Sequence Alignment Practical

Presented by

Kirill Bessonov

Nov 12, 2013

Université de Liège

Talk Structure

- Introduction to sequence alignments
- Methods / Logistics
  - Global Alignment: Needleman-Wunsch
  - Local Alignment: Smith-Waterman
- Illustrations of two types of alignments
  - step by step local alignment
- Computational implementation of alignment
  - Retrieval of sequences using R
  - Alignment of sequences using R
- Homework – HW2

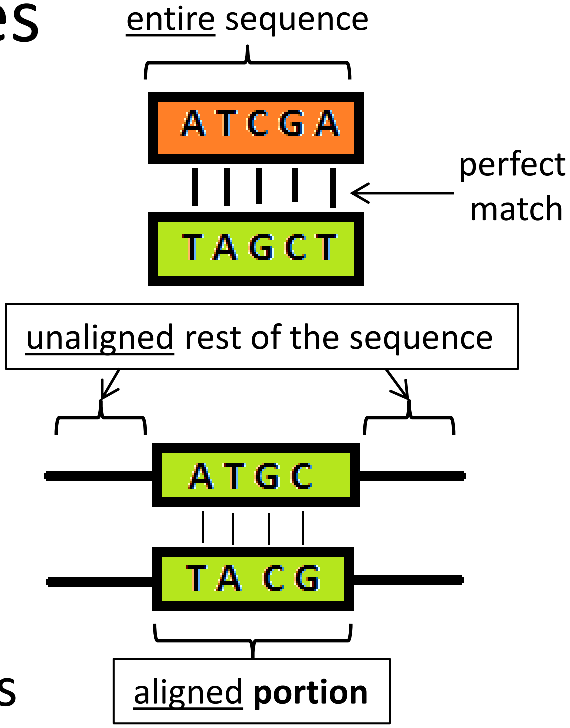
Sequence Alignments

Comparing two objects is intuitive. Likewise sequence pairwise alignments provide info on:

- evolutionary distance between species (e.g. homology)
- new functional motifs / regions
- genetic manipulation (e.g. alternative splicing)
- new functional roles of unknown sequence
- identification of binding sites of primers / TFs
- *de novo* genome assembly
  - alignment of the short “reads” from high-throughput sequencer (e.g. Illumina or Roche platforms)

Comparing two sequences

- There are two ways of pairwise comparison
  - Global using **Needleman-Wunsch algorithm (NW)**
  - Local using **Smith-Waterman algorithm (SW)**
- Both approaches use similar methodology, but have completely different objectives
  - Global alignment (NW)
    - tries to align the "whole" sequence
    - more restrictive than local alignment
  - Local alignment (SW)
    - tries to align portions (e.g. motifs) of given sequences
    - more flexible as considers "parts" of the sequence
    - works well on highly divergent sequences



Global alignment (NW)

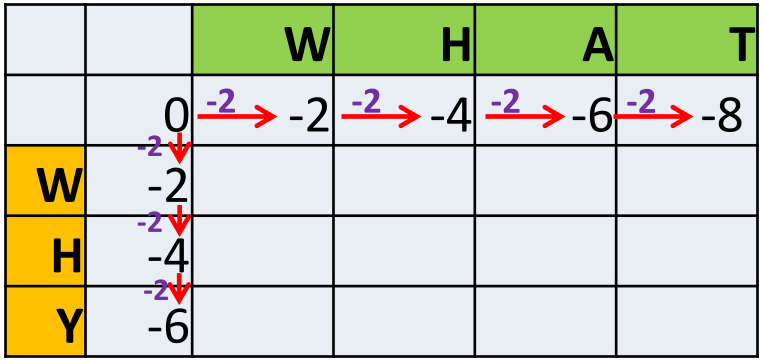
- Sequences are aligned end-to-end along their **entire** length
- Many possible alignments are produced
  - The alignment with the highest score is chosen
- Naïve algorithm is very inefficient (O^{exp})
  - To align sequence of length 15, need to consider
    - Possibilities # = (insertion, deletion, gap)^{15} = 3^{15} = $1,4*10^{7}$
  - Impractical for sequences of length >20 nt
- Used to analyze homology/similarity of entire:
  - genes and proteins
  - assess gene/protein overall homology between species

Methodology of global alignment (1 of 4)

- Define scoring scheme for each event
  - mismatch between *a_i* and *b_j*
    - $s(a_i, b_j) = -1$ if $a_i \neq b_j$
  - gap (insertion or deletion)
    - $s(a_i, -) = s(-, b_j) = -2$
  - match between *a_i* and *b_j*
    - $s(a_i, b_j) = +2$ if $a_i = b_j$
- Provide no restrictions on minimal score
- Start completing the alignment `MxN` matrix

Methodology of global alignment (2 of 4)

- The matrix should have extra column and row
  - – M+1 columns , where M is the length sequence M
  - – N+1 rows, where N is the length of sequence N
- Initialize the matrix by introducing **gap penalty** at every **initial** position along rows and columns
- Scores at each cell are **cumulative**

