

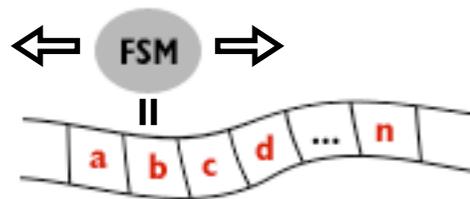
G4120: Introduction to Computational Biology

Oliver Jovanovic, Ph.D.
Columbia University
Department of Microbiology & Immunology

The Digital Language of Computers

Modern computers are digital machines, which means their basic function involves using discrete symbols from a finite set.

In 1936, Alan Turing proved that a finite state machine (FSM) moving up or down a tape of symbols, reading or writing one symbol at a time, could solve any computable problem, and serve as a universal machine.



Universal Turing Machine

The most basic level of information in nearly all current computers represents only one of two possibilities: **0 (off) or 1 (on)**. A signal that can carry one of two possible messages (**0** or **1**) is called a binary signal, or a **bit**, so these computers are binary machines.

Binary Computing and DNA

Binary Units

0 or 1 = 1 bit

8 bits = 1 byte

1,024 bits = 1 kilobit

1,024 bytes = 1 kilobyte (K)

1,024 kilobytes = 1 megabyte (M)

1,024 megabytes = 1 gigabyte (G)

1,024 gigabytes = 1 terabyte (T)

1 bit =	0 or 1	= 2 possibilities
2 bits =	0 0 or or 1 1	= $2 \times 2 = 4$ possibilities
3 bits =	0 0 0 or or or 1 1 1	= $2 \times 2 \times 2 = 8$ possibilities
4 bits =	0 0 0 0 or or or or 1 1 1 1	= $2 \times 2 \times 2 \times 2 = 16$ possibilities
5 bits =	0 0 0 0 0 or or or or or 1 1 1 1 1	= $2 \times 2 \times 2 \times 2 \times 2 = 32$ possibilities
6 bits =	0 0 0 0 0 0 or or or or or or 1 1 1 1 1 1	= $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$ possibilities
7 bits =	0 0 0 0 0 0 0 or or or or or or or 1 1 1 1 1 1 1	= $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 128$ possibilities
8 bits =	0 0 0 0 0 0 0 0 or or or or or or or or 1 1 1 1 1 1 1 1	= $2 \times 2 = 256$ possibilities

DNA has only four possibilities (so can be represented 2 bits)

G = 00

C = 11

A = 01

T = 10

Complementation (with intelligent choice of representation)

G C C A = 00 11 11 01

C G G T = 11 00 00 10

ASCII Representation of DNA

American Standard Code for Information Interchange (ASCII)

- For practical purposes, DNA and RNA is generally represented in ASCII code, using the upper or lower case letters A, C, G, and T or A, C, G and U.
- Each ASCII character occupies one byte, and thus has 256 possibilities, including all upper and lower case letters of the English alphabet, the ten Arabic numerals, punctuation, and special characters, such as @.
- Thus, a kilobase of DNA (1,000 base pairs) occupies just under a kilobyte (1 K = 1,024 bytes) of storage in ASCII. An entire human genome, roughly 3 billion base pairs (3 gigabases), occupies just under 3 gigabytes of storage in ASCII.

Transcription

- Transcription is computationally trivial. One need only substitute a U for a T if dealing with a sense strand, or complement, then transcribe if dealing with the antisense strand.

Translation

- Translation is also computationally trivial. A computer can refer to a species appropriate translation table to translate DNA or RNA into the appropriate protein sequence.

AUA	I	Isoleucine
AUC	I	Isoleucine
AUG	M	Methionine start
AUU	I	Isoleucine
etc.		

Alternate Representation

- Can readily convert an ASCII representation of DNA into other forms, such as graphics, or even music.

Algorithms in Computational Biology

Algorithm

- An algorithm is simply a series of steps used to solve a problem. One of a computer's great strengths is its ability to rapidly and accurately repeat recursive steps in an algorithm.

Consensus

- Early algorithms for searching sequence data depended on consensus sequences. Thus, to find a prokaryotic promoter, one would try to find something that matched a consensus -10 sequence (TATAAT), not too far downstream of a consensus -35 sequence (TTGACA).
- It rapidly became clear that biologically significant sequences rarely perfectly matched a consensus, and more sophisticated approaches were adopted, including the use of matrices, Markov chains and hidden Markov models.

Matrices

- Matrices take into account the distribution of every possible nucleotide (or amino acid) at a position in a set of known sequences. Searching with a matrix is therefore more sensitive than searching with a consensus, and can find biological features that a strict consensus approach would miss.

Markov chains and hidden Markov models (HMMs)

- Markov chains and hidden Markov models are probabilistic models of sequences, and have proven useful in database searching, gene finding and multiple sequence alignment.
- A first-order Markov chain is a finite state automaton (a restricted Turing machine which only moves left to right) with probabilities for each transition to a new state (symbol) based on its current state. Higher order Markov chains take into account one or more previous states.
- A hidden Markov model is a Markov chain in which only the output can be observed (its current state is hidden).

