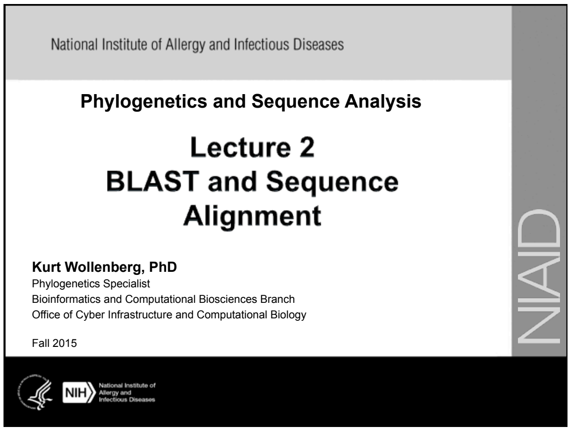


National Institute of Allergy and Infectious Diseases

Phylogenetics and Sequence Analysis

Lecture 2 BLAST and Sequence Alignment

Kurt Wollenberg, PhD
Phylogenetics Specialist
Bioinformatics and Computational Biosciences Branch
Office of Cyber Infrastructure and Computational Biology

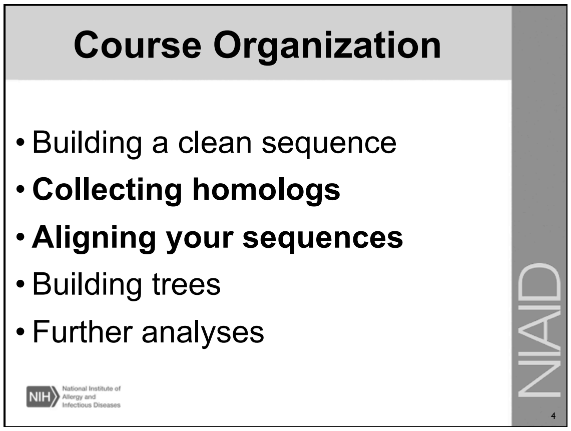
Fall 2015

NIAID

We Are BCBB!

- Group of 37
  - Bioinformatics Software Developers
  - Computational Biologists
  - Project Management & Analysis Professionals

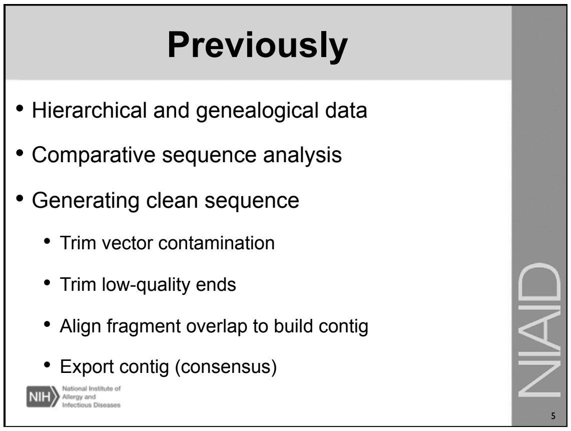
NIAID



Course Organization

- Building a clean sequence
- **Collecting homologs**
- **Aligning your sequences**
- Building trees
- Further analyses

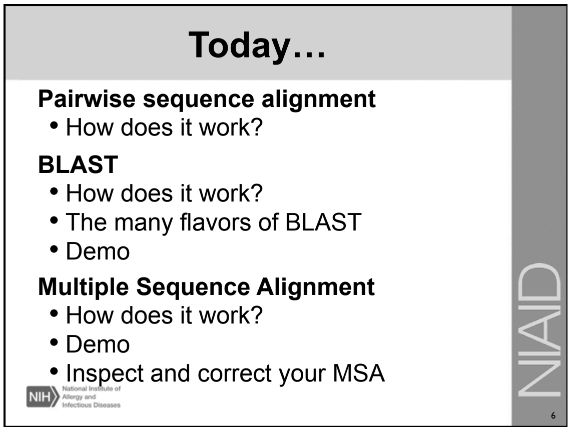
NIAID



Previously

- Hierarchical and genealogical data
- Comparative sequence analysis
- Generating clean sequence
  - Trim vector contamination
  - Trim low-quality ends
  - Align fragment overlap to build contig
  - Export contig (consensus)

5



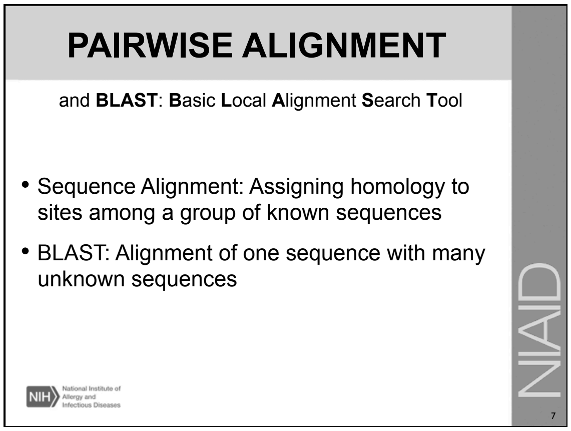
Today…

Pairwise sequence alignment
- How does it work?

BLAST
- How does it work?
- The many flavors of BLAST
- Demo

Multiple Sequence Alignment
- How does it work?
- Demo
- Inspect and correct your MSA

6

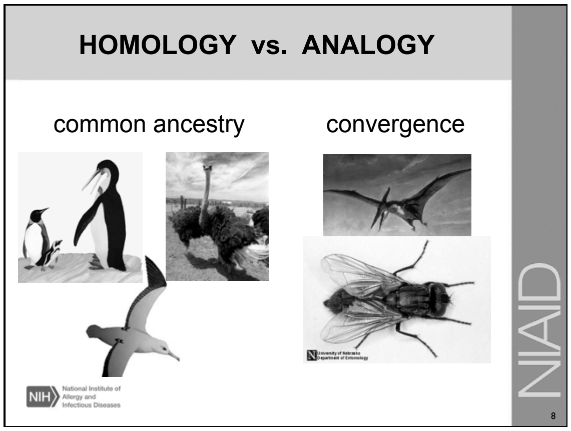


PAIRWISE ALIGNMENT

and **BLAST**: **B**asic **L**ocal **A**lignment **S**earch **T**ool

- Sequence Alignment: Assigning homology to sites among a group of known sequences
- BLAST: Alignment of one sequence with many unknown sequences

7



HOMOLOGY vs. ANALOGY

common ancestry

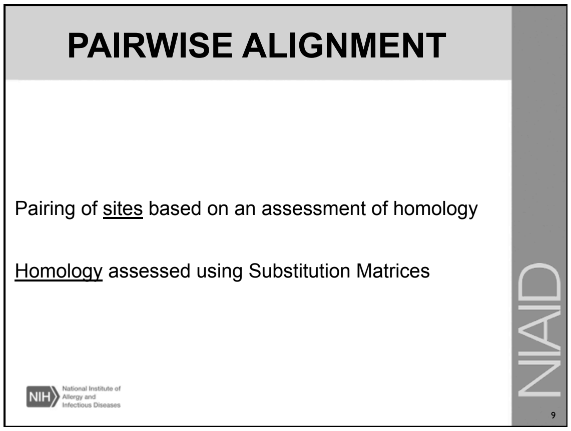
convergence

University of Nebraska Department of Entomology

National Institute of Allergy and Infectious Diseases

NIAID

8



PAIRWISE ALIGNMENT

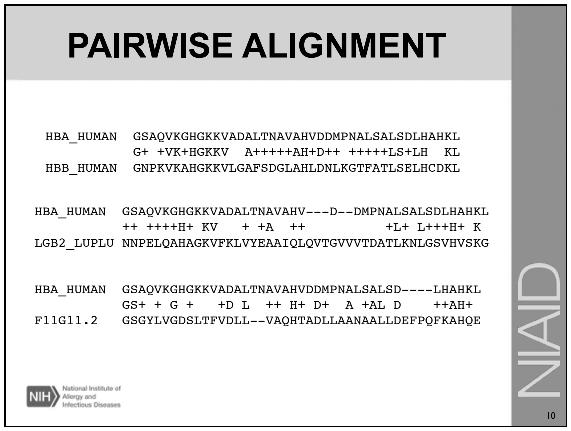
Pairing of sites based on an assessment of homology

