

BLAST basics

Shuji Shigenobu / 重信秀治

Aim

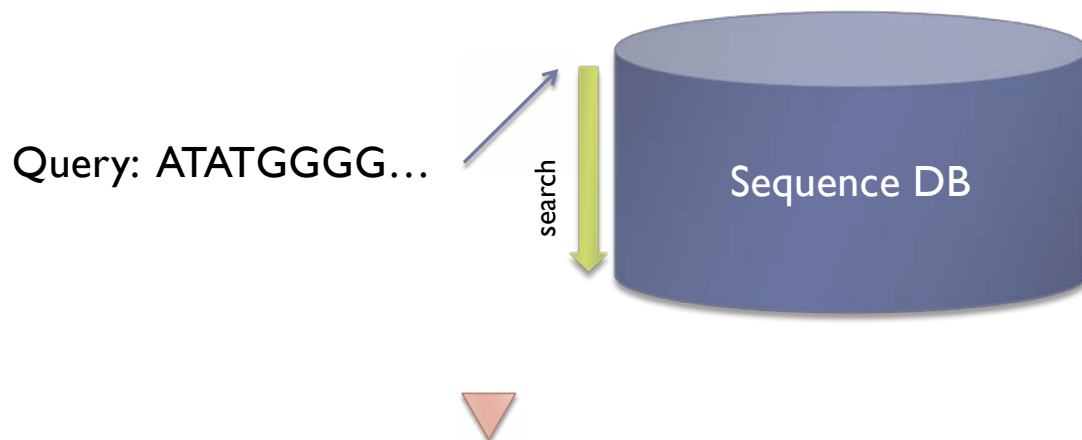
- BLASTの基本を学ぶ
- CUI環境でBLASTを実行する技術を習得する

What's **BLAST**

Basic Local Alignment Search Tool

- ▶ BLAST is the tool most frequently used for calculating sequence similarity
- ▶ BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families

Sequence Similarity Search (Database Search)



Find similar genes to get insights into:

- function
- evolution etc.

The screenshot shows the NCBI BLAST website interface. At the top, there are logos for NIH and NCBI, and a navigation bar with links like Home, Recent Results, Saved Strategies, and Help. The main content area is titled "Basic Local Alignment Search Tool" and describes the tool's function. Below this, there's a "Web BLAST" section with four options: "Nucleotide BLAST" (nucleotide to nucleotide), "blastx" (translated nucleotide to protein), "tblastn" (protein to translated nucleotide), and "Protein BLAST" (protein to protein). There's also a "BLAST Genomes" section with a search bar and buttons for Human, Mouse, Rat, and Microbes. At the bottom, there's a "Standalone and API BLAST" section with links for "Download BLAST", "Use BLAST API", and "Use BLAST in the cloud". Finally, there's a "Specialized searches" section with links for "SmartBLAST", "Primer-BLAST", "Global Align", and "CD-search". A URL box at the bottom right contains the text: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

BLAST search in GUI

BLASTに関する宿題

コースではコマンドライン上でCUIでBLAST解析を行います。予習として、NCBIのホームページ上でのGUIでのBLAST解析をやってみましょう。

accession no.: XM_002716705 はウサギのオートファジー遺伝子ATG3です。この遺伝子のマウスホモログ候補を探索する事を目的として、BLAST解析をやってみましょう。

- [NCBI BLAST website](#)

1. BLASTのトップヒットのマウスタンパク質は何ですか？
2. トップヒットのタンパク質とqueryとの、アミノ酸のペアワイズアライメントを示しなさい。

(コースでは同じ解析をCUIで行い、GUIでの結果、操作と比較します。accession#から配列を取得するステップ、複数のBLASTプログラムから適切な物を選択するステップ、BLASTを実行するステップ、BLASTの結果を解釈するステップ、これら4つのステップ全てが宿題です。)



Standard BLAST Programs

<u>Query</u>	<u>DB</u>	<u>Program</u>
▶ DNA	DNA	BLASTN
▶ Protein	Protein	BLASTP
▶ DNA translation	Protein	BLASTX
▶ Protein	DNA translation	TBLASTN
▶ DNA translation	DNA translation	TBLASTX

