTGAACCAGCGGCGGGAGCGGCGGTCGCTGGCGTTC
TTCCTGTGCCCGCGGGAGGACAGGGTGGTGCGGCCGCCGCCGAGCGCCGCCACGCCGCAG
CACTACCCGGACTTCACCTGGGCCGACCTCATGCGCTTCACGCAGCGCCACTACCGCGCC
GACACCCGCACGCTCGACGCCTTCACGCGCTGGCTCGCGCCGCCGGCCGCCGACGCCGCC
GCGACGGCGCAGGTCGAGGCGGCCAGCTGA
```

Protein sequence (translated ORF)

```
>mRNA_1.1 Protein sequence
MVAEHPTPPQPHQPPPMDSTAGSGIAAPAAAAVCDLRMEPKIPEPFVWPNGDARPASAAE
LDMPVVDVGVLRDGDAEGLRRAAAQVAAACATHGFFQVSEHGVDAALARAALDGASDFFR
LPLAEKRRARRVPGTVSGYTSAHADRFASKLPWKETLSFGFHDRAAAPVVADYFSSTLGP
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVERGYYREFFADSSSIMRCNYYPPCPEP
```



5'UTR	47	130	84 bp
CDS	131	687	557 bp
CDS	790	1111	322 bp
CDS	2583	2873	291 bp

```
>mRNA_1.1 position=query:47..2873 (+ strand)
CTGCACACAC ACACACACTC ACACTCACAC ACGCTCTCAA CTCACTCCCG CTCAACACAG CGCTCACTTC TCATCTCCAA
TCTCATGGTG GCCGAGCACC CCACGCCACC ACAGCCGCAC CAACCACCGC CCATGGACTC CACCGCCGGC TCTGGCATTG
CCGCCCCGGC GGCGGCGGCG GTGTGCGACC TGAGGATGGA GCCCAAGATC CCGGAGCCAT TCGTGTGGCC GAACGGCGAC
GCGAGGCCGG CGTCGGCGGC GGAGCTGGAC ATGCCCGTGG TCGACGTGGG CGTGCTCCGC GACGGCGACG CCGAGGGGCT
GCGCCGCGCC GCGGCGCAGG TGGCCGCCGC GTGCGCCACG CACGGGTTCT TCCAGGTGTC CGAGCACGGC GTCGACGCCG
CTCTGGCGCG CGCCGCGCTC GACGGCGCCA GCGACTTCTT CCGCCTCCCG CTCGCCGAGA AGCGCCGCGC GCGCCGCGTC
CCGGGCACCG TGTCCGGCTA CACCAGCGCC CACGCCGACC GCTTCGCCTC CAAGCTCCCA TGGAAGGAGA CCCTCTCCTT
CGGCTTCCAC GACCGCGCCG CCGCCCCCGT CGTCGCCGAC TACTTCTCCA GCACCCTCGG CCCCGACTTC GCGCCAATGG
GGTAATTAAA ACGATGGTGG ACGACATTGC ATTTCAAATT CAAAACAAAT TCAAAACACA CCGACCGAGA TTATGCTGAA
TTCAAACGCG TTTGTGCGCG CAGGAGGGTG TACCAGAAGT ACTGCGAGGA GATGAAGGAG CTGTCGCTGA CGATCATGGA
ACTCCTGGAG CTGAGCCTGG GCGTGGAGCG AGGCTACTAC AGGGAGTTCT TCGCGGACAG CAGCTCAATC ATGCGGTGCA
ACTACTACCC GCCATGCCCG GAGCCGGAGC GGACGCTCGG CACGGGCCCG CACTGCGACC CCACCGCCCT CACCATCCTC
CTCCAGGACG ACGTCGGCGG CCTCGAGGTC CTCGTCGACG GCGAATGGCG CCCCGTCAGC CCCGTCCCCG GCGCCATGGT
CATCAACATC GGCGACACCT TCATGGTAAA CCATCTCCTA TTCTCCTCTC CTCTGTTCTC CTCTGCTTCG AAGCAACAGA
ACAAGTAATT CAAGCTTTTT TTTCTCTCTC GCGCGAAATT GACGAGAAAA ATAAGATCGT GGTAGGGGCG GGGCTTTCAG
CTGAAAGCGG GAAGAAACCG ACCTGACGTG ATTTCTCTGT TCCAATCACA AACAATGGAA TGCCCCACTC CTCCATGTGT
TATGATTTAT CTCACATCTT ATAGTTAATA GGAGTAAGTA ACAAGCTACT TTtttcatat tatagttcgt ttgatttttt
tttttaaag ttttttagt tttatccaaa tttattgaaa aacttagcaa cgtttataat accaaattag tctcatttag
tttaatattg tatatatttt gataatatat ttatgttata ttaaaaatat tactatattt ttctataaac attattaaaa
gccatttata atataaaatg gaaggagtaA TTAATATGGA TCTCCCCCGA CATGAGAATA TTTTCCGATG GTGTGACGAC
GCCATGTAAG CTTCGGTGGG CCTGGACGGC CAGAGGTGCC AACAGCCACG TCCAACAACC CCTGGGTCCC CCCCTAACAC
TCCAAACAGT AGTGAGTAGT GTCTCGTCGC GTTTTAGTAT TTGATGACAA ACAAAGTGTG AGTTGAGTTA GCCACCACCA
ACTTGCACAC GAGCACATAC ATTTGTGTCC ATTCTCGCCA GTCACTTCCA TCTCTAGTCC TAACTCCTAT CTAGCGATGT
AAGCGGATAA TTTCATCATC CGTATATAAA CCTGTTTGTT ATAGTTAATT TCCTATATAA TACTATAACA GTATACATTT
TAAAAGAAAA CAAAATTAGG ATAAACAGGC CCTGCTCCTA TCCATCCATG GCACTTGGAA GGACCAGACT CGGTCATGCC
ATGCCAAGCC AAGATATGGG TTATGGAAGA GTAGAGAAGA GGAGAGATGA GAGATAAGCA TGCGTTCTCC TCCTCGTTGG
ATGTGTATTT TGGAGGGATT TGTGTAGTAG TAGCAGCGGC GCCGCGGGGA CGGATGCGGA TGGTGGCGCT TTCGGTGGCG
TTTTCCCGGG GGGGTTTTGG TTTGGCGCTT GGGGGGGATG GCATGGCGCG GCGTGCGGCT GCACGCCACA CACACGCGCG
CGCACGCACG TACGTCGTCG TCGCCGCGGG CGGACGGTAG CTTAGGGTGG TGTGTTCCGC GCGCGGGCGC GGATTGTTCC
ATGCCGATCG ATTTGGCGCC ACCCTCGCCG CGGCTCTTGT CGCGTCGTGC GCCTCTCTCG CGCGGTTTGT CCTTGTCGCG
TTGCTCAGCC GGCGACGGGG GCACGGACAT TGGCGATGTA GCCCTGCACG TGTCGGCCTC TCCGTTGATG AATGATGATG
TATGTATGTA TTTTTTTTTG TCTGAAGGAA TTTGTGGGGA ATTGTTGTGT GTGCAGGCGC TGTCGAACGG GAGGTATAAG
AGCTGCCTGC ACAGGGCGGT GGTGAACCAG CGGCGGGAGC GGCGGTCGCT GGCGTTCTTC CTGTGCCCGC GGGAGGACAG
GGTGGTGCGG CCGCCGCCGA GCGCCGCCAC GCCGCAGCAC TACCCGGACT TCACCTGGGC CGACCTCATG CGCTTCACGC
AGCGCCACTA CCGCGCCGAC ACCCGCACGC TCGACGCCTT CACGCGCTGG CTCGCGCCGC CGGCCGCCGA CGCCGCCGCG
ACGGCGCAGG TCGAGGCGGC CAGCTGA
```

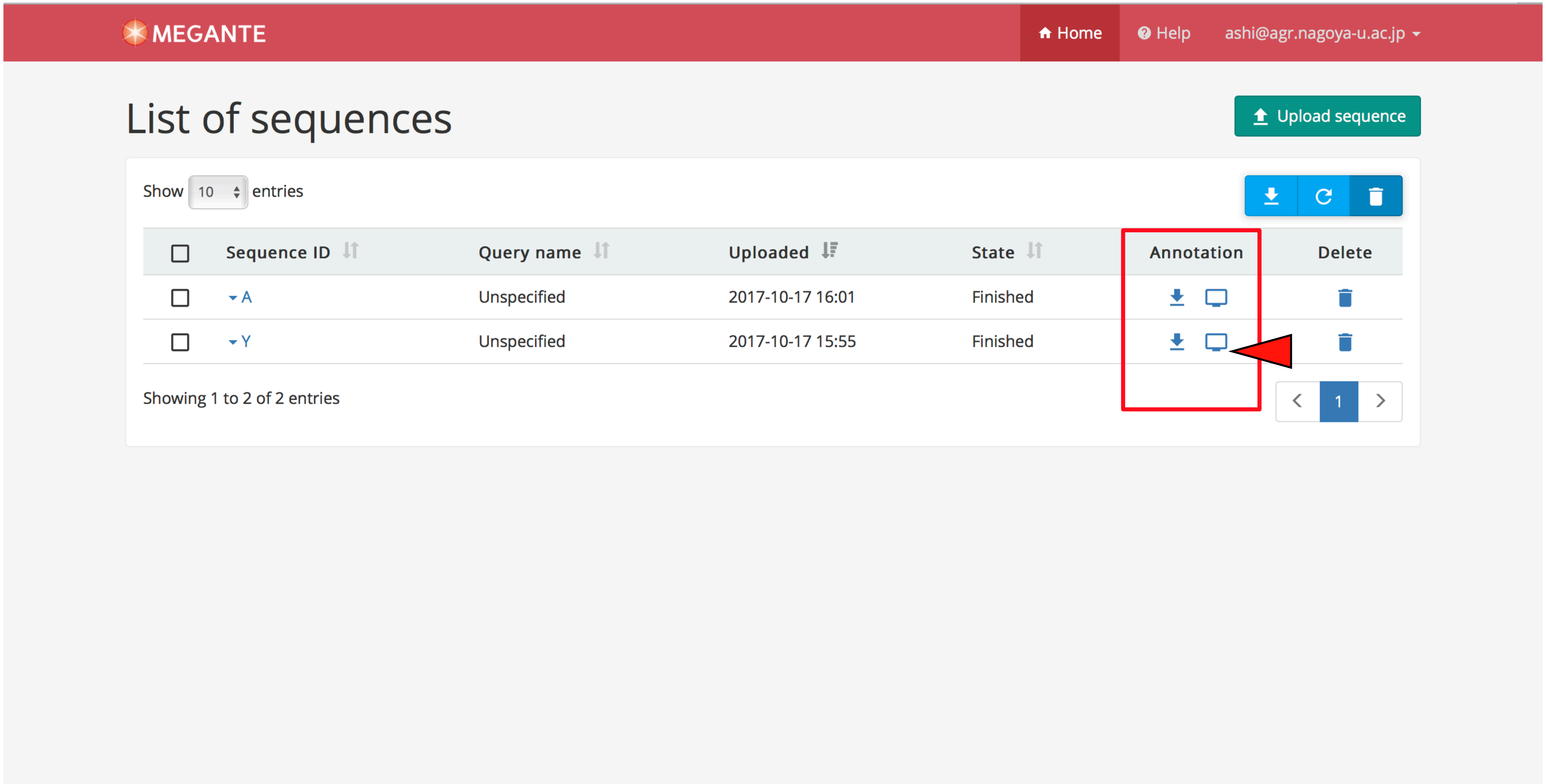
Sequences

ORF sequence

```
>mRNA_1.1 ORF sequence
ATGGTGGCCGAGCACCCCACGCCACCACAGCCGCACCAACCACCGCCCATGGACTCCACC
GCCGGCTCTGGCATTGCCGCCCCGGCGGCGGCGGCGGTGTGCGACCTGAGGATGGAGCCC
AAGATCCCGGAGCCATTCGTGTGGCCGAACGGCGACGCGAGGCCGGCGTCGGCGGCGGAG
CTGGACATGCCCGTGGTCGACGTGGGCGTGCTCCGCGACGGCGACGCCGAGGGGCTGCGC
CGCGCCGCGGCGCAGGTGGCCGCCGCGTGCGCCACGCACGGGTTCTTCCAGGTGTCCGAG
CACGGCGTCGACGCCGCTCTGGCGCGCGCCGCGCTCGACGGCGCCAGCGACTTCTTCCGC
CTCCCGCTCGCCGAGAAGCGCCGCGCGCGCCGCGTCCCGGGCACCGTGTCCGGCTACACC
AGCGCCCACGCCGACCGCTTCGCCTCCAAGCTCCCATGGAAGGAGACCCTCTCCTTCGGC
TTCCACGACCGCGCCGCCGCCCCCGTCGTCGCCGACTACTTCTCCAGCACCCTCGGCCCC
GACTTCGCGCCAATGGGGAGGGTGTACCAGAAGTACTGCGAGGAGATGAAGGAGCTGTCG
CTGACGATCATGGAACTCCTGGAGCTGAGCCTGGGCGTGGAGCGAGGCTACTACAGGGAG
TTCTTCGCGGACAGCAGCTCAATCATGCGGTGCAACTACTACCCGCCATGCCCGGAGCCG
GAGCGGACGCTCGGCACGGGCCCGCACTGCGACCCCACCGCCCTCACCATCCTCCTCCAG
GACGACGTCGGCGGCCTCGAGGTCCTCGTCGACGGCGAATGGCGCCCCGTCAGCCCCGTC
CCCGGCGCCATGGTCATCAACATCGGCGACACCTTCATGGCGCTGTCGAACGGGAGGTAT
AAGAGCTGCCTGCACAGGGCGGTGGTGAACCAGCGGCGGGAGCGGCGGTCGCTGGCGTTC
TTCCTGTGCCCGCGGGAGGACAGGGTGGTGCGGCCGCCGCCGAGCGCCGCCACGCCGCAG
CACTACCCGGACTTCACCTGGGCCGACCTCATGCGCTTCACGCAGCGCCACTACCGCGCC
GACACCCGCACGCTCGACGCCTTCACGCGCTGGCTCGCGCCGCCGGCCGCCGACGCCGCC
GCGACGGCGCAGGTCGAGGCGGCCAGCTGA
```

Protein sequence (translated ORF)

```
>mRNA_1.1 Protein sequence
MVAEHPTPPQPHQPPPMDSTAGSGIAAPAAAAVCDLRMEPKIPEPFVWPNGDARPASAAE
LDMPVVDVGVLRDGDAEGLRRAAAQVAAACATHGFFQVSEHGVDAALARAALDGASDFFR
LPLAEKRRARRVPGTVSGYTSAHADRFASKLPWKETLSFGFHDRAAAPVVADYFSSTLGP
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVERGYYREFFADSSSIMRCNYYPPCPEP
ERTLGTGPHCDPTALTILLQDDVGGLEVLVDGEWRPVSPVPGAMVINIGDTFMALSNGRY
KSCLHRAVVNQRRERRSLAFFLCPREDRVVRPPPSAATPQHYPDFTWADLMRFTQRHYRA
DTRTLDAFTRWLAPPAADAAATAQVEAAS*
```



List of sequences

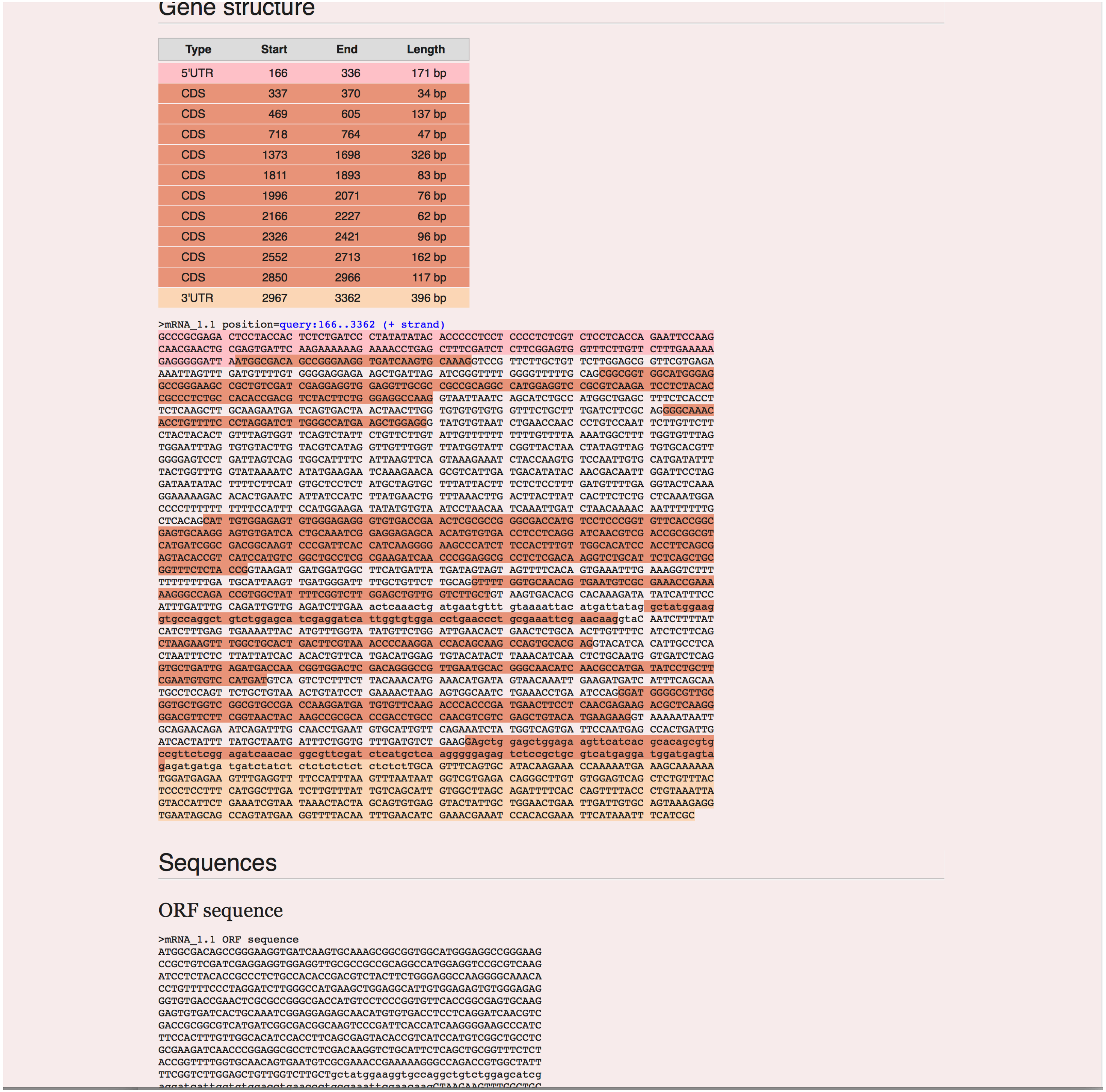
☐	Sequence ID	Query name	Uploaded	State	Annotation	Delete
☐	A	Unspecified	2017-10-17 16:01	Finished		
☐	Y	Unspecified	2017-10-17 15:55	Finished		

Showing 1 to 2 of 2 entries

Y: 3.42 kbp の範囲を query から表示、塩基番号 1 から 3,420

ランドマークまたは領域: query:1..3,420

データソース: Y



Gene structure

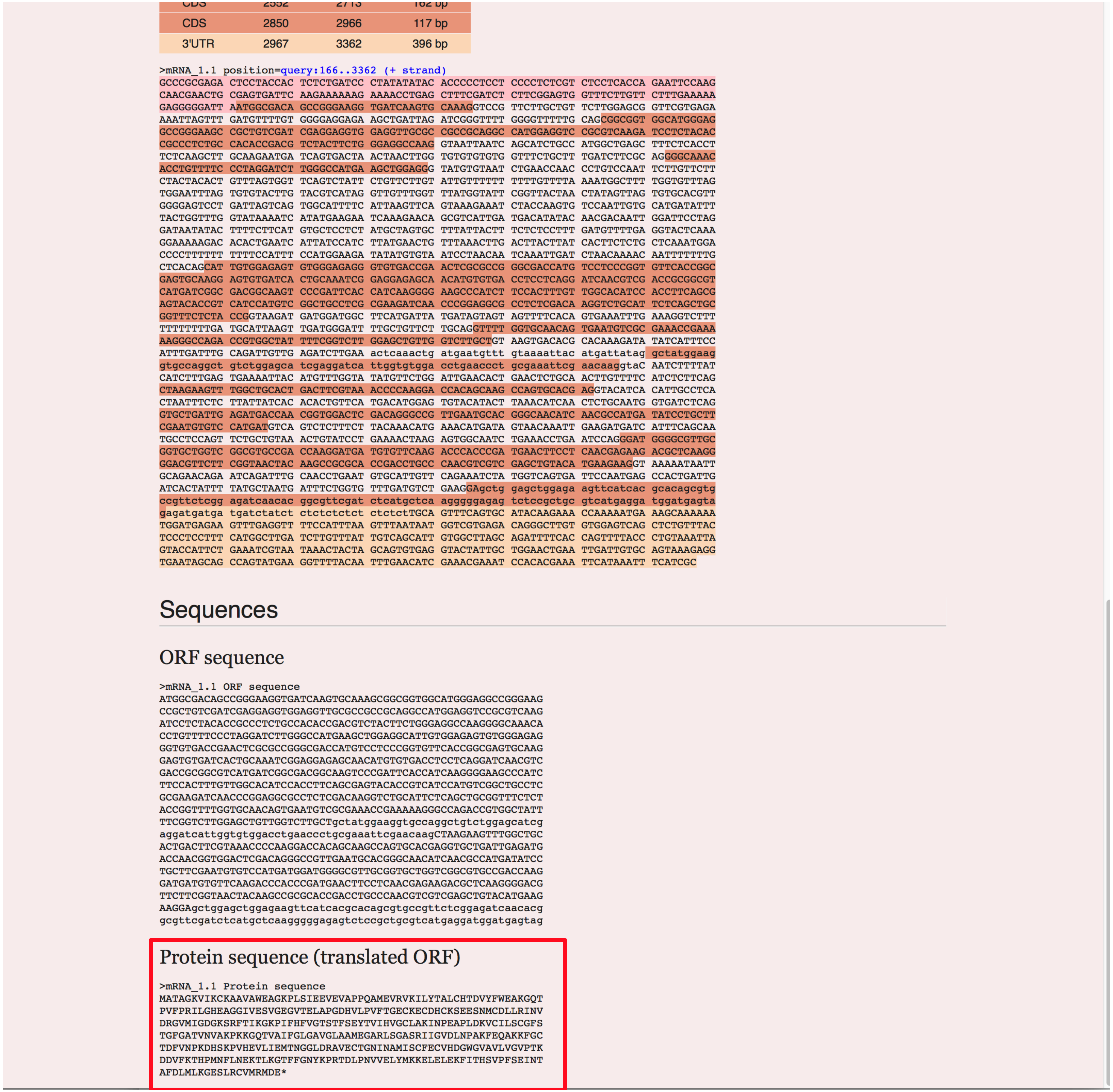
Type	Start	End	Length
5'UTR	166	336	171 bp
CDS	337	370	34 bp
CDS	469	605	137 bp
CDS	718	764	47 bp
CDS	1373	1698	326 bp
CDS	1811	1893	83 bp
CDS	1996	2071	76 bp
CDS	2166	2227	62 bp
CDS	2326	2421	96 bp
CDS	2552	2713	162 bp
CDS	2850	2966	117 bp
3'UTR	2967	3362	396 bp