Homology assessed using Substitution Matrices

National Institute of Allergy and Infectious Diseases

NIAID

9



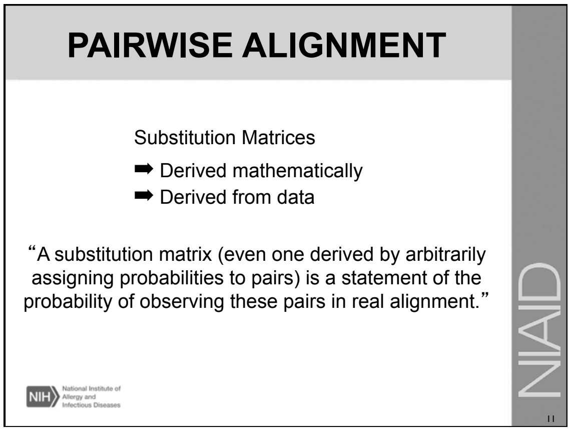
PAIRWISE ALIGNMENT

```
 HBA_HUMAN  GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKL
            G+ +VK+HGKKV  A+++++AH+D++ +++++LS+LH  KL
 HBB_HUMAN  GNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKL

HBA_HUMAN  GSAQVKGHGKKVADALTNAVAHV---D--DMPNALSALSDLHAHKL
           ++ ++++H+ KV   + +A  ++           +L+ L+++H+ K
LGB2_LUPLU NNPELQAHAGKVFKLVYEAAIQLQVTGVVVTDATLKNLGSVHVSKG

HBA_HUMAN  GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSD----LHAHKL
           GS+ + G +   +D L  ++ H+ D+  A +AL D    ++AH+
F11G11.2   GSGYLVGDSLTFVDLL--VAQHTADLLAANAALLDEFPQFKAHQE
```

National Institute of Allergy and Infectious Diseases

NIAID

10



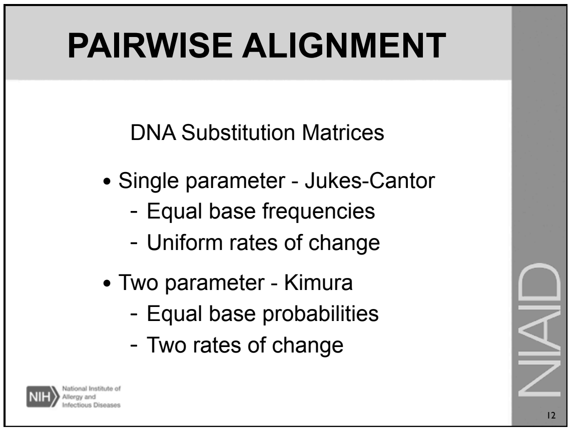
PAIRWISE ALIGNMENT

Substitution Matrices

➡ Derived mathematically
➡ Derived from data

“A substitution matrix (even one derived by arbitrarily assigning probabilities to pairs) is a statement of the probability of observing these pairs in real alignment.”

11

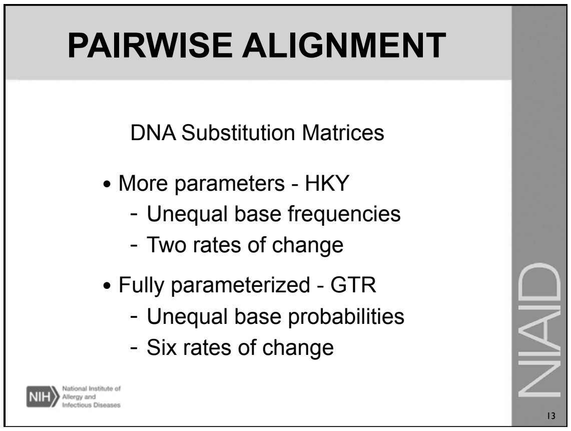


PAIRWISE ALIGNMENT

DNA Substitution Matrices

- Single parameter - Jukes-Cantor
  - Equal base frequencies
  - Uniform rates of change
- Two parameter - Kimura
  - Equal base probabilities
  - Two rates of change

12

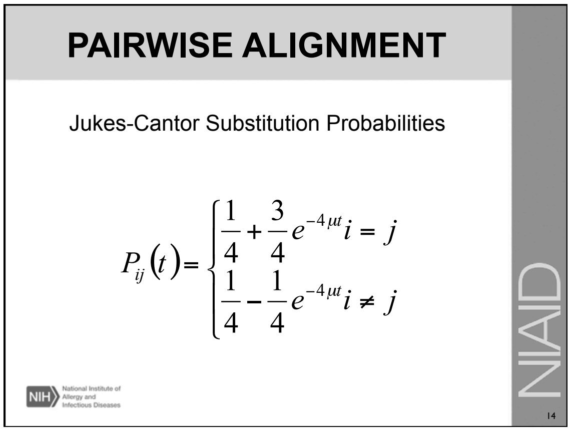


PAIRWISE ALIGNMENT

DNA Substitution Matrices

- More parameters - HKY
  - Unequal base frequencies
  - Two rates of change
- Fully parameterized - GTR
  - Unequal base probabilities
  - Six rates of change

13

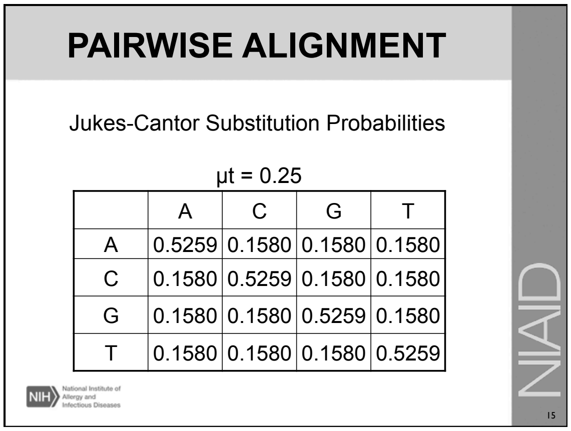


PAIRWISE ALIGNMENT

Jukes-Cantor Substitution Probabilities

$$P_{ij}(t) = \begin{cases} \frac{1}{4} + \frac{3}{4} e^{-4\mu t} & i = j \\ \frac{1}{4} - \frac{1}{4} e^{-4\mu t} & i \neq j \end{cases}$$

14



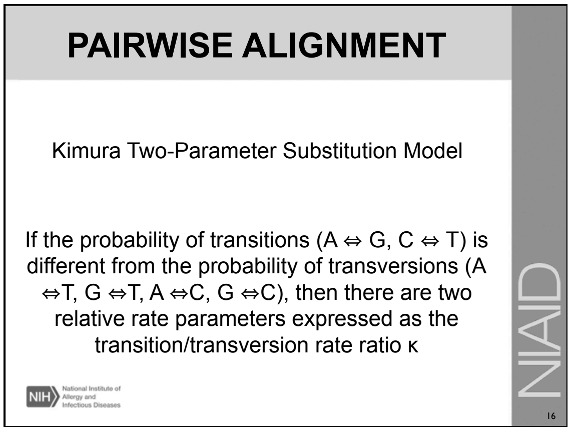
PAIRWISE ALIGNMENT

Jukes-Cantor Substitution Probabilities

μt = 0.25

	A	C	G	T
A	0.5259	0.1580	0.1580	0.1580
C	0.1580	0.5259	0.1580	0.1580
G	0.1580	0.1580	0.5259	0.1580
T	0.1580	0.1580	0.1580	0.5259

15

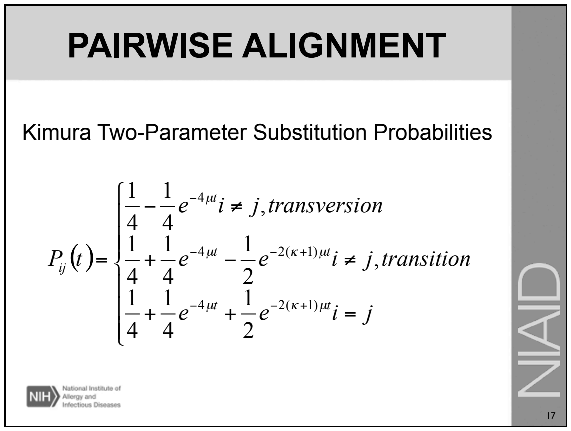


PAIRWISE ALIGNMENT

Kimura Two-Parameter Substitution Model

If the probability of transitions (A ⇔ G, C ⇔ T) is different from the probability of transversions (A ⇔T, G ⇔T, A ⇔C, G ⇔C), then there are two relative rate parameters expressed as the transition/transversion rate ratio κ