Quick start: command line BLAST search

ウサギのあるcDNA断片データを得た。これが何の遺伝子をコードしているのかを、マウスタンパク質データベースに対してsimilarity searchする事により明らかにする。


- ▶ Query: XM_002716705.3.fasta
- ▶ DB: mouse_proteins.pep.fasta

データファイルは、
~/data/SS 以下にある

1. Format DB

```
$makeblastdb -in mouse_proteins.pep.fasta -dbtype prot -parse_seqids
```

2. Search

```
$  -db mouse_proteins.pep.fasta -query XM_002716705.3.fasta
```

 適切なblastコマンドは？

Quick start: command line BLAST search

ウサギのあるcDNA断片データを得た。これが何の遺伝子をコードしているのかを、マウスタンパク質データベースに対してsimilarity searchする事により明らかにする。

- ▶ Query: XM_002716705.3.fasta
- ▶ DB: mouse_proteins.pep.fasta

データファイルは、
~/data/SS 以下にある

1. Format DB

```
$makeblastdb -in mouse_proteins.pep.fasta -dbtype prot -parse_seqids
```

2. Search

```
$blastx -db mouse_proteins.pep.fasta -query XM_002716705.3.fasta
```

```

BLASTX 2.5.0+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A.
Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J.
Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of
protein database search programs", Nucleic Acids Res. 25:3389-3402.

Database: mouse_proteins.pep.fasta
49,878 sequences; 21,467,191 total letters

Query= XM_002716705.3 PREDICTED: Oryctolagus cuniculus autophagy related 3
(ATG3), mRNA
Length=1516

Sequences producing significant alignments:

Score E
(Bits) Value
Q9CPX6 Ubiquitin-like-conjugating enzyme ATG3 OS=Mus musculus GN=Atg3 598 0.0
Q8R1P4 Ubiquitin-like-conjugating enzyme ATG10 OS=Mus musculus G... 35.0 0.13
AA084J029 Autophagy-related 10 (Yeast), isoform CRA_a OS=Mus mu... 34.7 0.14
AA08X1K62 Negative elongation factor B OS=Mus musculus, GN=Nelfb... 30.4 5.0
Q8V722 Coiled-coil domain-containing protein 138 OS=Mus musculus... 30.0 6.9
Q8C4Y3 Negative elongation factor B OS=Mus musculus, GN=Nelfb PE... 30.0 7.1
Q9JK38 Glucosamine 6-phosphate N-acetyltransferase OS=Mus muscu... 29.3 8.3
Q8CSK5 Uncharacterized protein CXorf38 homolog OS=Mus musculus P... 29.6 8.8

>Q9CPX6 Ubiquitin-like-conjugating enzyme ATG3 OS=Mus musculus GN=Atg3
PE=1 SV=1
Length=314

Score = 598 bits (1542), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 309/314 (98%), Positives = 313/314 (99%), Gaps = 0/314 (0%)
Frame = +1

Query 418 MNQVINTVKGKALEVAEYLTPVLKESKFKEGTGVTPEEFVAAGDHLVHCPTQWATGEE 597
Sbjct 1 MNQVINTVKGKALEVAEYLTPVLKESKFKEGTGVTPEEFVAAGDHLVHCPTQWATGEE 60
MNQVINTVKGKALEVAEYLTPVLKESKFKEGTGVTPEEFVAAGDHLVHCPTQWATGEE 60

Query 598 LKVKAYLPTGKQFLVTQWPCYKRCOMEYSDELEATIEEDGGGWDVTHNTGITGIT 777
Sbjct 61 LKVKAYLPTGKQFLVTQWPCYKRCOMEYSDELEATIEEDGGGWDVTHNTGITGIT 120
LKVKAYLPTGKQFLVTQWPCYKRCOMEYSDELEATIEEDGGGWDVTHNTGITGIT 120

Query 778 EAVKEITLENKDSIKLQDCSALCEEEDEGEAAMDEEYEESSGLLETDEATLDRKIVE 957
Sbjct 121 EAVKEITLENKDSIKLQDCSALCEEEDEGEAAMDEEYEESSGLLETDEATLDRKIVE 180
EAVKEITLENKDSIKLQDCSALCEEEDEGEAAMDEEYEESSGLLETDEATLDRKIVE 180

Query 958 ACKAKADAGGEDAIIQTRTYDLYITYKYOTPRLWFGYDEORQPLTVEHMYEDISODH 1137
Sbjct 181 ACKAKADAGGEDAIIQTRTYDLYITYKYOTPRLWFGYDEORQPLTVEHMYEDISODH 240
ACKAKADAGGEDAIIQTRTYDLYITYKYOTPRLWFGYDEORQPLTVEHMYEDISODH 240

Query 1138 VKKTYTIEHHPHLPPPPMCSVHPCRHAEMMKIETVAEGGDELGVHMYLLIFLKFOAV 1317
Sbjct 241 VKKTYTIEHHPHLPPPPMCSVHPCRHAEMMKIETVAEGGDELGVHMYLLIFLKFOAV 300
VKKTYTIEHHPHLPPPPMCSVHPCRHAEMMKIETVAEGGDELGVHMYLLIFLKFOAV 300

Query 1318 IPTIEYDYTRHFTM 1359
Sbjct 301 IPTIEYDYTRHFTM 314
IPTIEYDYTRHFTM 314

>Q8R1P4 Ubiquitin-like-conjugating enzyme ATG10 OS=Mus musculus GN=Atg10
PE=1 SV=1
Length=215

Score = 35.0 bits (79), Expect = 0.13, Method: Compositional matrix adjust.
Identities = 25/116 (22%), Positives = 52/116 (45%), Gaps = 11/116 (9%)
Frame = +1

Query 994 AILQITRYDLYITYKYOTPRLWFGYDEORQPLTVEHMYEDISODHVKK-----TV 1152
Sbjct 88 AVAEVTKHEYHVLVYCSYQVPLVFRASFLDGRPLALFDINFGVHCYKPLRLOGPWDTI 147
AVAEVTKHEYHVLVYCSYQVPLVFRASFLDGRPLALFDINFGVHCYKPLRLOGPWDTI 147

