Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

BLAST Scores and Practical BLASTing

Lec˙07˙slides

Outline for today

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1. Review
   - Local versus global alignment
2. Practical BLASTing
   - Protein searches
   - Types of BLAST
   - E-value
   - Key steps
3. Summary
4. Reading for next class

Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

Review: Local versus global alignment

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

Use **local alignment** (BLAST) when:

- Searching databases for biologically related sequences

```
Human catggccctgtggatgcgcctcctgcccctgctggcgctgctggccc
      |||||||||||||||||| |||||||||||||| |||||   ||
Mouse catggccctgtggatgcgcttcctgcccctgctggccctgctcttcc
```

- Analyzing multidomain proteins

Use **multiple alignment** (CLUSTALW, etc) when:

- Aligning several members of a protein family (soon...)

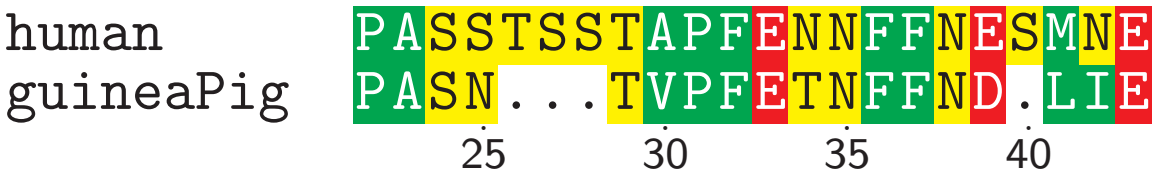
```
P1   GLGIEIIGTLQLVLCVLATTDR.RRRDLGG
P2   AVTVELFLTMQLVLCIFASTDE.RRGDNLG
P3   GFFDQFIGTAALIVCVLAIVDPYNNPVPRG
P4   GLLVELIITFQLVFTIFASCDS.KRTDVTG
```

```
P1   SAPL
P2   SPAL
P3   LEAF
P4   SVAL
```

Review: Calculating alignment score

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

+	identical positions
−	penalties from amino acid substitution matrix
−	gap opening penalties
−	gap extension penalties
=	**alignment score**



X acidic (−)
X basic (+)
X polar uncharged
X hydrophobic nonpolar

Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

BLAST acronym

BLAST =

Basic:

Local: versus global. BLAST align fragments of query

Alignment: like, uh, why we're here

Search: BLAST aligns your query against large database, looks for best alignments

Tool:

When do we use BLAST?

- Find homologues (orthologs & paralogs)
- Find other proteins with conserved domains
- Find repeating domains in same protein (like a dotplot)

Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

BLAST works on DNA or protein sequences

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

- Blast can search by DNA or translated amino acid sequence.
- **Protein searching is usually better**
- Deep question: Why?
  - Hint 1: Think about genetic code and degeneracy
  - Hint 2: Think about substitution matrices and amino acid properties
  - Hint 3: Think about evolution of DNA sequences
- When would you use a DNA BLAST search instead?

Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

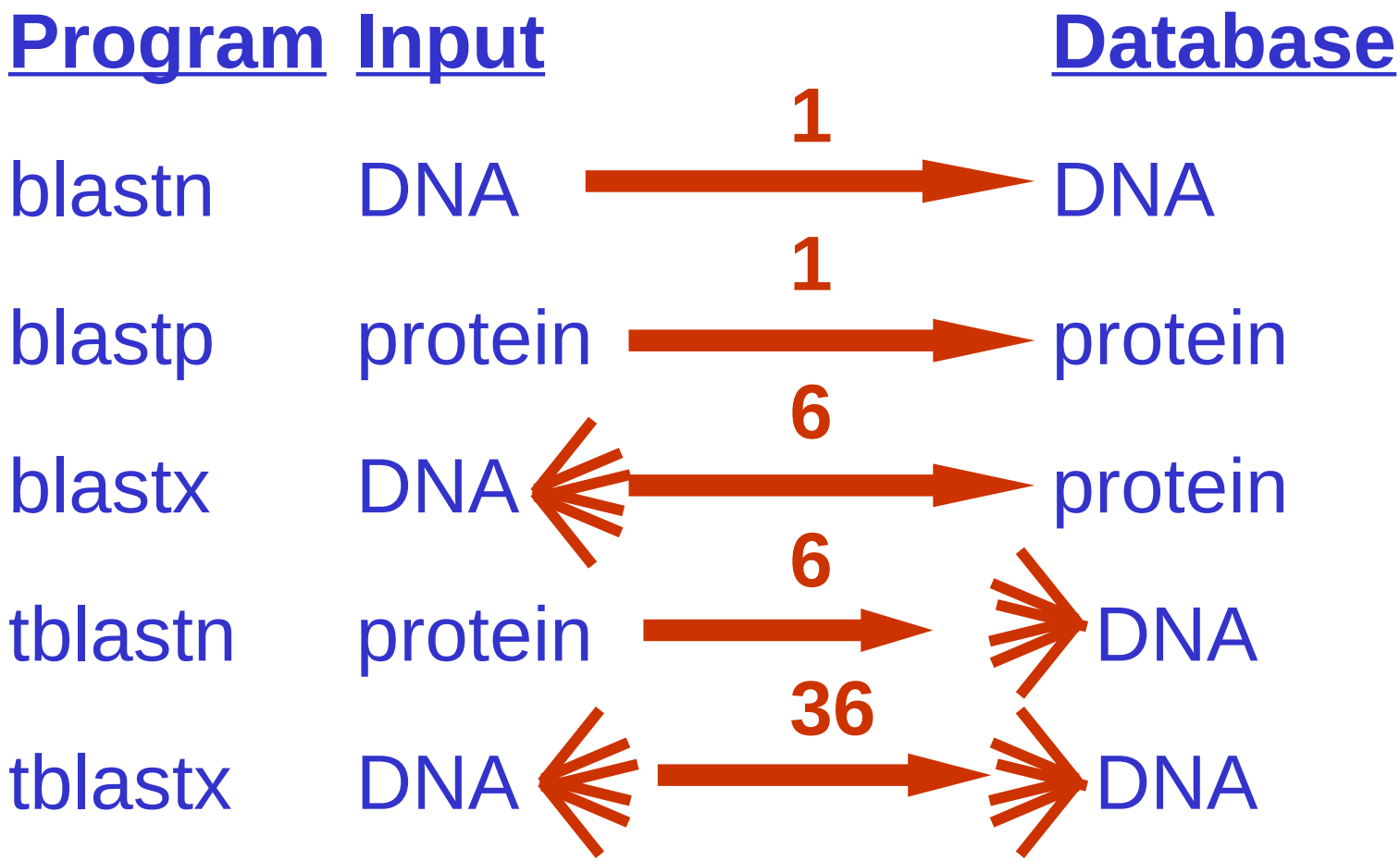
Types of BLAST database searches

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

Type	Query	Database	Remarks
blastn	DNA	DNA	
blastp	protein	protein	
blastx	DNA	protein	Tries all 6 reading frames in DNA query
tblastn	protein	DNA	Tries all 6 DNA reading frames in database
tblastx	DNA	DNA	All $6 \times 6 = 36$ combinations of reading frames (slow!!)

When would you use each of these?

BLAST programs illustrated



BLAST Demo

`http://blast.ncbi.nlm.nih.gov/Blast.cgi`

Things we'll try:

1. Get BLAST help
2. Simple protein blast against NR database (NP_001612 aryl hydrocarbon receptor)
3. Look at taxonomy results
4. Look at main results table & E-values
   - Click on **Description** column to see actual alignment
   - Click on **Accession** to see BLAST hit in GenBank
5. Examine E-values and bit scores
6. Edit and resubmit BLAST parameters

Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

How do we evaluate a BLAST hit?

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

E-value

sp|P32871|P11A_BOVIN	PHOSPHATIDYLINOSITOL 3-KINASE CATALYTI...	680	0.0
sp|P42336|P11A_HUMAN	PHOSPHATIDYLINOSITOL 3-KINASE CATALYT...	676	0.0
sp|P42337|P11A_MOUSE	PHOSPHATIDYLINOSITOL 3-KINASE CATALYT...	674	0.0
sp|P42338|P11B_HUMAN	PHOSPHATIDYLINOSITOL 3-KINASE CATALYT...	338	9e-93
sp|O35904|P11D_MOUSE	PHOSPHATIDYLINOSITOL 3-KINASE CATALYT...	332	7e-91
sp|O00329|P11D_HUMAN	PHOSPHATIDYLINOSITOL 3-KINASE CATALYT...	331	2e-90
sp|P47473|RIR1_MYCGE	RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE A...	34	0.59