16

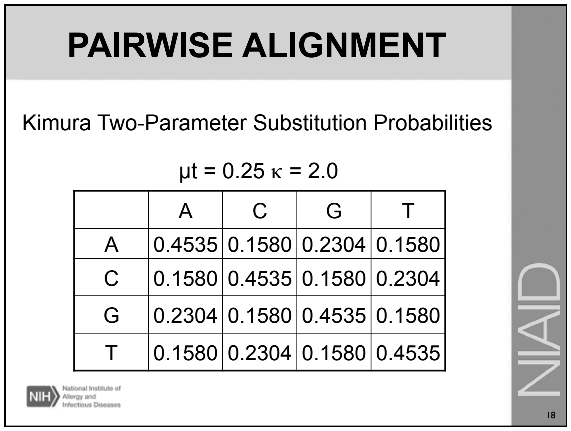


PAIRWISE ALIGNMENT

Kimura Two-Parameter Substitution Probabilities

$$P_{ij}(t) = \begin{cases} \frac{1}{4} - \frac{1}{4}e^{-4\mu t} i \neq j, transversion \\ \frac{1}{4} + \frac{1}{4}e^{-4\mu t} - \frac{1}{2}e^{-2(\kappa+1)\mu t} i \neq j, transition \\ \frac{1}{4} + \frac{1}{4}e^{-4\mu t} + \frac{1}{2}e^{-2(\kappa+1)\mu t} i = j \end{cases}$$

17



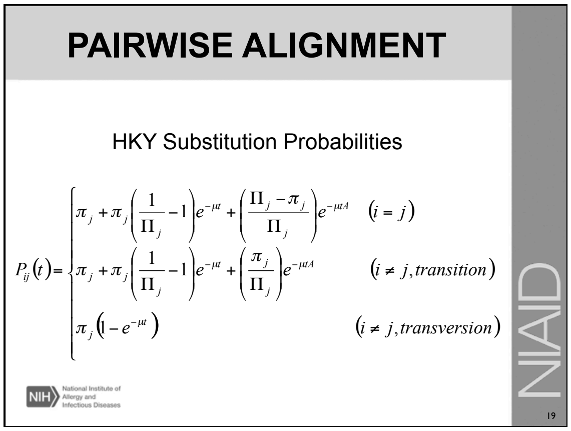
PAIRWISE ALIGNMENT

Kimura Two-Parameter Substitution Probabilities

μt = 0.25 κ = 2.0

	A	C	G	T
A	0.4535	0.1580	0.2304	0.1580
C	0.1580	0.4535	0.1580	0.2304
G	0.2304	0.1580	0.4535	0.1580
T	0.1580	0.2304	0.1580	0.4535

18

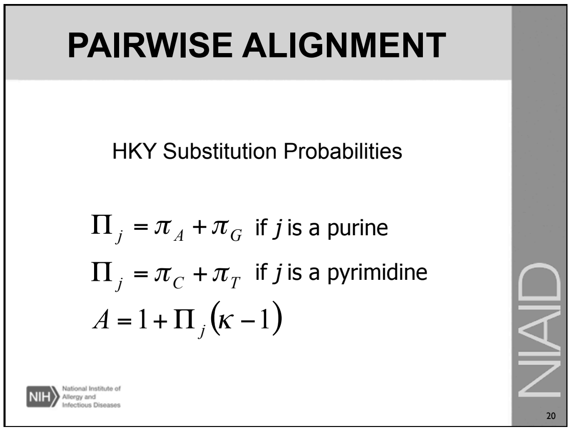


PAIRWISE ALIGNMENT

HKY Substitution Probabilities

$$P_{ij}(t) = \begin{cases} \pi_j + \pi_j\left(\frac{1}{\Pi_j} - 1\right)e^{-\mu t} + \left(\frac{\Pi_j - \pi_j}{\Pi_j}\right)e^{-\mu tA} & (i = j) \\ \pi_j + \pi_j\left(\frac{1}{\Pi_j} - 1\right)e^{-\mu t} + \left(\frac{\pi_j}{\Pi_j}\right)e^{-\mu tA} & (i \neq j, transition) \\ \pi_j\left(1 - e^{-\mu t}\right) & (i \neq j, transversion) \end{cases}$$

19



PAIRWISE ALIGNMENT

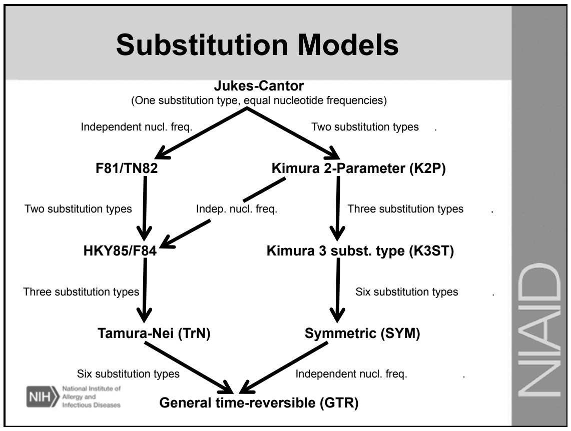
HKY Substitution Probabilities

$$\Pi_j = \pi_A + \pi_G \text{ if } j \text{ is a purine}$$

$$\Pi_j = \pi_C + \pi_T \text{ if } j \text{ is a pyrimidine}$$

$$A = 1 + \Pi_j(\kappa - 1)$$

20



Substitution Models

Jukes-Cantor

(One substitution type, equal nucleotide frequencies)

Independent nucl. freq. → **F81/TN82**

Two substitution types → **Kimura 2-Parameter (K2P)**

F81/TN82 — Two substitution types → **HKY85/F84**

Kimura 2-Parameter (K2P) — Indep. nucl. freq. → **HKY85/F84**

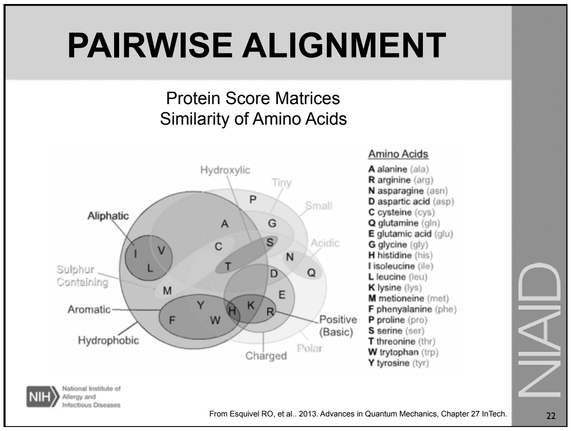
Kimura 2-Parameter (K2P) — Three substitution types → **Kimura 3 subst. type (K3ST)**

HKY85/F84 — Three substitution types → **Tamura-Nei (TrN)**

Kimura 3 subst. type (K3ST) — Six substitution types → **Symmetric (SYM)**

Tamura-Nei (TrN) — Six substitution types → **General time-reversible (GTR)**

Symmetric (SYM) — Independent nucl. freq. → **General time-reversible (GTR)**



PAIRWISE ALIGNMENT

Protein Score Matrices

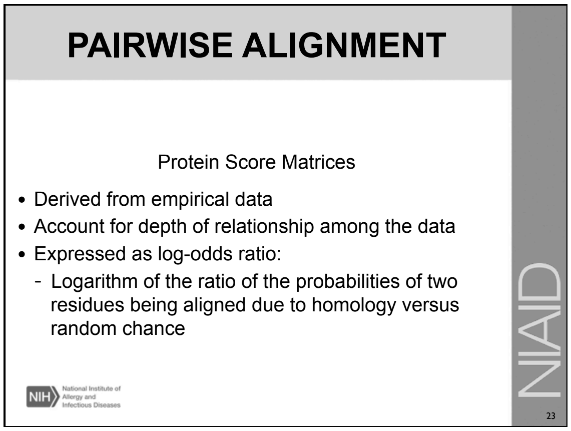
Similarity of Amino Acids

Amino Acids

A alanine (ala)
R arginine (arg)
N asparagine (asn)
D aspartic acid (asp)
C cysteine (cys)
Q glutamine (gln)
E glutamic acid (glu)
G glycine (gly)
H histidine (his)
I isoleucine (ile)
L leucine (leu)
K lysine (lys)
M metioneine (met)
F phenyalanine (phe)
P proline (pro)
S serine (ser)
T threonine (thr)
W trytophan (trp)
Y tyrosine (tyr)

From Esquivel RO, et al.. 2013. Advances in Quantum Mechanics, Chapter 27 InTech.