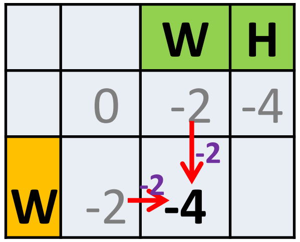


		W	H	A	T
	0	-2	-4	-6	-8
W	-2				
H	-4				
Y	-6				

Methodology of global alignment (3 of 4)

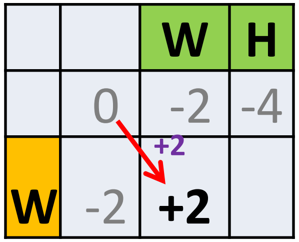
- For each cell consider all three possibilities

1)Gap (horiz/vert)



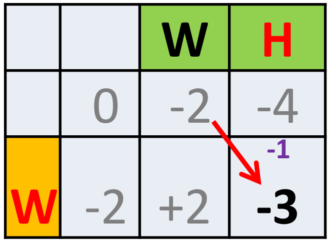
		W	H
	0	-2	-4
W	-2	-4	

2)Match (W-W diag.)



		W	H
	0	-2	-4
W	-2	+2	

3)Mismatch(W-H diag)



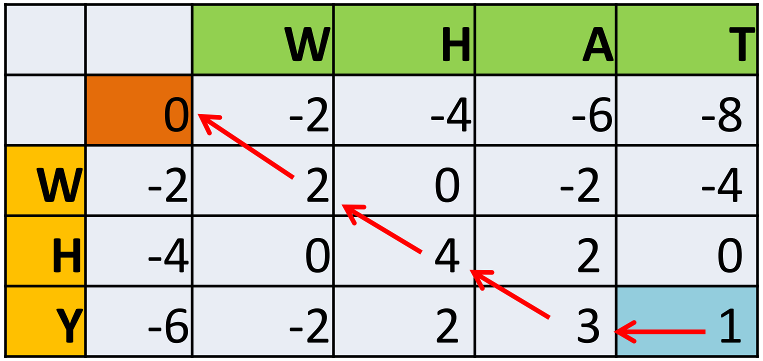
		W	H
	0	-2	-4
W	-2	+2	-3

- Select **the maximum** score for each cell and fill the matrix

		W	H	A	T
	0	-2	-4	-6	-8
W	-2	2	0	-2	-4
H	-4	0	4	2	0
Y	-6	-2	2	3	1

Methodology of global alignment (4 of 4)

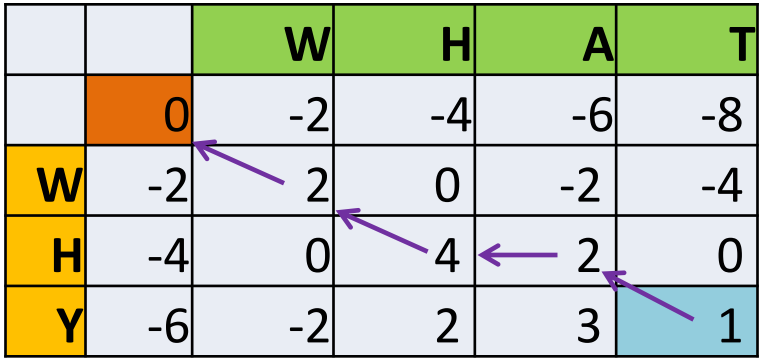
- Select the most **very bottom right** cell
- Consider different path(s) going to **very top left cell**
  - Path is constructed by finding **the source cell** w.r.t. the current cell
  - How the current cell value was generated? From where?



		W	H	A	T
	0	-2	-4	-6	-8
W	-2	2	0	-2	-4
H	-4	0	4	2	0
Y	-6	-2	2	3	1

```
WHAT
WHY-
```

Overall score = 1



		W	H	A	T
	0	-2	-4	-6	-8
W	-2	2	0	-2	-4
H	-4	0	4	2	0
Y	-6	-2	2	3	1

```
WHAT
WH-Y
```

Overall score = 1

- Select the best alignment(s)

Local alignment (SW)

- Sequences are aligned to find **regions** where **the best** alignment occurs (i.e. highest score)
- Assumes a **local** context (aligning parts of seq.)
- Ideal for finding short motifs, DNA binding sites
  - **helix**-**loop**-**helix (bHLH**) - motif
  - TATAAT box (a famous promoter region) – DNA binding site
- Works well on highly divergent sequences
  - Sequences with highly variable introns but highly conserved and sparse exons

Methodology of local alignment (1 of 4)

- The scoring system is similar with one exception
  - The **minimum** possible score in the matrix is **zero**
  - **There are no negative scores in the matrix**
- Let's define the same scoring system as in global

1) mismatch between a_i and b_j

$s(a_i, b_j) = \mathbf{-1}$ if $a_i \neq b_j$

2) gap (insertion or deletion)