```
>mRNA_1.1 position=query:166..3362 (+ strand)
GCCCGCGAGA CTCCTACCAC TCTCTGATCC CTATATATAC ACCCCCTCCT CCCCTCTCGT CTCCTCACCA GAATTCCAAG
CAACGAACTG CGAGTGATTC AAGAAAAAAG AAAACCTGAG CTTTCGATCT CTTCGGAGTG GTTTCTTGTT CTTTGAAAAA
GAGGGGGATT AATGGCGACA GCCGGGAAGG TGATCAAGTG CAAAGGTCCG TTCTTGCTGT TCTTGGAGCG GTTCGTGAGA
AAATTAGTTT GATGTTTTGT GGGGAGGAGA AGCTGATTAG ATCGGGTTTT GGGGTTTTTG CAGCGGCGGT GGCATGGGAG
GCCGGGAAGC CGCTGTCGAT CGAGGAGGTG GAGGTTGCGC CGCCGCAGGC CATGGAGGTC CGCGTCAAGA TCCTCTACAC
CGCCCTCTGC CACACCGACG TCTACTTCTG GGAGGCCAAG GTAATTAATC AGCATCTGCC ATGGCTGAGC TTTCTCACCT
TCTCAAGCTT GCAAGAATGA TCAGTGACTA ACTAACTTGG TGTGTGTGTG GTTTCTGCTT TGATCTTCGC AGGGGCAAAC
ACCTGTTTTC CCTAGGATCT TGGGCCATGA AGCTGGAGGG TATGTGTAAT CTGAACCAAC CCTGTCCAAT TCTTGTTCTT
CTACTACACT GTTTAGTGGT TCAGTCTATT CTGTTCTTGT ATTGTTTTTT TTTTGTTTTA AAATGGCTTT TGGTGTTTAG
TGGAATTTAG TGTGTACTTG TACGTCATAG GTTGTTTGGT TTATGGTATT CGGTTACTAA CTATAGTTAG TGTGCACGTT
GGGGAGTCCT GATTAGTCAG TGGCATTTTC ATTAAGTTCA GTAAAGAAAT CTACCAAGTG TCCAATTGTG CATGATATTT
TACTGGTTTG GTATAAAATC ATATGAAGAA TCAAAGAACA GCGTCATTGA TGACATATAC AACGACAATT GGATTCCTAG
GATAATATAC TTTTCTTCAT GTGCTCCTCT ATGCTAGTGC TTTATTACTT TCTCTCCTTT GATGTTTTGA GGTACTCAAA
GGAAAAAGAC ACACTGAATC ATTATCCATC TTATGAACTG TTTAAACTTG ACTTACTTAT CACTTCTCTG CTCAAATGGA
CCCCTTTTTT TTTTCCATTT CCATGGAAGA TATATGTGTA ATCCTAACAA TCAAATTGAT CTAACAAAAC AATTTTTTTG
CTCACAGCAT TGTGGAGAGT GTGGGAGAGG GTGTGACCGA ACTCGCGCCG GGCGACCATG TCCTCCCGGT GTTCACCGGC
GAGTGCAAGG AGTGTGATCA CTGCAAATCG GAGGAGAGCA ACATGTGTGA CCTCCTCAGG ATCAACGTCG ACCGCGGCGT
CATGATCGGC GACGGCAAGT CCCGATTCAC CATCAAGGGG AAGCCCATCT TCCACTTTGT TGGCACATCC ACCTTCAGCG
AGTACACCGT CATCCATGTC GGCTGCCTCG CGAAGATCAA CCCGGAGGCG CCTCTCGACA AGGTCTGCAT TCTCAGCTGC
GGTTTCTCTA CCGGTAAGAT GATGGATGGC TTCATGATTA TGATAGTAGT AGTTTTCACA GTGAAATTTG AAAGGTCTTT
TTTTTTTTGA TGCATTAAGT TGATGGGATT TTGCTGTTCT TGCAGGTTTT GGTGCAACAG TGAATGTCGC GAAACCGAAA
AAGGGCCAGA CCGTGGCTAT TTTCGGTCTT GGAGCTGTTG GTCTTGCTGT AAGTGACACG CACAAAGATA TATCATTTCC
ATTTGATTTG CAGATTGTTG AGATCTTGAA actcaaactg atgaatgttt gtaaaattac atgattatag gctatggaag
gtgccaggct gtctggagca tcgaggatca ttggtgtgga cctgaaccct gcgaaattcg aacaaggtaC AATCTTTTAT
CATCTTTGAG TGAAAATTAC ATGTTTGGTA TATGTTCTGG ATTGAACACT GAACTCTGCA ACTTGTTTTC ATCTCTTCAG
CTAAGAAGTT TGGCTGCACT GACTTCGTAA ACCCCAAGGA CCACAGCAAG CCAGTGCACG AGGTACATCA CATTGCCTCA
CTAATTTCTC TTATTATCAC ACACTGTTCA TGACATGGAG TGTACATACT TAAACATCAA CTCTGCAATG GTGATCTCAG
GTGCTGATTG AGATGACCAA CGGTGGACTC GACAGGGCCG TTGAATGCAC GGGCAACATC AACGCCATGA TATCCTGCTT
CGAATGTGTC CATGATGTCA GTCTCTTTCT TACAAACATG AAACATGATA GTAACAAATT GAAGATGATC ATTTCAGCAA
TGCCTCCAGT TCTGCTGTAA ACTGTATCCT GAAAACTAAG AGTGGCAATC TGAAACCTGA ATCCAGGGAT GGGGCGTTGC
GGTGCTGGTC GGCGTGCCGA CCAAGGATGA TGTGTTCAAG ACCCACCCGA TGAACTTCCT CAACGAGAAG ACGCTCAAGG
GGACGTTCTT CGGTAACTAC AAGCCGCGCA CCGACCTGCC CAACGTCGTC GAGCTGTACA TGAAGAAGGT AAAAATAATT
GCAGAACAGA ATCAGATTTG CAACCTGAAT GTGCATTGTT CAGAAATCTA TGGTCAGTGA TTCCAATGAG CCACTGATTG
ATCACTATTT TATGCTAATG ATTTCTGGTG TTTGATGTCT GAAGGAgctg gagctggaga agttcatcac gcacagcgtg
ccgttctcgg agatcaacac ggcgttcgat ctcatgctca agggggagag tctccgctgc gtcatgagga tggatgagta
gagatgatga tgatctatct ctctctctct ctctctTGCA GTTTCAGTGC ATACAAGAAA CCAAAAATGA AAGCAAAAAA
TGGATGAGAA GTTTGAGGTT TTCCATTTAA GTTTAATAAT GGTCGTGAGA CAGGGCTTGT GTGGAGTCAG CTCTGTTTAC
TCCCTCCTTT CATGGCTTGA TCTTGTTTAT TGTCAGCATT GTGGCTTAGC AGATTTTCAC CAGTTTTACC CTGTAAATTA
GTACCATTCT GAAATCGTAA TAAACTACTA GCAGTGTGAG GTACTATTGC TGGAACTGAA TTGATTGTGC AGTAAAGAGG
TGAATAGCAG CCAGTATGAA GGTTTTACAA TTTGAACATC GAAACGAAAT CCACACGAAA TTCATAAATT TCATCGC
```

Sequences

ORF sequence

```
>mRNA_1.1 ORF sequence
ATGGCGACAGCCGGGAAGGTGATCAAGTGCAAAGCGGCGGTGGCATGGGAGGCCGGGAAG
CCGCTGTCGATCGAGGAGGTGGAGGTTGCGCCGCCGCAGGCCATGGAGGTCCGCGTCAAG
ATCCTCTACACCGCCCTCTGCCACACCGACGTCTACTTCTGGGAGGCCAAGGGGCAAACA
CCTGTTTTCCCTAGGATCTTGGGCCATGAAGCTGGAGGCATTGTGGAGAGTGTGGGAGAG
GGTGTGACCGAACTCGCGCCGGGCGACCATGTCCTCCCGGTGTTCACCGGCGAGTGCAAG
GAGTGTGATCACTGCAAATCGGAGGAGAGCAACATGTGTGACCTCCTCAGGATCAACGTC
GACCGCGGCGTCATGATCGGCGACGGCAAGTCCCGATTCACCATCAAGGGGAAGCCCATC
TTCCACTTTGTTGGCACATCCACCTTCAGCGAGTACACCGTCATCCATGTCGGCTGCCTC
GCGAAGATCAACCCGGAGGCGCCTCTCGACAAGGTCTGCATTCTCAGCTGCGGTTTCTCT
ACCGGTTTTGGTGCAACAGTGAATGTCGCGAAACCGAAAAAGGGCCAGACCGTGGCTATT
TTCGGTCTTGGAGCTGTTGGTCTTGCTgctatggaaggtgccaggctgtctggagcatcg
aggatcattggtgtggacctgaaccctgcgaaattcgaacaagCTAAGAAGTTTGGCTGC
```



CDS	2552	2713	162 bp
CDS	2850	2966	117 bp
3'UTR	2967	3362	396 bp

```
>mRNA_1.1 position=query:166..3362 (+ strand)
GCCCGCGAGA CTCCTACCAC TCTCTGATCC CTATATATAC ACCCCCTCCT CCCCTCTCGT CTCCTCACCA GAATTCCAAG
CAACGAACTG CGAGTGATTC AAGAAAAAAG AAAACCTGAG CTTTCGATCT CTTCGGAGTG GTTTCTTGTT CTTTGAAAAA
GAGGGGGATT AATGGCGACA GCCGGGAAGG TGATCAAGTG CAAAGGTCCG TTCTTGCTGT TCTTGGAGCG GTTCGTGAGA
AAATTAGTTT GATGTTTTGT GGGGAGGAGA AGCTGATTAG ATCGGGTTTT GGGGTTTTTG CAGCGGCGGT GGCATGGGAG
GCCGGGAAGC CGCTGTCGAT CGAGGAGGTG GAGGTTGCGC CGCCGCAGGC CATGGAGGTC CGCGTCAAGA TCCTCTACAC
CGCCCTCTGC CACACCGACG TCTACTTCTG GGAGGCCAAG GTAATTAATC AGCATCTGCC ATGGCTGAGC TTTCTCACCT
TCTCAAGCTT GCAAGAATGA TCAGTGACTA ACTAACTTGG TGTGTGTGTG GTTTCTGCTT TGATCTTCGC AGGGGCAAAC
ACCTGTTTTC CCTAGGATCT TGGGCCATGA AGCTGGAGGG TATGTGTAAT CTGAACCAAC CCTGTCCAAT TCTTGTTCTT
CTACTACACT GTTTAGTGGT TCAGTCTATT CTGTTCTTGT ATTGTTTTTT TTTTGTTTTA AAATGGCTTT TGGTGTTTAG
TGGAATTTAG TGTGTACTTG TACGTCATAG GTTGTTTGGT TTATGGTATT CGGTTACTAA CTATAGTTAG TGTGCACGTT
GGGGAGTCCT GATTAGTCAG TGGCATTTTC ATTAAGTTCA GTAAAGAAAT CTACCAAGTG TCCAATTGTG CATGATATTT
TACTGGTTTG GTATAAAATC ATATGAAGAA TCAAAGAACA GCGTCATTGA TGACATATAC AACGACAATT GGATTCCTAG
GATAATATAC TTTTCTTCAT GTGCTCCTCT ATGCTAGTGC TTTATTACTT TCTCTCCTTT GATGTTTTGA GGTACTCAAA
GGAAAAAGAC ACACTGAATC ATTATCCATC TTATGAACTG TTTAAACTTG ACTTACTTAT CACTTCTCTG CTCAAATGGA
CCCCTTTTTT TTTTCCATTT CCATGGAAGA TATATGTGTA ATCCTAACAA TCAAATTGAT CTAACAAAAC AATTTTTTTG
CTCACAGCAT TGTGGAGAGT GTGGGAGAGG GTGTGACCGA ACTCGCGCCG GGCGACCATG TCCTCCCGGT GTTCACCGGC
GAGTGCAAGG AGTGTGATCA CTGCAAATCG GAGGAGAGCA ACATGTGTGA CCTCCTCAGG ATCAACGTCG ACCGCGGCGT
CATGATCGGC GACGGCAAGT CCCGATTCAC CATCAAGGGG AAGCCCATCT TCCACTTTGT TGGCACATCC ACCTTCAGCG
AGTACACCGT CATCCATGTC GGCTGCCTCG CGAAGATCAA CCCGGAGGCG CCTCTCGACA AGGTCTGCAT TCTCAGCTGC
GGTTTCTCTA CCGGTAAGAT GATGGATGGC TTCATGATTA TGATAGTAGT AGTTTTCACA GTGAAATTTG AAAGGTCTTT
TTTTTTTTGA TGCATTAAGT TGATGGGATT TTGCTGTTCT TGCAGGTTTT GGTGCAACAG TGAATGTCGC GAAACCGAAA
AAGGGCCAGA CCGTGGCTAT TTTCGGTCTT GGAGCTGTTG GTCTTGCTGT AAGTGACACG CACAAAGATA TATCATTTCC
ATTTGATTTG CAGATTGTTG AGATCTTGAA actcaaactg atgaatgttt gtaaaattac atgattatag gctatggaag
gtgccaggct gtctggagca tcgaggatca ttggtgtgga cctgaaccct gcgaaattcg aacaaggtaC AATCTTTTAT
CATCTTTGAG TGAAAATTAC ATGTTTGGTA TATGTTCTGG ATTGAACACT GAACTCTGCA ACTTGTTTTC ATCTCTTCAG
CTAAGAAGTT TGGCTGCACT GACTTCGTAA ACCCCAAGGA CCACAGCAAG CCAGTGCACG AGGTACATCA CATTGCCTCA
CTAATTTCTC TTATTATCAC ACACTGTTCA TGACATGGAG TGTACATACT TAAACATCAA CTCTGCAATG GTGATCTCAG
GTGCTGATTG AGATGACCAA CGGTGGACTC GACAGGGCCG TTGAATGCAC GGGCAACATC AACGCCATGA TATCCTGCTT
CGAATGTGTC CATGATGTCA GTCTCTTTCT TACAAACATG AAACATGATA GTAACAAATT GAAGATGATC ATTTCAGCAA
TGCCTCCAGT TCTGCTGTAA ACTGTATCCT GAAAACTAAG AGTGGCAATC TGAAACCTGA ATCCAGGGAT GGGGCGTTGC
GGTGCTGGTC GGCGTGCCGA CCAAGGATGA TGTGTTCAAG ACCCACCCGA TGAACTTCCT CAACGAGAAG ACGCTCAAGG
GGACGTTCTT CGGTAACTAC AAGCCGCGCA CCGACCTGCC CAACGTCGTC GAGCTGTACA TGAAGAAGGT AAAAATAATT
GCAGAACAGA ATCAGATTTG CAACCTGAAT GTGCATTGTT CAGAAATCTA TGGTCAGTGA TTCCAATGAG CCACTGATTG
ATCACTATTT TATGCTAATG ATTTCTGGTG TTTGATGTCT GAAGGAgctg gagctggaga agttcatcac gcacagcgtg
ccgttctcgg agatcaacac ggcgttcgat ctcatgctca aggggagag tctccgctgc gtcatgagga tggatgagta
gagatgatga tgatctatct ctctctctct ctctctTGCA GTTTCAGTGC ATACAAGAAA CCAAAAATGA AAGCAAAAAA
TGGATGAGAA GTTTGAGGTT TTCCATTTAA GTTTAATAAT GGTCGTGAGA CAGGGCTTGT GTGGAGTCAG CTCTGTTTAC
TCCCTCCTTT CATGGCTTGA TCTTGTTTAT TGTCAGCATT GTGGCTTAGC AGATTTTCAC CAGTTTTACC CTGTAAATTA
GTACCATTCT GAAATCGTAA TAAACTACTA GCAGTGTGAG GTACTATTGC TGGAACTGAA TTGATTGTGC AGTAAAGAGG
TGAATAGCAG CCAGTATGAA GGTTTTACAA TTTGAACATC GAAACGAAAT CCACACGAAA TTCATAAATT TCATCGC
```

Sequences

ORF sequence

```
>mRNA_1.1 ORF sequence
ATGGCGACAGCCGGGAAGGTGATCAAGTGCAAAGCGGCGGTGGCATGGGAGGCCGGGAAG
CCGCTGTCGATCGAGGAGGTGGAGGTTGCGCCGCCGCAGGCCATGGAGGTCCGCGTCAAG
ATCCTCTACACCGCCCTCTGCCACACCGACGTCTACTTCTGGGAGGCCAAGGGGCAAACA
CCTGTTTTCCCTAGGATCTTGGGCCATGAAGCTGGAGGCATTGTGGAGAGTGTGGGAGAG
GGTGTGACCGAACTCGCGCCGGGCGACCATGTCCTCCCGGTGTTCACCGGCGAGTGCAAG
GAGTGTGATCACTGCAAATCGGAGGAGAGCAACATGTGTGACCTCCTCAGGATCAACGTC
GACCGCGGCGTCATGATCGGCGACGGCAAGTCCCGATTCACCATCAAGGGGAAGCCCATC
TTCCACTTTGTTGGCACATCCACCTTCAGCGAGTACACCGTCATCCATGTCGGCTGCCTC
GCGAAGATCAACCCGGAGGCGCCTCTCGACAAGGTCTGCATTCTCAGCTGCGGTTTCTCT
ACCGGTTTTGGTGCAACAGTGAATGTCGCGAAACCGAAAAAGGGCCAGACCGTGGCTATT
TTCGGTCTTGGAGCTGTTGGTCTTGCTgctatggaaggtgccaggctgtctggagcatcg
aggatcattggtgtggacctgaaccctgcgaaattcgaacaagCTAAGAAGTTTGGCTGC
ACTGACTTCGTAAACCCCAAGGACCACAGCAAGCCAGTGCACGAGGTGCTGATTGAGATG
ACCAACGGTGGACTCGACAGGGCCGTTGAATGCACGGGCAACATCAACGCCATGATATCC
TGCTTCGAATGTGTCCATGATGGATGGGGCGTTGCGGTGCTGGTCGGCGTGCCGACCAAG
GATGATGTGTTCAAGACCCACCCGATGAACTTCCTCAACGAGAAGACGCTCAAGGGGACG
TTCTTCGGTAACTACAAGCCGCGCACCGACCTGCCCAACGTCGTCGAGCTGTACATGAAG
AAGGAgctggagctggagaagttcatcacgcacagcgtgccgttctcggagatcaacacg
gcgttcgatctcatgctcaaggggagagtctccgctgcgtcatgaggatggatgagtag
```

Protein sequence (translated ORF)

```
>mRNA_1.1 Protein sequence
MATAGKVIKCKAAVAWEAGKPLSIEEVEVAPPQAMEVRVKILYTALCHTDVYFWEAKGQT
PVFPRILGHEAGGIVESVGEGVTELAPGDHVLPVFTGECKECDHCKSEESNMCDLLRINV
DRGVMIGDGKSRFTIKGKPIFHFVGTSTFSEYTVIHVGCLAKINPEAPLDKVCILSCGFS
TGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGARLSGASRIIGVDLNPAKFEQAKKFGC
TDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGNINAMISCFECVHDGWGVAVLVGVPTK
DDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNVVELYMKKELELEKFITHSVPFSEINT
AFDLMLKGESLRCVMRMDE*
```

DNA配列の中から遺伝子を予測する

遺伝子予想ソフト

MEGANTE : https://megante.dna.affrc.go.jp/

Softberry/FGENESH: http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind

GENSCAN : http://genes.mit.edu/GENSCAN.html

-2（11/15）演習

第９回目-1（11/22）KEGGを用いた解析

連鎖解析

出席と宿題の提出　提出方法

リンク集

データベース検索：

1. PubMed: 論文検索
2. Entrez：総合データベース
3. Google：何でも検索
4. Google Scholar：科学に関する検索
5. 特許データベース

ホモロジー検索：

1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. MEGANTE ：遺伝子の予測
8. Softberry/FGENESH ：遺伝子の予測 ◀
9. GeneMark ：遺伝子の予測
10. GENESCAN ：遺伝子の予測
11. clustalW： アライメント
12. InterProScan： ドメイン解析
13. PSORT： 細胞内局在性予測
14. GEO：遺伝子発現解析
15. Gene Ontology：用語を用いた検索
16. SSPN：スプライシングの予測
17. ORF Finder：ORFの予測
18. TFSEARCH ：転写因子結合部位の予測
19. MOTIF ：タンパク質モチーフ検索
20. pI/Mw： 等電点、分子量の予測
21. RAP-DB：イネのゲノムデータベース

FGENESH

Reference: Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. ***Genome Biol.*** 2006,7, Suppl 1: P. 10.1-10.12.

HMM-based gene structure prediction (multiple genes, both chains). The Fgenesh gene-finder was selected as the most accurate program for plant gene identification. Plant Molecular Biology (2005), 57, 3, 445-460: "Five ab initio programs (FGENESH, GeneMark.hmm, GENSCAN, GlimmerR and Grail) were evaluated for their accuracy in predicting maize genes. FGENESH yielded the most accurate and GeneMark.hmm the second most accurate predictions" (FGENESH identified 11% more correct gene models than GeneMark on a set of 1353 test genes).

Paste nucleotide sequence here:

```
agaggtgaatagcagccagtatgaaggttttacaatttgaacatcgaaacgaaatccacacgaa
attcataaatttcatcgcatcttgctgaagcttaccagattctatcgattcgatctagctccaatag gtc
gatggg
```

ある遺伝子YのGenomicDNA

Alternatively, load a local file with sequence in Fasta format:

Local file name: ファイル未選択

Select organism specific gene-finding parameters :

Total 353 genome-specific parameters are available for genefinders of FGENESH suite

Oryza sativa Indica Group(long-grained rice)

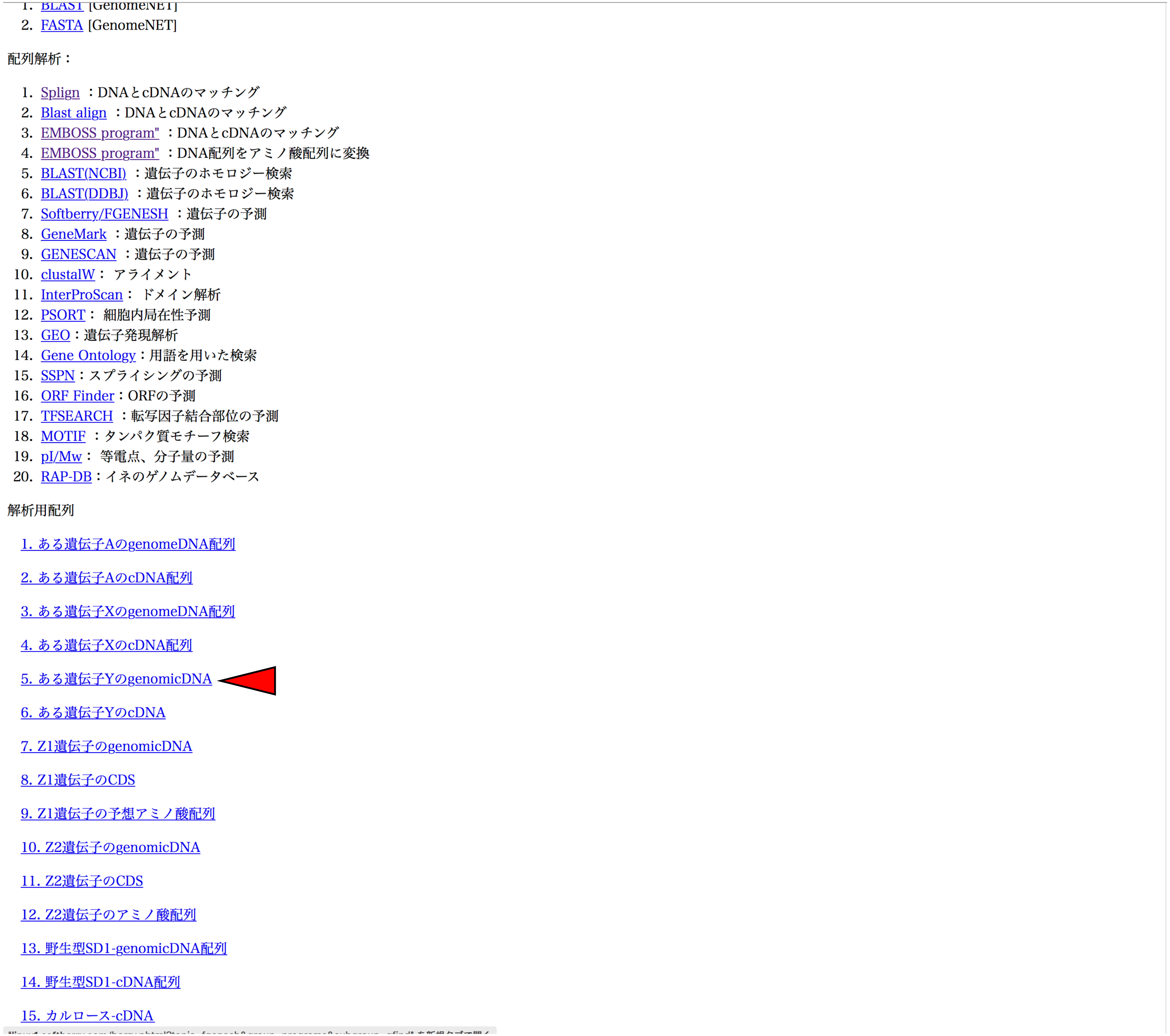
[Help] [Show advanced options]
[Example: Homo sapiens genomic beta globin region (HBB@) on chromosome 11]
[Example: Search in -chain]

Return to page with other programs of group: Gene finding

Most gene finding parameters presented here were trained by Softberry for its own use and distribution, using proprietary and publicly available data. Some of the parameters were created for our academic customers, including Broad Institute/MIT, Washington University, University of Minnesota and The Institute for Genomic Research (TIGR).

Your use of Softberry programs signifies that you accept Terms of Use

Last modification date: 12 Dec 2013



1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. Softberry/FGENESH ：遺伝子の予測
8. GeneMark ：遺伝子の予測
9. GENESCAN ：遺伝子の予測
10. clustalW： アライメント
11. InterProScan： ドメイン解析
12. PSORT： 細胞内局在性予測
13. GEO：遺伝子発現解析
14. Gene Ontology：用語を用いた検索
15. SSPN：スプライシングの予測
16. ORF Finder：ORFの予測
17. TFSEARCH ：転写因子結合部位の予測
18. MOTIF ：タンパク質モチーフ検索
19. pI/Mw： 等電点、分子量の予測
20. RAP-DB：イネのゲノムデータベース

解析用配列

1. ある遺伝子AのgenomeDNA配列

2. ある遺伝子AのcDNA配列

3. ある遺伝子XのgenomeDNA配列

4. ある遺伝子XのcDNA配列

5. ある遺伝子YのgenomicDNA

6. ある遺伝子YのcDNA

7. Z1遺伝子のgenomicDNA

8. Z1遺伝子のCDS

9. Z1遺伝子の予想アミノ酸配列

10. Z2遺伝子のgenomicDNA

11. Z2遺伝子のCDS

12. Z2遺伝子のアミノ酸配列

13. 野生型SD1-genomicDNA配列

14. 野生型SD1-cDNA配列

15. カルロース-cDNA

FGENESH

Reference: Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. ***Genome Biol.*** 2006,7, Suppl 1: P. 10.1-10.12.

HMM-based gene structure prediction (multiple genes, both chains). The Fgenesh gene-finder was selected as the most accurate program for plant gene identification. Plant Molecular Biology (2005), 57, 3, 445-460: "Five ab initio programs (FGENESH, GeneMark.hmm, GENSCAN, GlimmerR and Grail) were evaluated for their accuracy in predicting maize genes. FGENESH yielded the most accurate and GeneMark.hmm the second most accurate predictions" (FGENESH identified 11% more correct gene models than GeneMark on a set of 1353 test genes).

Paste nucleotide sequence here:

```
agaggtgaatagcagccagtatgaaggttttacaatttgaacatcgaaacgaaatccacacgaa
attcataaatttcatcgcatcttgctgaagcttaccagattctatcgattcgatctagctccaataggtc
gatggg
```

ある遺伝子YのGenomicDNA

Alternatively, load a local file with sequence in Fasta format:

Local file name: ファイルを選択 ファイル未選択

Select organism specific gene-finding parameters :

Total 353 genome-specific parameters are available for genefinders of FGENESH suite

Oryza sativa Indica Group(long-grained rice) ▾

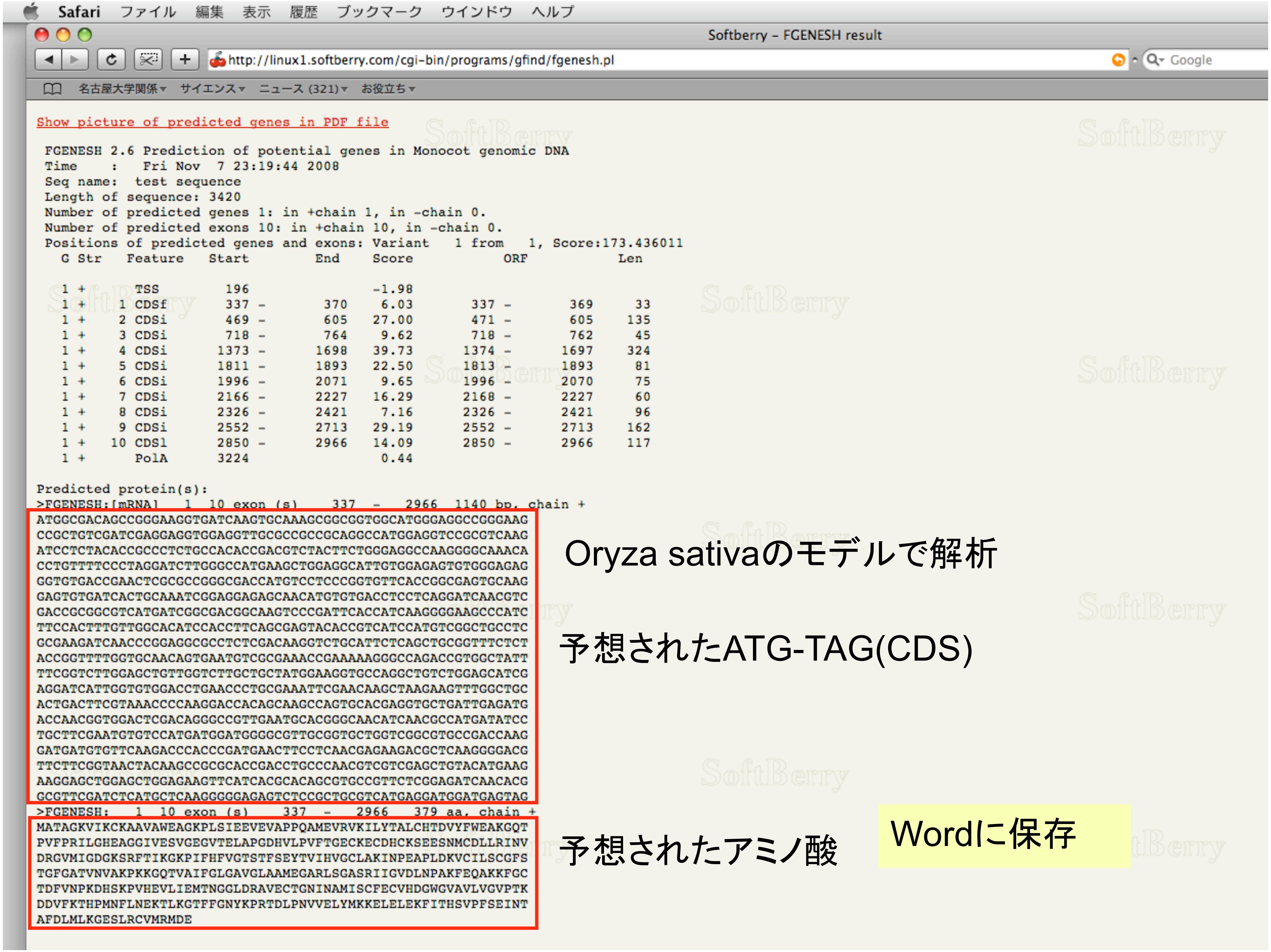
[Help] [Show advanced options]
[Example: Homo sapiens genomic beta globin region (HBB@) on chromosome 11]
[Example: Search in -chain]

Return to page with other programs of group: Gene finding

Most gene finding parameters presented here were trained by Softberry for its own use and distribution, using proprietary and publicly available data. Some of the parameters were created for our academic customers, including Broad Institute/MIT, Washington University, University of Minnesota and The Institute for Genomic Research (TIGR).

Your use of Softberry programs signifies that you accept Terms of Use

Last modification date: 12 Dec 2013



```
 FGENESH 2.6 Prediction of potential genes in Monocot genomic DNA
 Time    :   Fri Nov  7 23:19:44 2008
 Seq name:  test sequence
 Length of sequence: 3420
 Number of predicted genes 1: in +chain 1, in -chain 0.
 Number of predicted exons 10: in +chain 10, in -chain 0.
 Positions of predicted genes and exons: Variant   1 from   1, Score:173.436011
   G Str   Feature    Start        End    Score           ORF           Len

   1 +       TSS       196                -1.98
   1 +    1 CDSf       337 -       370     6.03       337 -       369     33
   1 +    2 CDSi       469 -       605    27.00       471 -       605    135
   1 +    3 CDSi       718 -       764     9.62       718 -       762     45
   1 +    4 CDSi      1373 -      1698    39.73      1374 -      1697    324
   1 +    5 CDSi      1811 -      1893    22.50      1813 -      1893     81
   1 +    6 CDSi      1996 -      2071     9.65      1996 -      2070     75
   1 +    7 CDSi      2166 -      2227    16.29      2168 -      2227     60
   1 +    8 CDSi      2326 -      2421     7.16      2326 -      2421     96
   1 +    9 CDSi      2552 -      2713    29.19      2552 -      2713    162
   1 +   10 CDSl      2850 -      2966    14.09      2850 -      2966    117
   1 +       PolA     3224                 0.44

Predicted protein(s):
>FGENESH:[mRNA]   1  10 exon (s)     337  -   2966  1140 bp, chain +
ATGGCGACAGCCGGGAAGGTGATCAAGTGCAAAGCGGCGGTGGCATGGGAGGCCGGGAAG
CCGCTGTCGATCGAGGAGGTGGAGGTTGCGCCGCCGCAGGCCATGGAGGTCCGCGTCAAG
ATCCTCTACACCGCCCTCTGCCACACCGACGTCTACTTCTGGGAGGCCAAGGGGCAAACA
CCTGTTTTCCCTAGGATCTTGGGCCATGAAGCTGGAGGCATTGTGGAGAGTGTGGGAGAG
GGTGTGACCGAACTCGCGCCGGGCGACCATGTCCTCCCGGTGTTCACCGGCGAGTGCAAG
GAGTGTGATCACTGCAAATCGGAGGAGAGCAACATGTGTGACCTCCTCAGGATCAACGTC
GACCGCGGCGTCATGATCGGCGACGGCAAGTCCCGATTCACCATCAAGGGGAAGCCCATC
TTCCACTTTGTTGGCACATCCACCTTCAGCGAGTACACCGTCATCCATGTCGGCTGCCTC
GCGAAGATCAACCCGGAGGCGCCTCTCGACAAGGTCTGCATTCTCAGCTGCGGTTTCTCT
ACCGGTTTTGGTGCAACAGTGAATGTCGCGAAACCGAAAAAGGGCCAGACCGTGGCTATT
TTCGGTCTTGGAGCTGTTGGTCTTGCTGCTATGGAAGGTGCCAGGCTGTCTGGAGCATCG
AGGATCATTGGTGTGGACCTGAACCCTGCGAAATTCGAACAAGCTAAGAAGTTTGGCTGC
ACTGACTTCGTAAACCCCAAGGACCACAGCAAGCCAGTGCACGAGGTGCTGATTGAGATG
ACCAACGGTGGACTCGACAGGGCCGTTGAATGCACGGGCAACATCAACGCCATGATATCC
TGCTTCGAATGTGTCCATGATGGATGGGGCGTTGCGGTGCTGGTCGGCGTGCCGACCAAG
GATGATGTGTTCAAGACCCACCCGATGAACTTCCTCAACGAGAAGACGCTCAAGGGGACG
TTCTTCGGTAACTACAAGCCGCGCACCGACCTGCCCAACGTCGTCGAGCTGTACATGAAG
AAGGAGCTGGAGCTGGAGAAGTTCATCACGCACAGCGTGCCGTTCTCGGAGATCAACACG
GCGTTCGATCTCATGCTCAAGGGGGAGAGTCTCCGCTGCGTCATGAGGATGGATGAGTAG
>FGENESH:   1  10 exon (s)     337  -   2966   379 aa, chain +
MATAGKVIKCKAAVAWEAGKPLSIEEVEVAPPQAMEVRVKILYTALCHTDVYFWEAKGQT
PVFPRILGHEAGGIVESVGEGVTELAPGDHVLPVFTGECKECDHCKSEESNMCDLLRINV
DRGVMIGDGKSRFTIKGKPIFHFVGTSTFSEYTVIHVGCLAKINPEAPLDKVCILSCGFS
TGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGARLSGASRIIGVDLNPAKFEQAKKFGC
TDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGNINAMISCFECVHDGWGVAVLVGVPTK
DDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNVVELYMKKELELEKFITHSVPFSEINT
AFDLMLKGESLRCVMRMDE
```

Oryza sativaのモデルで解析

予想されたATG-TAG(CDS)

予想されたアミノ酸

Wordに保存

Organism; Oryza sativa

```
MATAGKVIKCKAAVAWEAGKPLSIEEVEVAPPQAMEVRVKILYTALCHTDVYFWEAKGQT
PVFPRILGHEAGGIVESVGEGVTELAPGDHVLPVFTGECKECDHCKSEESNMCDLLRINV
DRGVMIGDGKSRFTIKGKPIFHFVGTSTFSEYTVIHVGCLAKINPEAPLDKVCILSCGFS
TGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGARLSGASRIIGVDLNPAKFEQAKKFGC
TDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGNINAMISCFECVHDGWGVAVLVGVPTK
DDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNVVELYMKKELELEKFITHSVPFSEINT
AFDLMLKGESLRCVMRMDE
```

This is outdated Softberry web archive,
use the current Softberry Web Site (www.softberry.com)

FGENESH

Reference: Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. ***Genome Biol.*** 2006,7, Suppl 1: P. 10.1-10.12.

HMM-based gene structure prediction (multiple genes, both chains). The Fgenesh gene-finder was selected as the most accurate program for plant gene identification. Plant Molecular Biology (2005), 57, 3, 445-460: "Five ab initio programs (FGENESH, GeneMark.hmm, GENSCAN, GlimmerR and Grail) were evaluated for their accuracy in predicting maize genes. FGENESH yielded the most accurate and GeneMark.hmm the second most accurate predictions" (FGENESH identified 11% more correct gene models than GeneMark on a set of 1353 test genes).

Paste nucleotide sequence here:

```
agaggtgaatagcagccagtatgaaggttttacaatttgaacatcgaaacgaaatccacacgaa
attcataaatttcatcgcatcttgctgaagcttaccagattctatcgattcgatctagctccaataggtc
gatggg
```

ある遺伝子YのGenomicDNA

Alternatively, load a local file with sequence in Fasta format:

Local file name: ファイルを選択 ファイル未選択

Select organism specific gene-finding parameters :

Total 353 genome-specific parameters are available for genefinders of FGENESH suite

Human (Homo sapiens)

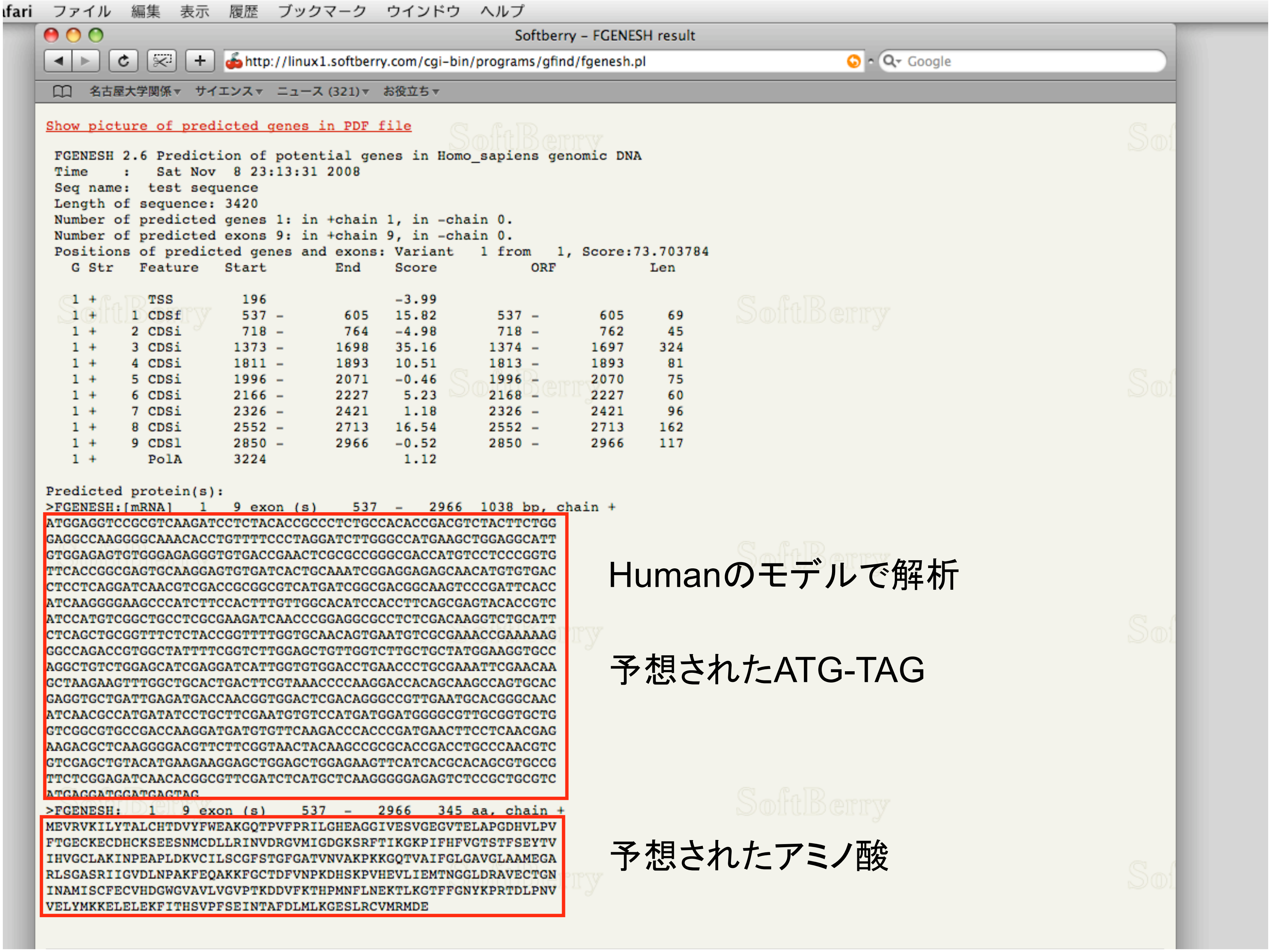
[Help] [Show advanced options]
[Example: Homo sapiens genomic beta globin region (HBB@) on chromosome 11]
[Example: Search in -chain]

Return to page with other programs of group: Gene finding

Most gene finding parameters presented here were trained by Softberry for its own use and distribution, using proprietary and publicly available data. Some of the parameters were created for our academic customers, including Broad Institute/MIT, Washington University, University of Minnesota and The Institute for Genomic Research (TIGR).

Your use of Softberry programs signifies that you accept Terms of Use

Last modification date: 12 Dec 2013



Show picture of predicted genes in PDF file

```
 FGENESH 2.6 Prediction of potential genes in Homo_sapiens genomic DNA
 Time    :   Sat Nov  8 23:13:31 2008
 Seq name:  test sequence
 Length of sequence: 3420
 Number of predicted genes 1: in +chain 1, in -chain 0.
 Number of predicted exons 9: in +chain 9, in -chain 0.
 Positions of predicted genes and exons: Variant   1 from   1, Score:73.703784
   G Str   Feature   Start        End    Score           ORF           Len

   1 +      TSS       196                -3.99
   1 +    1 CDSf      537 -       605    15.82       537 -       605     69
   1 +    2 CDSi      718 -       764    -4.98       718 -       762     45
   1 +    3 CDSi     1373 -      1698    35.16      1374 -      1697    324
   1 +    4 CDSi     1811 -      1893    10.51      1813 -      1893     81
   1 +    5 CDSi     1996 -      2071    -0.46      1996 -      2070     75
   1 +    6 CDSi     2166 -      2227     5.23      2168 -      2227     60
   1 +    7 CDSi     2326 -      2421     1.18      2326 -      2421     96
   1 +    8 CDSi     2552 -      2713    16.54      2552 -      2713    162
   1 +    9 CDSl     2850 -      2966    -0.52      2850 -      2966    117
   1 +      PolA     3224                 1.12