22



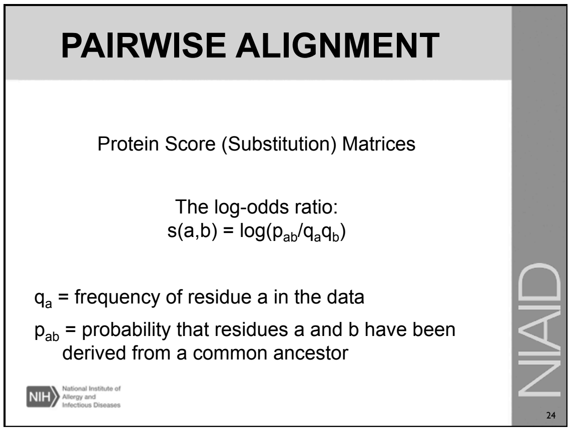
PAIRWISE ALIGNMENT

Protein Score Matrices

- Derived from empirical data
- Account for depth of relationship among the data
- Expressed as log-odds ratio:
  - Logarithm of the ratio of the probabilities of two residues being aligned due to homology versus random chance

NIAID

23



PAIRWISE ALIGNMENT

Protein Score (Substitution) Matrices

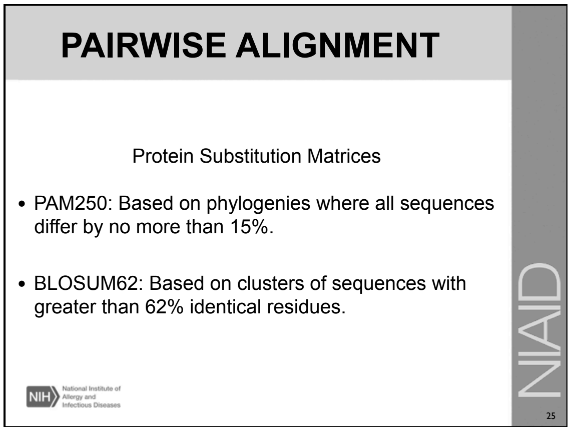
The log-odds ratio:
$s(a,b) = \log(p_{ab}/q_aq_b)$

q_a = frequency of residue a in the data

p_{ab} = probability that residues a and b have been derived from a common ancestor

NIAID

24



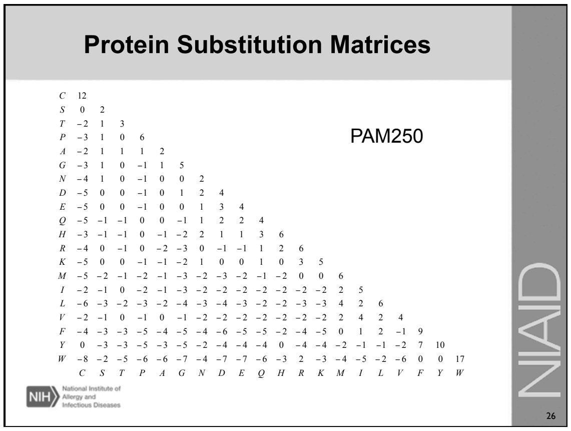
PAIRWISE ALIGNMENT

Protein Substitution Matrices

- PAM250: Based on phylogenies where all sequences differ by no more than 15%.
- BLOSUM62: Based on clusters of sequences with greater than 62% identical residues.

NIAID

25



Protein Substitution Matrices

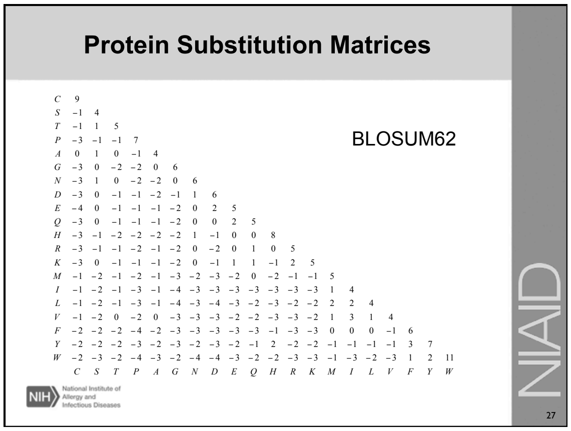
PAM250

C	12																			
S	0	2																		
T	−2	1	3																	
P	−3	1	0	6																
A	−2	1	1	1	2															
G	−3	1	0	−1	1	5														
N	−4	1	0	−1	0	0	2													
D	−5	0	0	−1	0	1	2	4												
E	−5	0	0	−1	0	0	1	3	4											
Q	−5	−1	−1	0	0	−1	1	2	2	4										
H	−3	−1	−1	0	−1	−2	2	1	1	3	6									
R	−4	0	−1	0	−2	−3	0	−1	−1	1	2	6								
K	−5	0	0	−1	−1	−2	1	0	0	1	0	3	5							
M	−5	−2	−1	−2	−1	−3	−2	−3	−2	−1	−2	0	0	6						
I	−2	−1	0	−2	−1	−3	−2	−2	−2	−2	−2	−2	−2	2	5					
L	−6	−3	−2	−3	−2	−4	−3	−4	−3	−2	−2	−3	−3	4	2	6				
V	−2	−1	0	−1	0	−1	−2	−2	−2	−2	−2	−2	−2	2	4	2	4			
F	−4	−3	−3	−5	−4	−5	−4	−6	−5	−5	−2	−4	−5	0	1	2	−1	9		
Y	0	−3	−3	−5	−3	−5	−2	−4	−4	−4	0	−4	−4	−2	−1	−1	−2	7	10	
W	−8	−2	−5	−6	−6	−7	−4	−7	−7	−6	−3	2	−3	−4	−5	−2	−6	0	0	17
	C	*S*	*T*	*P*	*A*	*G*	*N*	*D*	*E*	*Q*	*H*	*R*	*K*	*M*	*I*	*L*	*V*	*F*	*Y*	*W*

NIH National Institute of Allergy and Infectious Diseases

NIAID

26



Protein Substitution Matrices

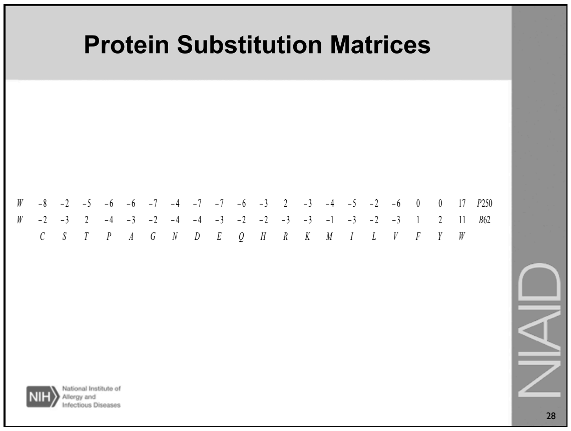
BLOSUM62

C	9																			
S	−1	4																		
T	−1	1	5																	
P	−3	−1	−1	7																
A	0	1	0	−1	4															
G	−3	0	−2	−2	0	6														
N	−3	1	0	−2	−2	0	6													
D	−3	0	−1	−1	−2	−1	1	6												
E	−4	0	−1	−1	−1	−2	0	2	5											
Q	−3	0	−1	−1	−1	−2	0	0	2	5										
H	−3	−1	−2	−2	−2	−2	1	−1	0	0	8									
R	−3	−1	−1	−2	−1	−2	0	−2	0	1	0	5								
K	−3	0	−1	−1	−1	−2	0	−1	1	1	−1	2	5							
M	−1	−2	−1	−2	−1	−3	−2	−3	−2	0	−2	−1	−1	5						
I	−1	−2	−1	−3	−1	−4	−3	−3	−3	−3	−3	−3	−3	1	4					
L	−1	−2	−1	−3	−1	−4	−3	−4	−3	−2	−3	−2	−2	2	2	4				
V	−1	−2	0	−2	0	−3	−3	−3	−2	−2	−3	−3	−2	1	3	1	4			
F	−2	−2	−2	−4	−2	−3	−3	−3	−3	−3	−1	−3	−3	0	0	0	−1	6		
Y	−2	−2	−2	−3	−2	−3	−2	−3	−2	−1	2	−2	−2	−1	−1	−1	−1	3	7	
W	−2	−3	−2	−4	−3	−2	−4	−4	−3	−2	−2	−3	−3	−1	−3	−2	−3	1	2	11
	C	*S*	*T*	*P*	*A*	*G*	*N*	*D*	*E*	*Q*	*H*	*R*	*K*	*M*	*I*	*L*	*V*	*F*	*Y*	*W*

NIH National Institute of Allergy and Infectious Diseases

NIAID

27



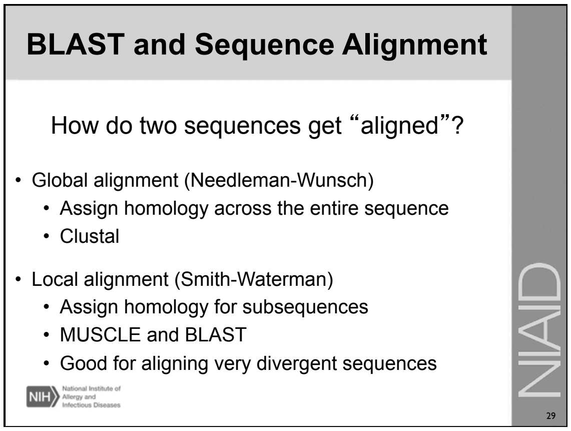
Protein Substitution Matrices

W	−8	−2	−5	−6	−6	−7	−4	−7	−7	−6	−3	2	−3	−4	−5	−2	−6	0	0	17	*P250*
W	−2	−3	2	−4	−3	−2	−4	−4	−3	−2	−2	−3	−3	−1	−3	−2	−3	1	2	11	*B62*
	C	*S*	*T*	*P*	*A*	*G*	*N*	*D*	*E*	*Q*	*H*	*R*	*K*	*M*	*I*	*L*	*V*	*F*	*Y*	*W*	