Consensus versus Matrix

E. coli Promoter Consensus

-35 Region -10 Region
 TTGACA.....TATAAT

E. coli Promoter Matrix

	-35 Region															
	T	T	G	A	C	A										
A	11	8	8	7	8	7	3	5	5	0	1	0	14	5	9	5
C	3	4	2	4	4	3	5	2	8	1	1	2	3	11	2	5
G	3	2	4	2	4	5	5	5	5	2	1	17	1	2	3	3
T	4	7	7	8	5	6	8	9	3	17	18	2	4	3	7	9

	Spacer Region						
Length	9	10	11	12	13	14	15
	1	6	14	6	1	1	1

	-10 Region													
	T	A	T	A	A	T								
A	4	5	3	4	4	0	20	5	12	11	0	7	4	6
C	5	4	5	4	5	2	0	3	3	4	1	2	7	6
G	2	5	5	8	7	2	0	3	3	3	0	6	5	6
T	10	6	8	5	6	17	1	9	3	4	20	6	5	4

Macintosh OS X and Sequence Analysis

SeqMatrix *E. coli* promoter output:

DNA Location: 3,075

Spacer Length: 11

Similarity Score: 55.29

CGACATTGCTTGACCC <11> GCGTGTTCAATTTCG
(TTGACA.....TATAAT)

Position	Score	Predicted promoter sequence (-35 < gap> -10)	Name
2313	52.94	GTTAATTGCTTTTCGA <10> TTAGCTAAACTTTC	
3075	55.29	CGACATTGCTTGACCC <11> GCGTGTTCAATTTCG	<i>korE</i>
3772	55.29	GCGTCATGCTTGAAAA <11> TGGCGTGCAATCAG	
5552	51.17	AAAAGGATCTTCACCT <10> AAATTAATAAATGAA	
5585	53.52	AAATGAAGTTTAAAT <15> ATGAGTAAACTTGG	
5695	52.35	CATAGTTGCCTGACTC <12> ATAACACGATACG	
6133	54.11	ACTGCATAATTCTCTT <11> ATCCGTAAGATGCT	
6478	56.47	ACGGAATGTTGAATA <11> CCTTTTCAATATT	
6511	53.52	TCAATATTATTGAAGC <12> TATTGTCTCATGAG	
6532	52.35	TCAGGGTTATTGTCTC <11> CATATTTGAATGTA	
6618	57.05	AAGTGCCACCTGACGT <10> ATTATTATCATGAC	
7292	59.41	ACGATTTAATGGACAC <11> CGTTTTACTATGTC	<i>tnpA</i>
8080	51.17	ATGAAGCAATTGAACG <11> TGAACGATTTTGGC	
8248	50.58	CCTTTGTCCTTGCATG <11> CGTTGGATGATGCA	
8482	58.23	AAAAGCTGGCTGAAAT <11> TCCGTGAAATGGCC	
11698	51.76	ATTATGCCCTAGCCTG <10> TTAGCTAAACTATG	
13937	60.58	TTGTCATGCTTGACAC <12> AAACATAATATGTC	<i>tetA</i>
17967	53.52	ACCGCTATATCGAAAA <10> CTTGTTAGAATTGC	
17999	57.64	AGAATTGCCATGACGT <11> ACGGGTAAGATTAC	<i>trbA</i>
18729	53.52	ATTAGCTGTTTGTCTT <14> TTCGGTATATCGTT	<i>trbB</i>
26625	50.58	ACCAGGCGTTTACTA <9> AGGAGTAACTTATG	
35054	43.52	CTCGCGCTGTAGCCTC <11> TGTGCTAATGTGGT	<i>parD</i>
38324	64.70	ATCGTGCGTTGACAA <11> CTGGCTACACTATG	<i>aphA</i>
40006	54.70	TCGTAGTTCTTGCCGA <11> TTCTCAAAGATGCC	
47326	52.35	ATCAGTTGCTTGATGC <11> TTGCTGACGTTGCG	
48938	54.70	CAAACGGTTTTGGCTT <12> TTTCGTCCAATGCG	
51306	57.64	GAAAAAGGATGGATAT <9> ATCGCTATAATGAC	<i>traK</i>
59051	54.70	TGTTTTTCTTGGCGT <11> TTCCGGACGATGTA	
2375c	66.47	CTAAAGGTGTTGACGA <12> TTAGCTAAACTTCT	<i>klaA</i>
3711c	52.94	ATTCTTGTTTTGAGGC <11> CCAGGTCAATTACC	<i>klec</i>
3745c	67.05	TAAAATTGCTTGACAA <12> TGCCCTATTCTTGT	<i>kleC</i>
3777c	52.94	GACGCCTCGCTGAATC <11> TTAGCTAAAATTGC	
4400c	65.88	TAAATTTCTTACTA <12> TGCCCTAATATAGC	<i>kleA</i>
5562c	53.52	AGATCCTTTTTGATAA <14> TCCCTAACGTGAG	
5619c	51.76	CATATATACTTTAGAT <12> TCATTTTAAATTA	
5647c	54.70	CATTGGTAACTGTCAG <11> TCATATATACTTTA	

Search Algorithms in Bioinformatics

Global Alignment Search

- **Needleman-Wunsch** algorithm, Needleman & Wunsch, 1970
- Finds the best complete alignment of two sequences that maximizes the number of matches and minimizes the number of gaps.