```

Anatomy of BLAST output (format 0)

header

one-line summaries

alignments

```

Sbjct 184 KKGGLL-PHEINIVDELVHIMKLOVETIAOKKFAELQKREQLAER-----EQL 234
KKGGLL-PHEINIVDELVHIMKLOVETIAOKKFAELQKREQLAER-----EQL 234

Query 682 EYSDE-----LEATIEEDGGGWDVTHNTGITTEAVKEITLENK 810
+5 E ++ EE + H T I +TEA+KE ENK
Sbjct 235 LFSHETALSKIKGVKEEVLTRFQILKEQGTETEHTEALKEKKNK 282
LFSHETALSKIKGVKEEVLTRFQILKEQGTETEHTEALKEKKNK 282

>Q8C4Y3 Negative elongation factor B OS=Mus musculus, GN=Nelfb PE=1 SV=2
Length=588

Score = 30.0 bits (66), Expect = 7.1, Method: Compositional matrix adjust.
Identities = 10/25 (40%), Positives = 16/25 (64%), Gaps = 1/25 (4%)
Frame = -2

Query 549 ISSCHKLWCYDTCFLKLFLEDRR 475
+ CHK WC D C ++ RF++ +R
Sbjct 256 VDPCHKLTNCLDAG--TRERVDSKR 279
VDPCHKLTNCLDAG--TRERVDSKR 279

>Q9JK38 Glucosamine 6-phosphate N-acetyltransferase OS=Mus musculus, GN=Gnpnat1
PE=1 SV=1
Length=184

Score = 29.3 bits (64), Expect = 8.3, Method: Compositional matrix adjust.
Identities = 20/73 (27%), Positives = 31/73 (42%), Gaps = 13/73 (18%)
Frame = +1

Query 337 WGRPTAQLLPGLPSPRRRAGRCSPRMONVINTVKGKALEVAEYLTPVLKESKFKEGTG 516
W + TA + P + S + G G P +N K L + ETVG
Sbjct 19 WSONTAIFSPAI--SPTHPGEGLVLRPLCTADLNKGFFKVL-----GQLTETGV 65
WSONTAIFSPAI--SPTHPGEGLVLRPLCTADLNKGFFKVL-----GQLTETGV 65

Query 517 ITPEEFVAAGDHL 555
++PE++ ++
Sbjct 66 VSQEFMKSFEHM 78
VSQEFMKSFEHM 78

>Q8CSK5 Uncharacterized protein CXorf38 homolog OS=Mus musculus PE=1
SV=1
Length=320

Score = 29.6 bits (65), Expect = 8.8, Method: Compositional matrix adjust.
Identities = 22/83 (27%), Positives = 37/83 (45%), Gaps = 2/83 (2%)
Frame = +1

Query 580 WATGEELKVKAYLPTGKQFLVTQWPCYKRCOMEYSDELEATIEEDGGGWDVTHNT 759
W ++K++ +L K + + V Y R +Q+ S D I EED+ DG +
Sbjct 177 WLRFQIKQNFVLEFKN--IPEIVAVYSRIEQLLTSQWAVHNPEDERDGCCEFGSYL 234
WLRFQIKQNFVLEFKN--IPEIVAVYSRIEQLLTSQWAVHNPEDERDGCCEFGSYL 234

Query 760 GITGITAEVKEITLENKDSIKLO 828
++ I E + E + LQ
Sbjct 235 SVSQINHEITELKEKLOEMYLO 257
SVSQINHEITELKEKLOEMYLO 257

Lambda K H a alpha
0.318 0.134 0.401 0.792 4.96

Gapped
Lambda K H a alpha sigma
0.267 0.0410 0.140 1.90 42.6 43.6

Effective search space used: 6420178198

Database: mouse_proteins.pep.fasta
