BIOS 477/877
Bioinformatics and Molecular Evolution

Instructor: Etsuko Moriyama
(School of Biological Sciences)

Spring 2026 **Lecture 18**

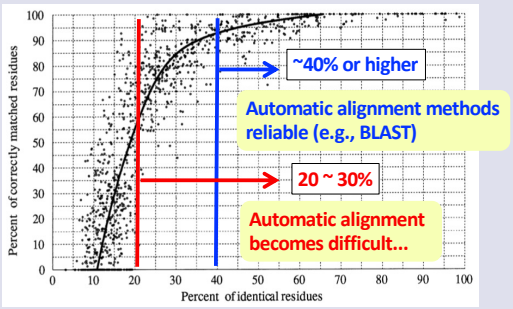
1

Today's topics

- Pattern and profile (PSSM)
- Hidden Markov model and profile HMM
- PSI-BLAST

2

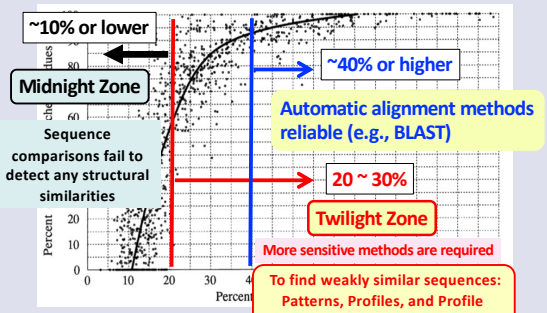
Sequence similarity and search sensitivity



Vogt *et al.* (1995)

3

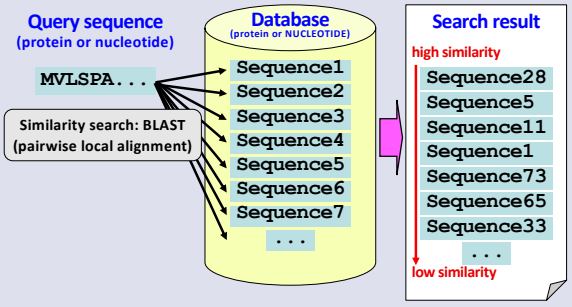
Sequence similarity and search sensitivity



Vogt *et al.* (1995)

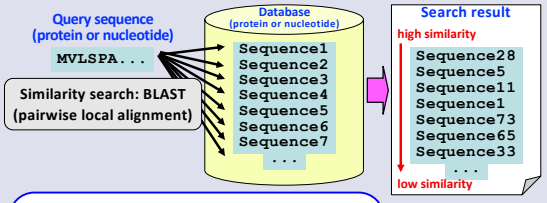
4

Similarity search



5

Finding remote similarity

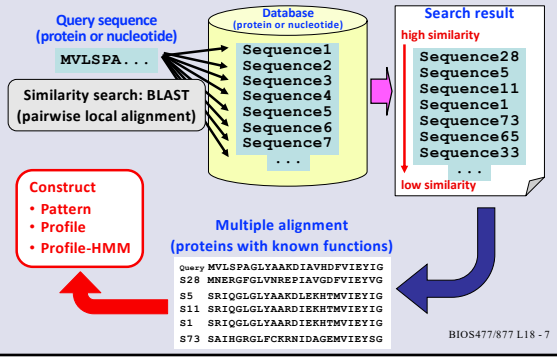


Finding more remotely (weakly) similar sequences helps:

- identify distant members of the protein family (how proteins have evolved)
- identify functions (unknown or not very obvious from the query sequence)