Local Alignment Search

- **Smith-Waterman** algorithm, Smith & Waterman, 1981
- Makes an optimal alignment of the best segment of similarity between two sequences.
- Often better for comparing sequences of different lengths, or when looking at a particular region of interest.

Heuristic Approximations to Smith-Waterman

- **FASTA**, Pearson, 1988
- **BLAST**, Altschul, 1990
- **BLAST 2** (aka Gapped BLAST), Altschul, 1997
- **BLAT**, Kent, 2002

Global Alignment (Needleman-Wunsch)

Gap, from the GCG Wisconsin Package, uses the algorithm of Needleman and Wunsch to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps.

GAP RK2_ssb x Ecoli_ssb January 29, 2003 00:07

```
1  ..MSHNQFQFIGNLTRDTEVRHGNSNKPQAIIFDIAVNEEWRNDA.GDKQE 47
      |.  :||| .| |||: .  |  :| .| ||. | | : .|
1  ASRGVNVKVLVGNLQDPEVRYMPNGGAVANITLATESWRDKATGEMKE 50
48  RTDFFRIKCFGSQAEAHGKYLKGKSLVFVQGKIRNTKY.EKDGQTVYGTD 96
      .|:. | : ||  ||  .|| ||| |::|:| | : :. || | | :
51  QTEWHRVVLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTQSGQDRYTE 100
97  FIAD...KVDYLDTKAPGGSNQE..... 116
      :.  .  |  :  ||.
101 VVVNVGGTMQMLGGRQGGGAPAGGNIIGGGQPQGGWGPQQPQGGNQFSGG 150
      .....
151 AQSRPQQSAPAAPSNEPPMDFDDDIPF 177
```

Matrix: blosum62
Gap Penalties: default
Length: 177
Percent Similarity: 45.690
Percent Identity: 32.759

Local Alignment (Smith-Waterman)

BestFit, from the GCG Wisconsin Package, makes an optimal alignment of the best segment of similarity between two sequences. Optimal alignments are found by inserting gaps to maximize the number of matches using the local homology algorithm of Smith and Waterman.

BESTFIT RK2_ssb x Ecoli_ssb January 29, 2003 00:08

```

      .           .           .           .
4 NQFQFIGNLTRDTEVRHGNSNKPQAIFFDIAVNEEWRNDA.GDKQERTDFF 52
  |   :||| .| |||: .   |   :| .| ||. | | : .|.|.:.
6 NKVILVGNLQDPEVRYMPNGGAVANITLATSESWRDKATGEMKEQTEWH 55

      .           .           .           .
53 RIKCFGSQAEAHGKYLKGKSLVFVQVKIRNTKY.EKDGQTVYGTDFIAD 100
  |:  ||  ||  .|| ||| |::|.:.| | : .  ||  | | : .
56 RVVLFGLAEVASEYLRKGSQVYIEGQLRTRKWTQSGQDRYTTEVVVN 104
```

Matrix: blosum62
Gap Penalties: default
Length: 99
Percent Similarity: 50.515
Percent Identity: 36.082

Global (Needleman-Wunsch) versus Local (Smith-Waterman) Alignment

```
1  ..MSHNQFQFIGNLTRDTEVRHGNSNKPQAIIFDIAVNEEWRNDA.GDKQE 47
      |.  :|||. | |||:  .  |  :| .| ||. | | : .|
1  ASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKATGEMKE 50
      .
48  RTDFFRIKCFGSQAEAHGKYLKGKSLVFVQ GKIRNTKY.EKDGQTVYGTD 96
      .|:. | : || ||  .|| ||| |::|.:| | : :. || | | :
51  QTEWHRVVLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYTTE 100
      .
97  FIAD...KVDYLDTKAPGGSNQE..... 116
      :.  . | : ||.
101 VVVNVGGTMQMLGGRQGGGAPAGGNIGGGQPQGGWGQPQPQGGNQFSGG 150
      .
      .....
151 AQRPPQSAPAAPSNEPPMDFDDDIPF 177
```