Predicted protein(s):
>FGENESH:[mRNA]   1   9 exon (s)    537  -   2966  1038 bp, chain +
ATGGAGGTCCGCGTCAAGATCCTCTACACCGCCCTCTGCCACACCGACGTCTACTTCTGG
GAGGCCAAGGGGCAAACACCTGTTTTCCCTAGGATCTTGGGCCATGAAGCTGGAGGCATT
GTGGAGAGTGTGGGAGAGGGTGTGACCGAACTCGCGCCGGGCGACCATGTCCTCCCGGTG
TTCACCGGCGAGTGCAAGGAGTGTGATCACTGCAAATCGGAGGAGAGCAACATGTGTGAC
CTCCTCAGGATCAACGTCGACCGCGGCGTCATGATCGGCGACGGCAAGTCCCGATTCACC
ATCAAGGGGAAGCCCATCTTCCACTTTGTTGGCACATCCACCTTCAGCGAGTACACCGTC
ATCCATGTCGGCTGCCTCGCGAAGATCAACCCGGAGGCGCCTCTCGACAAGGTCTGCATT
CTCAGCTGCGGTTTCTCTACCGGTTTTGGTGCAACAGTGAATGTCGCGAAACCGAAAAAG
GGCCAGACCGTGGCTATTTTCGGTCTTGGAGCTGTTGGTCTTGCTGCTATGGAAGGTGCC
AGGCTGTCTGGAGCATCGAGGATCATTGGTGTGGACCTGAACCCTGCGAAATTCGAACAA
GCTAAGAAGTTTGGCTGCACTGACTTCGTAAACCCCAAGGACCACAGCAAGCCAGTGCAC
GAGGTGCTGATTGAGATGACCAACGGTGGACTCGACAGGGCCGTTGAATGCACGGGCAAC
ATCAACGCCATGATATCCTGCTTCGAATGTGTCCATGATGGATGGGGCGTTGCGGTGCTG
GTCGGCGTGCCGACCAAGGATGATGTGTTCAAGACCCACCCGATGAACTTCCTCAACGAG
AAGACGCTCAAGGGGACGTTCTTCGGTAACTACAAGCCGCGCACCGACCTGCCCAACGTC
GTCGAGCTGTACATGAAGAAGGAGCTGGAGCTGGAGAAGTTCATCACGCACAGCGTGCCG
TTCTCGGAGATCAACACGGCGTTCGATCTCATGCTCAAGGGGGAGAGTCTCCGCTGCGTC
ATGAGGATGGATGAGTAG
>FGENESH:   1   9 exon (s)    537  -   2966   345 aa, chain +
MEVRVKILYTALCHTDVYFWEAKGQTPVFPRILGHEAGGIVESVGEGVTELAPGDHVLPV
FTGECKECDHCKSEESNMCDLLRINVDRGVMIGDGKSRFTIKGKPIFHFVGTSTFSEYTV
IHVGCLAKINPEAPLDKVCILSCGFSTGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGA
RLSGASRIIGVDLNPAKFEQAKKFGCTDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGN
INAMISCFECVHDGWGVAVLVGVPTKDDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNV
VELYMKKELELEKFITHSVPFSEINTAFDLMLKGESLRCVMRMDE
```

Humanのモデルで解析

予想されたATG-TAG

予想されたアミノ酸

Modelによって予想されるORFは多少異なる

Organism; 植物（Oryza sativa）

MATAGKVIKCKAAVAWEAGKPLSIEEVEVAPPQAMEVRVKILYTALCHTDVYFWEAKGQT
PVFPRILGHEAGGIVESVGEGVTELAPGDHVLPVFTGECKECDHCKSEESNMCDLLRINV
DRGVMIGDGKSRFTIKGKPIFHFVGTSTFSEYTVIHVGCLAKINPEAPLDKVCILSCGFS
TGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGARLSGASRIIGVDLNPAKFEQAKKFGC
TDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGNINAMISCFECVHDGWGVAVLVGVPTK
DDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNVVELYMKKELELEKFITHSVPFSEINT
AFDLMLKGESLRCVMRMDE

Organism; 人間

MEVRVKILYTALCHTDVYFWEAKGQTPVFPRILGHEAGGIVESVGEGVTELAPGDHVLPV
FTGECKECDHCKSEESNMCDLLRINVDRGVMIGDGKSRFTIKGKPIFHFVGTSTFSEYTV
IHVGCLAKINPEAPLDKVCILSCGFSTGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGA
RLSGASRIIGVDLNPAKFEQAKKFGCTDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGN
INAMISCFECVHDGWGVAVLVGVPTKDDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNV
VELYMKKELELEKFITHSVPFSEINTAFDLMLKGESLRCVMRMDE

Organism; ショウジョウバエ

MFCGEEKLIRSGFGVFAAAVAWEAGKPLSIEEVEVAPPQAMEVRVKILYTALCHTDVYFW
EAKGQTPVFPRILGHEAGGIVESVGEGVTELAPGDHVLPVFTGECKECDHCKSEESNMCD
LLRINVDRGVMIGDGKSRFTIKGKPIFHFVGTSTFSEYTVIHVGCLAKINPEAPLDKVCI
LSCGFSTGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGARLSGASRIIGVDLNPAKFEQ
AKKFGCTDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGNINAMISCFECVHDGWGVAVL
VGVPTKDDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNVVELYMKKELELEKFITHSVP
FSEINTAFDLMLKGESLRCVMRMDE

生物情報工学II (担当　芦苅）遺伝子の配列解析

第６回目-1（11/1）遺伝子の予測と機能解析

-2（11/1）演習

第７回目-1（11/8）DNA配列から遺伝子を探す

-2（11/8）DNA、アミノ酸のアラインメント、系統樹解析

-3（11/8）演習

第８回目-1（11/15）Genetyxを用いた解析

-2（11/15）演習

第９回目-1（11/22）KEGGを用いた解析

連鎖解析

出席と宿題の提出　提出方法

リンク集

データベース検索：

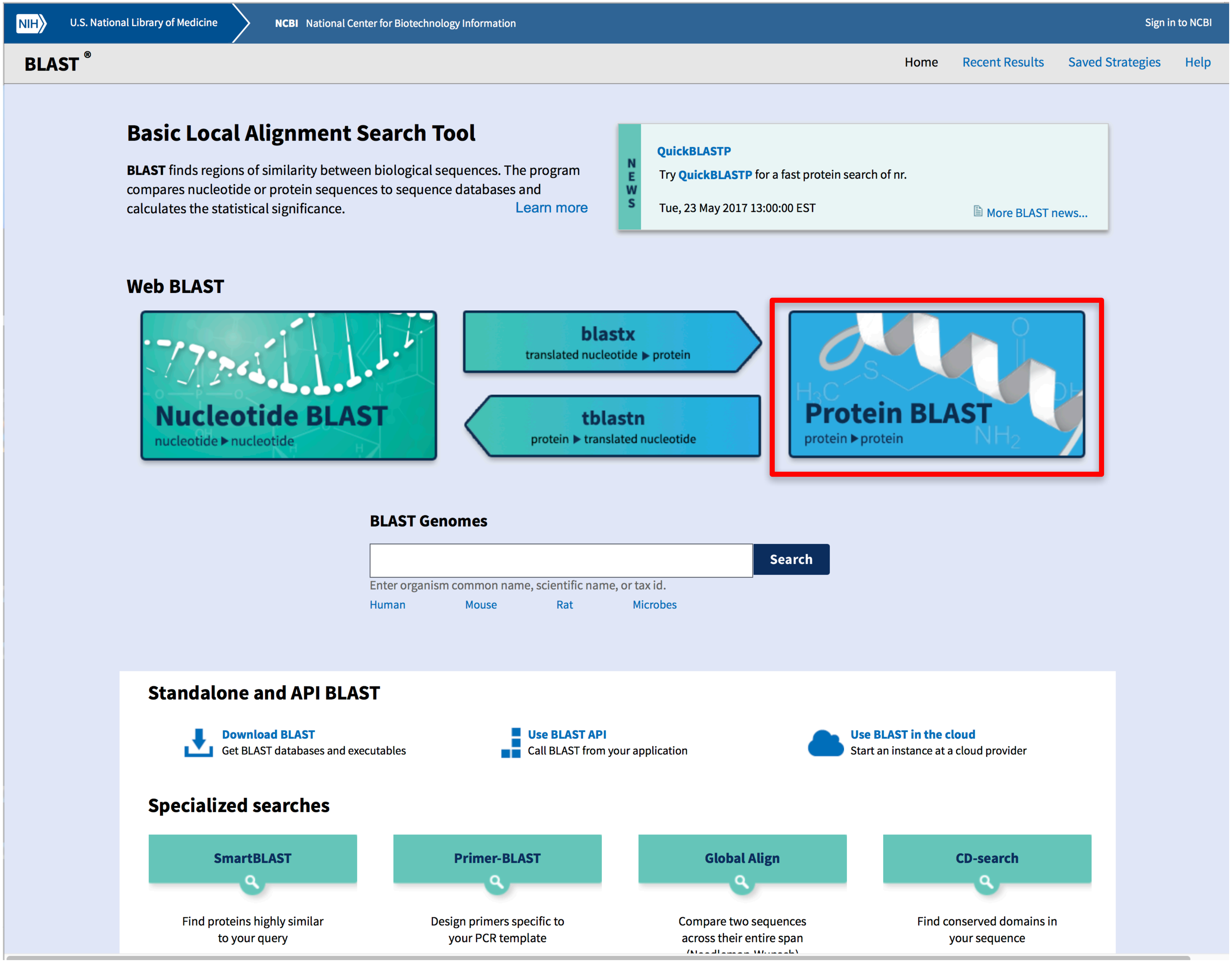
1. PubMed: 論文検索
2. Entrez：総合データベース
3. Google：何でも検索
4. Google Scholar：科学に関する検索
5. 特許データベース

ホモロジー検索：

1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. Softberry/FGENESH ：遺伝子の予測
8. GeneMark ：遺伝子の予測
9. GENESCAN ：遺伝子の予測
10. clustalW： アライメント
11. InterProScan： ドメイン解析
12. PSORT： 細胞内局在性予測
13. GEO：遺伝子発現解析
14. Gene Ontology：用語を用いた検索
15. SSPN：スプライシングの予測



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

QuickBLASTP

Try **QuickBLASTP** for a fast protein search of nr.

Tue, 23 May 2017 13:00:00 EST

More BLAST news...

Web BLAST

BLAST Genomes

Enter organism common name, scientific name, or tax id.

Standalone and API BLAST

Download BLAST
Get BLAST databases and executables

Use BLAST API
Call BLAST from your application

Use BLAST in the cloud
Start an instance at a cloud provider

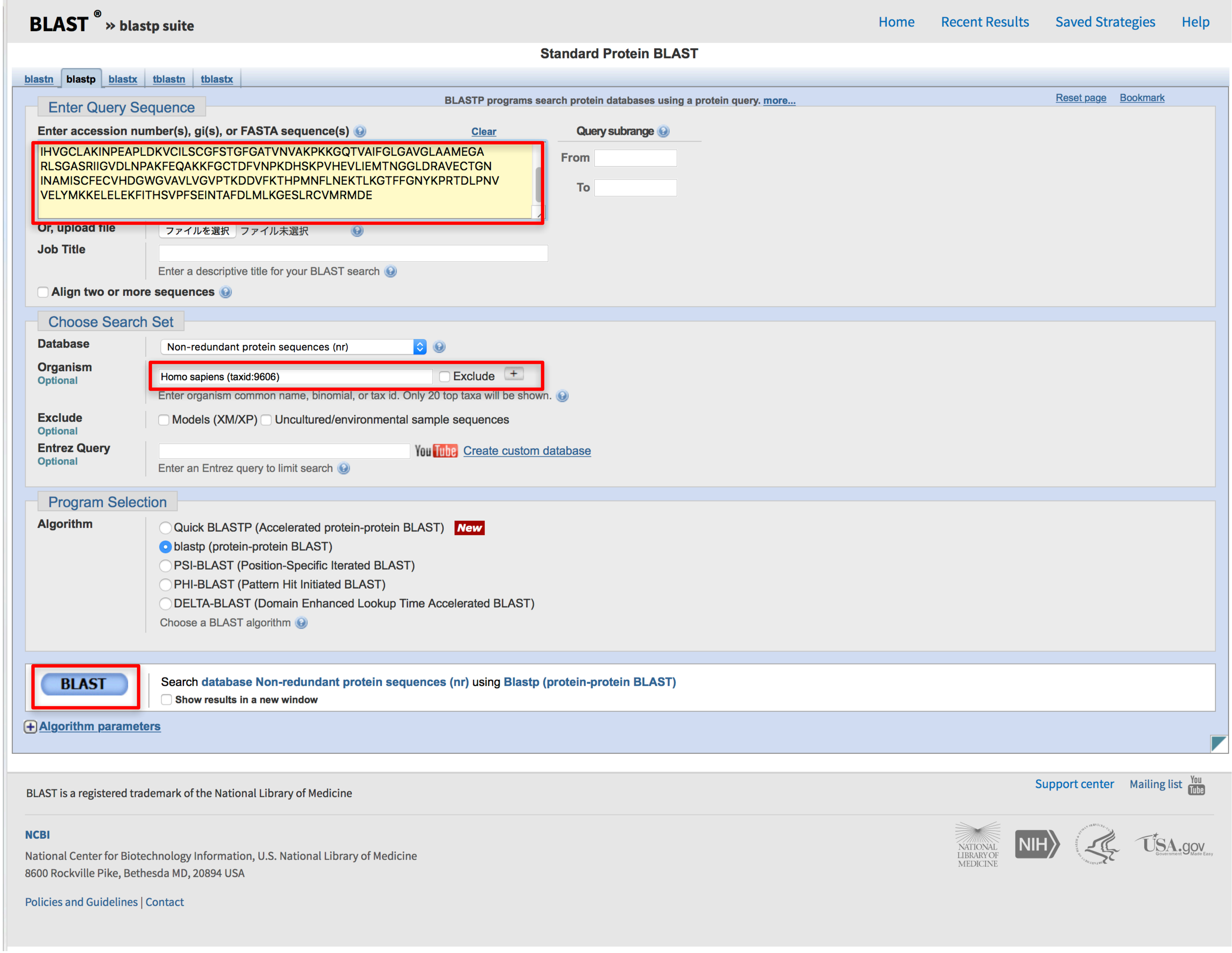
Specialized searches

SmartBLAST
Find proteins highly similar to your query

Primer-BLAST
Design primers specific to your PCR template

Global Align
Compare two sequences across their entire span

CD-search
Find conserved domains in your sequence



Standard Protein BLAST

BLASTP programs search protein databases using a protein query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