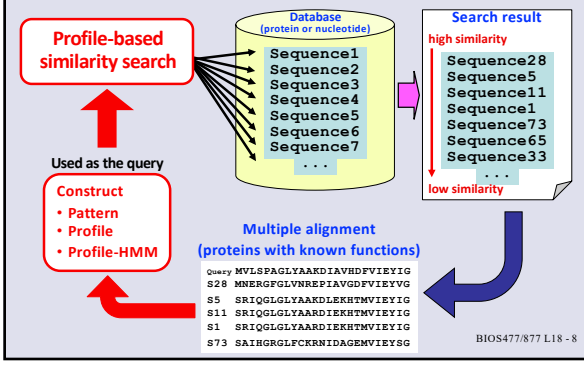
6

How remote similarity search is done



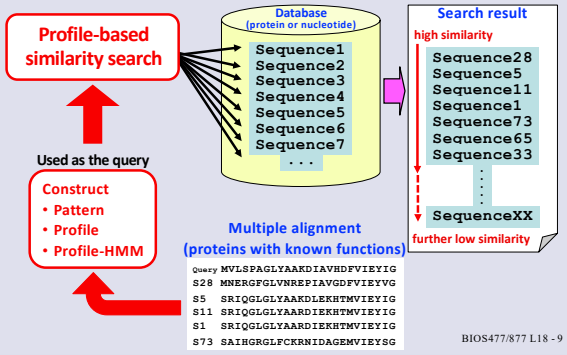
7

How remote similarity search is done



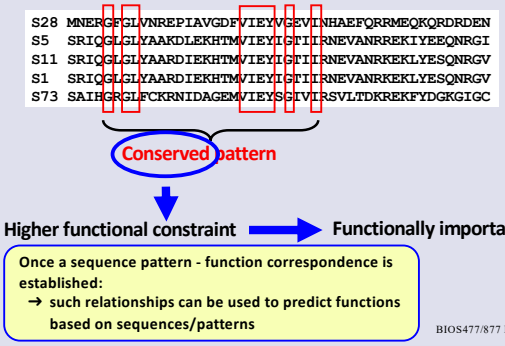
8

How remote similarity search is done



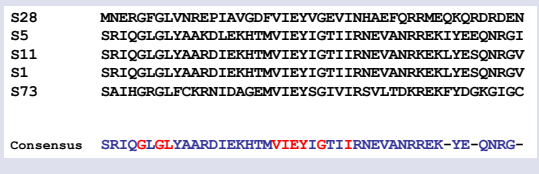
9

Alignments → patterns → functions



10

How to identify conserved patterns

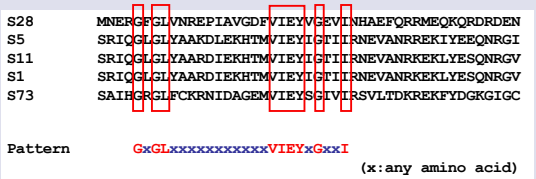


Majority-rule consensus sequence

Residues probably not important for functions can be included

11

Conserved pattern

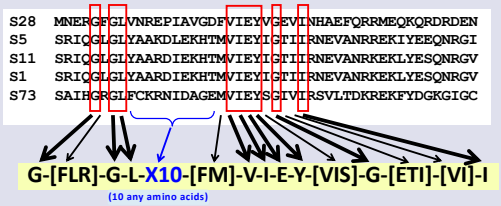


Conserved pattern including only identical sites

→ Very (too) strict

12

Reguler expression pattern



Regular expression

➜ More flexible than strict conserved pattern

13

PROSITE:
Database of protein domains, families and functional sites

https://prosite.expasy.org/

- PROSITE Documentation (entry)
  - → Biological information about protein families, domains, *etc.*

PROSITE documentation PDOC00210
G-protein coupled receptors family 1 signature and profile

- Two types of motif descriptors
  - Patterns (regular expressions)

G_PROTEIN_RECEP_F1_1, PS00237; G-protein coupled receptors family 1 signature (PATTERN)

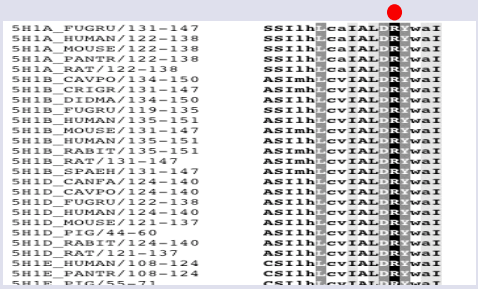
  - Profiles (position-specific scoring matrix)

G_PROTEIN_RECEP_F1_2, PS50262; G-protein coupled receptors family 1 profile (MATRIX)

14

PROSITE pattern

PROSITE: PS00237 (G-protein coupled receptors family 1 signature)
[GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)- [LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)- [LIVM]

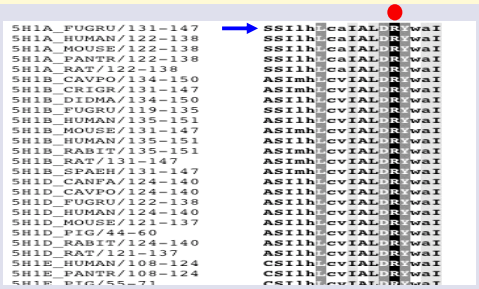


PROSITE pattern syntax: https://prosite.expasy.org/prosuser.html - meth1

15

PROSITE pattern

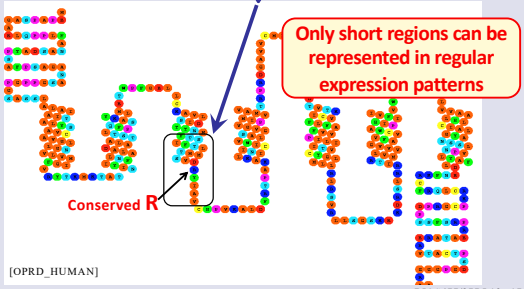
PROSITE: PS00237 (G-protein coupled receptors family 1 signature)
[GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)- [LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)- [LIVM]



16

PROSITE pattern

PROSITE: PS00237 (G-protein coupled receptors family 1 signature)
[GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)- [LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)- [LIVM]



17

Profile

- Position Specific Scoring Matrix (PSSM)
  - → Constructed from multiple alignments
    - Short conserved domains (BLOCKS, PRINTS)
    - Protein families (PROSITE)
    - Results of similarity search (PSI-BLAST)
- More flexible than simple patterns
  - → PSSM lists amino acid frequencies for each alignment position
- Profiles (PSSMs) can be used for database search to identify remote similarities

18

How to build a profile:
(a simple method)

- EMBOSS "Protein Profile" tools — EMBOSS website
  - PROPHECY: creates profiles from multiple alignment [option]
    - Simple amino acid "frequency"

```
# Columns are amino acid counts A->Z
# Rows are alignment positions 1->n
Simple
Name            mymatrix
Length          5
Maximum score   11
Thresh          75
Consensus       RCEGH
1 0 0 0 0 0 0 0 1 0 ... 0 ... 0 2 ...
2 0 0 3 0 0 0 0 0 0 ... 0 ... 0 0 ...
3 0 0 0 0 2 0 0 0 0 ... 0 ... 1 0 ...
4 1 0 0 0 0 0 2 0 0 ... 0 ... 0 0 ...
5 0 0 0 0 0 0 0 2 0 ... 1 ... 0 0 ...
  A B C D E F G H I ... N ... Q R ...
```

```
Seq1 RCQAH
Seq2 HCEGH
Seq3 RCEGN
```

B: Asx = Asn & Asp

19

How to build a profile:
(a simple method)

- EMBOSS "Protein Profile" tools — EMBOSS website
  - PROPHECY: creates profiles from multiple alignment [option]
    - Simple amino acid "frequency"

```
# Columns are amino acid counts A->Z
# Rows are alignment positions 1->n
Simple
Name            mymatrix
Length          5
Maximum score   11
Thresh          75
Consensus       RCEGH
1 0 0 0 0 0 0 0 1 0 ... 0 ... 0 2 ...
2 0 0 3 0 0 0 0 0 0 ... 0 ... 0 0 ...
3 0 0 0 0 2 0 0 0 0 ... 0 ... 1 0 ...
4 1 0 0 0 0 0 2 0 0 ... 0 ... 0 0 ...
5 0 0 0 0 0 0 0 2 0 ... 1 ... 0 0 ...
  A B C D E F G H I ... N ... Q R ...
```

```
Seq1 RCQAH
Seq2 HCEGH
Seq3 RCEGN
```

Lots of 0s!
Doe it cause any problem?