Posted date: Nov 18, 2016 4:07 PM
Number of letters in database: 21,467,191
Number of sequences in database: 49,878

Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Neighboring words threshold: 12
Window for multiple hits: 40

```

Anatomy of BLAST output (format 0)

alignments

Hierarchical structure of alignment part

- alignment #1
 - HSP #1
 - HSP #2
 - ...
- alignment #2
- alignment #3
- ...

footer

Let's try

- ▶ help を表示させる
- ▶ -evalue オプションを使ってみる。E-value cutoff = $1.0e-8$ を指定してみよう
- ▶ 他のフォーマットで出力してみよう
 - ▶ format 6: table

What's BLAST

Basic Local Alignment Search Tool

- ▶ BLAST is the tool most frequently used for calculating sequence similarity
- ▶ Finds best local alignments
- ▶ Heuristic approach based on Smith Waterman algorithm
- ▶ Provides statistical significance
- ▶ BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Why and when **command-line BLAST** is useful?

- ▶ Automated
- ▶ Many queries
- ▶ Use local database (ex. Unpublished data)
- ▶ Procedures can be recorded as a shell script

NCBI BLAST+ command line tools

- ▶ Provided as BLAST+ suite by NCBI
 - ▶ (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download)
- ▶ Functionality offered by BLAST+ suite
 - ▶ 1) Search tools
 - ▶ 2) BLAST database tools
 - ▶ 3) sequence filtering tools

Installation and configuration

▶ Installation

- ▶ OSごとにコンパイル済みのバイナリが配布されている。ダウンロードして解凍するだけでよい。
 - ▶ <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>
- ▶ Manuals
 - ▶ <https://www.ncbi.nlm.nih.gov/books/NBK279671/>
 - ▶ <https://www.ncbi.nlm.nih.gov/books/NBK52640/>

▶ Setup and configuration

- ▶ インストールした blast programs にパスを通す
- ▶ 環境変数 BLASTDB を設定する

```
export BLASTDB=$HOME/myblastdb
```

(コース用Macおよびbias5はすでに環境設定済み)

Run command line BLAST program

1. Format DB

```
$makeblastdb -in mouse_proteins.pep.fasta -dbtype prot -parse_seqids
```

2. Search

```
$blastx -db mouse_proteins.pep.fasta -query XM_002716705.3.fasta \  
-outformat 6
```


Output formats

▶ Popular formats

- ▶ format 0: human friendly
- ▶ format 6: table
- ▶ format 7: table with comment lines
- ▶ format 5: XML
- ▶ format 11: BLAST archive (ASN.1)

▶ Others

- ▶ JSON, SAM etc.