```
IHVGCLAKINPEAPLDKVCILSCGFSTGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGA
RLSGASRIIGVDLNPAKFEQAKKFGCTDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGN
INAMISCFECVHDGWGVAVLVGVPTKDDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNV
VELYMKKELELEKFITHSVPFSEINTAFDLMLKGESLRCVMRMDE
```

Query subrange

From

To

Or, upload file ファイルを選択 ファイル未選択

Job Title

Enter a descriptive title for your BLAST search

☐ **Align two or more sequences**

Choose Search Set

Database: Non-redundant protein sequences (nr)

Organism Optional: Homo sapiens (taxid:9606) ☐ Exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional: ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query Optional

Create custom database

Enter an Entrez query to limit search

Program Selection

Algorithm

○ Quick BLASTP (Accelerated protein-protein BLAST) New

● blastp (protein-protein BLAST)

○ PSI-BLAST (Position-Specific Iterated BLAST)

○ PHI-BLAST (Pattern Hit Initiated BLAST)

○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

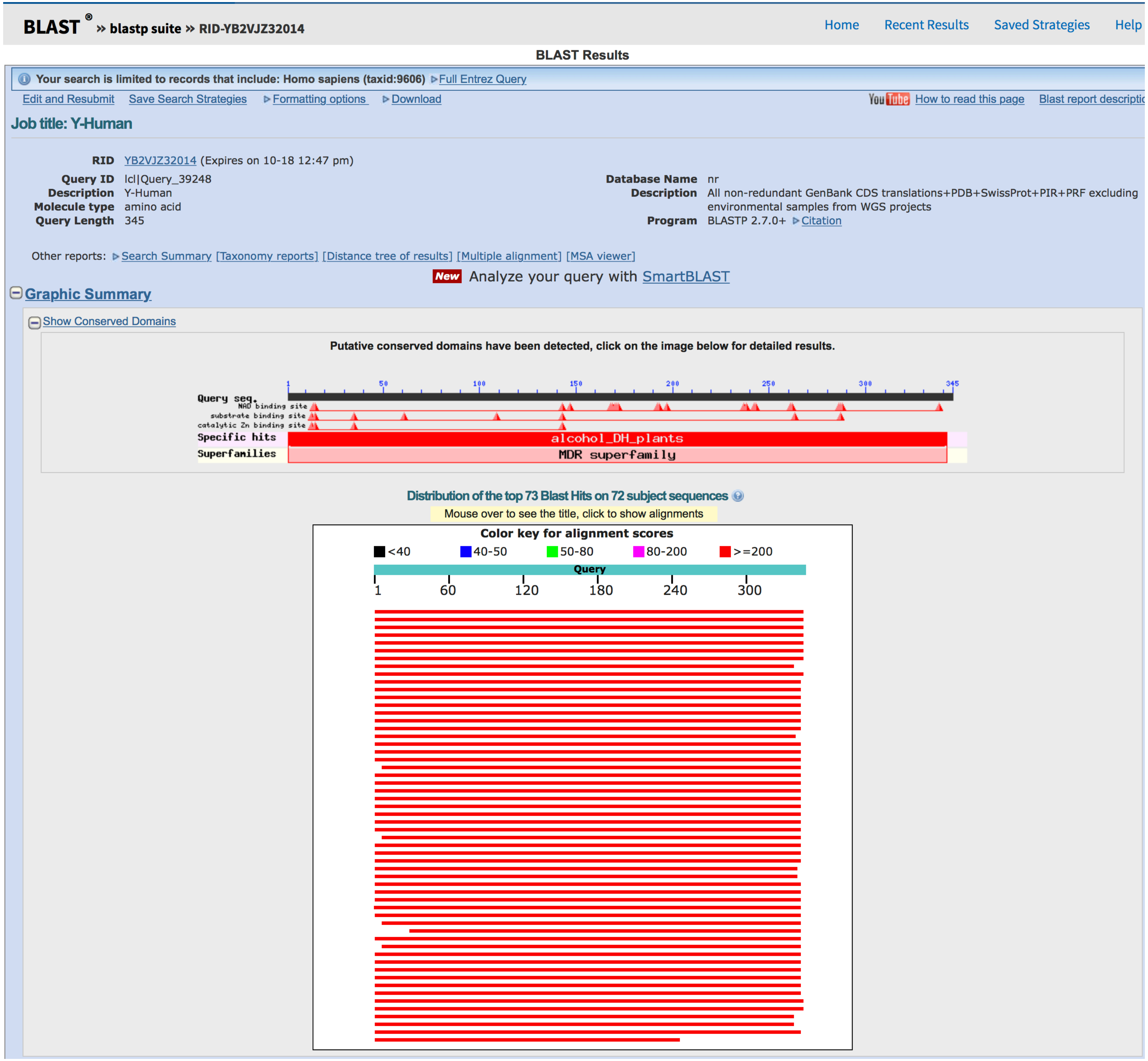
☐ **Show results in a new window**

BLAST is a registered trademark of the National Library of Medicine

NCBI

National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA

Policies and Guidelines | Contact



BLAST Results

Your search is limited to records that include: Homo sapiens (taxid:9606) ▷Full Entrez Query

Job title: Y-Human

RID YB2VJZ32014 (Expires on 10-18 12:47 pm)

Query ID lcl|Query_39248
Description Y-Human
Molecule type amino acid
Query Length 345

Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.7.0+ ▷Citation

Other reports: ▷Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment] [MSA viewer]

New Analyze your query with SmartBLAST

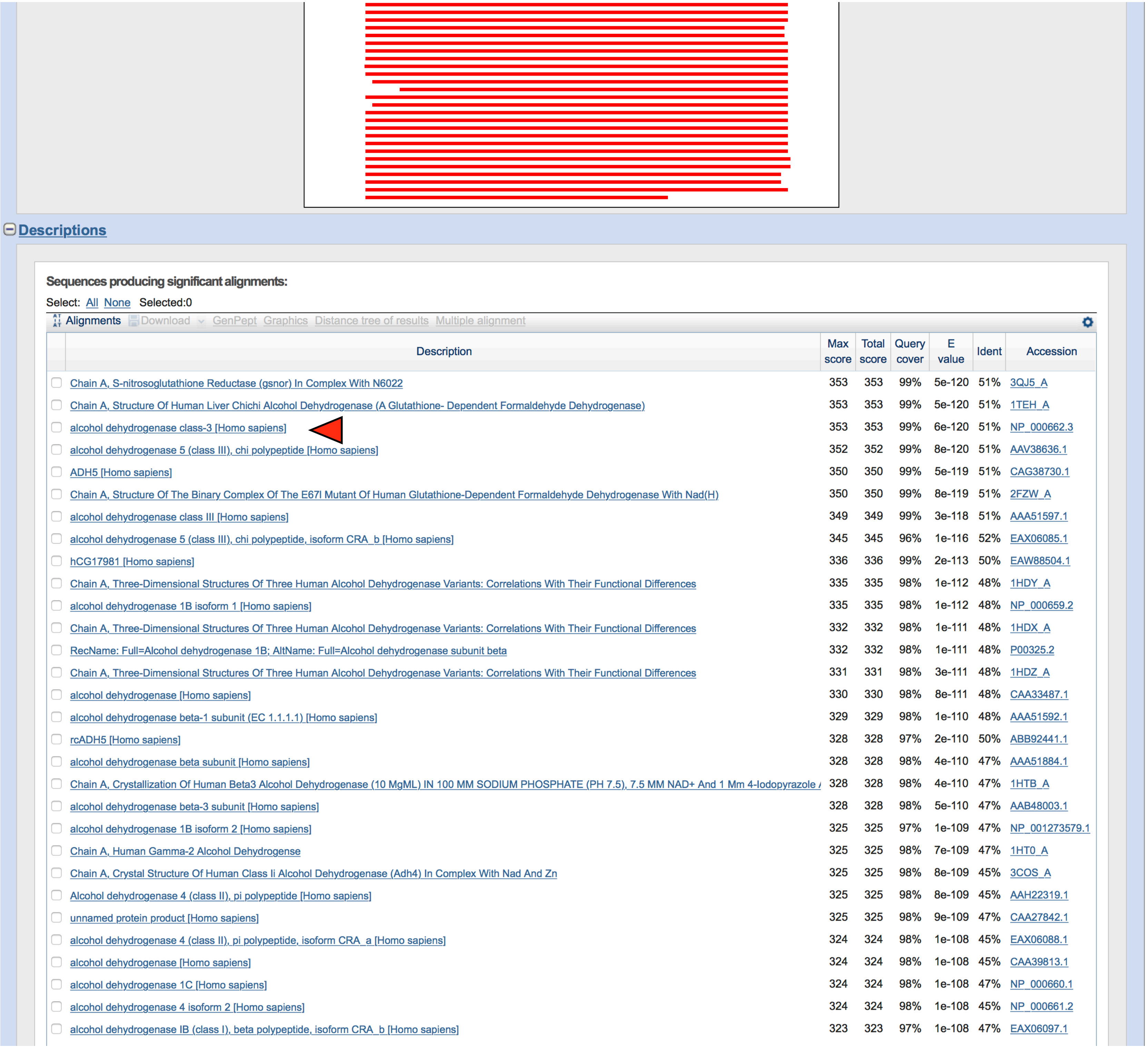
Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 73 Blast Hits on 72 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

⬛ <40 🟦 40-50 🟩 50-80 🟪 80-200 🟥 >=200



Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	Chain A, S-nitrosoglutathione Reductase (gsnor) In Complex With N6022	353	353	99%	5e-120	51%	3QJ5_A
☐	Chain A, Structure Of Human Liver Chichi Alcohol Dehydrogenase (A Glutathione- Dependent Formaldehyde Dehydrogenase)	353	353	99%	5e-120	51%	1TEH_A
☐	alcohol dehydrogenase class-3 [Homo sapiens]	353	353	99%	6e-120	51%	NP_000662.3
☐	alcohol dehydrogenase 5 (class III), chi polypeptide [Homo sapiens]	352	352	99%	8e-120	51%	AAV38636.1
☐	ADH5 [Homo sapiens]	350	350	99%	5e-119	51%	CAG38730.1
☐	Chain A, Structure Of The Binary Complex Of The E67I Mutant Of Human Glutathione-Dependent Formaldehyde Dehydrogenase With Nad(H)	350	350	99%	8e-119	51%	2FZW_A
☐	alcohol dehydrogenase class III [Homo sapiens]	349	349	99%	3e-118	51%	AAA51597.1
☐	alcohol dehydrogenase 5 (class III), chi polypeptide, isoform CRA_b [Homo sapiens]	345	345	96%	1e-116	52%	EAX06085.1
☐	hCG17981 [Homo sapiens]	336	336	99%	2e-113	50%	EAW88504.1
☐	Chain A, Three-Dimensional Structures Of Three Human Alcohol Dehydrogenase Variants: Correlations With Their Functional Differences	335	335	98%	1e-112	48%	1HDY_A
☐	alcohol dehydrogenase 1B isoform 1 [Homo sapiens]	335	335	98%	1e-112	48%	NP_000659.2
☐	Chain A, Three-Dimensional Structures Of Three Human Alcohol Dehydrogenase Variants: Correlations With Their Functional Differences	332	332	98%	1e-111	48%	1HDX_A
☐	RecName: Full=Alcohol dehydrogenase 1B; AltName: Full=Alcohol dehydrogenase subunit beta	332	332	98%	1e-111	48%	P00325.2
☐	Chain A, Three-Dimensional Structures Of Three Human Alcohol Dehydrogenase Variants: Correlations With Their Functional Differences	331	331	98%	3e-111	48%	1HDZ_A
☐	alcohol dehydrogenase [Homo sapiens]	330	330	98%	8e-111	48%	CAA33487.1
☐	alcohol dehydrogenase beta-1 subunit (EC 1.1.1.1) [Homo sapiens]	329	329	98%	1e-110	48%	AAA51592.1
☐	rcADH5 [Homo sapiens]	328	328	97%	2e-110	50%	ABB92441.1
☐	alcohol dehydrogenase beta subunit [Homo sapiens]	328	328	98%	4e-110	47%	AAA51884.1
☐	Chain A, Crystallization Of Human Beta3 Alcohol Dehydrogenase (10 MgML) IN 100 MM SODIUM PHOSPHATE (PH 7.5), 7.5 MM NAD+ And 1 Mm 4-Iodopyrazole	328	328	98%	4e-110	47%	1HTB_A
☐	alcohol dehydrogenase beta-3 subunit [Homo sapiens]	328	328	98%	5e-110	47%	AAB48003.1
☐	alcohol dehydrogenase 1B isoform 2 [Homo sapiens]	325	325	97%	1e-109	47%	NP_001273579.1
☐	Chain A, Human Gamma-2 Alcohol Dehydrogense	325	325	98%	7e-109	47%	1HT0_A
☐	Chain A, Crystal Structure Of Human Class Ii Alcohol Dehydrogenase (Adh4) In Complex With Nad And Zn	325	325	98%	8e-109	45%	3COS_A
☐	Alcohol dehydrogenase 4 (class II), pi polypeptide [Homo sapiens]	325	325	98%	8e-109	45%	AAH22319.1
☐	unnamed protein product [Homo sapiens]	325	325	98%	9e-109	47%	CAA27842.1
☐	alcohol dehydrogenase 4 (class II), pi polypeptide, isoform CRA_a [Homo sapiens]	324	324	98%	1e-108	45%	EAX06088.1
☐	alcohol dehydrogenase [Homo sapiens]	324	324	98%	1e-108	45%	CAA39813.1
☐	alcohol dehydrogenase 1C [Homo sapiens]	324	324	98%	1e-108	47%	NP_000660.1
☐	alcohol dehydrogenase 4 isoform 2 [Homo sapiens]	324	324	98%	1e-108	45%	NP_000661.2
☐	alcohol dehydrogenase IB (class I), beta polypeptide, isoform CRA_b [Homo sapiens]	323	323	97%	1e-108	47%	EAX06097.1

DNA配列の中から遺伝子を予測する

遺伝子予想ソフト

MEGANTE : https://megante.dna.affrc.go.jp/

Softberry/FGENESH: http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind

GENSCAN : http://genes.mit.edu/GENSCAN.html

-2（11/15）演習

第９回目-1（11/22）KEGGを用いた解析

連鎖解析

出席と宿題の提出　提出方法

リンク集

データベース検索：

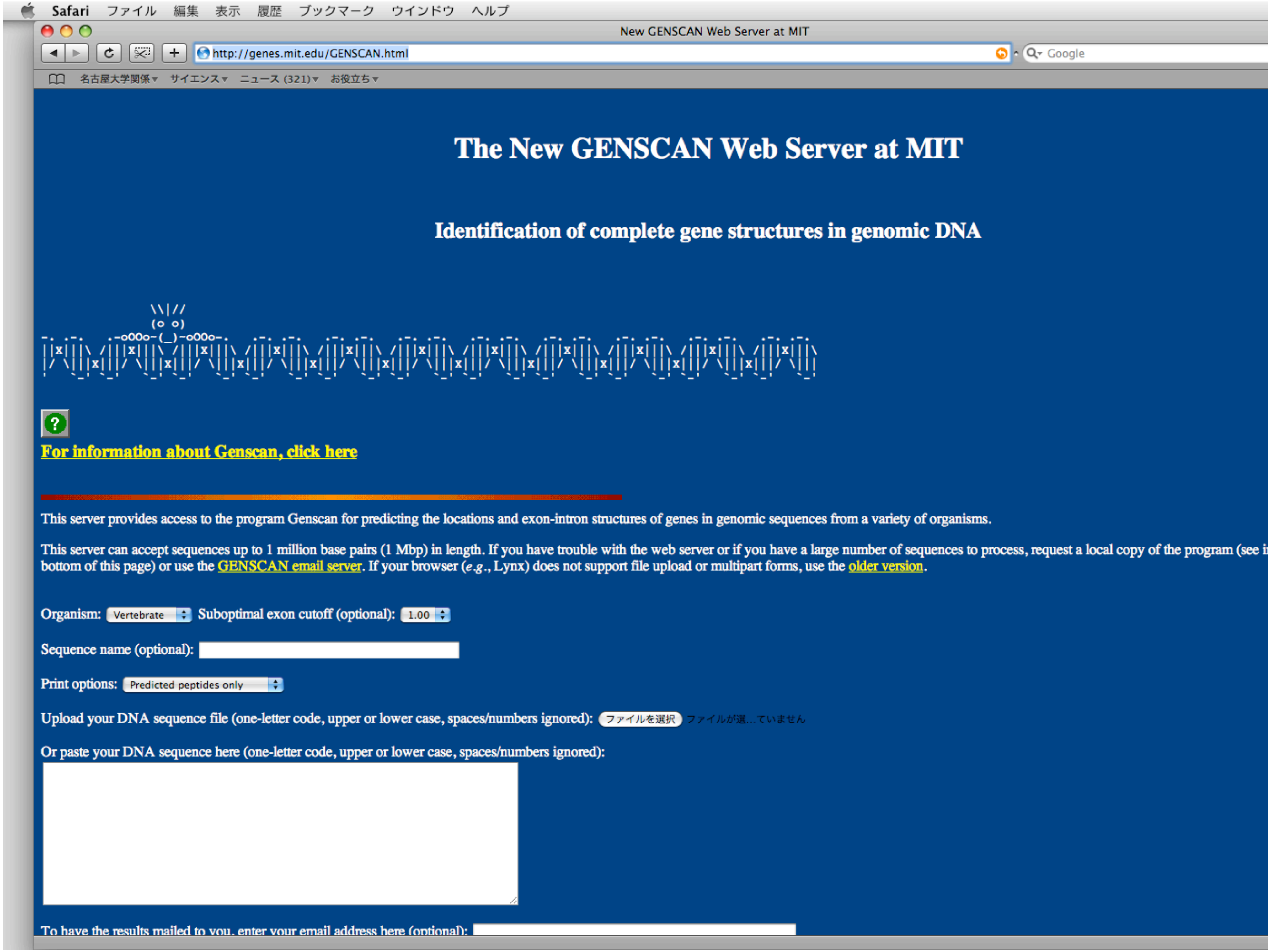
1. PubMed: 論文検索
2. Entrez：総合データベース
3. Google：何でも検索
4. Google Scholar：科学に関する検索
5. 特許データベース

ホモロジー検索：

1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. MEGANTE ：遺伝子の予測
8. Softberry/FGENESH ：遺伝子の予測
9. GeneMark ：遺伝子の予測
10. GENESCAN ：遺伝子の予測 ◀
11. clustalW： アライメント
12. InterProScan： ドメイン解析
13. PSORT： 細胞内局在性予測
14. GEO：遺伝子発現解析
15. Gene Ontology：用語を用いた検索
16. SSPN：スプライシングの予測
17. ORF Finder：ORFの予測
18. TFSEARCH ：転写因子結合部位の予測
19. MOTIF ：タンパク質モチーフ検索
20. pI/Mw： 等電点、分子量の予測
21. RAP-DB：イネのゲノムデータベース



The New GENSCAN Web Server at MIT

Identification of complete gene structures in genomic DNA

For information about Genscan, click here

This server provides access to the program Genscan for predicting the locations and exon-intron structures of genes in genomic sequences from a variety of organisms.

This server can accept sequences up to 1 million base pairs (1 Mbp) in length. If you have trouble with the web server or if you have a large number of sequences to process, request a local copy of the program (see i bottom of this page) or use the GENSCAN email server. If your browser (*e.g.*, Lynx) does not support file upload or multipart forms, use the older version.

Organism: Vertebrate Suboptimal exon cutoff (optional): 1.00

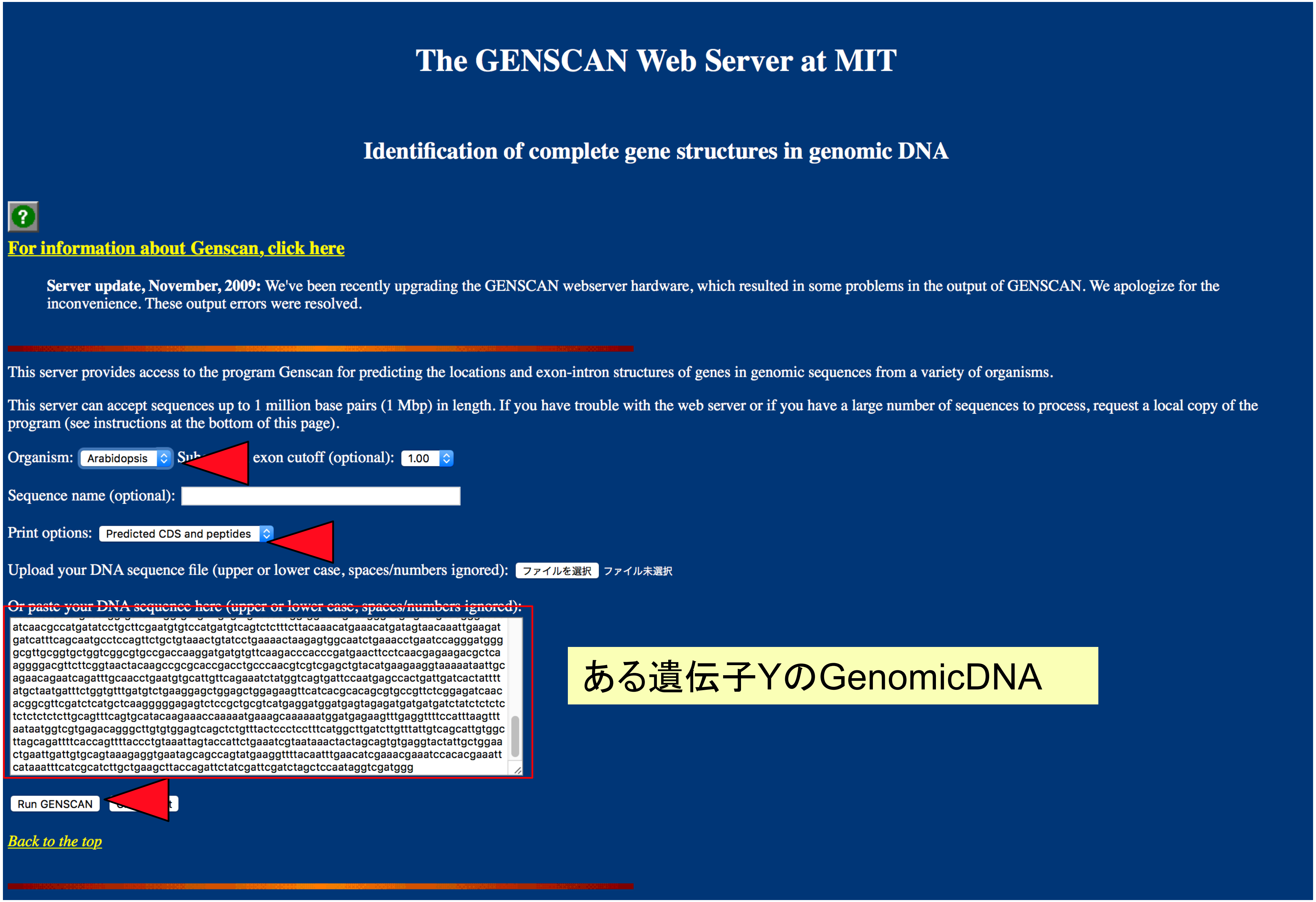
Sequence name (optional):

Print options: Predicted peptides only

Upload your DNA sequence file (one-letter code, upper or lower case, spaces/numbers ignored): ファイルを選択 ファイルが選...ていません

Or paste your DNA sequence here (one-letter code, upper or lower case, spaces/numbers ignored):

To have the results mailed to you, enter your email address here (optional):



The GENSCAN Web Server at MIT

Identification of complete gene structures in genomic DNA

For information about Genscan, click here

Server update, November, 2009: We've been recently upgrading the GENSCAN webserver hardware, which resulted in some problems in the output of GENSCAN. We apologize for the inconvenience. These output errors were resolved.

This server provides access to the program Genscan for predicting the locations and exon-intron structures of genes in genomic sequences from a variety of organisms.

This server can accept sequences up to 1 million base pairs (1 Mbp) in length. If you have trouble with the web server or if you have a large number of sequences to process, request a local copy of the program (see instructions at the bottom of this page).

Organism: Arabidopsis Sub exon cutoff (optional): 1.00

Sequence name (optional):

Print options: Predicted CDS and peptides

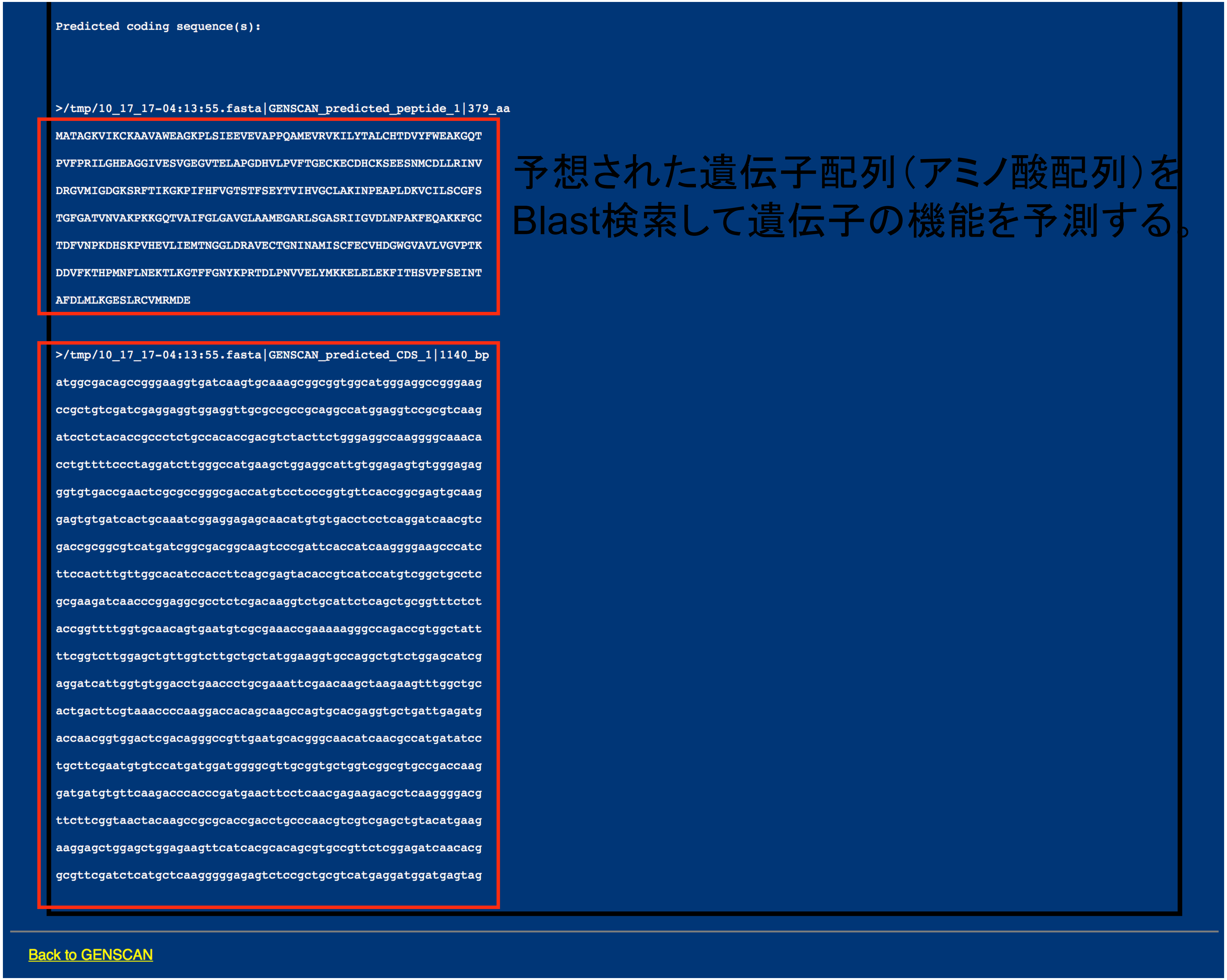
Upload your DNA sequence file (upper or lower case, spaces/numbers ignored): ファイルを選択 ファイル未選択

Or paste your DNA sequence here (upper or lower case, spaces/numbers ignored):

```
atcaacgccatgatatcctgcttcgaatgtgtccatgatgtcagtctctttcttacaaacatgaaacatgatagtaacaaattgaagat
gatcatttcagcaatgcctccagttctgctgtaaactgtatcctgaaaactaagagtggcaatctgaaacctgaatccagggatggg
gcgttgcggtgctggtcggcgtgccgaccaaggatgatgtgttcaagacccacccgatgaacttcctcaacgagaagacgctca
aggggacgttcttcggtaactacaagccgcgcaccgacctgcccaacgtcgtcgagctgtacatgaagaaggtaaaaataattgc
agaacagaatcagatttgcaacctgaatgtgcattgttcagaaatctatggtcagtgattccaatgagccactgattgatcactatttt
atgctaatgatttctggtgtttgatgtctgaaggagctggagctggagaagttcatcacgcacagcgtgccgttctcggagatcaac
acggcgttcgatctcatgctcaagggggagagtctccgctgcgtcatgaggatggatgagtagagatgatgatgatctatctctctc
tctctctcttgcagtttcagtgcatacaagaaaccaaaaatgaaagcaaaaaatggatgagaagtttgaggttttccatttaagttt
aataatggtcgtgagacagggcttgtgtggagtcagctctgtttactccctcctttcatggcttgatcttgtttattgtcagcattgtggc
ttagcagattttcaccagttttaccctgtaaattagtaccattctgaaatcgtaataaactactagcagtgtgaggtactattgctggaa
ctgaattgattgtgcagtaaagaggtgaatagcagccagtatgaaggttttacaatttgaacatcgaaacgaaatccacacgaaatt
cataaatttcatcgcatcttgctgaagcttaccagattctatcgattcgatctagctccaataggtcgatggg
```

Back to the top

ある遺伝子YのGenomicDNA



```
Predicted coding sequence(s):



>/tmp/10_17_17-04:13:55.fasta|GENSCAN_predicted_peptide_1|379_aa
MATAGKVIKCKAAVAWEAGKPLSIEEVEVAPPQAMEVRVKILYTALCHTDVYFWEAKGQT
PVFPRILGHEAGGIVESVGEGVTELAPGDHVLPVFTGECKECDHCKSEESNMCDLLRINV
DRGVMIGDGKSRFTIKGKPIFHFVGTSTFSEYTVIHVGCLAKINPEAPLDKVCILSCGFS
TGFGATVNVAKPKKGQTVAIFGLGAVGLAAMEGARLSGASRIIGVDLNPAKFEQAKKFGC
TDFVNPKDHSKPVHEVLIEMTNGGLDRAVECTGNINAMISCFECVHDGWGVAVLVGVPTK
DDVFKTHPMNFLNEKTLKGTFFGNYKPRTDLPNVVELYMKKELELEKFITHSVPFSEINT
AFDLMLKGESLRCVMRMDE