20

Over-fitting problem in profile

```
Seq1 RDA
Seq2 REA
Seq3 REG
```

Position 1 has only Arg (R)
Position 2 has only Asp (D) & Glu (E)
Position 3 has only Ala (A) & Gly (G)

Can we find XEA?

A simple profile (frequency)

Pos	A	R	N	D	C	Q	E	G	H	...	K	T	...
1	0	1.0	0	0	0	0	0	0	0	...	0	0	...
2	0	0	0	0.3	0	0	0.6	0	0	...	0	0	...
3	0.6	0	0	0	0	0	0	0.3	0	...	0	0	...

Lots of 0s!

21

Over-fitting problem in profile

```
Seq1 RDA
Seq2 REA
Seq3 REG
```

Position 1 has only Arg (R)
Position 2 has only Asp (D) & Glu (E)
Position 3 has only Ala (A) & Gly (G)

Can we find XEA?
How about RE?
No flexibility!

A simple profile (frequency)

Pos	A	R	N	D	C	Q	E	G	H	...	K	T	...
1	0	1.0	0	0	0	0	0	0	0	...	0	0	...
2	0	0	0	0.3	0	0	0.6	0	0	...	0	0	...
3	0.6	0	0	0	0	0	0	0.3	0	...	0	0	...

0 should be avoided → Use something else instead

22

Pseudocount methods

- Simplest method
  - Add a small constant (C) to all the counts:

$$W(b) = \{n(b) + C\} / (N + C),$$

Without peudocount: $W(b) = n(b) / N$

```
S1 R
S2 H
S3 R
```

where $W(b)$: Frequency of amino acid b
$n(b)$: Number of amino acid b
N: Number of sequences

  - Pseudocount (C): something small but not 0 (*e.g.*, $C = 0.1$)
- Substitution matrix dependent
  - Proportional to scores, S_{ij}
- Dirichlet mixtures (Sjölander *et al.* 1996)
  - Mixture of different types of pseudocounts
  - Representing various context of protein sequences (*e.g.*, loop region, hydrophobic region)
  - Used in profile HMM

23

Gribskov profile (PSSM)

Gribskov *et al.* (1987)

- Used in PROSITE profile
- Weighted scoring matrix for each position
  - → Generated from:
    - Multiple alignment
    - Scoring matrix (*e.g.*, PAM250)
    - Scoring matrix is weighted with amino acid frequencies at each position

$$M(p,a) = \sum_{b=1}^{20} W(p,b) * Y(a,b)$$

$W(p,b)$: Frequency of amino acid b in position p
$Y(a,b)$: Value in the scoring matrix for AA pair a and b

24

Building a Gribskov profile

- EMBOSS "Protein Profile" tools — EMBOSS website
  - PROPHECY: creates profiles from multiple alignment [option]
    - Gribskov protein profile $M(p,a) = \sum_{b=1}^{20} W(p,b) * Y(a,b)$

```
# Gribskov Protein Profile
# Columns are amino acids A->Z
# Last column is indel penalty
# Rows are alignment positions 1->n
...
Consensus       RCEGH
1 -2.33 0.00 -2.33  1.33  1.33  -3.67 -2.67  8.33 ... 11.67
2  3.00 0.00 15.00 -5.00 -6.00  -1.00  2.00 -1.00 ... -3.00
3  2.67 0.00 -6.00  8.67 12.00  -6.67  4.00  4.67 ...  1.33
4  9.00 0.00  2.33  5.00  4.33  -5.67 12.00 -1.67 ... -3.00
5  0.00 0.00 -1.67  4.67  4.33  -2.33  0.00 11.67 ...  3.67
     A    B    C    D    E     F     G     H   ...   R
```

Seq1 RCQAH
Seq2 HCEGH
Seq3 RCEGN

Pseudocounts are used ➜ No 0 count!

25

PROSITE profile entry

```
ID   G_PROTEIN_RECEP_F1_2; MATRIX.
AC   PS50262;
DT   DEC-2001 (CREATED); DEC-2001 (DATA UPDATE); FEB-2004 (INFO UPDATE).
DE   G-protein coupled receptors family 1 profile.
MA   /GENERAL_SPEC: ALPHABET='ABCDEFGHIKLMNPQRSTVWYZ'; LENGTH=259;
MA   /DISJOINT: DEFINITION=PROTECT; N1=6; N2=254;
MA   /NORMALIZATION: MODE=1; FUNCTION=LINEAR; R1=1.9359; R2=0.02006056; TEXT='-LogE';
MA   /CUT_OFF: LEVEL=0; SCORE=327; N_SCORE=8.5; MODE=1; TEXT='!';
MA   /CUT_OFF: LEVEL=-1; SCORE=227; N_SCORE=6.5; MODE=1; TEXT='?';
MA   /DEFAULT: D=-4; I=-4;
MA   /I: B1=0; B2=-23;
  Pos1 M: SY=-G- M=-1,11,24,-13,-15,-19,30,-19,-20,-18,-15,-11,-5,-19,-16,-18,-2,-11,-15,-22,-20, 16;
  Pos2 M: SY=-N- M=-9,33,-19,15,-2,-18,-2,8,-18,-2,-26,-18,51,-20,-1,-2,9,1,-26,-38,-17,-2;
  Pos3 M: SY=-I- M=-1,-21,-16,-26,-21,0,-16,-22,10,-22,8,4,-17,-22,-19,-20,-8,-3,10,-21,-7,-20;
  Pos4 M: SY=-L- M=-6,-24,-17,-28,-21,8,-25,-20,14,-24,23,12,-21,-25,-20,-19,-17,-5,11,-17,0,-20;
  Pos5 M: SY=-V- M=-
  Pos6 M: SY=-I-
  Pos7 M: SY=-I-
  Pos8 M: SY=-V-
  Pos9 M: SY=-I-
 Pos10 M: SY=-F-
 Pos11 M: SY=-R-
 Pos12 M: SY=-R-
 Pos13 M: SY=-K-
 Pos14 M: SY=-R- M=-8,-7,-25,-7,0,-19,-16,-6,-19,11,-18,-7,-3,-5,2,13,-3,-4,-15,-23,-10,0;
 Pos15 M: SY=-R- M=-8,-7,-21,-9,-3,-16,-16,-4,-14,7,-12,-1,-3,-14,3,11,-4,-4,-11,-23,-8,-1;
 Pos16 I: I=-4; MD=-23;
 Pos17 M: SY=-L- M=-7,-17,-20,-19,-11,-2,-20,-11,1,-8,11,8,-14,-19,-8,-1,-14,-7,0,-18,-3,-10; D=-4;
  ...  I: I=-4; MI=0; MD=-23; IM=0; DM=-23;
```