NIH National Institute of Allergy and Infectious Diseases

NIAID

28

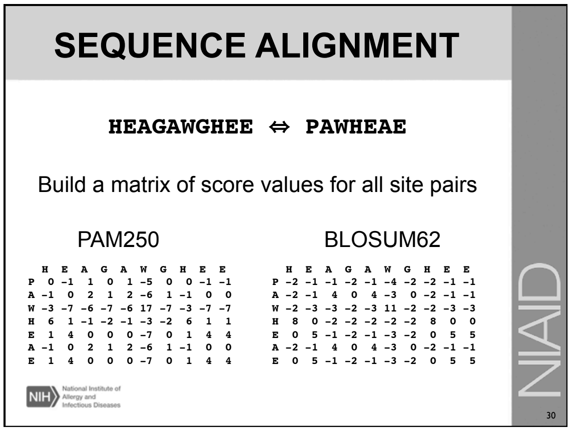


BLAST and Sequence Alignment

How do two sequences get “aligned”?

- Global alignment (Needleman-Wunsch)
  - Assign homology across the entire sequence
  - Clustal
- Local alignment (Smith-Waterman)
  - Assign homology for subsequences
  - MUSCLE and BLAST
  - Good for aligning very divergent sequences

NIAID



SEQUENCE ALIGNMENT

HEAGAWGHEE ⇔ PAWHEAE

Build a matrix of score values for all site pairs

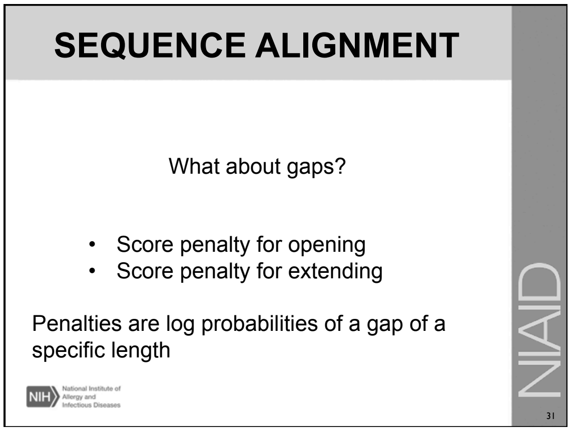
PAM250

	H	E	A	G	A	W	G	H	E	E
P	0	−1	1	0	1	−5	0	0	−1	−1
A	−1	0	2	1	2	−6	1	−1	0	0
W	−3	−7	−6	−7	−6	17	−7	−3	−7	−7
H	6	1	−1	−2	−1	−3	−2	6	1	1
E	1	4	0	0	0	−7	0	1	4	4
A	−1	0	2	1	2	−6	1	−1	0	0
E	1	4	0	0	0	−7	0	1	4	4

BLOSUM62

	H	E	A	G	A	W	G	H	E	E
P	−2	−1	−1	−2	−1	−4	−2	−2	−1	−1
A	−2	−1	4	0	4	−3	0	−2	−1	−1
W	−2	−3	−3	−2	−3	11	−2	−2	−3	−3
H	8	0	−2	−2	−2	−2	−2	8	0	0
E	0	5	−1	−2	−1	−3	−2	0	5	5
A	−2	−1	4	0	4	−3	0	−2	−1	−1
E	0	5	−1	−2	−1	−3	−2	0	5	5

NIAID



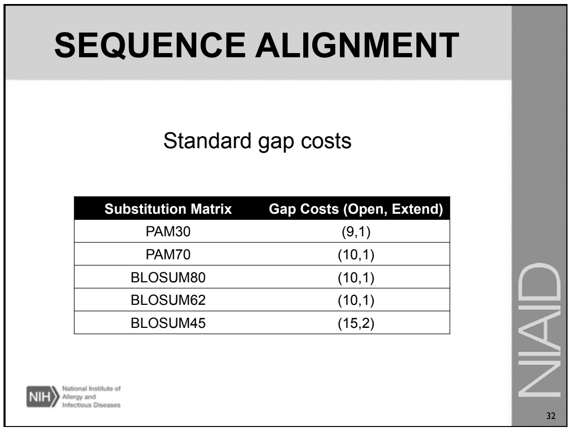
SEQUENCE ALIGNMENT

What about gaps?

- Score penalty for opening
- Score penalty for extending

Penalties are log probabilities of a gap of a specific length

NIAID



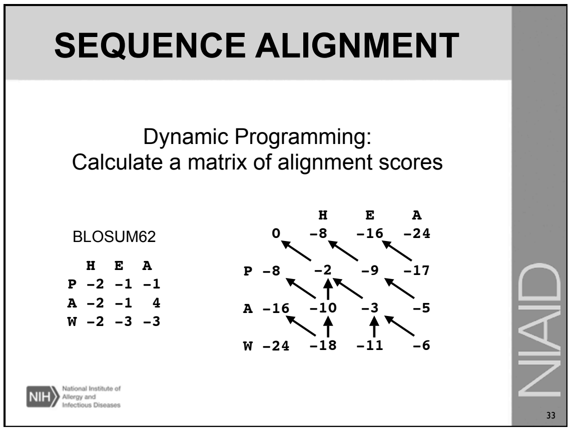
SEQUENCE ALIGNMENT

Standard gap costs

Substitution Matrix	Gap Costs (Open, Extend)
PAM30	(9,1)
PAM70	(10,1)
BLOSUM80	(10,1)
BLOSUM62	(10,1)
BLOSUM45	(15,2)

NIAID

32



SEQUENCE ALIGNMENT

Dynamic Programming:
Calculate a matrix of alignment scores

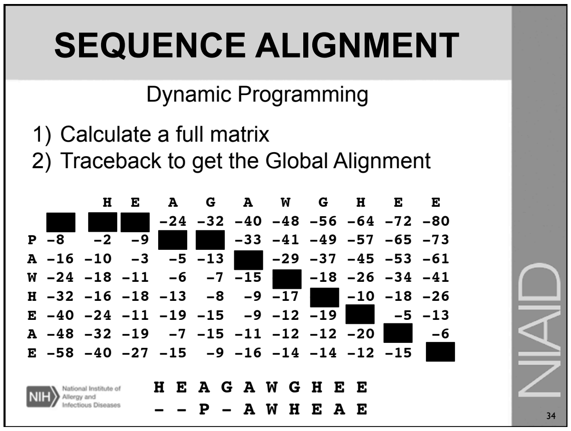
BLOSUM62

	H	E	A
P	-2	-1	-1
A	-2	-1	4
W	-2	-3	-3

		H	E	A
	0	-8	-16	-24
P	-8	-2	-9	-17
A	-16	-10	-3	-5
W	-24	-18	-11	-6

NIAID

33



SEQUENCE ALIGNMENT

Dynamic Programming

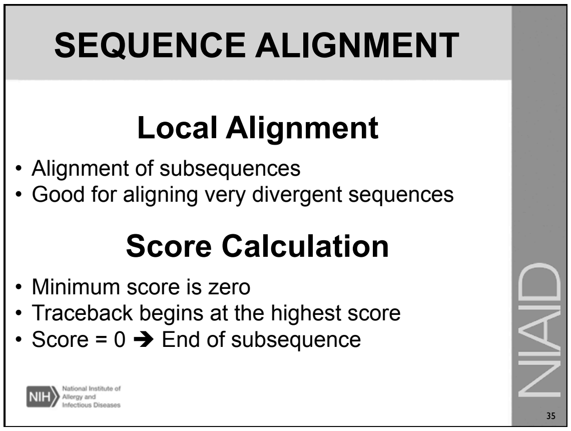
1) Calculate a full matrix
2) Traceback to get the Global Alignment

		H	E	A	G	A	W	G	H	E	E
	[redacted]	[redacted]	[redacted]	-24	-32	-40	-48	-56	-64	-72	-80
P	-8	-2	-9	[redacted]	[redacted]	-33	-41	-49	-57	-65	-73
A	-16	-10	-3	-5	-13	[redacted]	-29	-37	-45	-53	-61
W	-24	-18	-11	-6	-7	-15	[redacted]	-18	-26	-34	-41
H	-32	-16	-18	-13	-8	-9	-17	[redacted]	-10	-18	-26
E	-40	-24	-11	-19	-15	-9	-12	-19	[redacted]	-5	-13
A	-48	-32	-19	-7	-15	-11	-12	-12	-20	[redacted]	-6
E	-58	-40	-27	-15	-9	-16	-14	-14	-12	-15	[redacted]

```
H E A G A W G H E E
- - P - A W H E A E
```