Anatomy of BLAST output (format 6)

format 6/7 はcustomize可能

query	subject	%identity	align-len	mismatch	gap_open	q_start	q_end	s_start	s_end	bit-score	evaluate
spo:SPAC212.11	sce:YMR190C	27.297	370	238	12	1169	1525	653	1004	7.08e-23	105
spo:SPAC212.11	sce:YDR021W	38.053	113	64	3	1417	1529	267	373	1.42e-14	75.9
spo:SPAC212.11	sce:YNL112W	25.076	331	225	12	1204	1525	148	464	1.06e-12	70.9
spo:SPAC212.11	sce:YBR237W	24.403	377	237	17	1190	1532	276	638	1.68e-12	70.5
spo:SPAC212.11	sce:YOR204W	28.505	214	128	8	1352	1559	339	533	2.80e-12	69.7
spo:SPAC212.11	sce:YOR046C	28.205	234	140	8	1310	1530	227	445	1.65e-11	66.6
spo:SPAC212.11	sce:YDR243C	23.512	336	227	10	1208	1524	216	540	7.40e-11	65.1
spo:SPAC212.11	sce:YGL078C	22.689	357	239	11	1188	1527	130	466	7.62e-11	64.7
spo:SPAC212.11	sce:YPL119C	27.619	210	129	8	1352	1556	351	542	2.84e-10	63.2
spo:SPAC212.11	sce:YGL064C	22.195	410	216	16	1215	1541	166	555	8.25e-10	61.6
spo:SPAC212.11	sce:YDL084W	24.424	217	156	5	1308	1524	201	409	7.84e-07	51.6
spo:SPAC212.11	sce:YHR169W	27.273	143	93	4	1418	1555	257	393	1.09e-06	51.2
spo:SPAC212.11	sce:YLR276C	26.667	210	120	10	1189	1377	38	234	2.85e-06	50.1
spo:SPAC212.11	sce:YLR276C	34.146	41	27	0	1484	1524	386	426	0.12	35.0
spo:SPAC212.11	sce:YLL008W	21.965	346	233	13	1195	1524	258	582	4.71e-06	49.3
spo:SPAC212.11	sce:YPL082C	24.074	108	81	1	1417	1524	1648	1754	0.002	40.8
spo:SPAC212.11	sce:YGL070C	37.838	37	23	0	1583	1619	52	88	2.3	29.3
spo:SPBPB10D8.05C	sce:YPL092W	30.730	397	224	9	10	356	11	406	3.04e-47	165
spo:SPBPB10D8.05C	sce:YGL195W	31.034	58	37	2	140	195	2014	2070	0.37	30.8
spo:SPBPB10D8.05C	sce:YDR283C	51.613	31	14	1	221	250	589	619	0.66	30.0
spo:SPBPB10D8.05C	sce:YBR028C	35.135	37	24	0	229	265	127	163	0.84	29.6
spo:SPBPB10D8.05C	sce:YPL027W	27.957	93	54	3	138	221	53	141	3.1	27.3
spo:SPBPB10D8.05C	sce:YDR161W	28.571	49	35	0	318	366	102	150	3.2	27.7
spo:SPBPB10D8.05C	sce:YER166W	27.451	51	35	1	50	98	1273	1323	7.0	26.9
spo:SPBPB10D8.05C	sce:YGR040W	23.256	86	55	3	214	299	3	77	7.8	26.6
spo:SPBPB10D8.05C	sce:YAR019C	44.444	18	10	0	235	252	30	47	9.5	26.2
spo:SPBC359.02	sce:YOR368W	29.885	87	56	2	53	138	7	89	0.88	29.3
spo:SPCC330.07C	sce:YHR197W	26.761	71	39	3	65	130	287	349	2.0	28.9
spo:SPCC330.07C	sce:YGR224W	23.881	67	50	1	181	247	152	217	5.1	27.7