A B C D E F G H I K L M N P Q R S T V W Y Z

For each position, the likelihood of having each amino acid is listed
+ score: these amino acids are more likely to appear
- score: these amino acids are less likely to appear

Consensus sequence

for each position

26

Profile hidden Markov models

- Probabilistic models of multiple alignments (Krogh *et al.*, 1994; Krogh 1998)
  - ➜ Closely related to standard profiles (PSSMs) introduced by Gribskov
  - ➜ Used in protein domain/family databases (*e.g.*, PFAM, SMART, Superfamily, Panther)

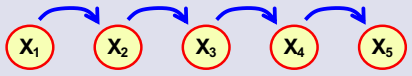
What is a hidden Markov model?
What is hidden?

Eddy (2004) "What is a hidden Markov model?"
[DNA regulatory element example]

27

Markov models: Definitions

- Markov chain: a stochastic process from one state to the next that exhibits the Markov property (*e.g.*, States for DNA: A, T, G, C)



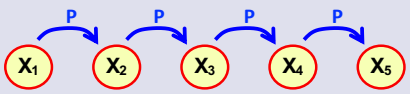
  - Markov property: The next state of the system depends only on the present state of the system [No memory]

28

Markov models: Definitions

- Markov chain: a stochastic process from one state to the next that exhibits the Markov property (*e.g.*, States for DNA: A, T, G, C)

[P: Transition probability]



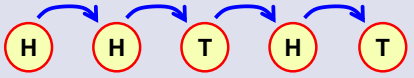
  - Markov property: The next state of the system depends only on the present state of the system [No memory]
  - Transition probabilities are independent of time [Time homogeneous]

• "Markov chain" if the state space is discrete
• "Markov process" if the state space is continuous

29

Markov models: A simple example

- Coin tossing [2 states: Head and Tail]

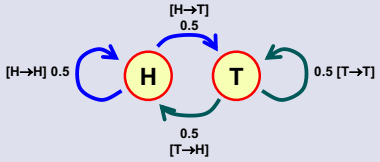


- P(Head) = 0.5, P(Tail) = 0.5
  - ➜ P(HH) = P(H) x P(H) = 0.5^2
  - ➜ P(HHT) = P(HH) x P(T) = P(H) x P(H) x P(T) = 0.5^3
  - ➜ P(HHTH) = P(HHT) x P(H) = P(H) x P(H) x P(T) x P(H) = 0.5^4
  - ➜ P(HHTHT) = P(HHTH) x P(T) = 0.5^5
  - ...

30

Markov models:
A simple example (diagram)

- **Coin tossing (fair coin)**



- States: Head or Tail
- Transition probability ⇒

At time t+1

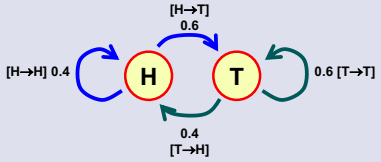
At time t	H	T
H	0.5	0.5
T	0.5	0.5

Row sum = 1

31

Markov models:
A simple example (diagram)

- **Coin tossing (loaded coin: more Tail than Head)**



- States: Head or Tail
- Transition probability ⇒

At time t+1

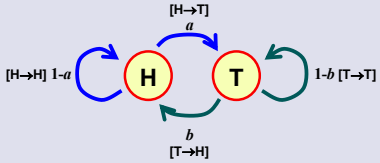
At time t	H	T
H	0.4	0.6
T	0.4	0.6

Row sum = 1

32

Markov models:
A simple example (diagram)