>/tmp/10_17_17-04:13:55.fasta|GENSCAN_predicted_CDS_1|1140_bp
atggcgacagccgggaaggtgatcaagtgcaaagcggcggtggcatgggaggccgggaag
ccgctgtcgatcgaggaggtggaggttgcgccgccgcaggccatggaggtccgcgtcaag
atcctctacaccgccctctgccacaccgacgtctacttctgggaggccaaggggcaaaca
cctgttttccctaggatcttgggccatgaagctggaggcattgtggagagtgtgggagag
ggtgtgaccgaactcgcgccgggcgaccatgtcctcccggtgttcaccggcgagtgcaag
gagtgtgatcactgcaaatcggaggagagcaacatgtgtgacctcctcaggatcaacgtc
gaccgcggcgtcatgatcggcgacggcaagtcccgattcaccatcaaggggaagcccatc
ttccactttgttggcacatccaccttcagcgagtacaccgtcatccatgtcggctgcctc
gcgaagatcaacccggaggcgcctctcgacaaggtctgcattctcagctgcggtttctct
accggttttggtgcaacagtgaatgtcgcgaaaccgaaaaagggccagaccgtggctatt
ttcggtcttggagctgttggtcttgctgctatggaaggtgccaggctgtctggagcatcg
aggatcattggtgtggacctgaaccctgcgaaattcgaacaagctaagaagtttggctgc
actgacttcgtaaaccccaaggaccacagcaagccagtgcacgaggtgctgattgagatg
accaacggtggactcgacagggccgttgaatgcacgggcaacatcaacgccatgatatcc
tgcttcgaatgtgtccatgatggatggggcgttgcggtgctggtcggcgtgccgaccaag
gatgatgtgttcaagacccacccgatgaacttcctcaacgagaagacgctcaaggggacg
ttcttcggtaactacaagccgcgcaccgacctgcccaacgtcgtcgagctgtacatgaag
aaggagctggagctggagaagttcatcacgcacagcgtgccgttctcggagatcaacacg
gcgttcgatctcatgctcaaggggagagtctccgctgcgtcatgaggatggatgagtag
```

予想された遺伝子配列（アミノ酸配列）を
Blast検索して遺伝子の機能を予測する。

Back to GENSCAN

ＤＮＡとアミノ酸の
アラインメント

ＤＮＡの突然変異と形質の違い

アルコールの代謝

アルコール脱水素酵素

アセトアルデヒド脱水素酵素
（Aldehyde dehydrogenase2）

$$CH_3CH_2OH \xrightarrow{ADH} CH_3CHO \xrightarrow{ALDH} CH_3COOH$$

エタノール

アセトアルデヒド
毒性

Z1遺伝子

Z2遺伝子

酢酸

生物情報工学II (担当　芦苅）遺伝子の配列解析

第６回目-1（11/1）遺伝子の予測と機能解析

-2（11/1）演習

第７回目-1（11/8）DNA配列から遺伝子を探す

-2（11/8）DNA、アミノ酸のアラインメント、系統樹解析

-3（11/8）演習

第８回目-1（11/15）Genetyxを用いた解析

-2（11/15）演習

第９回目-1（11/22）KEGGを用いた解析

連鎖解析

出席と宿題の提出　提出方法

リンク集

データベース検索：

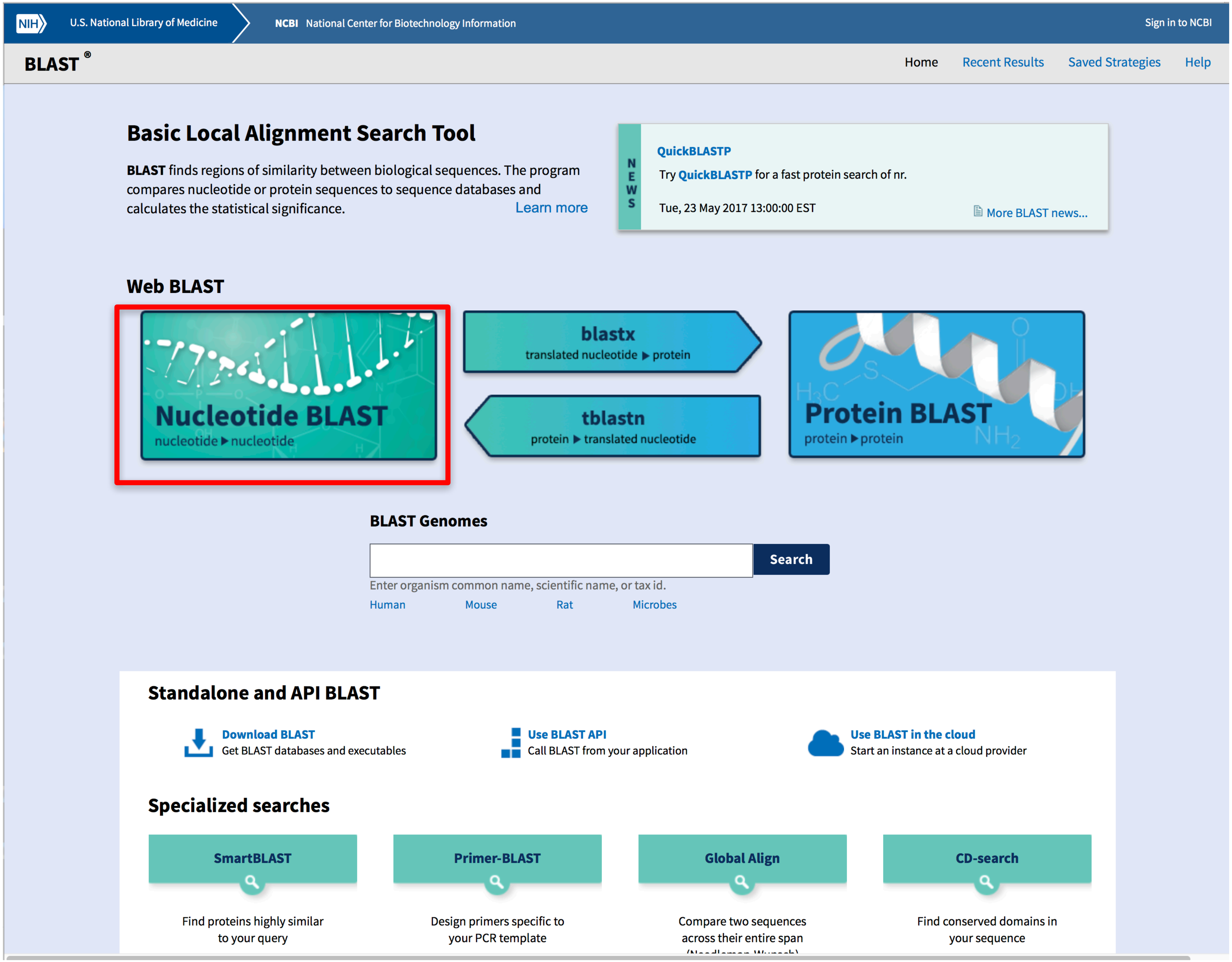
1. PubMed: 論文検索
2. Entrez：総合データベース
3. Google：何でも検索
4. Google Scholar：科学に関する検索
5. 特許データベース

ホモロジー検索：

1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. Softberry/FGENESH ：遺伝子の予測
8. GeneMark ：遺伝子の予測
9. GENESCAN ：遺伝子の予測
10. clustalW： アライメント
11. InterProScan： ドメイン解析
12. PSORT： 細胞内局在性予測
13. GEO：遺伝子発現解析
14. Gene Ontology：用語を用いた検索
15. SSPN：スプライシングの予測



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

NEWS

QuickBLASTP

Try **QuickBLASTP** for a fast protein search of nr.

Tue, 23 May 2017 13:00:00 EST

More BLAST news...

Web BLAST

BLAST Genomes

Enter organism common name, scientific name, or tax id.

Standalone and API BLAST

Download BLAST
Get BLAST databases and executables

Use BLAST API
Call BLAST from your application

Use BLAST in the cloud
Start an instance at a cloud provider

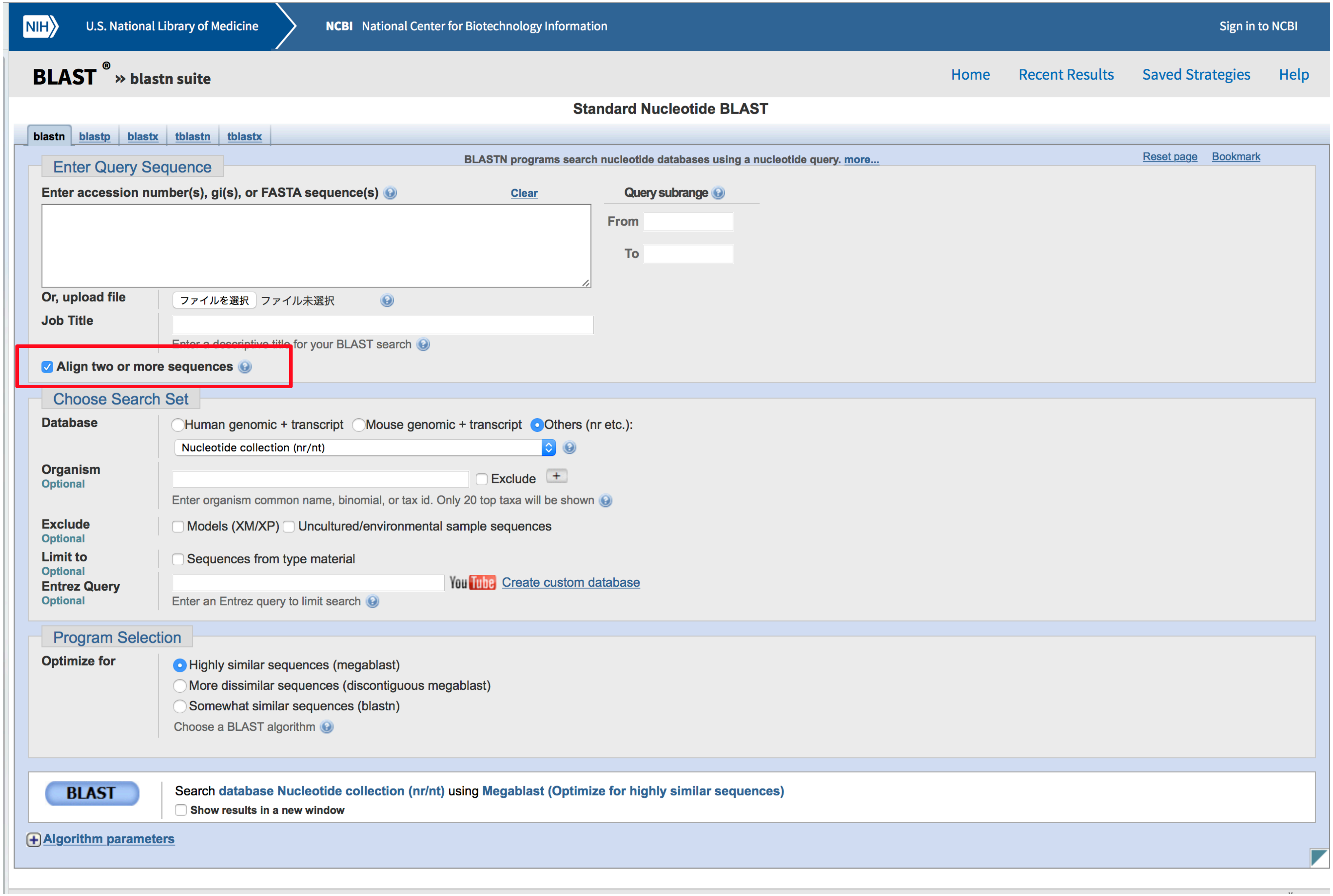
Specialized searches

SmartBLAST
Find proteins highly similar to your query

Primer-BLAST
Design primers specific to your PCR template

Global Align
Compare two sequences across their entire span (Needleman-Wunsch)

CD-search
Find conserved domains in your sequence



BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Database: ○ Human genomic + transcript ○ Mouse genomic + transcript ● Others (nr etc.):

Nucleotide collection (nr/nt)

Organism: ☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude: ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to: ☐ Sequences from type material

Enter an Entrez query to limit search

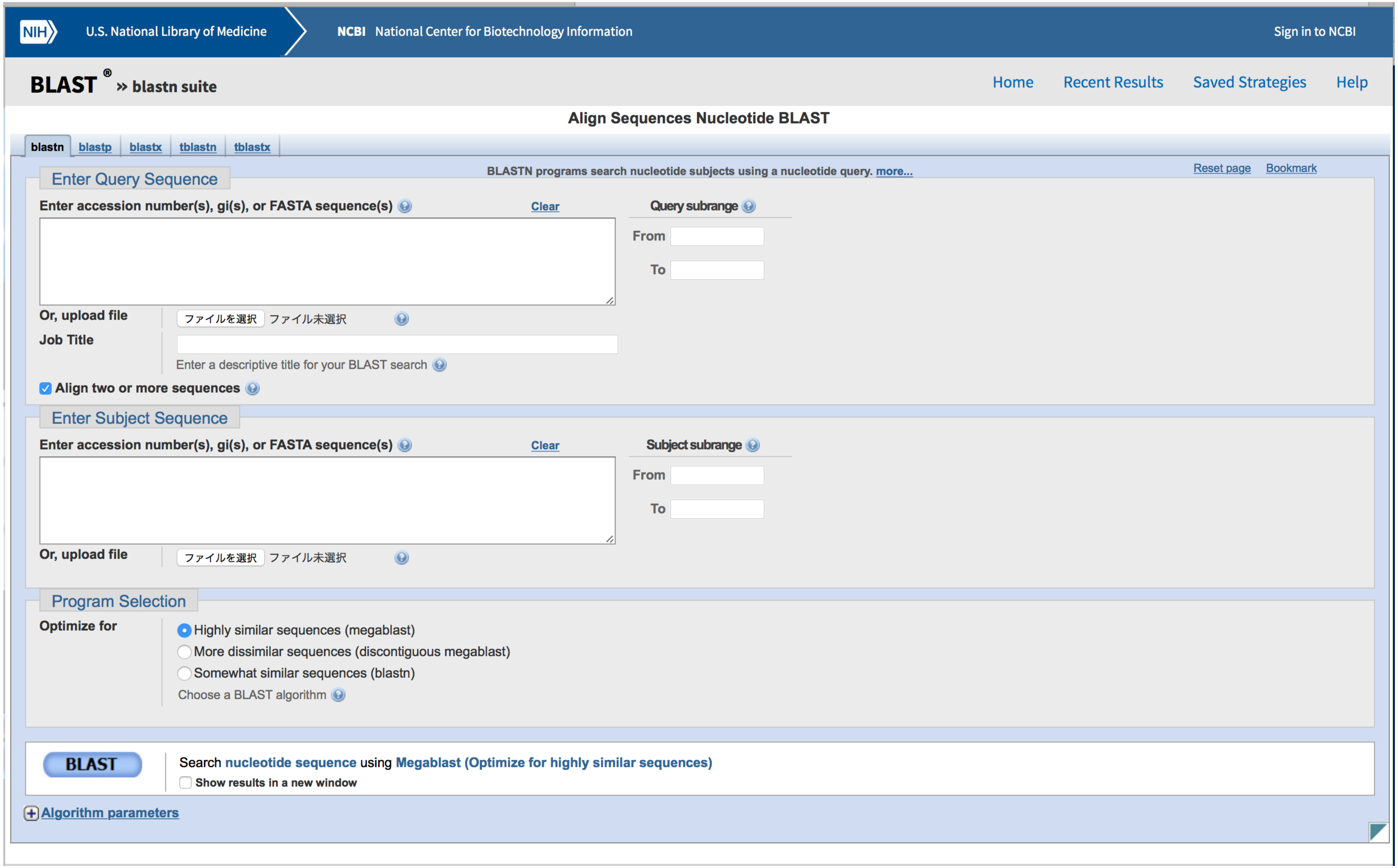
Optimize for:

● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window



Align Sequences Nucleotide BLAST

BLASTN programs search nucleotide subjects using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Query subrange: From / To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Subject subrange: From / To

Or, upload file

Program Selection

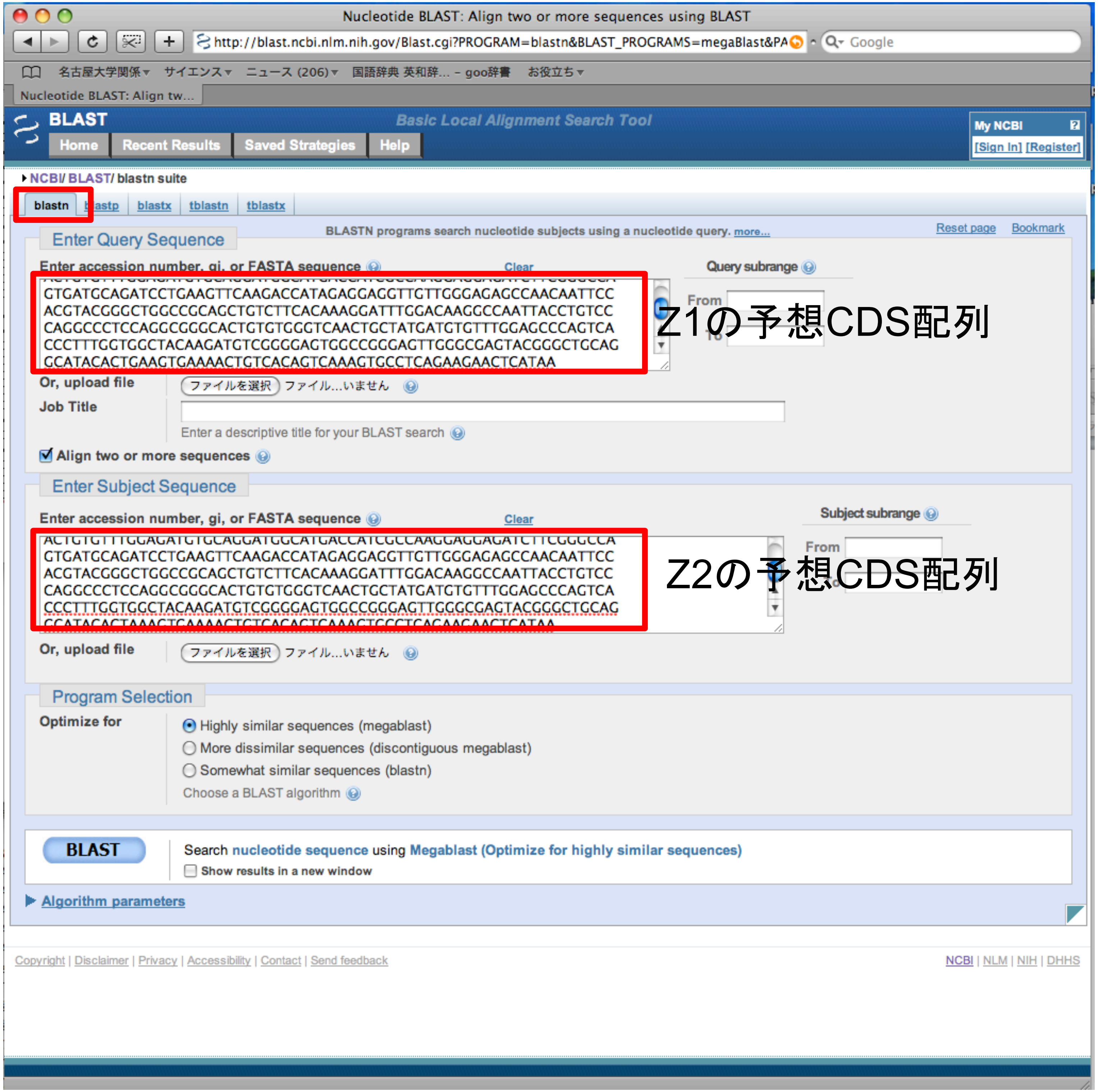
Optimize for

● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window



Enter Query Sequence

BLASTN programs search nucleotide subjects using a nucleotide query. more...

Enter accession number, gi, or FASTA sequence

```
GTGATGCAGATCCTGAAGTTCAAGACCATAGAGGAGGTTGTTGGGAGAGCCAACAATTCC
ACGTACGGGCTGGCCGCAGCTGTCTTCACAAAGGATTTGGACAAGGCCAATTACCTGTCC
CAGGCCCTCCAGGCGGGCACTGTGTGGGTCAACTGCTATGATGTGTTTGGAGCCCAGTCA
CCCTTTGGTGGCTACAAGATGTCGGGGAGTGGCCGGGAGTTGGGCGAGTACGGGCTGCAG
GCATACACTGAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA
```

Z1の予想CDS配列

Or, upload file: ファイル...いません

Job Title:

Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence

```
ACTGTGTTTGGAGATGTGCAGGATGGCATGACCATCGCCAAGGAGGAGATCTTCGGGCCA
GTGATGCAGATCCTGAAGTTCAAGACCATAGAGGAGGTTGTTGGGAGAGCCAACAATTCC
ACGTACGGGCTGGCCGCAGCTGTCTTCACAAAGGATTTGGACAAGGCCAATTACCTGTCC
CAGGCCCTCCAGGCGGGCACTGTGTGGGTCAACTGCTATGATGTGTTTGGAGCCCAGTCA
CCCTTTGGTGGCTACAAGATGTCGGGGAGTGGCCGGGAGTTGGGCGAGTACGGGCTGCAG
GCATACACTAAAGTCAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA
```

Z2の予想CDS配列

Or, upload file: ファイル...いません

Program Selection

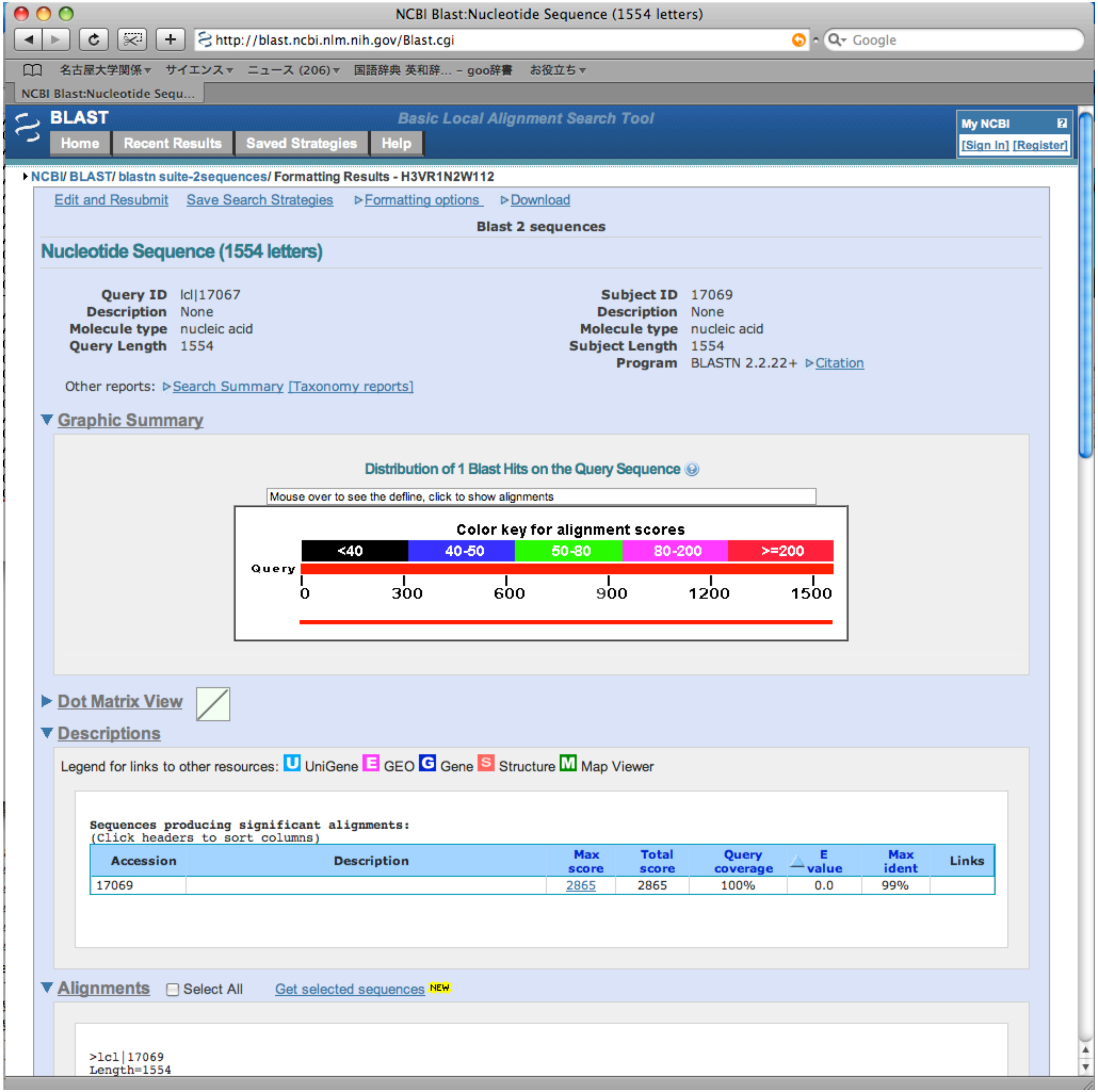
Optimize for

● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window



NCBI/ BLAST/ blastn suite-2sequences/ Formatting Results - H3VR1N2W112

Edit and Resubmit Save Search Strategies ▷Formatting options ▷Download

Blast 2 sequences

Nucleotide Sequence (1554 letters)

Query ID lcl|17067
Description None
Molecule type nucleic acid
Query Length 1554

Subject ID 17069
Description None
Molecule type nucleic acid
Subject Length 1554
Program BLASTN 2.2.22+ ▷Citation

Other reports: ▷Search Summary [Taxonomy reports]

▼ **Graphic Summary**

Distribution of 1 Blast Hits on the Query Sequence

Mouse over to see the defline, click to show alignments

▶ **Dot Matrix View**

▼ **Descriptions**

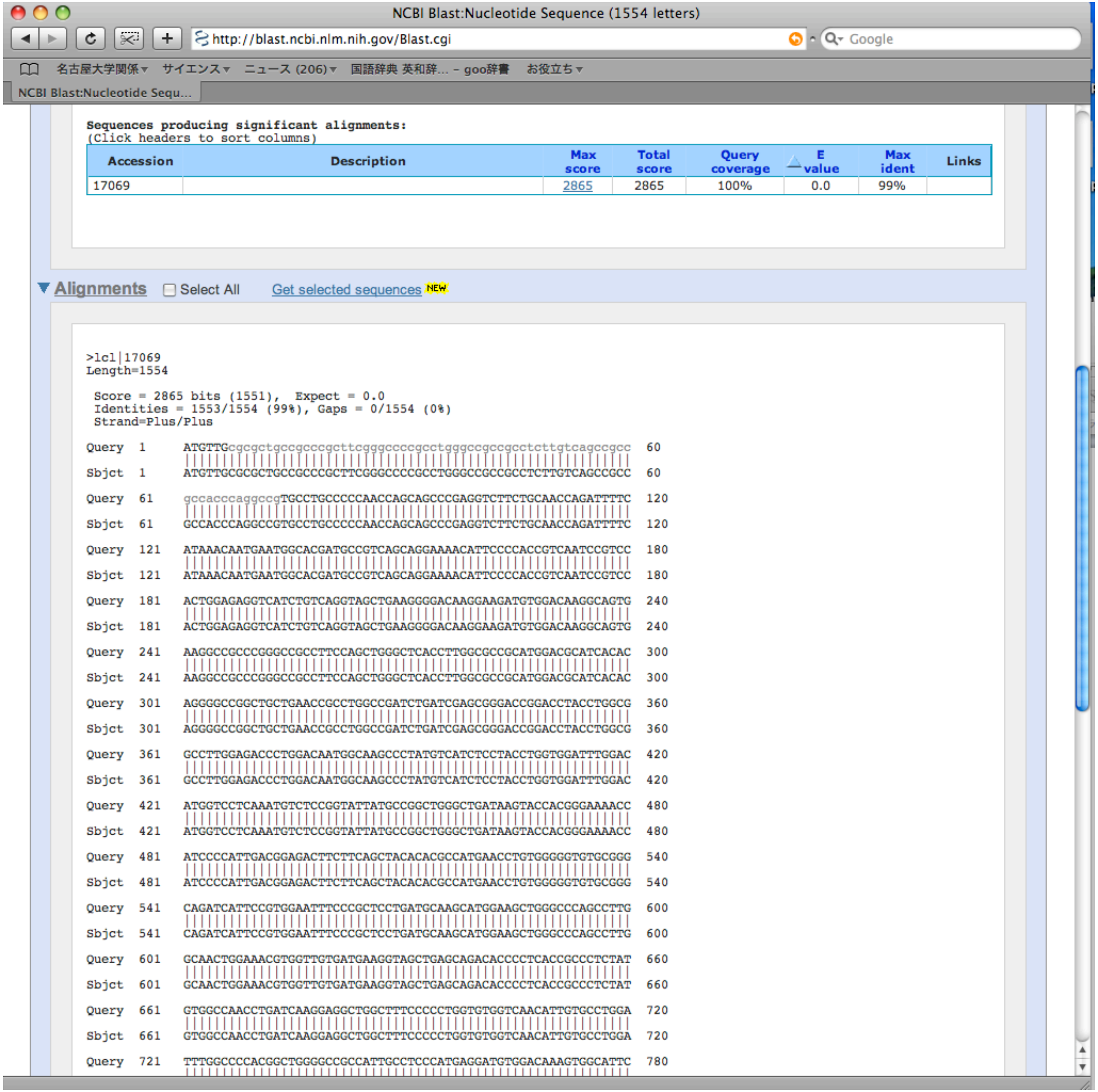
Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
17069		2865	2865	100%	0.0	99%	

▼ **Alignments** ☐ Select All Get selected sequences NEW