Optimal versus Heuristic Algorithms

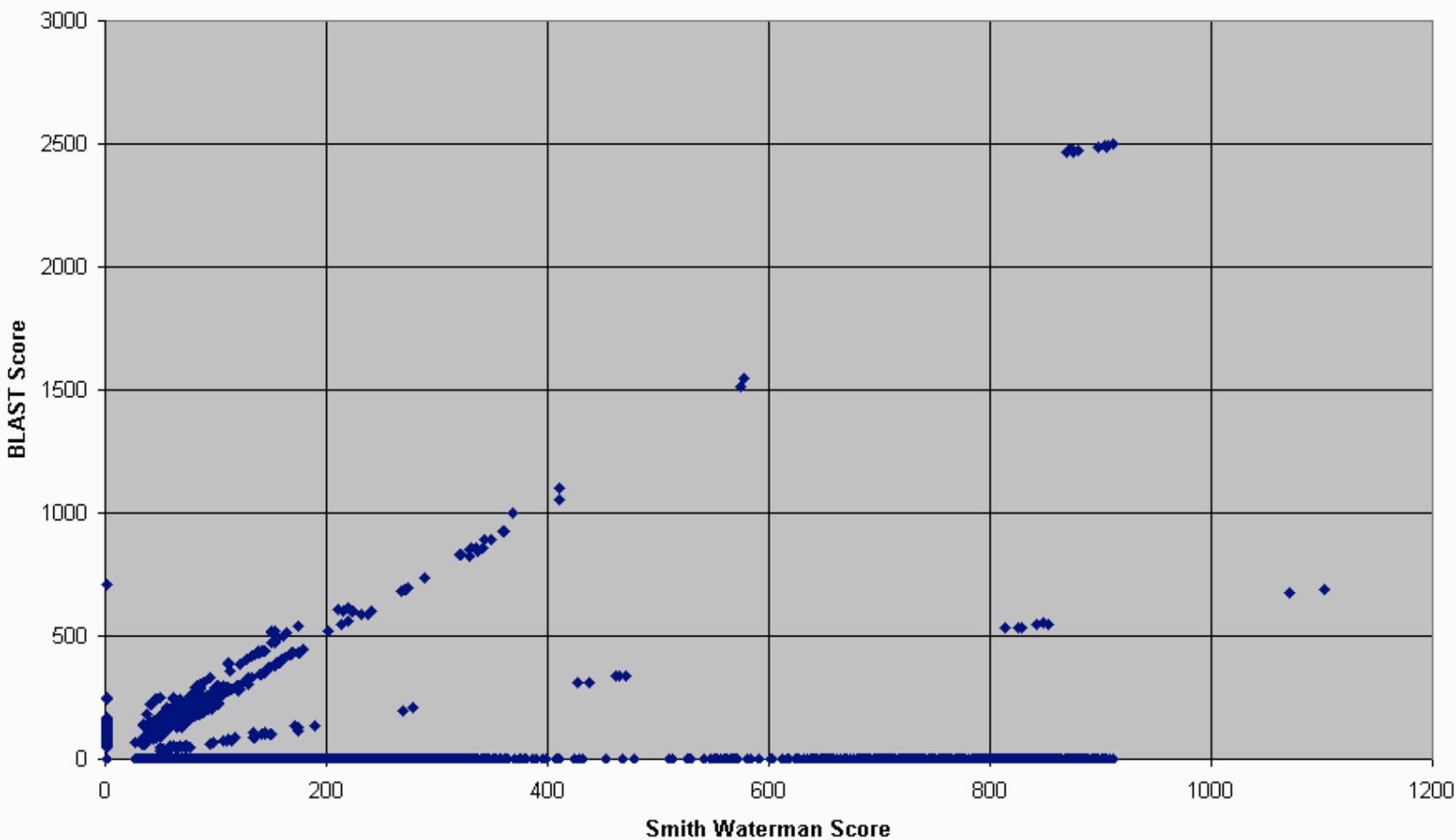
Optimal Algorithm

- Finds the optimal solution to a problem.
- Often uses an approach called dynamic programming, which solves the problem by breaking it into smaller subproblems, which are separately solved, then sequentially reassembled to solve the entire problem.
- This approach was first applied to solving biological sequence comparison problems by Saul Needleman and Christian Wunsch in 1970, and can be used to solve either a global comparison problem (Needleman-Wunsch) or a local comparison problem (Smith-Waterman).

Heuristic Algorithm

- Solves a problem by using rules of thumb to reach a solution. The solution is not guaranteed to be an optimal solution, but is generally arrived at far faster than using an optimal solution approach such as dynamic programming.
- In sequence comparison, well known heuristic approaches include the BLAST algorithm, developed by Stephen Altschul in 1990, and the FASTA algorithm, developed by William Pearson in 1988.
- Although a heuristic algorithm such as BLAST may be 100 to 1,000 times faster than an optimal algorithm such as Smith-Waterman, it may miss matches found by the optimal algorithm.

BLAST (Heuristic) versus Smith-Waterman (Optimal)



Search Algorithm Variables

Gaps

- When scoring an alignment, penalties are assigned for creating and extending gaps. The longer the gap, the greater the penalty.
- Can vary the gap penalties. This will likely change your results, so take note.

Dayhoff Substitution Matrices

- For protein comparison, all of these search algorithms use Dayhoff substitution matrices which encode likelihoods of an amino acid substitution.
- When scoring an alignment, penalties are assigned for what the substitution matrix considers poor substitutions, the worse the substitution, the greater the penalty.
- Can get different results depending on which substitution matrix you use.