- **Coin tossing (general model)**



- States: Head or Tail
- Transition probability ⇒

At time t+1

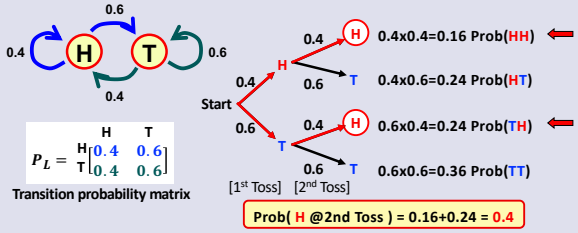
At time t	H	T
H	$1-a$	a
T	b	$1-b$

Row sum = 1

33

Markov models:
State paths

- **Loaded coin tossing [states: H and T]**



P_L =

	H	T
H	0.4	0.6
T	0.4	0.6

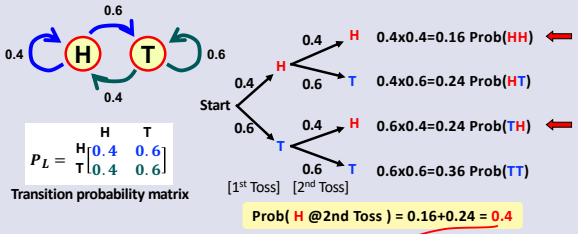
Transition probability matrix

Prob(H @2nd Toss) = 0.16+0.24 = 0.4

34

Markov models:
State paths

- **Loaded coin tossing [states: H and T]**



P_L =

	H	T
H	0.4	0.6
T	0.4	0.6

Transition probability matrix

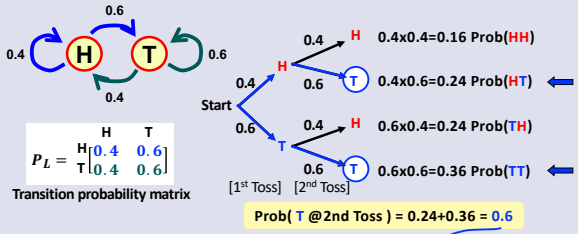
Prob(H @2nd Toss) = 0.16+0.24 = 0.4

H_1	T_1		H_2	T_2
[0.4	0.6]	×	[0.4	0.6]
P_1		P_L		P_2

35

Markov models:
State paths

- **Loaded coin tossing [states: H and T]**



P_L =

	H	T
H	0.4	0.6
T	0.4	0.6

Transition probability matrix

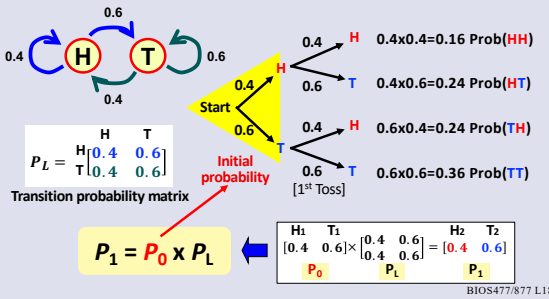
Prob(T @2nd Toss) = 0.24+0.36 = 0.6

H_1	T_1		H_2	T_2
[0.4	0.6]	×	[0.4	0.6]
P_1		P_L		P_2

36

Markov models:
State paths

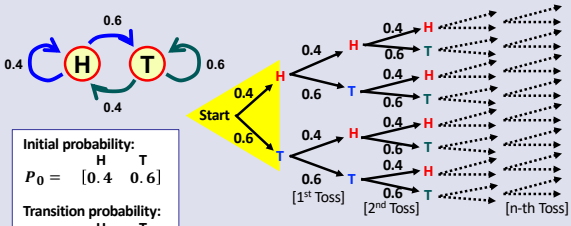
- Loaded coin tossing [states: H and T]



37

Markov models:
State paths

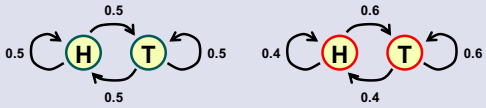
- Loaded coin tossing [states: H and T]



38

What is hidden Markov model?

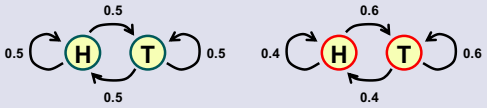
- Fair vs. loaded coin [states: H and T]



39

What is hidden Markov model?

- Fair vs. loaded coin [states: H and T]

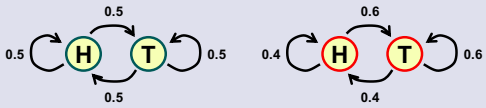


Try tossing the coin several times.
Observe what happens!

40

Hidden Markov model
(observed states)

- Fair vs. loaded coin [states: H and T]



Observation: H, H, H

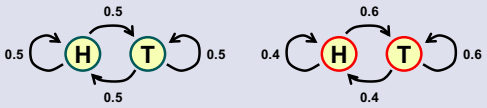
P(HHH| Fair coin) = ? P(HHH| Loaded coin) = ?

(P_{F0}=P_{L0}, so we can ignore the initial probability)

41

Hidden Markov model
(observed states)

- Fair vs. loaded coin [states: H and T]



Observation: H, H, H
P(HHH| Fair coin) > P(HHH| Loaded coin)
Based on the observed sequence: HHH ➜ Fair coin is more likely

44

Hidden Markov model
(observed vs. hidden states)

- If fair and loaded coins are mixed!

$P_{F0} = \begin{matrix} & H & T \\ & [0.5 & 0.5] \end{matrix}$

$P_F = \begin{matrix} & H & T \\ H & [0.5 & 0.5 \\ T & 0.5] & 0.5] \end{matrix}$

We have to consider these probabilities, too!

$P_{L0} = \begin{matrix} & H & T \\ & [0.5 & 0.5] \end{matrix}$

$P_L = \begin{matrix} & H & T \\ H & [0.4 & 0.6 \\ T & 0.4] & 0.6] \end{matrix}$

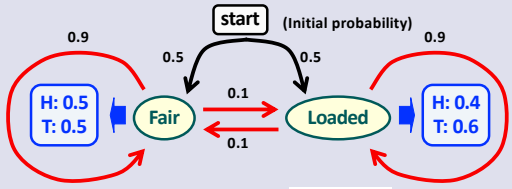
If the coin can be changed?

For observed sequence: H, H, H;
$P(H_F H_F H_F)$ or $P(H_L H_L H_L)$ or $P(H_F H_F H_L)$ or $P(H_F H_L H_L)$ or … ?

45

Hidden Markov model
(observed vs. hidden states)

- Mixed coins: fair and loaded



- States: Fair and Loaded
- Transition probability matrix → $\begin{matrix} & F & L \\ F & [0.9 & 0.1 \\ L & 0.1] & 0.9] \end{matrix}$
- Each state emits symbols H and T with certain emission probabilities

46

Hidden Markov model
(hidden state sequence)

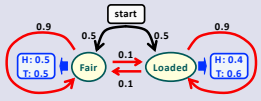
- State sequence is unknown (hidden): Fair or Loaded

Observation: H T H T T T

Hidden state: F or L | F or L | F or L | ? | F or L | F or L | F or L

Can we guess the hidden state sequence?