NIAID

34



SEQUENCE ALIGNMENT

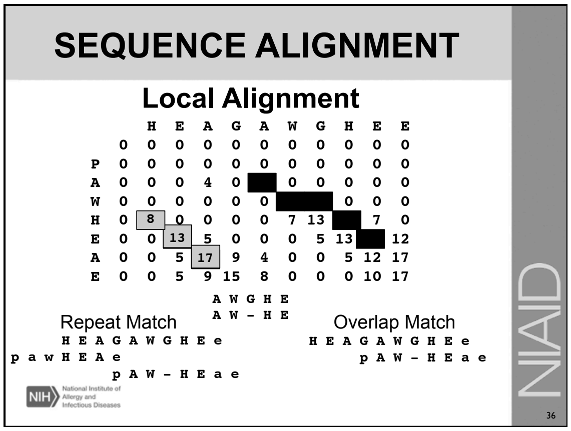
Local Alignment

- Alignment of subsequences
- Good for aligning very divergent sequences

Score Calculation

- Minimum score is zero
- Traceback begins at the highest score
- Score = 0 ➔ End of subsequence

NIH National Institute of Allergy and Infectious Diseases
NIAID
35



SEQUENCE ALIGNMENT

Local Alignment

		H	E	A	G	A	W	G	H	E	E
	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	4	0	[redacted]	0	0	0	0	0
W	0	0	0	0	0	0	[redacted]	[redacted]	0	0	0
H	0	8	0	0	0	0	7	13	[redacted]	7	0
E	0	0	13	5	0	0	0	5	13	[redacted]	12
A	0	0	5	17	9	4	0	0	5	12	17
E	0	0	5	9	15	8	0	0	0	10	17

```
A W G H E
A W - H E
```

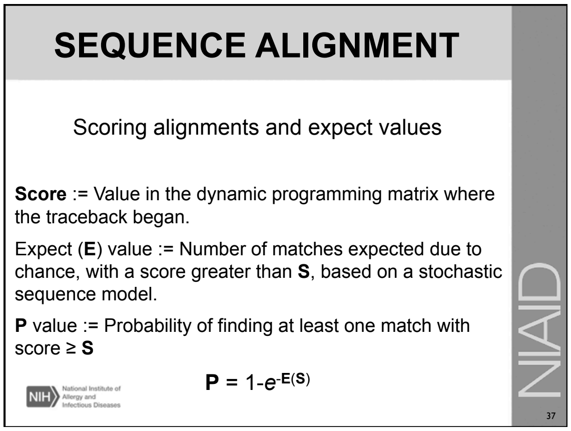
Repeat Match

```
      H E A G A W G H E e
p a w H E A e
          p A W - H E a e
```

Overlap Match

```
H E A G A W G H E e
    p A W - H E a e
```

NIH National Institute of Allergy and Infectious Diseases
NIAID
36



SEQUENCE ALIGNMENT

Scoring alignments and expect values

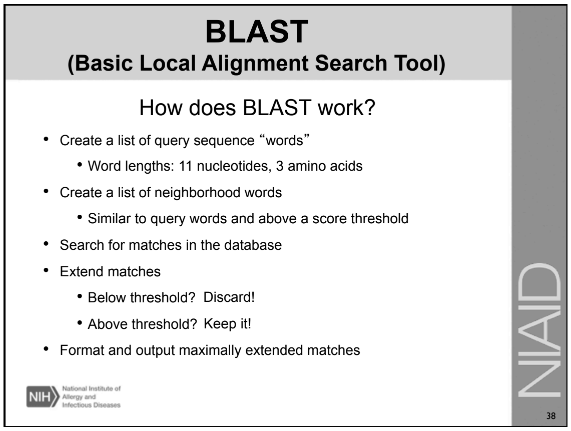
Score := Value in the dynamic programming matrix where the traceback began.

Expect (**E**) value := Number of matches expected due to chance, with a score greater than **S**, based on a stochastic sequence model.

P value := Probability of finding at least one match with score ≥ **S**

$$\mathbf{P} = 1\text{-}e^{-\mathbf{E(S)}}$$

NIH National Institute of Allergy and Infectious Diseases
NIAID
37

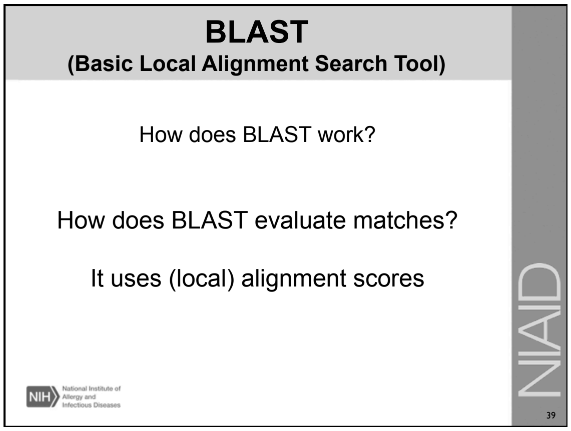


BLAST
(Basic Local Alignment Search Tool)

How does BLAST work?

- Create a list of query sequence “words”
  - Word lengths: 11 nucleotides, 3 amino acids
- Create a list of neighborhood words
  - Similar to query words and above a score threshold
- Search for matches in the database
- Extend matches
  - Below threshold? Discard!
  - Above threshold? Keep it!
- Format and output maximally extended matches

38



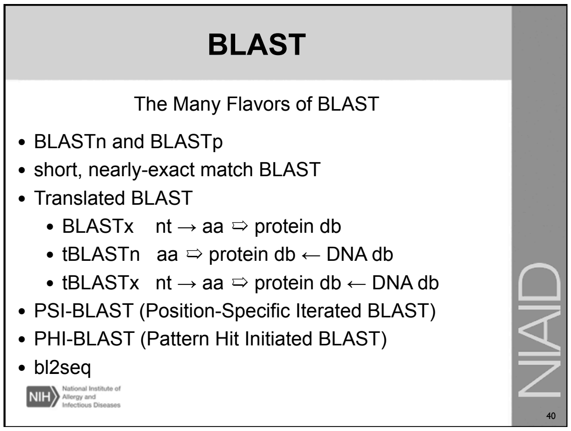
BLAST
(Basic Local Alignment Search Tool)

How does BLAST work?

How does BLAST evaluate matches?

It uses (local) alignment scores

39

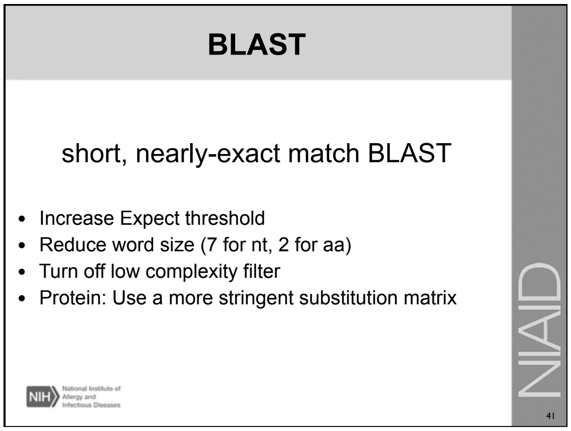


BLAST

The Many Flavors of BLAST

- BLASTn and BLASTp
- short, nearly-exact match BLAST
- Translated BLAST
  - BLASTx nt → aa ⇨ protein db
  - tBLASTn aa ⇨ protein db ← DNA db
  - tBLASTx nt → aa ⇨ protein db ← DNA db
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- bl2seq