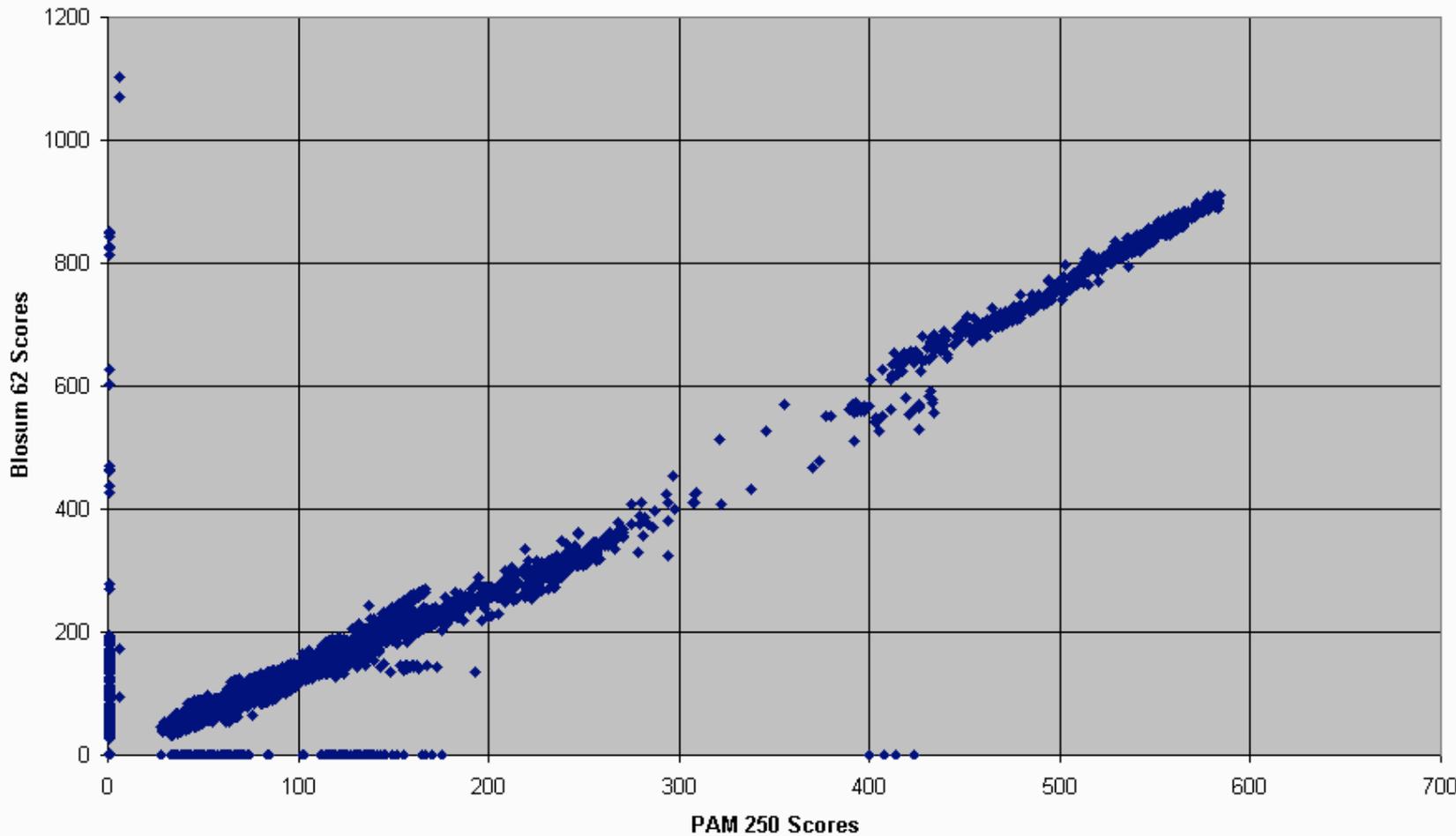
Blosum 62

Blosum 62 is the default matrix for NCBI BLAST protein comparison. It is optimized for known **close** homologies.

PAM 250

PAM 250 is a matrix which is optimized for known **distant** homologies.

Blosum 62 (Close) versus PAM 250 (Distant)



National Center for Biotechnology Information (NCBI)

NCBI

<http://www.ncbi.nlm.nih.gov>

- PubMed (over 19 million citations), PubMed Central (over 1.5 million full text articles from over 450 journals), textbooks, other reference material
- GenBank, RefSeq, CDD, MMDB and other sequence and structure databases
- Over 213 billion base pairs (213 gigabases) from over 300,000 species
- Prokaryotic genome data and browsers (over 950 complete microbial genomes, over 3,500 viruses, over 1,900 plasmids, environmental samples and additional sequences)
- Eukaryotic genome data and browsers (24 complete eukaryotic genomes, over 1,000 complete chromosomes, over 1,900 organelles, hundreds of genomes in progress)
- BLAST, PSI-BLAST and VAST search tools