What we know
- Transition probabilities: $\begin{matrix} & F & L \\ F & [0.9 & 0.1 \\ L & 0.1] & 0.9] \end{matrix}$
- Initial probabilities: (0.5, 0.5)
- Emission probabilities: Fair: $\begin{matrix} H & [0.5] \\ T & 0.5] \end{matrix}$ Loaded: $\begin{matrix} H & [0.4] \\ T & 0.6] \end{matrix}$



47

Hidden Markov model
(hidden state sequence)

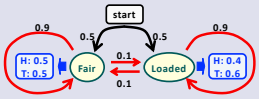
- State sequence is unknown (hidden): Fair or Loaded

Observation: H T H T T T
(emission) 0.5 0.5 0.5 0.5 0.5 0.5

State
Fair: F_H F_T F_H F_T F_T F_T
start
Loaded: L_H L_T L_H L_T L_T L_T

Probability of this path?

Prob=0.0046



[Prob(path)
= [initial prob. X emission prob. X transition prob.]

Prob = 0.5x0.5x0.9x0.5x0.9x0.5x0.9x0.5x0.9x0.5x0.9x0.5
$= (0.5)^7 x (0.9)^5 = 0.0046$

48

Hidden Markov model
(hidden state sequence)

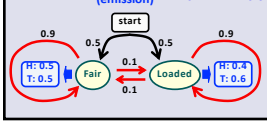
- State sequence is unknown (hidden): Fair or Loaded

Observation: H T H T T T
(emission) 0.5 0.5 0.5 0.5 0.5 0.5

State
Fair: F_H F_T F_H F_T F_T F_T
start
Loaded: L_H L_T L_H L_T L_T L_T
(emission) 0.4 0.6 0.4 0.6 0.6 0.6

Probability of this path?

Prob=0.0046



[Prob(path)
= [initial prob. X emission prob. X transition prob.]

Prob = 0.5x0.4x0.9x0.6x0.9x0.4x0.9x0.6x0.9x0.6x0.9x0.6
$= 0.5x(0.4)^2 x (0.6)^4 x (0.9)^5 = 0.0061$

49

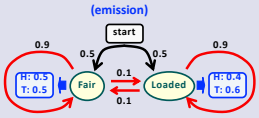
Hidden Markov model
(hidden state sequence)

- State sequence is unknown (hidden): Fair or Loaded

Observation: H T H T T T
(emission) 0.5 0.5 0.5

State
Fair: F_H F_T F_H F_T F_T F_T
start
Loaded: L_H L_T L_H L_T L_T L_T
(emission) 0.4 0.6 0.6

Prob=0.000007



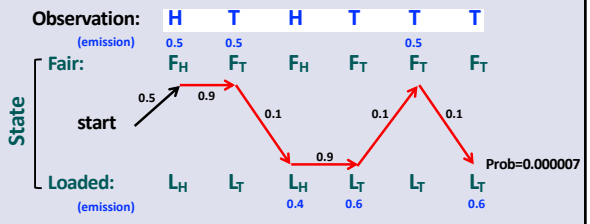
[Prob(path)
= [initial prob. X emission prob. X transition prob.]

Prob = 0.5x0.5x0.9x0.5x0.1x0.4x0.9x0.6x0.1x0.5x0.1x0.6
$= (0.5)^4 x (0.9)^2 x (0.1)^3 x 0.4x(0.6)^2 = 0.000007$

50

Hidden Markov model
(hidden state sequence)

- State sequence is unknown (hidden): Fair or Loaded

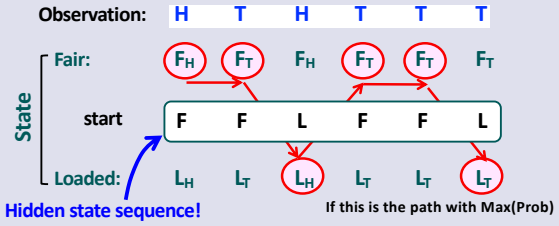


The path with the maximum probability = the most likely path

51

Hidden Markov model
(hidden state sequence)

- State sequence is unknown (hidden): Fair or Loaded



Path with the maximum probability shows the most likely hidden state sequence

52

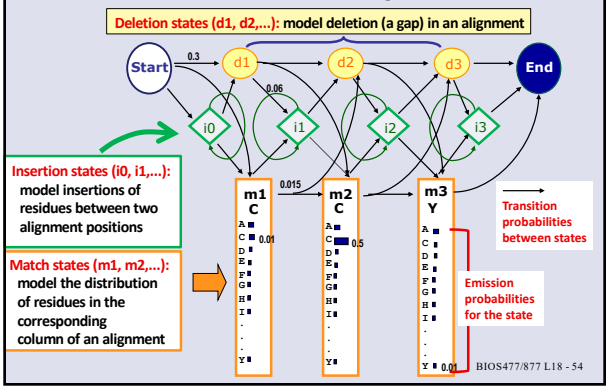
Profile hidden Markov models

- Probabilistic models of multiple alignments
  (Krogh *et al.*, 1994; Krogh 1998)
  - → Closely related to standard profiles (PSSMs) introduced by Gribskov
  - → Used in protein domain/family databases (*e.g.*, PFAM, SMART, Superfamily, Panther)

- States: Insertion, Deletion, and Match
- Transition probabilities: between states
- Emission probabilities:
  for 20 amino acids or 4 nucleotides