Anatomy of BLAST output (format 7)

```
out.bl7.txt — Edited v
# BLASTX 2.5.0+
# Query: XM_002716705.3 PREDICTED: Oryctolagus cuniculus autophagy related 3 (ATG3), mRNA
# Database: mouse_proteins.pep.fasta
# Fields: query acc., subject acc., % identity, alignment length, mismatches, gap opens, q. start,
q. end, s. start, s. end, evalue, bit score
# 8 hits found
XM_002716705.3      Q9CPX6 98.408 314    5    0    418    1359    1    314    0.0    598
XM_002716705.3      Q8R1P4 21.552 116    80    3    994    1320    88    199    0.13   35.0
XM_002716705.3      Q0VF22 27.778 108    64    3    502    810    184    282    6.9    30.0
XM_002716705.3      Q8C4Y3 40.000 25    14    1    549    475    256    279    7.1    30.0
XM_002716705.3      Q9JK38 27.397 73    40    2    337    555    19    78    8.3    29.3
XM_002716705.3      Q8C5K5 26.506 83    59    1    580    828    177    257    8.8    29.6
# BLAST processed 1 queries
```

Concise.

Useful for large scale analysis.

Tab-delimited text is easy to parse.

You can modified the format. (explain later)

Multiple queries

ウサギのあるcDNA断片データを100個得た。これが何の遺伝子をコードしているのかを、マウスタンパク質データベースに対してsimilarity searchする事により明らかにする。多数の配列を解析したいときにこそCUIの本領発揮。

▶ Query: rabbit_100mRNAs.nuc.fasta

データファイルは、
~/data/SS 以下にある

▶ DB: mouse_proteins.pep.fasta

▶ options:

▶ evalue < 1.0e-5; format 7; 高速化のためにCPUを4コア使う。

1. Format DB

新たに作成する必要はない。

2. Search

自分で考えてみよう

Multiple queries

ウサギのあるcDNA断片データを100個得た。これが何の遺伝子をコードしているのかを、マウスタンパク質データベースに対してsimilarity searchする事により明らかにする。多数の配列を解析したいときにこそCUIの本領発揮。

- ▶ Query: rabbit_100mRNAs.nuc.fasta
- ▶ DB: mouse_proteins.pep.fasta

1. Format DB

新たに作成する必要はない。

2. Search

```
blastx -db mouse_proteins.pep.fasta -query rabbit_100mRNAs.nuc.fasta  
-evalue 1.0e-5 -num_threads 4 -outfmt 7 > practice.out.txt
```

Advanced usage

- ▶ Pairwise alignment by using blast commands
- ▶ Advanced parameters
- ▶ Customize output format
- ▶ blastdbcmdを使った配列データのmanipulation – Day 2
- ▶ advanced blast search tools – Day 2

Pairwise BLAST

BLASTプログラムを使って、2つの配列のローカルアライメントをとることも可能である。blastdbcmdでデータベースを作製する必要はない。以下は、ウサギとマウスのAtg3タンパク質の配列である。blastpを使ってpairwiseアライメントをとってみよう。

- ▶ Query 1: OcAtg3.aa.fa
- ▶ Query 2: MmAtg3.aa.fa

Alignment

```
$blastp -query OcAtg3.aa.fa -subject MmAtg3.aa.fa
```

-dbのかわりに-subjectを使う以外は、通常のBLASTとほぼ同じ。Database formatは不要。マルチスレッドは効かない。

BLASTN –task option

- ▶ **-task**
 - ▶ **megablast** (default): for very similar sequences
 - ▶ **dc-megablast**: typically used for inter-species comparisons
 - ▶ **blastn**: traditional program used for inter-species comparisons
 - ▶ **blastn-short**: optimized for sequences less than 30 nucleotides