NCBI BLAST

<http://blast.ncbi.nlm.nih.gov/>

NCBI BLAST Tutorials and Information

http://www.ncbi.nlm.nih.gov/Class/BLAST/blast_course.short.html

http://www.ncbi.nlm.nih.gov/Class/BLAST/blast_course.html

http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=ProgSelectionGuide

PubMed and EndNote

PubMed

<http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/>

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed>

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=books>

EndNote and PubMed

- EndNote can directly connect to PubMed and search and retrieve references from it:
Edit > Connection Files > Open Connection Manager... > PubMed (NLM) > click in **Favorites** and close window, then **Tools > Connect > PubMed (NLM)**
- Alternately, the PubMed Clipboard (use the entrez URL above for PubMed) can be used to collect references, export them as a text file in MEDLINE format, then import them into EndNote:
In PubMed, select **Clipboard** from **Send to > Clipboard** tab > select **Display MEDLINE** and **File** from **Send to** (the MEDLINE format PubMed text file will typically be named **pubmed-result.txt** and be saved on your Desktop or in your Downloads folder)
Then, in EndNote, **Edit > Import Filters > Open Filter Manager... > PubMed (NLM) >** click in **Favorites** and close window, then with a library open, **File > Import... > Import Options: PubMed (NLM) >** select file to import > **Import**

EndNote and PDF Files

- Can also use EndNote to organize PDF files, which are otherwise easy to lose track of:
With a reference open, **References > File Attachments > Attach File** inserts a link to the PDF file you select. Use **References > File Attachments > Open File** to open the linked PDF (or double click the file in the File Attachments field). To remove, just select and delete the linked file. Papers is another excellent option for PDFs (<http://mekentosj.com/papers/>).

Bioinformatics Resources

Books

Fundamental Concepts of Bioinformatics by Dan E. Krane & Michael L. Rayme

Developing Bioinformatics Computer Skills by Cynthia Gibas & Per Jambeck

BLAST: An Essential Guide to the BASIC Local Alignment Search Tool by Ian Korf, Mark Yandell & Joseph Bedell

Sequence Analysis in a Nutshell: A Guide to Common Tools and Databases by Scott Markel & Darryl Leon

Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Third Edition by A.D. Baxevanis & B.F.F. Ouellette

Introduction to Computational Biology: Maps, Sequences and Genomes by Michael S. Waterman

Bioinformatics: Sequence and Genome Analysis by David W. Mount

Biological Sequence Analysis by Richard Durbin, et al.

Journals

Bioinformatics (formerly CABIOS)

Journal of Computational Biology

Comparative and Functional Genomics

Biotechnology Software & Internet Journal

Nucleic Acids Research

Websites

<http://www.ncbi.nlm.nih.gov> National Center for Biotechnology Information

<http://www.sanger.ac.uk> Wellcome Trust Sanger Institute

<http://www.ebi.ac.uk> European Bioinformatics Institute

<http://www.iscb.org> International Society for Computational Biology

<http://www.google.com> Google

Macintosh OS X History

Mac OS X 10.0 aka “Cheetah” released March 24, 2001

Introduced Darwin Unix core, preemptive multitasking, protected memory.

Mac OS X 10.1 aka “Puma” released September 25, 2001

Introduced performance enhancements and increased stability.

Mac OS X 10.2 aka “Jaguar” released August 24, 2002

Introduced journaled file system, increased speed and stability. Begins replacing OS 9.

Mac OS X 10.3 aka “Panther” released October 24, 2003

User interface improved, increased speed, Safari default browser, Classic supports OS 9.

Mac OS X 10.4 aka “Tiger” released April 29, 2005

Supports Intel and PowerPC architectures, increased performance, Dashboard, launchd.

Mac OS X 10.5 aka “Leopard” released October 26, 2007

User interface improvements, Boot Camp, Spaces, Time Machine, removed Classic support.

Mac OS X 10.6 aka “Snow Leopard” released August 28, 2009

Intel architecture only, smaller footprint, 64 bit, performance and user interface improvements.

Macintosh OS X Startup

Startup Sequence

BootROM POST → Open Firmware → BootX → Kernel extensions → System and kernel initialization → launchd → startup daemons, scripts and items → loginwindow

Useful Startup Keys

Hold **C** key at startup: boots from **CD** drive (if a bootable CD or DVD is in it)

Hold **Mouse button** at startup: ejects disk from CD/DVD drive

Hold **T** key at startup: boots into FireWire **T**arget Mode

Hold **Option** key at startup: select startup disk or system to boot from (if more than one)

Hold **Command, Option, P** and **R** keys at startup: reset PRAM

Hold **Shift** key at startup: boots into **S**afe Boot mode

Preventive Maintenance

OS X features a journaled file system which automatically corrects problems and defragments files, so file system maintenance is generally not required. Certain daily, weekly and monthly maintenance tasks are scheduled to automatically run between 3:15 and 5:30 A.M. These tasks cannot run if the computer is off or asleep at these times, so it good to occasionally leave it plugged in and turned on overnight (set to never sleep in the Energy Saver pane in System Preferences).