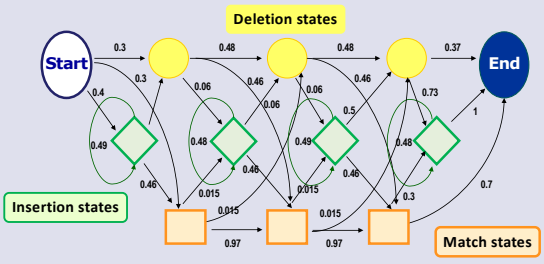
53

Basic architecture of a profile HMM



54

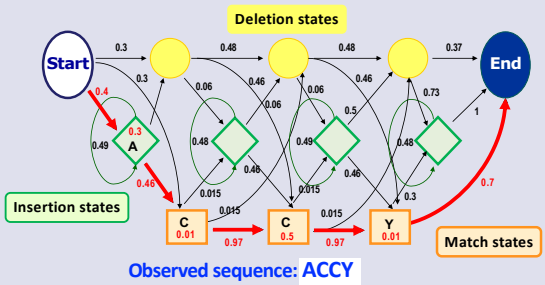
Basic architecture of a profile HMM



- Any sequence can be represented by a path through the model
- Multiple paths are possible to model a sequence

55

Basic architecture of a profile HMM



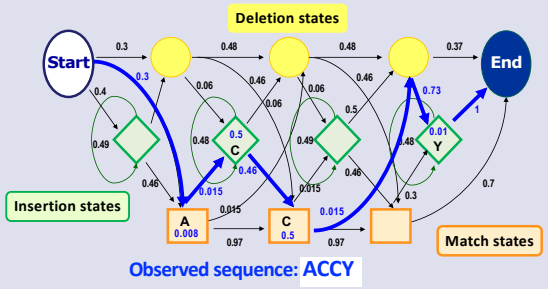
Observed sequence: ACCY

P(sequence | path): the product of emission and transition probabilities

P(→) = 0.4x0.3x0.46x0.01x0.97x0.5x0.97x0.01x0.7

56

Basic architecture of a profile HMM

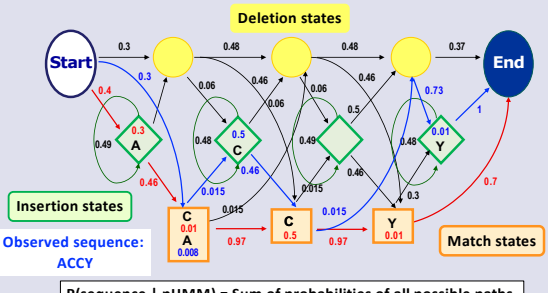


P(sequence | path): the product of emission and transition probabilities

P(→) = 0.3x0.008x0.015x0.5x0.46x0.5x0.015x0.73x0.01x1.0

57

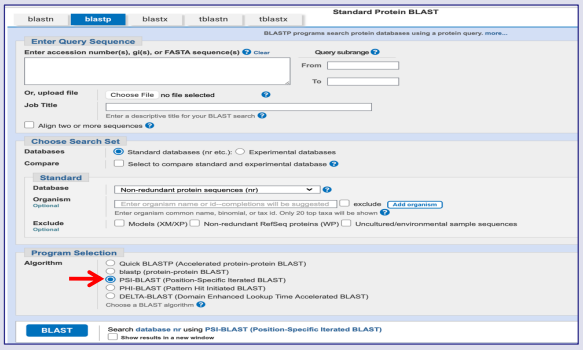
Basic architecture of a profile HMM



P(sequence | pHMM) = Sum of probabilities of all possible paths
= P(sequence | path1) + P(sequence | path2) + ...
→ Probability of the sequence given the model

58

PSI-BLAST: Position-Specific Iterated BLAST



59

PSI-BLAST: Position-Specific Iterated BLAST

https://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-2.html
(PSI-BLAST introduction)

https://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-3.html
(PSI-BLAST statistics)

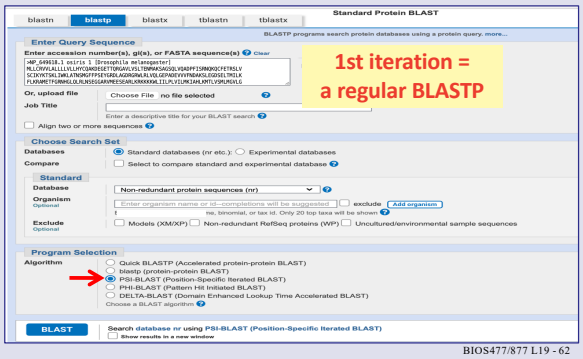
60

PSI-BLAST: Position-Specific Iterated BLAST

- **PSI-BLAST** https://blast.ncbi.nlm.nih.gov/Blast.cgi
  - → 1st iteration: a regular **BLASTP** search
    - Uses a scoring matrix (e.g., BLOSUM62)
  - → 2nd iteration
    - Multiple alignment is constructed from the highly similar hits
    - Positive-specific scoring matrix (PSSM) is constructed
    - Similarity search using the PSSM instead of the single query sequence
  - → 3rd, ..., iterations
    - Stop when no more new hit (or anytime)