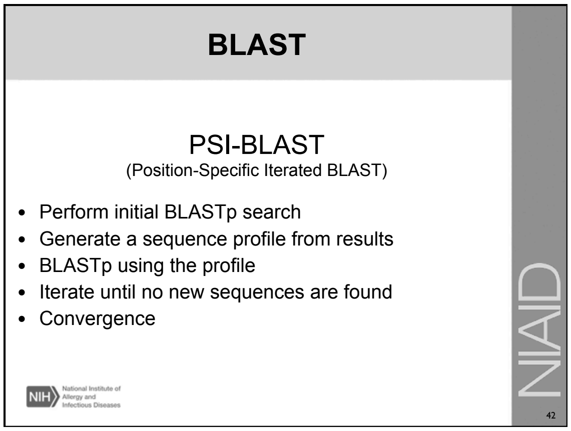
40



BLAST

short, nearly-exact match BLAST

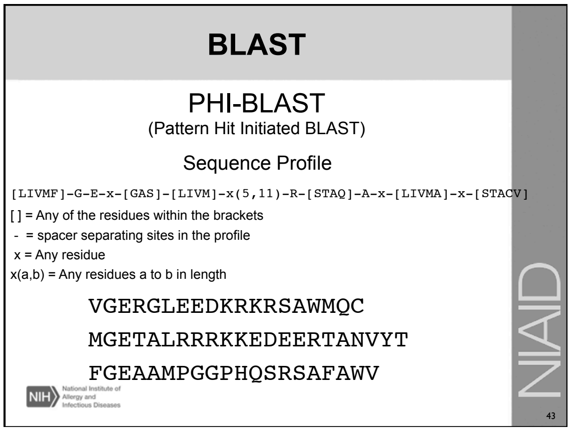
- Increase Expect threshold
- Reduce word size (7 for nt, 2 for aa)
- Turn off low complexity filter
- Protein: Use a more stringent substitution matrix



BLAST

PSI-BLAST
(Position-Specific Iterated BLAST)

- Perform initial BLASTp search
- Generate a sequence profile from results
- BLASTp using the profile
- Iterate until no new sequences are found
- Convergence



BLAST

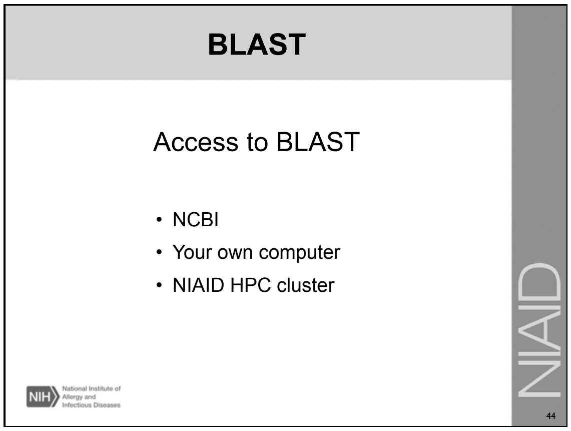
PHI-BLAST
(Pattern Hit Initiated BLAST)

Sequence Profile

```
[LIVMF]-G-E-x-[GAS]-[LIVM]-x(5,11)-R-[STAQ]-A-x-[LIVMA]-x-[STACV]
```

[ ] = Any of the residues within the brackets
 - = spacer separating sites in the profile
 x = Any residue
x(a,b) = Any residues a to b in length

```
VGERGLEEDKRKRSAWMQC
MGETALRRRKKEDEERTANVYT
FGEAAMPGGPHQSRSAFAWV
```

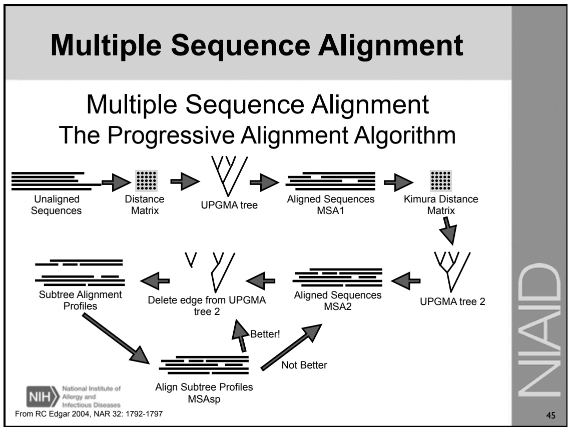


BLAST

Access to BLAST

- NCBI
- Your own computer
- NIAID HPC cluster

44



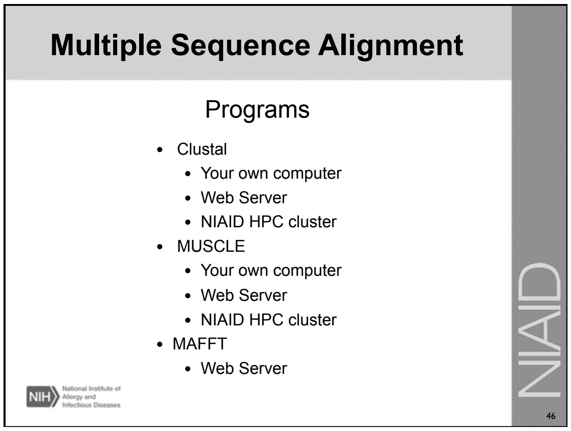
Multiple Sequence Alignment

Multiple Sequence Alignment
The Progressive Alignment Algorithm

Unaligned Sequences
Distance Matrix
UPGMA tree
Aligned Sequences MSA1
Kimura Distance Matrix
UPGMA tree 2
Aligned Sequences MSA2
Delete edge from UPGMA tree 2
Subtree Alignment Profiles
Align Subtree Profiles MSAsp
Better!
Not Better

From RC Edgar 2004, NAR 32: 1792-1797

45

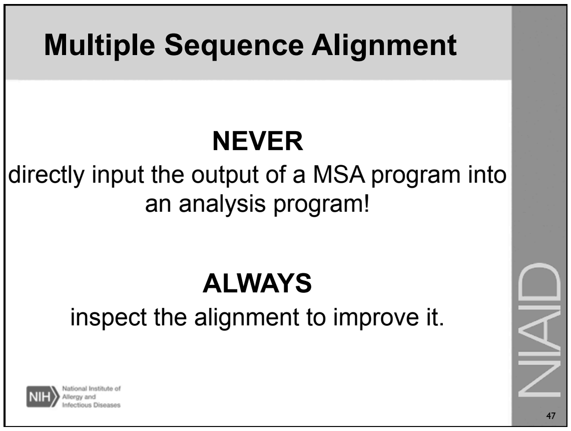


Multiple Sequence Alignment

Programs

- Clustal
  - Your own computer
  - Web Server
  - NIAID HPC cluster
- MUSCLE
  - Your own computer
  - Web Server
  - NIAID HPC cluster
- MAFFT
  - Web Server

46

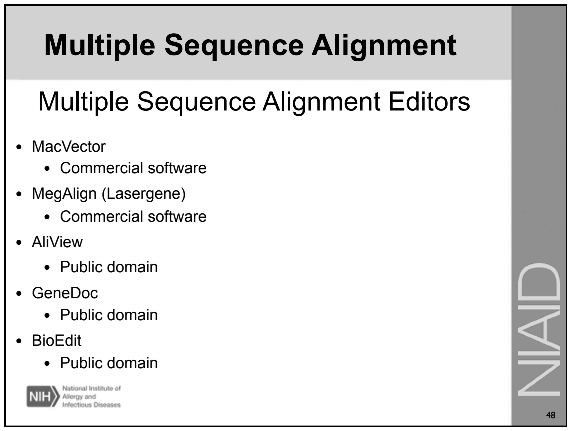


Multiple Sequence Alignment

NEVER
directly input the output of a MSA program into an analysis program!

ALWAYS
inspect the alignment to improve it.

NIAID
NIH National Institute of Allergy and Infectious Diseases
47

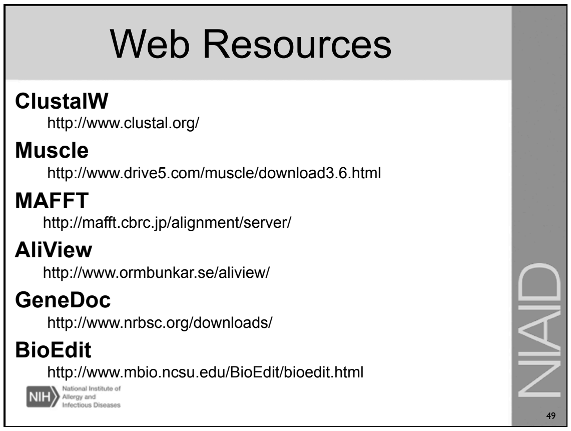


Multiple Sequence Alignment

Multiple Sequence Alignment Editors

- MacVector
  - Commercial software
- MegAlign (Lasergene)
  - Commercial software
- AliView
  - Public domain
- GeneDoc
  - Public domain
- BioEdit
  - Public domain

NIAID
NIH National Institute of Allergy and Infectious Diseases
48



Web Resources

ClustalW
http://www.clustal.org/

Muscle
http://www.drive5.com/muscle/download3.6.html

MAFFT
http://mafft.cbrc.jp/alignment/server/

AliView
http://www.ormbunkar.se/aliview/

GeneDoc
http://www.nrbsc.org/downloads/

BioEdit
http://www.mbio.ncsu.edu/BioEdit/bioedit.html

NIAID
NIH National Institute of Allergy and Infectious Diseases
49

Recapitulation

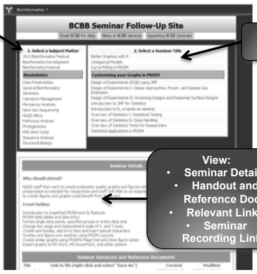
- BLAST search for contig0001 homologs
- Download selected sequence records
- Align sequence records with Clustal2