Definition: E-value

The number of hits we'd expect by chance.
Small E-value $\rightarrow$ unlikely by chance $\rightarrow$ likely homology

Example: Database with 1000 DNA sequences:

Query	**E-value**
T	$\sim$ 1000
TA	fewer
TACCGATGTTGGACGTATTAGCTTAGGAC	$\sim$ 0

Informally defining E-value

Slightly simplified but shows the key idea:

$$E = m \times n \times 2^{-S}$$

Where:

- E expected number of hits
- m database size
- n query size
- S alignment score

S	2^{-S}
1	0.5
5	0.03
10	0.001
50	9×10^{-16}
100	8×10^{-31}

Key idea: **High alignment score $\longrightarrow$ small E value**

Interpreting E-values

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

the best scores are:

					E(86391)
SW:P11A_BOVIN	P32871	PHOSPHATIDYLINOSITOL 3-KINAS	(1068)	2228 493	1.2e-138
SW:P11A_HUMAN	P42336	PHOSPHATIDYLINOSITOL 3-KINAS	(1068)	2216 490	7.4e-138
SW:P11A_MOUSE	P42337	PHOSPHATIDYLINOSITOL 3-KINAS	(1068)	2204 488	4.5e-137
SW:P11B_HUMAN	P42338	PHOSPHATIDYLINOSITOL 3-KINAS	(1070)	1126 254	1.1e-66

↓ other sequences

SW:ESR1_YEAST	P38111	ESR1 PROTEIN.	(2368)	144 41	0.028
SW:PRA2_USTMA	P31303	PHEROMONE RECEPTOR 2.	(346)	116 35	0.35
SW:TEL1_YEAST	P38110	TELOMER LENGTH REGULATION PR	(2787)	127 37	0.42
SW:YA51_METJA	Q58451	HYPOTHETICAL PROTEIN MJ1051.	(513)	112 34	0.91
SW:RIR1_MYCGE	P47473	RIBONUCLEOSIDE-DIPHOSPHATE R	(721)	106 33	3
SW:YAY1_SCHPO	Q10209	HYPOTHETICAL 44.8 KDA PROTEI	(392)	99 31	5.1
SW:PAFA_CAVPO	P70683	PLATELET-ACTIVATING FACTOR A	(436)	96 30	8.8
SW:KC47_ORYSA	P29620	CDC2+/CDC28-RELATED PROTEIN	(424)	95 30	9.9

- Shorthand for scientific notation: 1.2e-138 $= 1.2 \times 10^{-138}$ $= 0.000000000000000000000000000000000000 \ldots 12$
- Rule of thumb: consider E-values < 0.1 or 0.05 as interesting.
- But also consider annotations

Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

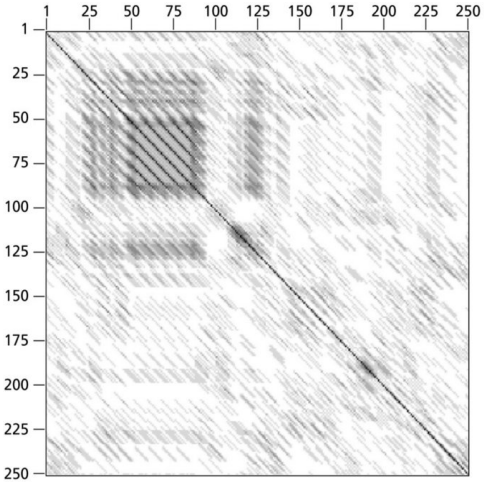
1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

Key steps in BLASTing

1. Choose a BLAST program (blastp, blastn, etc)
2. Enter the DNA or protein sequence
3. Choose a database
   - nr = non-redundant, default, not as good as RefSeq,
   - RefSeq
   - others
4. Adjust parameters (if necessary)
5. BLAST
6. Evaluate results (low E-values)
7. *Think like a biologist:*
   Follow interesting leads to PubMed, OMIM, ...