```
>lcl|17069
Length=1554
```

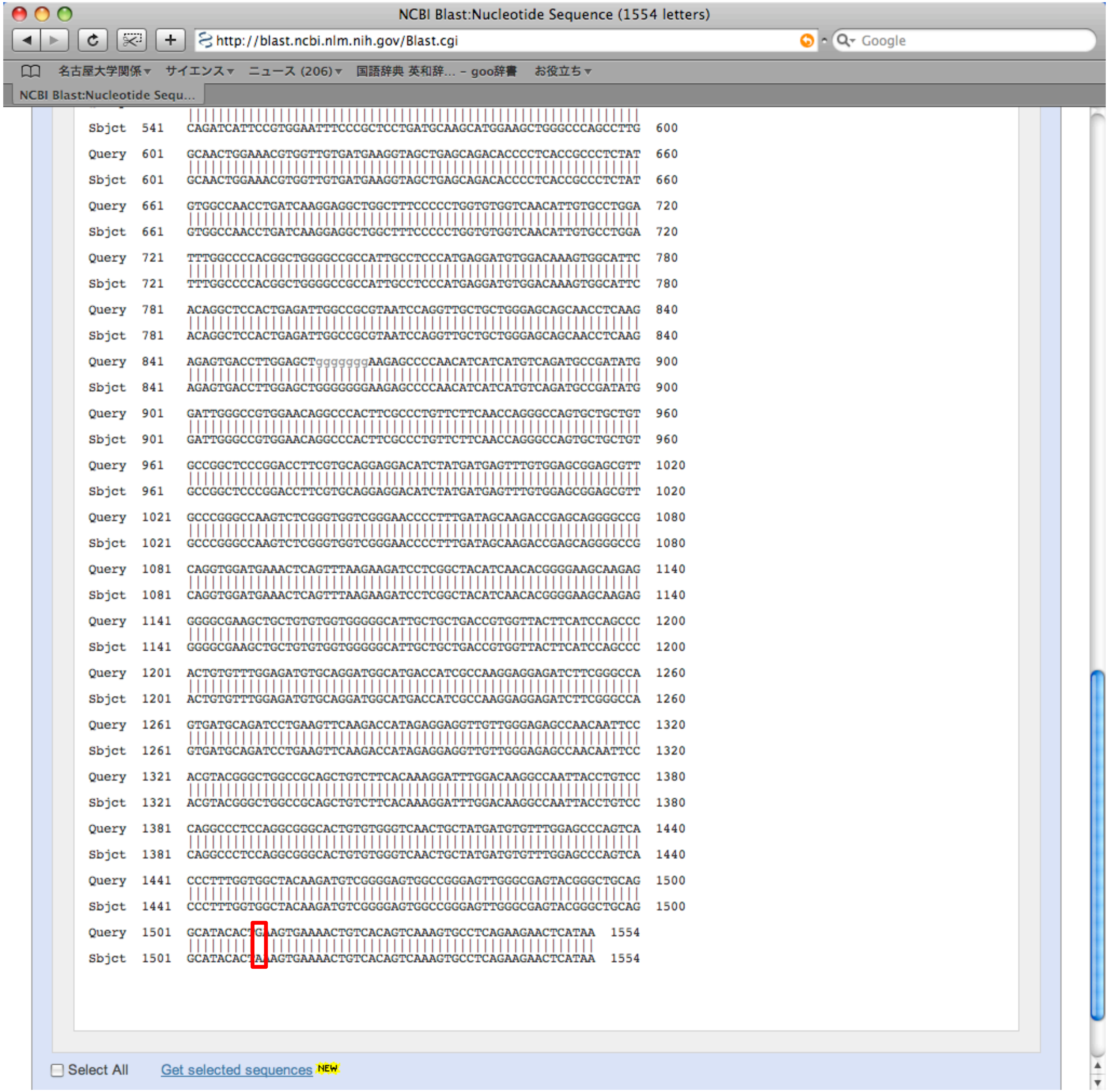


Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
17069		2865	2865	100%	0.0	99%	

Alignments ☐ Select All Get selected sequences NEW

```
>lcl|17069
Length=1554

 Score = 2865 bits (1551),  Expect = 0.0
 Identities = 1553/1554 (99%), Gaps = 0/1554 (0%)
 Strand=Plus/Plus

Query  1    ATGTTGcgcgctgccgcccgcttcggccccgcctgggccgccgcctcttgtcagccgcc  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGTTGCGCGCTGCCGCCCGCTTCGGCCCCGCCTGGGCCGCCGCCTCTTGTCAGCCGCC  60

Query  61   gccacccaggccgTGCCTGCCCCCAACCAGCAGCCCGAGGTCTTCTGCAACCAGATTTTC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCCACCCAGGCCGTGCCTGCCCCCAACCAGCAGCCCGAGGTCTTCTGCAACCAGATTTTC  120

Query  121  ATAAACAATGAATGGCACGATGCCGTCAGCAGGAAAACATTCCCCACCGTCAATCCGTCC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATAAACAATGAATGGCACGATGCCGTCAGCAGGAAAACATTCCCCACCGTCAATCCGTCC  180

Query  181  ACTGGAGAGGTCATCTGTCAGGTAGCTGAAGGGGACAAGGAAGATGTGGACAAGGCAGTG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  ACTGGAGAGGTCATCTGTCAGGTAGCTGAAGGGGACAAGGAAGATGTGGACAAGGCAGTG  240

Query  241  AAGGCCGCCCGGGCCGCCTTCCAGCTGGGCTCACCTTGGCGCCGCATGGACGCATCACAC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  AAGGCCGCCCGGGCCGCCTTCCAGCTGGGCTCACCTTGGCGCCGCATGGACGCATCACAC  300

Query  301  AGGGGCCGGCTGCTGAACCGCCTGGCCGATCTGATCGAGCGGGACCGGACCTACCTGGCG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AGGGGCCGGCTGCTGAACCGCCTGGCCGATCTGATCGAGCGGGACCGGACCTACCTGGCG  360

Query  361  GCCTTGGAGACCCTGGACAATGGCAAGCCCTATGTCATCTCCTACCTGGTGGATTTGGAC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GCCTTGGAGACCCTGGACAATGGCAAGCCCTATGTCATCTCCTACCTGGTGGATTTGGAC  420

Query  421  ATGGTCCTCAAATGTCTCCGGTATTATGCCGGCTGGGCTGATAAGTACCACGGGAAAACC  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATGGTCCTCAAATGTCTCCGGTATTATGCCGGCTGGGCTGATAAGTACCACGGGAAAACC  480

Query  481  ATCCCCATTGACGGAGACTTCTTCAGCTACACACGCCATGAACCTGTGGGGGTGTGCGGG  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  ATCCCCATTGACGGAGACTTCTTCAGCTACACACGCCATGAACCTGTGGGGGTGTGCGGG  540

Query  541  CAGATCATTCCGTGGAATTTCCCGCTCCTGATGCAAGCATGGAAGCTGGGCCCAGCCTTG  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  CAGATCATTCCGTGGAATTTCCCGCTCCTGATGCAAGCATGGAAGCTGGGCCCAGCCTTG  600

Query  601  GCAACTGGAAACGTGGTTGTGATGAAGGTAGCTGAGCAGACACCCCTCACCGCCCTCTAT  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601  GCAACTGGAAACGTGGTTGTGATGAAGGTAGCTGAGCAGACACCCCTCACCGCCCTCTAT  660

Query  661  GTGGCCAACCTGATCAAGGAGGCTGGCTTTCCCCCTGGTGTGGTCAACATTGTGCCTGGA  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661  GTGGCCAACCTGATCAAGGAGGCTGGCTTTCCCCCTGGTGTGGTCAACATTGTGCCTGGA  720

Query  721  TTTGGCCCCACGGCTGGGGCCGCCATTGCCTCCCATGAGGATGTGGACAAAGTGGCATTC  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541   CAGATCATTCCGTGGAATTTCCCGCTCCTGATGCAAGCATGGAAGCTGGGCCCAGCCTTG  600

Query  601   GCAACTGGAAACGTGGTTGTGATGAAGGTAGCTGAGCAGACACCCCTCACCGCCCTCTAT  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601   GCAACTGGAAACGTGGTTGTGATGAAGGTAGCTGAGCAGACACCCCTCACCGCCCTCTAT  660

Query  661   GTGGCCAACCTGATCAAGGAGGCTGGCTTTCCCCCTGGTGTGGTCAACATTGTGCCTGGA  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661   GTGGCCAACCTGATCAAGGAGGCTGGCTTTCCCCCTGGTGTGGTCAACATTGTGCCTGGA  720

Query  721   TTTGGCCCCACGGCTGGGGCCGCCATTGCCTCCCATGAGGATGTGGACAAAGTGGCATTC  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  721   TTTGGCCCCACGGCTGGGGCCGCCATTGCCTCCCATGAGGATGTGGACAAAGTGGCATTC  780

Query  781   ACAGGCTCCACTGAGATTGGCCGCGTAATCCAGGTTGCTGCTGGGAGCAGCAACCTCAAG  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  781   ACAGGCTCCACTGAGATTGGCCGCGTAATCCAGGTTGCTGCTGGGAGCAGCAACCTCAAG  840

Query  841   AGAGTGACCTTGGAGCTgggggggAAGAGCCCCAACATCATCATGTCAGATGCCGATATG  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  841   AGAGTGACCTTGGAGCTGGGGGGGAAGAGCCCCAACATCATCATGTCAGATGCCGATATG  900

Query  901   GATTGGGCCGTGGAACAGGCCCACTTCGCCCTGTTCTTCAACCAGGGCCAGTGCTGCTGT  960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  901   GATTGGGCCGTGGAACAGGCCCACTTCGCCCTGTTCTTCAACCAGGGCCAGTGCTGCTGT  960

Query  961   GCCGGCTCCCGGACCTTCGTGCAGGAGGACATCTATGATGAGTTTGTGGAGCGGAGCGTT  1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  961   GCCGGCTCCCGGACCTTCGTGCAGGAGGACATCTATGATGAGTTTGTGGAGCGGAGCGTT  1020

Query  1021  GCCCGGGCCAAGTCTCGGGTGGTCGGGAACCCCTTTGATAGCAAGACCGAGCAGGGGCCG  1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1021  GCCCGGGCCAAGTCTCGGGTGGTCGGGAACCCCTTTGATAGCAAGACCGAGCAGGGGCCG  1080

Query  1081  CAGGTGGATGAAACTCAGTTTAAGAAGATCCTCGGCTACATCAACACGGGGAAGCAAGAG  1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1081  CAGGTGGATGAAACTCAGTTTAAGAAGATCCTCGGCTACATCAACACGGGGAAGCAAGAG  1140

Query  1141  GGGGCGAAGCTGCTGTGTGGTGGGGGCATTGCTGCTGACCGTGGTTACTTCATCCAGCCC  1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1141  GGGGCGAAGCTGCTGTGTGGTGGGGGCATTGCTGCTGACCGTGGTTACTTCATCCAGCCC  1200

Query  1201  ACTGTGTTTGGAGATGTGCAGGATGGCATGACCATCGCCAAGGAGGAGATCTTCGGGCCA  1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1201  ACTGTGTTTGGAGATGTGCAGGATGGCATGACCATCGCCAAGGAGGAGATCTTCGGGCCA  1260

Query  1261  GTGATGCAGATCCTGAAGTTCAAGACCATAGAGGAGGTTGTTGGGAGAGCCAACAATTCC  1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1261  GTGATGCAGATCCTGAAGTTCAAGACCATAGAGGAGGTTGTTGGGAGAGCCAACAATTCC  1320

Query  1321  ACGTACGGGCTGGCCGCAGCTGTCTTCACAAAGGATTTGGACAAGGCCAATTACCTGTCC  1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1321  ACGTACGGGCTGGCCGCAGCTGTCTTCACAAAGGATTTGGACAAGGCCAATTACCTGTCC  1380

Query  1381  CAGGCCCTCCAGGCGGGCACTGTGTGGGTCAACTGCTATGATGTGTTTGGAGCCCAGTCA  1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1381  CAGGCCCTCCAGGCGGGCACTGTGTGGGTCAACTGCTATGATGTGTTTGGAGCCCAGTCA  1440

Query  1441  CCCTTTGGTGGCTACAAGATGTCGGGGAGTGGCCGGGAGTTGGGCGAGTACGGGCTGCAG  1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1441  CCCTTTGGTGGCTACAAGATGTCGGGGAGTGGCCGGGAGTTGGGCGAGTACGGGCTGCAG  1500

Query  1501  GCATACACTGAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA  1554
             ||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1501  GCATACACTAAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA  1554
```

☐ Select All

Get selected sequences NEW

```
Z1CDS     GCATACACTGAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA
Z2CDS     GCATACACTAAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA
          ********* ********************************************
```

BLASTP programs search protein subjects using a protein query. more...

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

```
MLRAAARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEWHDAVSRKTFPTVNPS
TGEVICQVAEGDKEDVDKAVKAARAAFQLGSPWRRMDASHRGRLLNRLADLIERDRTYLA
ALETLDNGKPYVISYLVDLDMVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCG
QIIPWNFPLLMQAWKLGPALATGNVVVMKVAEQTPLTALYVANLIKEAGFPPGVVNIVPG
FGPTAGAAIASHEDVDKVAFTGSTEIGRVIQVAAGSSNLKRVTLELGGKSPNIIMSDADM
```

Z1のアミノ酸配列

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence

```
MLRAAARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEWHDAVSRKTFPTVNPS
TGEVICQVAEGDKEDVDKAVKAARAAFQLGSPWRRMDASHRGRLLNRLADLIERDRTYLA
ALETLDNGKPYVISYLVDLDMVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCG
QIIPWNFPLLMQAWKLGPALATGNVVVMKVAEQTPLTALYVANLIKEAGFPPGVVNIVPG
FGPTAGAAIASHEDVDKVAFTGSTEIGRVIQVAAGSSNLKRVTLELGGKSPNIIMSDADM
```

Z2のアミノ酸配列

Or, upload file

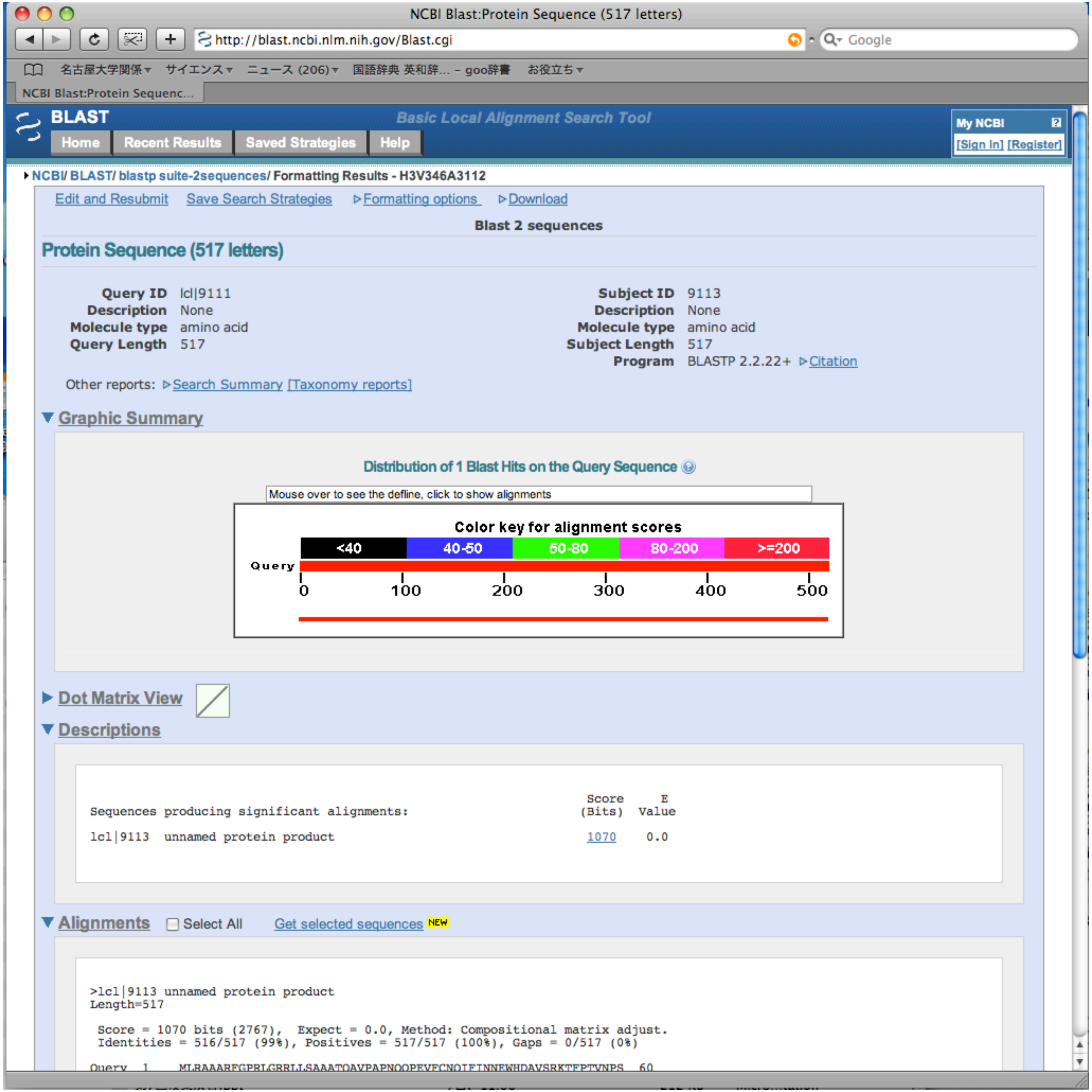
Program Selection

Algorithm ● blastp (protein-protein BLAST)

Choose a BLAST algorithm

Search protein sequence using Blastp (protein-protein BLAST)

☐ Show results in a new window



NCBI/ BLAST/ blastp suite-2sequences/ Formatting Results - H3V346A3112

Edit and Resubmit Save Search Strategies ▷Formatting options ▷Download

Blast 2 sequences

Protein Sequence (517 letters)

Query ID lcl|9111
Description None
Molecule type amino acid
Query Length 517

Subject ID 9113
Description None
Molecule type amino acid
Subject Length 517
Program BLASTP 2.2.22+ ▷Citation

Other reports: ▷Search Summary [Taxonomy reports]

▼ Graphic Summary

Distribution of 1 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

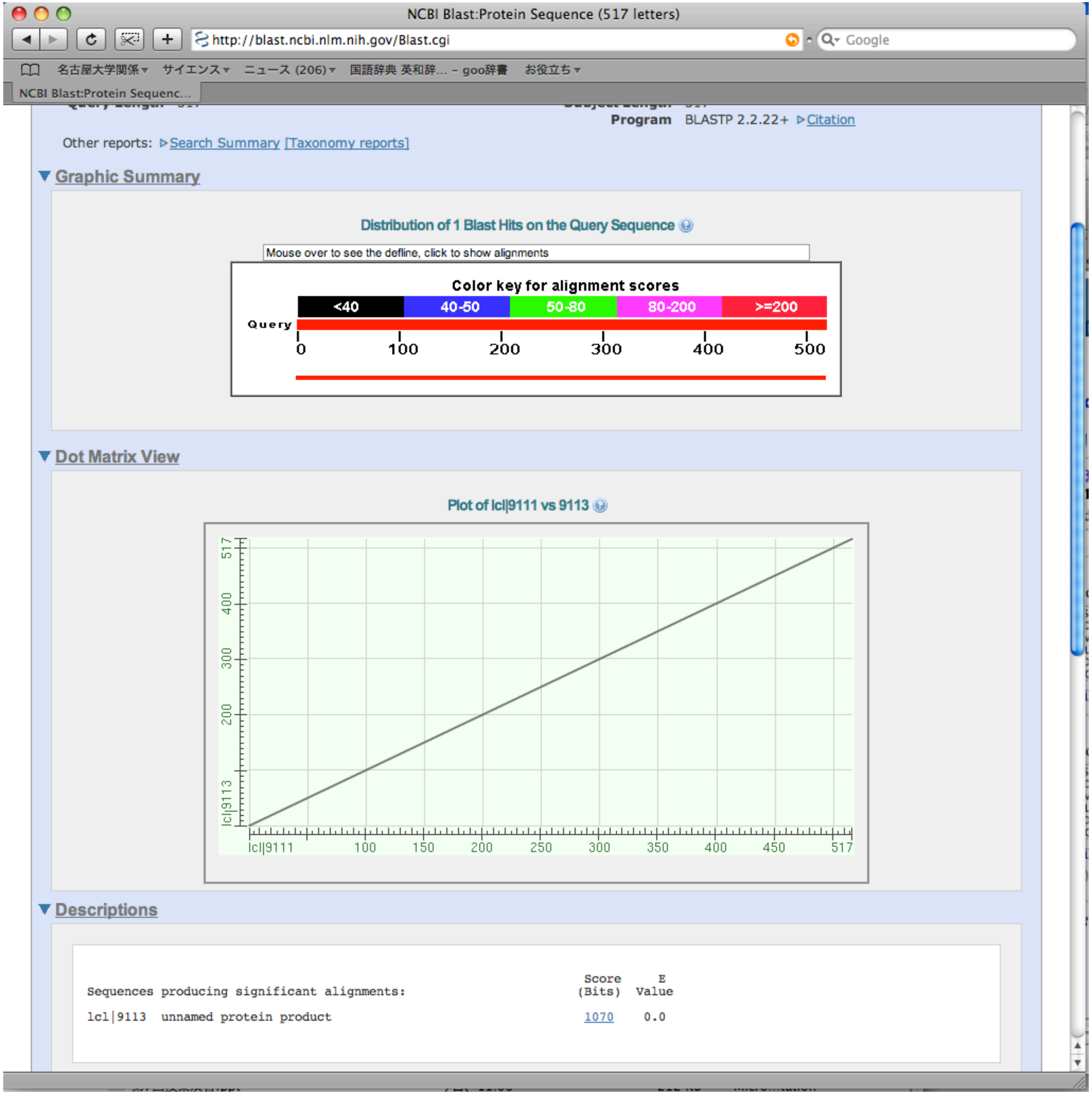
▶ Dot Matrix View

▼ Descriptions

```
Sequences producing significant alignments:                    Score     E
                                                               (Bits)  Value

lcl|9113  unnamed protein product                               1070    0.0
```

▼ Alignments ☐ Select All Get selected sequences NEW

```
>lcl|9113 unnamed protein product
Length=517

 Score = 1070 bits (2767),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 516/517 (99%), Positives = 517/517 (100%), Gaps = 0/517 (0%)