To head off possible file problems, run Disk Utility once a month (on the startup volume). Disk Utility is located in the Utilities folder inside the Applications folder. Use it to check S.M.A.R.T. Status, Repair Disk Permissions and Verify Disk.

In OS X 10.5 and beyond, **Time Machine** allows for automatic incremental backups.

Macintosh OS X Finder

Finder

Finder **Preferences...** in the Finder menu lets you alter the Finder's behavior: in General, try **Always open folders in a new window** and **Spring-loaded folders and windows**, in Advanced, try **Show all file extensions**.

Install all new applications in the **Applications** folder, save all other files in your **Home** folder.

Useful Finder Key Combinations

Hold **Option** key while **dragging** a file to **duplicate** it

Hold **Command** and **Option** keys down while **dragging** a file to create an **alias**

Hold **Option** while **double clicking a folder** to close the previous folder

Press **Shift** while **clicking** to select more than one item

In **List** view, press **Command** while **clicking** for discontinuous selection

Press **Control** while **clicking** to get a contextual menu

Press **Shift** and **Command** and **N** to create a **new folder**

Press **Command** and **I** with an item selected to **Show Info**

Press **Command** and **K** to **connect** to a server

Press **Command** and **space** to open a Spotlight search window

Help

Press **Command** and **?** to get **Mac Help**, or select it from the Help menu. It has extensive documentation on OS X, your computer, and various applications and **Shortcuts**.

The Toolbar, Sidebar and Dock

Red button to close windows, **Yellow** button to dock windows, **Green** button to resize windows (note that document windows with unsaved changes will have a **dot** in the **Red** button).

Toolbar

- Back/Forward Arrows; Action; Icon, List, Column Views; Shortcuts and Search
- Drag items to the Toolbar to add them (hold **Command** while **dragging** to rearrange or remove an item).
- To **hide the Toolbar**, click the **Clear** lozenge in the upper right corner (or show it again)
- Select **Customize Toolbar...** from the Finder View menu to modify the Toolbar, e.g. add Path.

Sidebar

- Drag folders into the sidebar to add them, out of the sidebar to remove them
- Eject CDs, DVDs and network mounted volumes by clicking on the eject icon next to them
- Burn CDs or DVDs by clicking on the burn icon next to them

Dock

- Drag applications or utilities on to add, off to remove, or drag left or right to rearrange
- Can add a folder or folders with aliases to your favorite applications after the dock separator
- Click and hold to get application options, or hierarchical submenus for a folder or hard drive item
- Press **Command** and **Option** and **click** an item to bring it forward and hide everything else (click in dock to make it visible again, or select Show All from the Finder menu to make all visible)
- To **hide the Dock**, press **Command** and **Option** and **D** (same keys to show it again)

Press and hold **Command**, then press and keep pressing **Tab** to cycle through all open applications in the dock one by one. To cycle backwards, press **Command, Shift** and **Tab**.

Dashboard and Exposé

Dashboard

Dashboard is a layer that displays useful tools called Widgets. By default, it is activated by pressing the **F12** key or **Dashboard** icon in the Dock, and cancelled by pressing it again, or clicking outside of any Widget.

- To control which Widgets are displayed, click on the **+** symbol in the lower left hand corner.
- Additional Widgets can be downloaded from sites such as <http://www.apple.com/downloads/dashboard/> or <http://www.dashboardwidgets.com/> and installed automatically, or manually by copying them to the Widgets folder inside the Library folder of your Home folder.

Exposé

Exposé is a utility for managing and accessing open windows. It can be particularly useful when you are working with a large number of windows on a small screen, such as that of a laptop. By default, the following keys trigger these Exposé functions:

- F9** Display all open windows in a temporarily rearranged and minimized format, so you can locate a particular window.
- F10** Display all open windows *in the current application* in a temporarily rearranged and minimized format, so you can locate a particular application window or document.
- F11** Display the Desktop by temporarily moving all open windows that would cover it offscreen.

Spotlight

Finding Information with Spotlight

Spotlight provides a systemwide searchable index of file information, including not only information such as file names, but also information inside a file (such as the text of a document), or metadata about a file (such as a description, keywords, or when a file was last opened).

To activate Spotlight, click in the upper right hand corner of the screen, or press **Command** and **Space**, then enter your query. It will return an expandable, categorized list of search results.

To modify the categories of results that will appear in a Spotlight search, or the order in which they will appear, use the Spotlight pane in System Preferences.

Finds and Smart Folders

A more specific search can be performed using the Find function of the Finder. This can be activated while in the Finder by pressing **Command** and **F**, pressing **Option, Command** and **N**, or by typing in a Toolbar search window. You can then specify exactly what to search (by selecting a listed scope or by clicking **Others...** and dragging in a folder or volume), and the criteria to search with (click **+** to add additional criteria to a search).