Low-complexity regions cause false hits

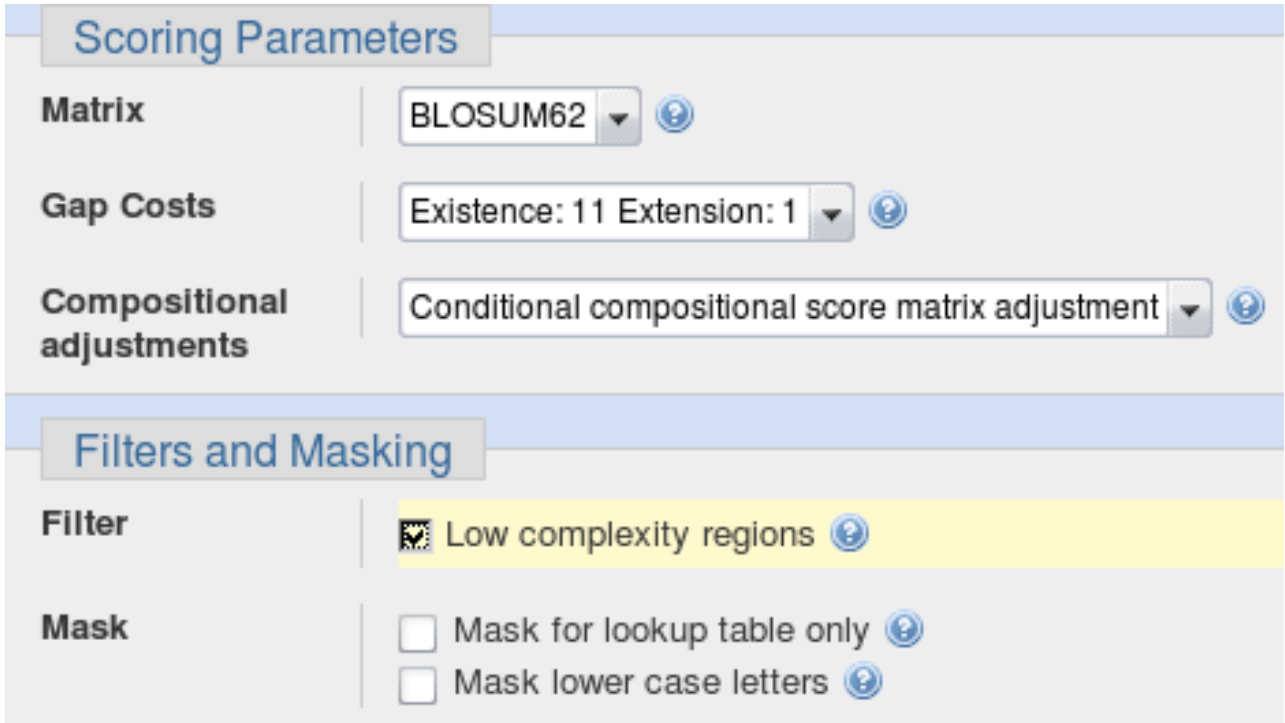
Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class




```
Score =  312 bits (792), Expect = 5e-85
Identities = 154/236 (65%), Positives = 154/236 (65%)

Query: 64   MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYXXXXXXXXXXX
            MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRY
Sbjct: 1    MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQP 23

Query: 124  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTHSQWNKPSKPKTNMKHMXXXXXXXX 0
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTHSQWNKPSKPKTNMKHM
Sbjct: 1    HGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMAGAAAAGA 83

Query: 184  XXXXXXXXXXXXXXXRPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCV 20
                         RPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCV
Sbjct: 121  VVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCV 43

Query: 244  NITIKQHXXXXXXXXXXXXXXDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFS 299 80
            NITIKQH               DVKMMERVVEQMCITQYERESQAYYQRGSSMVLFS
Sbjct: 181  NITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFS 236
```

BLAST can filter low-complexity regions

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class



Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

Outline for today

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1. Review
   - Local versus global alignment
2. Practical BLASTing
   - Protein searches
   - Types of BLAST
   - E-value
   - Key steps
3. Summary
4. Reading for next class

Outline

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

1 Review
Local versus global alignment
2 Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
3 Summary
4 Reading for next class

Review
Local versus global alignment
Practical BLASTing
Protein searches
Types of BLAST
E-value
Key steps
Summary
Reading for next class

Chapter 4

Section "Stand-alone BLAST"
to
Section "Using BLAST for gene discovery: find-a-gene"

Pages	Notes
136-155	Read