Query  1     MLRAAARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEWHDAVSRKTFPTVNPS  60
```



Program BLASTP 2.2.22+ ▷Citation

Other reports: ▷Search Summary [Taxonomy reports]

▼ Graphic Summary

Distribution of 1 Blast Hits on the Query Sequence

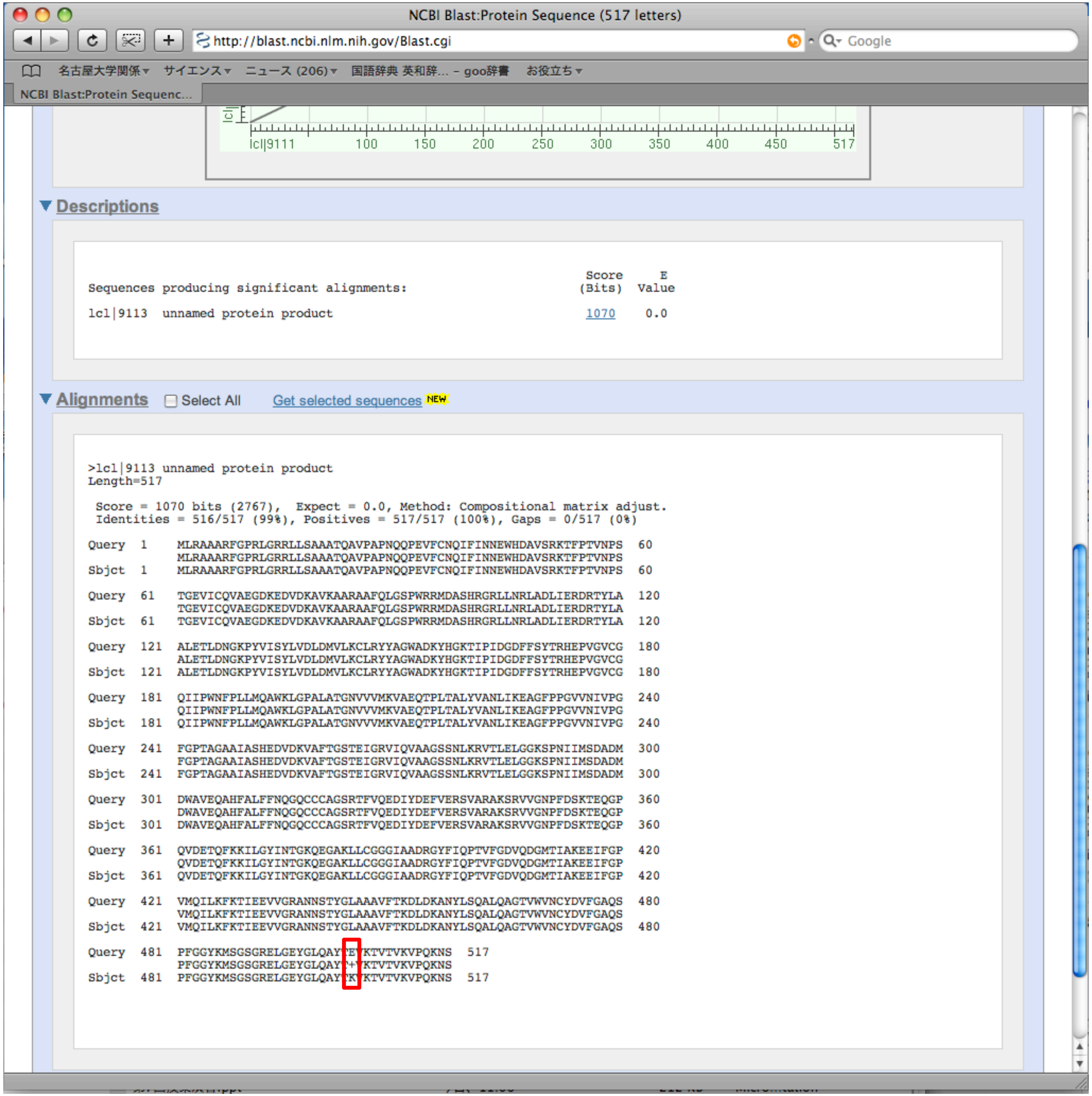
Mouse over to see the defline, click to show alignments

▼ Dot Matrix View

Plot of lcl|9111 vs 9113

▼ Descriptions

```
                                                               Score     E
Sequences producing significant alignments:                    (Bits)  Value

lcl|9113  unnamed protein product                               1070     0.0
```



Descriptions

```
Sequences producing significant alignments:                      Score    E
                                                                  (Bits)  Value

lcl|9113  unnamed protein product                                 1070    0.0
```

Alignments ☐ Select All Get selected sequences NEW

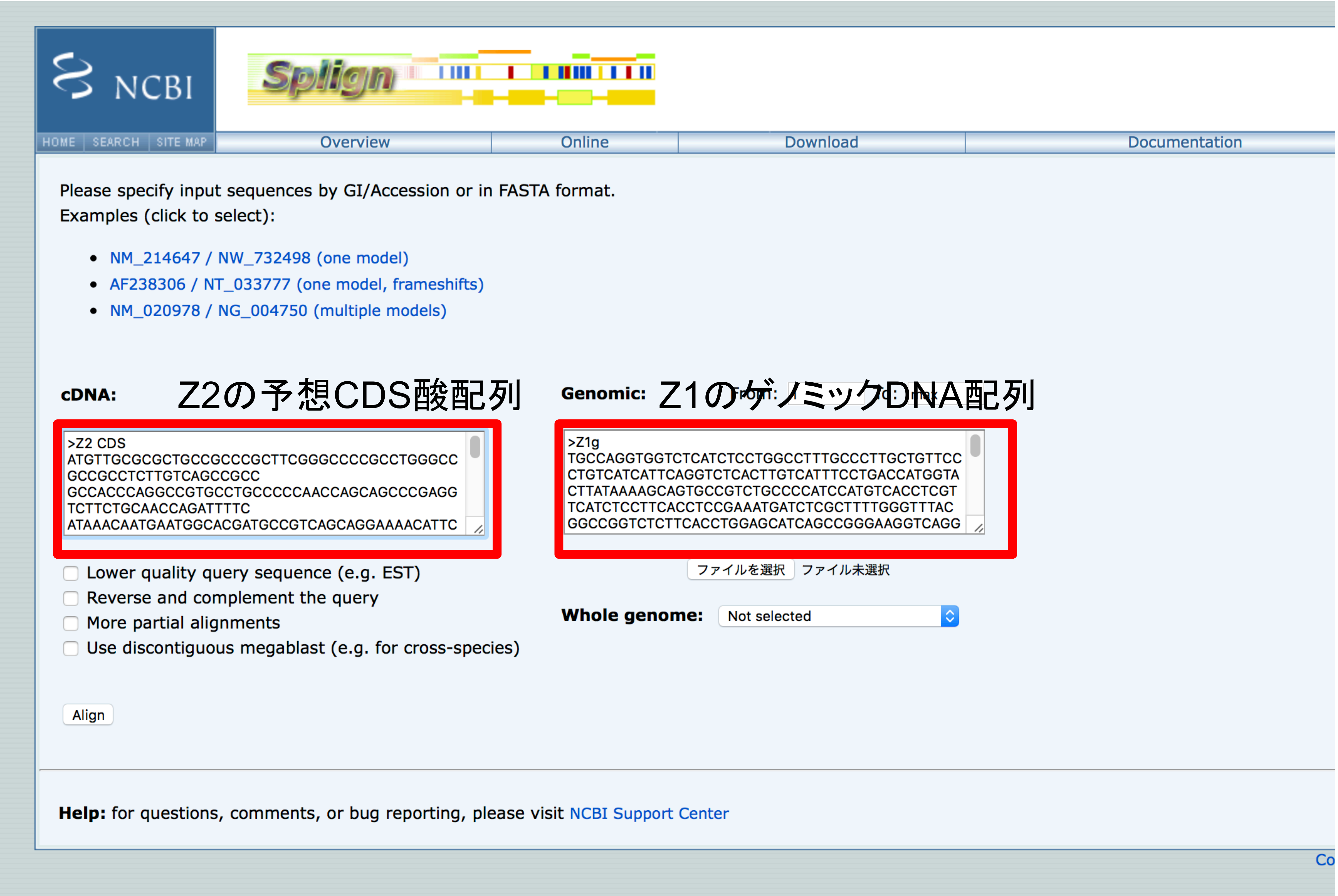
```
>lcl|9113 unnamed protein product
Length=517

 Score = 1070 bits (2767),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 516/517 (99%), Positives = 517/517 (100%), Gaps = 0/517 (0%)

Query  1    MLRAAARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEWHDAVSRKTFPTVNPS  60
            MLRAAARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEWHDAVSRKTFPTVNPS
Sbjct  1    MLRAAARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEWHDAVSRKTFPTVNPS  60

Query  61   TGEVICQVAEGDKEDVDKAVKAARAAFQLGSPWRRMDASHRGRLLNRLADLIERDRTYLA  120
            TGEVICQVAEGDKEDVDKAVKAARAAFQLGSPWRRMDASHRGRLLNRLADLIERDRTYLA
Sbjct  61   TGEVICQVAEGDKEDVDKAVKAARAAFQLGSPWRRMDASHRGRLLNRLADLIERDRTYLA  120

Query  121  ALETLDNGKPYVISYLVDLDMVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCG  180
            ALETLDNGKPYVISYLVDLDMVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCG
Sbjct  121  ALETLDNGKPYVISYLVDLDMVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCG  180

Query  181  QIIPWNFPLLMQAWKLGPALATGNVVVMKVAEQTPLTALYVANLIKEAGFPPGVVNIVPG  240
            QIIPWNFPLLMQAWKLGPALATGNVVVMKVAEQTPLTALYVANLIKEAGFPPGVVNIVPG
Sbjct  181  QIIPWNFPLLMQAWKLGPALATGNVVVMKVAEQTPLTALYVANLIKEAGFPPGVVNIVPG  240

Query  241  FGPTAGAAIASHEDVDKVAFTGSTEIGRVIQVAAGSSNLKRVTLELGGKSPNIIMSDADM  300
            FGPTAGAAIASHEDVDKVAFTGSTEIGRVIQVAAGSSNLKRVTLELGGKSPNIIMSDADM
Sbjct  241  FGPTAGAAIASHEDVDKVAFTGSTEIGRVIQVAAGSSNLKRVTLELGGKSPNIIMSDADM  300

Query  301  DWAVEQAHFALFFNQGQCCCAGSRTFVQEDIYDEFVERSVARAKSRVVGNPFDSKTEQGP  360
            DWAVEQAHFALFFNQGQCCCAGSRTFVQEDIYDEFVERSVARAKSRVVGNPFDSKTEQGP
Sbjct  301  DWAVEQAHFALFFNQGQCCCAGSRTFVQEDIYDEFVERSVARAKSRVVGNPFDSKTEQGP  360

Query  361  QVDETQFKKILGYINTGKQEGAKLLCGGGIAADRGYFIQPTVFGDVQDGMTIAKEEIFGP  420
            QVDETQFKKILGYINTGKQEGAKLLCGGGIAADRGYFIQPTVFGDVQDGMTIAKEEIFGP
Sbjct  361  QVDETQFKKILGYINTGKQEGAKLLCGGGIAADRGYFIQPTVFGDVQDGMTIAKEEIFGP  420

Query  421  VMQILKFKTIEEVVGRANNSTYGLAAAVFTKDLDKANYLSQALQAGTVWVNCYDVFGAQS  480
            VMQILKFKTIEEVVGRANNSTYGLAAAVFTKDLDKANYLSQALQAGTVWVNCYDVFGAQS
Sbjct  421  VMQILKFKTIEEVVGRANNSTYGLAAAVFTKDLDKANYLSQALQAGTVWVNCYDVFGAQS  480

Query  481  PFGGYKMSGSGRELGEYGLQAYTEVKTVTVKVPQKNS  517
            PFGGYKMSGSGRELGEYGLQAYT+VKTVTVKVPQKNS
Sbjct  481  PFGGYKMSGSGRELGEYGLQAYTKVKTVTVKVPQKNS  517
```

```
Z1CDS    GCATACACTGAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA
Z2CDS    GCATACACTAAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA
         ********* ********************************************
```

グルタミン酸

```
Z1Amionoacid   PFGGYKMSGSGRELGEYGLQAYTEVKTVTVKVPQKNS
Z2Amionoacid   PFGGYKMSGSGRELGEYGLQAYTKVKTVTVKVPQKNS
               ***********************:*************
```

リジン



Please specify input sequences by GI/Accession or in FASTA format.
Examples (click to select):

- NM_214647 / NW_732498 (one model)
- AF238306 / NT_033777 (one model, frameshifts)
- NM_020978 / NG_004750 (multiple models)

cDNA: Z2の予想CDS酸配列

```
>Z2 CDS
ATGTTGCGCGCTGCCGCCCGCTTCGGGCCCCGCCTGGGCC
GCCGCCTCTTGTCAGCCGCC
GCCACCCAGGCCGTGCCTGCCCCCAACCAGCAGCCCGAGG
TCTTCTGCAACCAGATTTTC
ATAAACAATGAATGGCACGATGCCGTCAGCAGGAAAACATTC
```

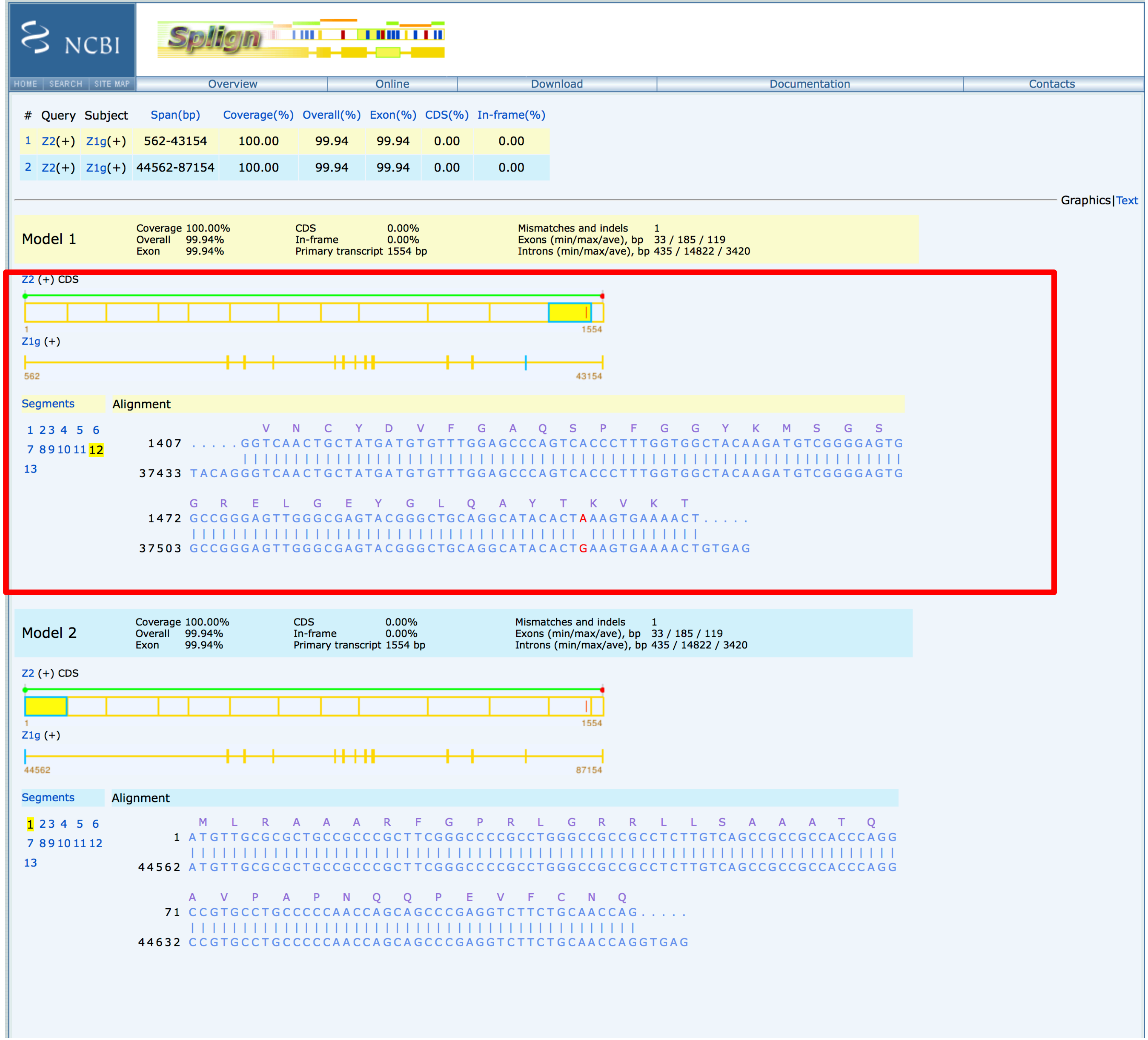
☐ Lower quality query sequence (e.g. EST)
☐ Reverse and complement the query
☐ More partial alignments
☐ Use discontiguous megablast (e.g. for cross-species)

Genomic: Z1のゲノミックDNA配列

```
>Z1g
TGCCAGGTGGTCTCATCTCCTGGCCTTTGCCCTTGCTGTTCC
CTGTCATCATTCAGGTCTCACTTGTCATTTCCTGACCATGGTA
CTTATAAAAGCAGTGCCGTCTGCCCCATCCATGTCACCTCGT
TCATCTCCTTCACCTCCGAAATGATCTCGCTTTTGGGTTTAC
GGCCGGTCTCTTCACCTGGAGCATCAGCCGGGAAGGTCAGG
```

Whole genome: Not selected

Help: for questions, comments, or bug reporting, please visit NCBI Support Center



#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	Z2(+)	Z1g(+)	562-43154	100.00	99.94	99.94	0.00	0.00
2	Z2(+)	Z1g(+)	44562-87154	100.00	99.94	99.94	0.00	0.00

Graphics|Text

Model 1

Coverage 100.00%
Overall 99.94%
Exon 99.94%

CDS 0.00%
In-frame 0.00%
Primary transcript 1554 bp

Mismatches and indels 1
Exons (min/max/ave), bp 33 / 185 / 119
Introns (min/max/ave), bp 435 / 14822 / 3420

Segments

1 2 3 4 5 6
7 8 9 10 11 12
13

Alignment

```
             V  N  C  Y  D  V  F  G  A  Q  S  P  F  G  G  Y  K  M  S  G  S
 1407 .....GGTCAACTGCTATGATGTGTTTGGAGCCCAGTCACCCTTTGGTGGCTACAAGATGTCGGGGAGTG
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
37433 TACAGGGTCAACTGCTATGATGTGTTTGGAGCCCAGTCACCCTTTGGTGGCTACAAGATGTCGGGGAGTG

      G  R  E  L  G  E  Y  G  L  Q  A  Y  T  K  V  K  T
 1472 GCCGGGAGTTGGGCGAGTACGGGCTGCAGGCATACACTAAAGTGAAACT.....
      |||||||||||||||||||||||||||||||||||||| ||||||||||
37503 GCCGGGAGTTGGGCGAGTACGGGCTGCAGGCATACACTGAAGTGAAACTGTGAG
```

Model 2

Coverage 100.00%
Overall 99.94%
Exon 99.94%

CDS 0.00%
In-frame 0.00%
Primary transcript 1554 bp

Mismatches and indels 1
Exons (min/max/ave), bp 33 / 185 / 119
Introns (min/max/ave), bp 435 / 14822 / 3420

Segments

1 2 3 4 5 6
7 8 9 10 11 12
13

Alignment

```
       M  L  R  A  A  A  R  F  G  P  R  L  G  R  R  L  L  S  A  A  A  T  Q
    1 ATGTTGCGCGCTGCCGCCCGCTTCGGGCCCCGCCTGGGCCGCCGCCTCTTGTCAGCCGCCGCCACCCAGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
44562 ATGTTGCGCGCTGCCGCCCGCTTCGGGCCCCGCCTGGGCCGCCGCCTCTTGTCAGCCGCCGCCACCCAGG

      A  V  P  A  P  N  Q  Q  P  E  V  F  C  N  Q
   71 CCGTGCCTGCCCCCAACCAGCAGCCCGAGGTCTTCTGCAACCAG.....
      ||||||||||||||||||||||||||||||||||||||||||||
44632 CCGTGCCTGCCCCCAACCAGCAGCCCGAGGTCTTCTGCAACCAGGTGAG
```

第12exon 第13exon

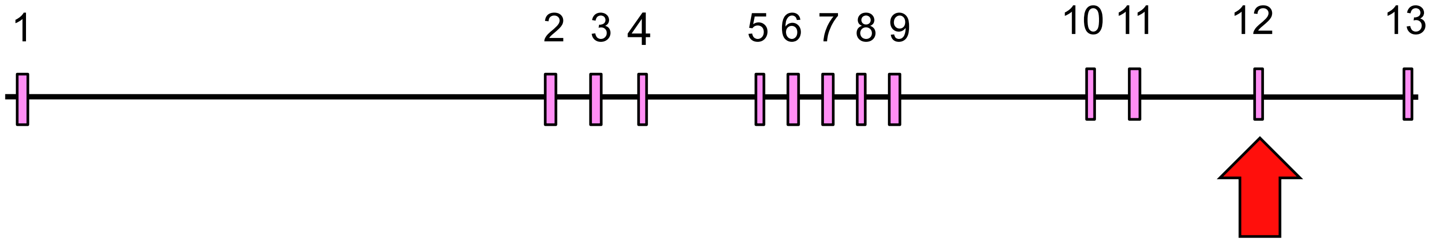
```
Z1CDS     GCATACACTGAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA
Z2CDS     GCATACACTAAAGTGAAAACTGTCACAGTCAAAGTGCCTCAGAAGAACTCATAA
          ********* ********************************************
```

グルタミン酸

第12exon 第13exon

```
Z1Amionoacid     PFGGYKMSGSGRELGEYGLQAYTEVKTVTVKVPQKNS
Z2Amionoacid     PFGGYKMSGSGRELGEYGLQAYTKVKTVTVKVPQKNS
                 ***********************:*************
```

リジン



アルコールの代謝

アルコール脱水素酵素

アセトアルデヒド脱水素酵素
（Aldehyde dehydrogenase2）

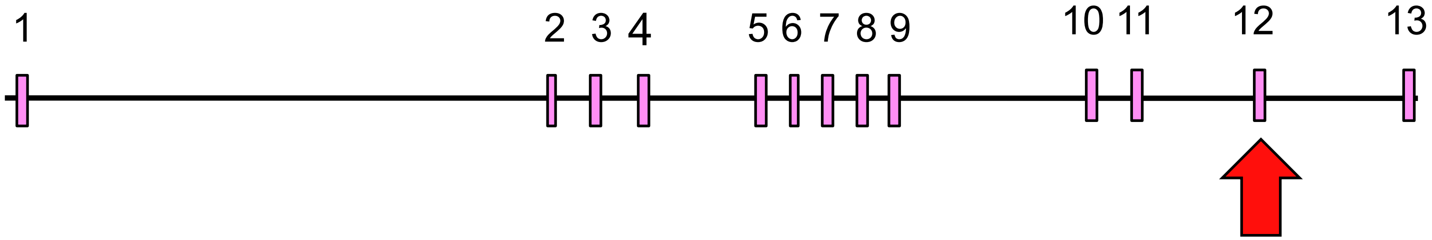
$$CH_3CH_2OH \xrightarrow{ADH} CH_3CHO \xrightarrow{ALDH} CH_3COOH$$

エタノール　アセトアルデヒド　酢酸

毒性

肝臓で働くアセトアルデヒド脱水素酵素には2種類ある。
ALDH1とALDH2

お酒に強い人、弱い人はALDH2遺伝子の1塩基の違いによる。
1塩基多型：SNP（Single Nucleotide Polymorphism）と呼ぶ



アルコールの代謝

アルコール脱水素酵素

アセトアルデヒド脱水素酵素
（Aldehyde dehydrogenase2）

$$CH_3CH_2OH \xrightarrow{\text{ADH}} CH_3CHO \xrightarrow{\xcancel{\text{ALDH}}} CH_3COOH$$

エタノール　アセトアルデヒド　酢酸

毒性

G型(1型)

CAG GCA TAC ACT GAA GAG AAA
--- G A Y T E V K --

A型 (2型)

CAG GCA TAC ACT AAA GAG AAA
--- G A Y T K V K --

酵素活性がない

アルコールの代謝

アルコール脱水素酵素

アセトアルデヒド脱水素酵素
（Aldehyde dehydrogenase2）

$$CH_3CH_2OH \xrightarrow{\text{ADH}} CH_3CHO \xrightarrow{\xcancel{\text{ALDH}}} CH_3COOH$$

エタノール　アセトアルデヒド　酢酸

毒性

G型(1型)

CAG GCA TAC ACT GAA GAG AAA
--- G A Y T E V K --

酸性アミノ酸

A型 (2型)

CAG GCA TAC ACT AAA GAG AAA
--- G A Y T K V K --

塩基性アミノ酸

酵素活性がない

GGtype：G型(1型)をホモにもつお酒に強い

AGtype：G型、A型をヘテロにもつ
お酒にほんの少し強い人（強い人の1/16）
（4量体を形成しAタイプが優性に活性を無力化するため）

AAtype：A型(2型)をホモにもつお酒に弱い