NIH National Institute of Allergy and Infectious Diseases

NIAID

Seminar Follow-Up Site

- For access to past recordings, handouts, slides visit this site <u>from the NIH network</u>: http://collab.niaid.nih.gov/sites/research/SIG/Bioinformatics/



Bioinformatics ›

BCBB Seminar Follow-Up Site

Email BCBB For Help | Menu of BCBB Services | Upcoming BCBB Seminars

1. Select a Subject Matter
- 2013 Bioinformatics Festival
- Bioinformatics Development
- Bioinformatics Festival
- **Biostatistics**
- Data Presentation
- General Bioinformatics
- Genomics
- Literature Management
- Microarray Analysis
- Next Gen Sequencing
- NIAID BRCs
- Pathways Analysis
- Phylogenetics
- RML Boot Camp
- Sequence Analysis
- Structural Biology

2. Select a Seminar Title
- Better Graphics with R
- Categorical Models
- Curve Fitting in PRISM
- **Customizing your Graphs in PRISM**
- Design of Experiments (DOE) using JMP
- Design of Experiments I: Classic Approaches, Power, and Sample Size Estimation
- Design of Experiments II: Screening Designs and Response Surface Designs
- Introduction to JMP for Statistics
- Introduction to R, a hands-on seminar
- Overview of Statistics I: Statistical Testing
- Overview of Statistics II: Data Handling
- Overview of Statistics Tests For Researchers
- Statistical Applications in PRISM

Seminar Details

Who should attend?

NIAID staff that want to create publication quality graphs and figures [illegible] presentation is intended for researchers and staff with little to no experi[illegible] to create figures and graphs could benefit from this event.

Event Outline:

Introduction to GraphPad PRISM and its features
PRISM data tables and data entry
Format single data points, specified groups or entire data sets
Change the range and measurement scale of X- and Y-axes
Create axis breaks, add error bars and insert special characters
Overlay one figure over another using PRISM Layouts
Create similar graphs using PRISM's MagicTool and clone figure option
Export graphs to MS Word, MS PowerPoint, and other options

Seminar Handouts and Reference Documents

File | Link to file (right click and select "Save As") | Created | Modified

1. Select a Subject Matter

2. Select a Topic

View:
- Seminar Details
- Handout and Reference Docs
- Relevant Links
- Seminar Recording Links

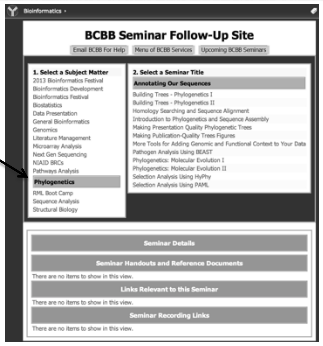
Recommended Browsers:
- IE for Windows,
- Safari for Mac (Firefox on a Mac is incompatible with NIH Authentication technology)

Login
- If prompted to log in use "NIH\" in front of your username

NIH National Institute of Allergy and Infectious Diseases

NIAID

Retrieving Slides/Handouts



Bioinformatics ›

BCBB Seminar Follow-Up Site

Email BCBB For Help | Menu of BCBB Services | Upcoming BCBB Seminars

1. Select a Subject Matter
- 2013 Bioinformatics Festival
- Bioinformatics Development
- Bioinformatics Festival
- Biostatistics
- Data Presentation
- General Bioinformatics
- Genomics
- Literature Management
- Microarray Analysis
- Next Gen Sequencing
- NIAID BRCs
- Pathways Analysis
- **Phylogenetics**
- RML Boot Camp
- Sequence Analysis
- Structural Biology

2. Select a Seminar Title
- **Annotating Our Sequences**
- Building Trees - Phylogenetics I
- Building Trees - Phylogenetics II
- Homology Searching and Sequence Alignment
- Introduction to Phylogenetics and Sequence Assembly
- Making Presentation Quality Phylogenetic Trees
- Making Publication-Quality Trees Figures
- More Tools for Adding Genomic and Functional Context to Your Data
- Pathogen Analysis Using BEAST
- Phylogenetics: Molecular Evolution I
- Phylogenetics: Molecular Evolution II
- Selection Analysis Using HyPhy
- Selection Analysis Using PAML

Seminar Details

Seminar Handouts and Reference Documents

There are no items to show in this view.

Links Relevant to this Seminar

There are no items to show in this view.

Seminar Recording Links

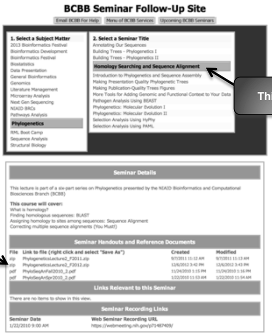
There are no items to show in this view.

This lecture series

NIH National Institute of Allergy and Infectious Diseases

NIAID

Retrieving Slides/Handouts



BCBB Seminar Follow-Up Site

1. Select a Subject Matter

- 2013 Bioinformatics Festival
- Bioinformatics Development
- Bioinformatics Festival
- Biostatistics
- Data Presentation
- General Bioinformatics
- Genomics
- Literature Management
- Microarray Analysis
- Next Gen Sequencing
- NIAID BRCs
- Pathways Analysis
- **Phylogenetics**
- RML Boot Camp
- Sequence Analysis
- Structural Biology

2. Select a Seminar Title

- Annotating Our Sequences
- Building Trees - Phylogenetics I
- Building Trees - Phylogenetics II
- **Homology Searching and Sequence Alignment**
- Introduction to Phylogenetics and Sequence Assembly
- Making Presentation Quality Phylogenetic Trees
- Making Publication-Quality Trees Figures
- More Tools for Adding Genomic and Functional Context to Your Data
- Pathogen Analysis Using BEAST
- Phylogenetics: Molecular Evolution I
- Phylogenetics: Molecular Evolution II
- Selection Analysis Using HyPhy
- Selection Analysis Using PAML

Seminar Details

This lecture is part of a six-part series on Phylogenetics presented by the NIAID Bioinformatics and Computational Biosciences Branch (BCBB)

This course will cover:
What is homology?
Finding homologous sequences: BLAST
Assigning homology to sites among sequences: Sequence Alignment
Correcting multiple sequence alignments (You Must!)

Seminar Handouts and Reference Documents

File	Link to file (right click and select "Save As")	Created	Modified
zip	PhylogeneticsLecture2_F2011.zip	9/7/2011 11:12 AM	9/7/2011 11:13 AM
zip	PhylogeneticsLecture2_F2012.zip	12/6/2012 3:42 PM	12/6/2012 3:43 PM
pdf	PhyloSeqAnFall2010_2.pdf	11/24/2010 1:15 PM	11/24/2010 1:16 PM
pdf	PhyloSeqAnSpr2010_2.pdf	1/22/2010 11:53 AM	1/22/2010 11:54 AM

Links Relevant to this Seminar

There are no items to show in this view.

Seminar Recording Links

Seminar Date	Web Seminar Recording URL
1/22/2010 9:00 AM	https://webmeeting.nih.gov/p71487409/

This lecture

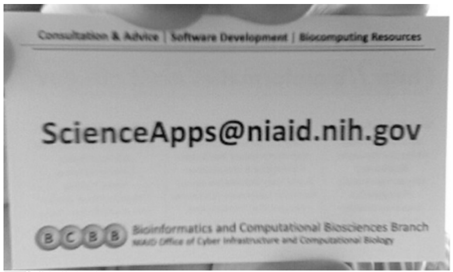
These slides

NIH National Institute of Allergy and Infectious Diseases

NIAID

53

Questions?



Consultation & Advice | Software Development | Biocomputing Resources

ScienceApps@niaid.nih.gov

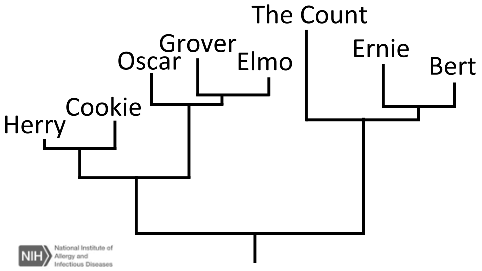
BCBB Bioinformatics and Computational Biosciences Branch
NIAID Office of Cyber Infrastructure and Computational Biology

NIH National Institute of Allergy and Infectious Diseases

NIAID

54

Next Lecture



NIH National Institute of Allergy and Infectious Diseases

NIAID

55