61

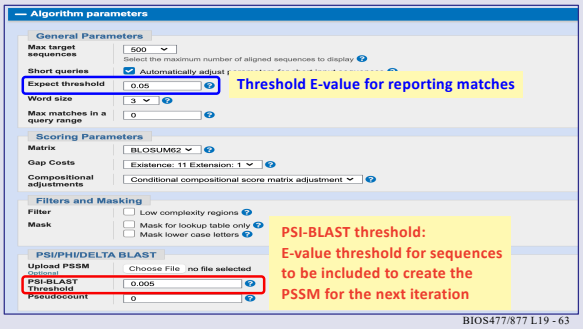
PSI-BLAST



62

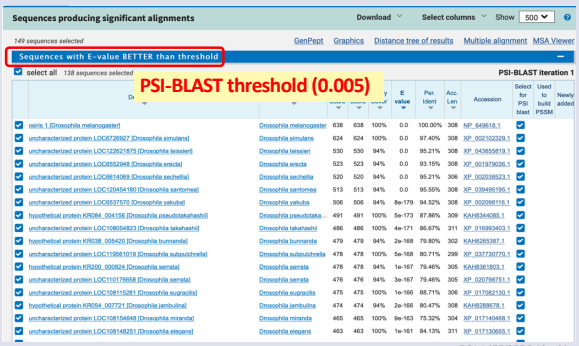
PSI-BLAST

➢ Two different E-value thresholds



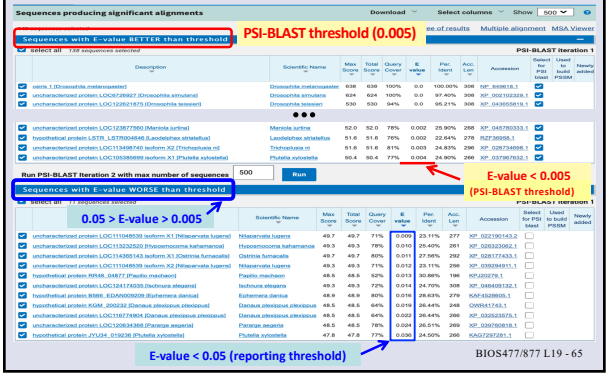
63

PSI-BLAST



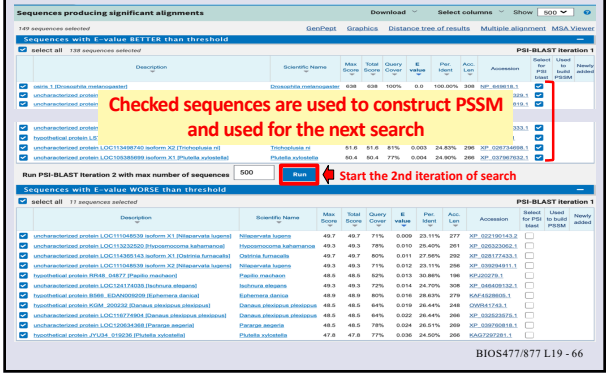
64

PSI-BLAST



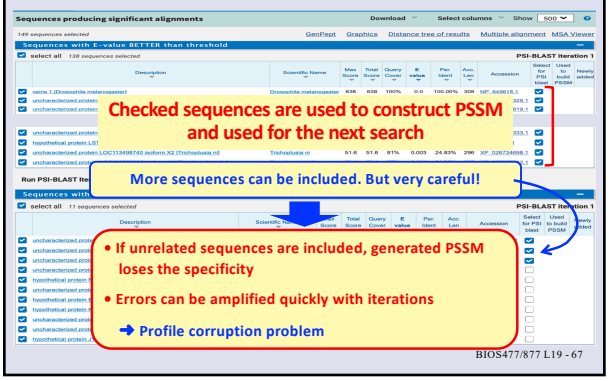
65

PSI-BLAST



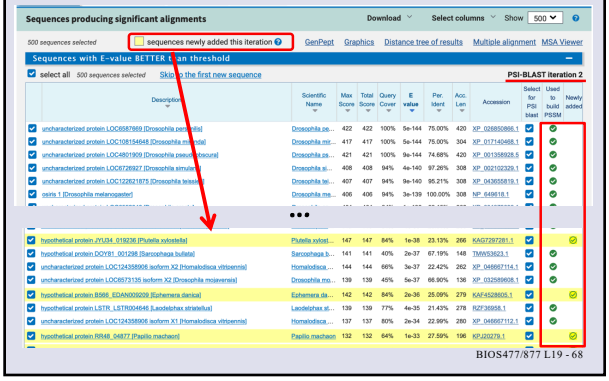
66

PSI-BLAST

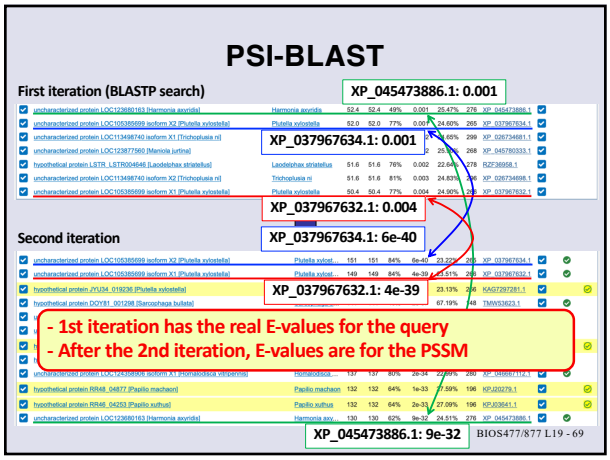


67

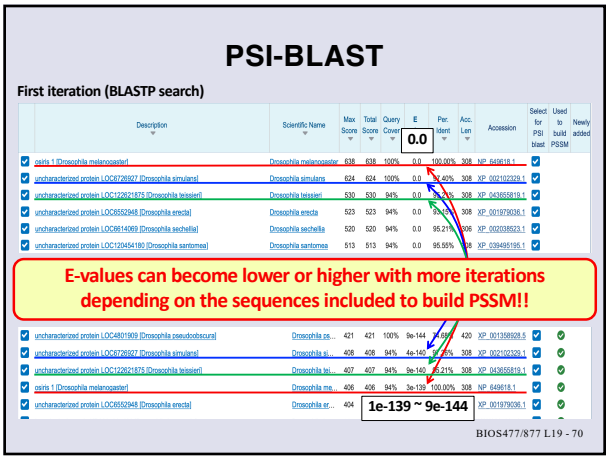
PSI-BLAST



68


PSI-BLAST
First iteration (BLASTP search)
XP_045473886.1: 0.001
XP_037967634.1: 0.001
XP_037967632.1: 0.004
Second iteration
XP_037967634.1: 6e-40
XP_037967632.1: 4e-39
- 1st iteration has the real E-values for the query
- After the 2nd iteration, E-values are for the PSSM
XP_045473886.1: 9e-32
BIOS477/877 L19 - 69

69


PSI-BLAST
First iteration (BLASTP search)
0.0
E-values can become lower or higher with more iterations depending on the sequences included to build PSSM!!
1e-139 ~ 9e-144
BIOS477/877 L19 - 70

70