BLASTN: -task

肺炎クラミジア *Chlamydomonas reinhardtii* のCVL029株と、GPIC株のゲノム全長をBLASTNで比較しなさい。-task で megablast (default) , blastn, dc-megablast を指定したときで結果がどのように変わるだろうか。

- ▶ Seq1: CpneCVL029.NC_000922.uc.genome.fa
- ▶ Seq2: CpneGPIC.AE015925.uc.genome.fa

Alignment

```
$blastn -task megablast -subject seq1 -query seq2 -outfmt 6
```

-task を blastn, dc-megablast に変えてみよう。

Customize output format

- ▶ Format 6/7 can be arranged by users

OcAtg3.aa.fa と MmAtg3.aa.fa のpairwise BLASTをカスタムフォーマットで出力する例。

```
$blastp -query OcAtg3.aa.fa -subject MmAtg3.aa.fa \
-outfmt "6 std qseq sseq"
```

標準の12カラム(qacc sacc pident length mismatch gapopen qstart qend sstart send eval evalue bitscore)に加えて、qseqとsseqを表示

Options 6, 7, 10 and 17 can be additionally configured to produce a custom format specified by space delimited format specifiers.

The supported format specifiers for options 6, 7 and 10 are:

- qseqid means Query Seq-id
- qgi means Query GI
- qacc means Query accession
- qaccver means Query accession.version
- qlen means Query sequence length
- sseqid means Subject Seq-id
- sallseqid means All subject Seq-id(s), separated by a ';'
- sgi means Subject GI
- sallgi means All subject GIs
- sacc means Subject accession
- saccver means Subject accession.version
- sallacc means All subject accessions
- slen means Subject sequence length
- qstart means Start of alignment in query
- qend means End of alignment in query
- sstart means Start of alignment in subject
- send means End of alignment in subject
- qseq means Aligned part of query sequence
- sseq means Aligned part of subject sequence
- evaluate means Expect value
- bitscore means Bit score
- score means Raw score
- length means Alignment length
- pident means Percentage of identical matches
- nident means Number of identical matches
- mismatch means Number of mismatches
- positive means Number of positive-scoring matches
- gapopen means Number of gap openings
- gaps means Total number of gaps

- positive means Number of positive-scoring matches
- gapopen means Number of gap openings
- gaps means Total number of gaps
- ppos means Percentage of positive-scoring matches
- frames means Query and subject frames separated by a '/'
- qframe means Query frame
- sframe means Subject frame
- btop means Blast traceback operations (BTOP)
- staxid means Subject Taxonomy ID
- ssciname means Subject Scientific Name
- scomname means Subject Common Name
- sblastname means Subject Blast Name
- sskingdom means Subject Super Kingdom
- staxids means unique Subject Taxonomy ID(s), separated by a ';'
- (in numerical order)
- sscinames means unique Subject Scientific Name(s), separated by a ';'
- scomnames means unique Subject Common Name(s), separated by a ';'
- sblastnames means unique Subject Blast Name(s), separated by a ';'
- (in alphabetical order)
- sskingdoms means unique Subject Super Kingdom(s), separated by a ';'
- (in alphabetical order)
- stitle means Subject Title salltitles means All Subject Title(s), separated by a '<>'
- sstrand means Subject Strand
- qcovs means Query Coverage Per Subject
- qcovhsp means Query Coverage Per HSP
- qcovus means Query Coverage Per Unique Subject (blastn only)

When not provided, the default value is: 'qaccver saccver pident length mismatch gapopen qstart qend sstart send evaluate bitscore', which is equivalent to the keyword 'std'

Exercises

- ▶ ex 1-1: (blastn) mapping cDNA/EST to a genome
- ▶ ex 1-2: (blastp) homolog candidate search
- ▶ ex 1-3: (blastx) annotate RNA-seq contigs
- ▶ ex 1-4: (tblastn) mapping proteins to genome
- ▶ ex 1-5: pairwise alignment by BLAST
- ▶ ex 1-6: BLASTN -task option
- ▶ ex 1-7: blastn: CRISPR off-target check
- ▶ ex 1-8: compare two closely related genome
- ▶ ex 1-9: mapping a PacBio read to a genome