To save a particular search as a Smart Folder, click the **Save** button, then select where you want the search saved. These search results can then be accessed at any time through the Smart Folder, and updated for new files that might have been added since the last search by clicking the **arrow** in the bottom right of the folder.

System Preferences

Setting System Preferences

Use **Show All** to view all preference panes. Can select **Organize by Categories** or **Organize Alphabetically** from the **View** menu.

Appearance: recommend Font smoothing style: Medium, can adjust appearance and interface

Dashboard & Exposé: set preferences for activating Dashboard, Exposé and Screen Saver

Dock: set Dock preferences

Spotlight: set preferences for display of search results and search shortcuts

Displays: allows you to adjust screen resolution and colors

Energy Saver: lets you adjust when your computer goes to sleep, with different settings for battery and power adapter use

Keyboard & Mouse: allows you to review and set custom Keyboard Shortcuts, and adjust mouse double click and tracking speed

Network: set TCP/IP and AppleTalk settings for Built-in Ethernet and AirPort. Can create multiple settings for different locations from the popup Location menu (select New Location... or Edit Locations...)

Sharing: turn various file sharing options on or off, activate/deactivate Firewall

Accounts: controls password, login picture and items to automatically start up at login

Classic: controls OS 9 emulation (Classic mode). Only in OS X 10.4 and prior.

Software Update: check at least once a month, can set to download important updates automatically

Startup Disk: controls how you will start up, including Target Disk Mode

Useful Applications

Dictionary

Built in dictionary and thesaurus. Can activate in many applications by **Command Control D** (can hover and activate).

Excel

The spreadsheet standard. Capable of producing excellent graphs and charts with some effort.

iPhoto

Can do limited file format conversion, can create libraries and custom albums useful for storing and organizing gel photos and other scientific data. If you keep each library under 650 MB you can easily back it up to a single CD-R.

iMovie

Useful for simple film editing, including time lapse photography, which can be useful for scientific presentations. More advanced users can upgrade to QuickTime Pro, or get Final Cut Express.

iCal

Calendering application, allows for sharing of schedules using a WebDAV server. Use **Option Return** to create a line break in an entry. Additional calendars for import (e.g. U.S. holidays) are available at www.apple.com/ical/library

Keynote and PowerPoint

Apple's Keynote provides an easier to use alternative to the presentation standard, Microsoft's PowerPoint. Keynote has fewer features but far better graphics and templates.

Word

The standard for word processing. Works well with EndNote to manage references in publications. If you are doing a lot of text or sequence manipulation, an application such as **TextWrangler** (free) or **BBEdit** (commercial) is better.

Useful Utilities

Disk Utility

- Can verify and repair permissions on your hard drive. Worth checking once a month
- Can verify your hard drive (and repair it if booted from another drive or CD). Check monthly
- Allows you to create disk images and encrypted disk images, which are virtual disks. It is one of the best ways to securely store or distribute Macintosh files
- To dismount a mounted disk image, click eject icon, or drag to trash (if refuses to eject, logout)
- Can copy CDs and DVDs, erase CD-RWs and DVD-RWs, and burn multisession CD-Rs (select **Burn... from Images** menu, click on triangle and select **Leave disc appendable**)

Print Center and PDF Printing

- Add new printers here, and select default printers
- Any OS X application can print to PDF file format by selecting **File > Print** and clicking on **Save As PDF...** To reduce the size of an image-heavy PDF select **Compress PDF** instead

Keychain Access

- Allows you to store and automatically access different passwords in a safe, encrypted file

Preview

- Allows for image file viewing and conversion (Adobe Acrobat is also good for viewing PDFs)

Calculator

- Has **Convert** menu for scientific conversions and Scientific Calculator mode.

Grab

- Useful for doing screen captures (can also use **Command-Shift-3** and **Command-Shift-4**)

Telecommunications Applications

Mail

- Apple's email client, features sophisticated sorting and junk mail filtering capabilities, works with IMAP (Internet Message Access Protocol) or POP (Post Office Protocol) mail servers
- Mark junk mail as such, and correct mistakes to improve the accuracy of the junk mail filter, then set **Mail > Preferences > Junk Mail** to **Automatic** to automatically sort it out

Address Book

- Stores contact information, linked to Mail
- To quickly create groups, select members by **Command clicking**, then select **New Group From Selection** command in **File** menu
- Can bring up maps from the **home** or **work** labels
- Backup by backing up entire Address Book application (**File > Back up Address Book...**)

Safari

- Apple's web browser, meant to replace Internet Explorer, which is no longer being developed for the Macintosh. Another good alternative is Firefox.
- Has a tabbed browsing mode which can be activated from **Safari > Preferences > Tabs**
- Can activate a **Debug** menu in Safari which has many useful options (including letting Safari appear to be Internet Explorer) with these Terminal commands (first quit Safari):

```
defaults write com.apple.Safari IncludeDebugMenu 1
exit
```

iChat

- Chat, audio chat, video chat, file transfer, chat rooms

Internet Connect

- Virtual Private Network/Point-to-Point Tunneling Protocol (VPN/PPTP) client