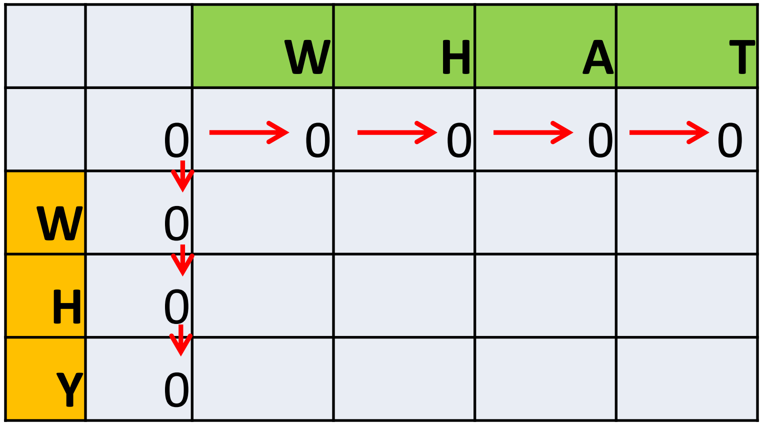
$s(a_i, -) = s(-, b_j) = \mathbf{-2}$

3) match between a_i and b_j

$s(a_i, b_j) = \mathbf{+2}$ if $a_i = b_j$

Methodology of local alignment (2 of 4)

- Construct the MxN alignment matrix with M+1 columns and N+1 rows
- Initialize the matrix by introducing **gap penalty** at 1st row and 1st column



		W	H	A	T
	0	0	0	0	0
W	0				
H	0				
Y	0				

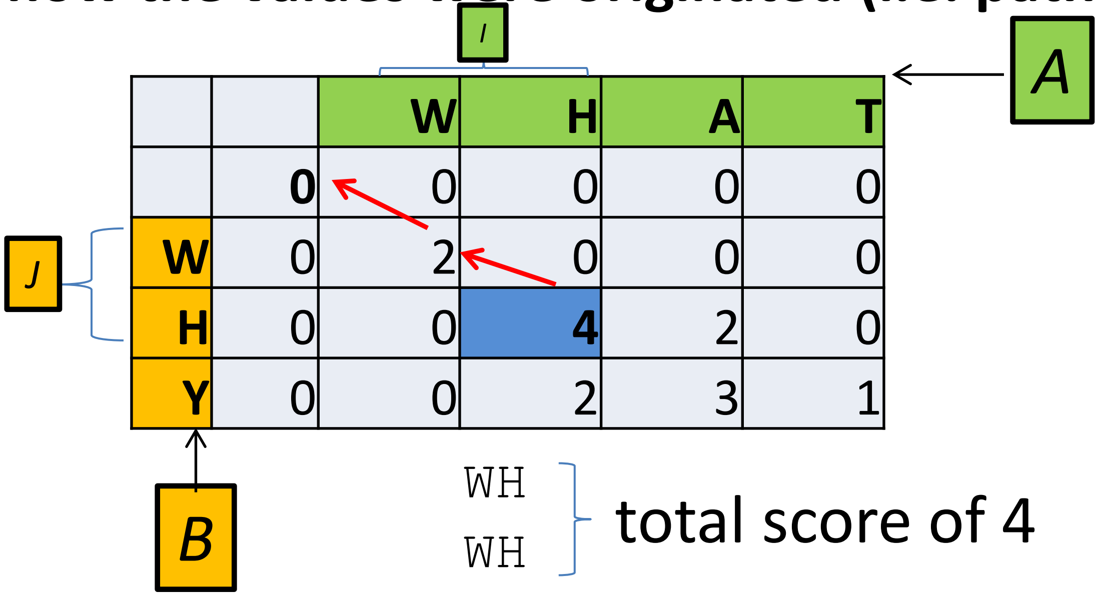
Methodology of local alignment (3 of 4)

- For each subsequent cell consider all possibilities (i.e. motions)
  1) Vertical 2)Horizontal 3)Diagonal
- For each cell select the highest score
  – If score is negative → assign **zero**

		W	H	A	T
	0	0	0	0	0
W	0	2	0	0	0
H	0	0	4	2	0
Y	0	0	2	3	1

Methodology of local alignment (4 of 4)

- Select the initial cell with the **highest score(s)**
- Consider different path(s) leading to score of **zero**
  - **Trace-back the cell values**
  - **Look how the values were originated (i.e. path)**



		W	H	A	T
	0	0	0	0	0
W	0	2	0	0	0
H	0	0	**4**	2	0
Y	0	0	2	3	1

```
WH
WH
```

total score of 4

- Mathematically $M(A,B) = \max\{S(I,J) : I \subset A, J \subset B\}$
  - where *S*(*I*, *J*) is the score for **sub-sequences** *I* and *J*

Local alignment illustration (1 of 2)

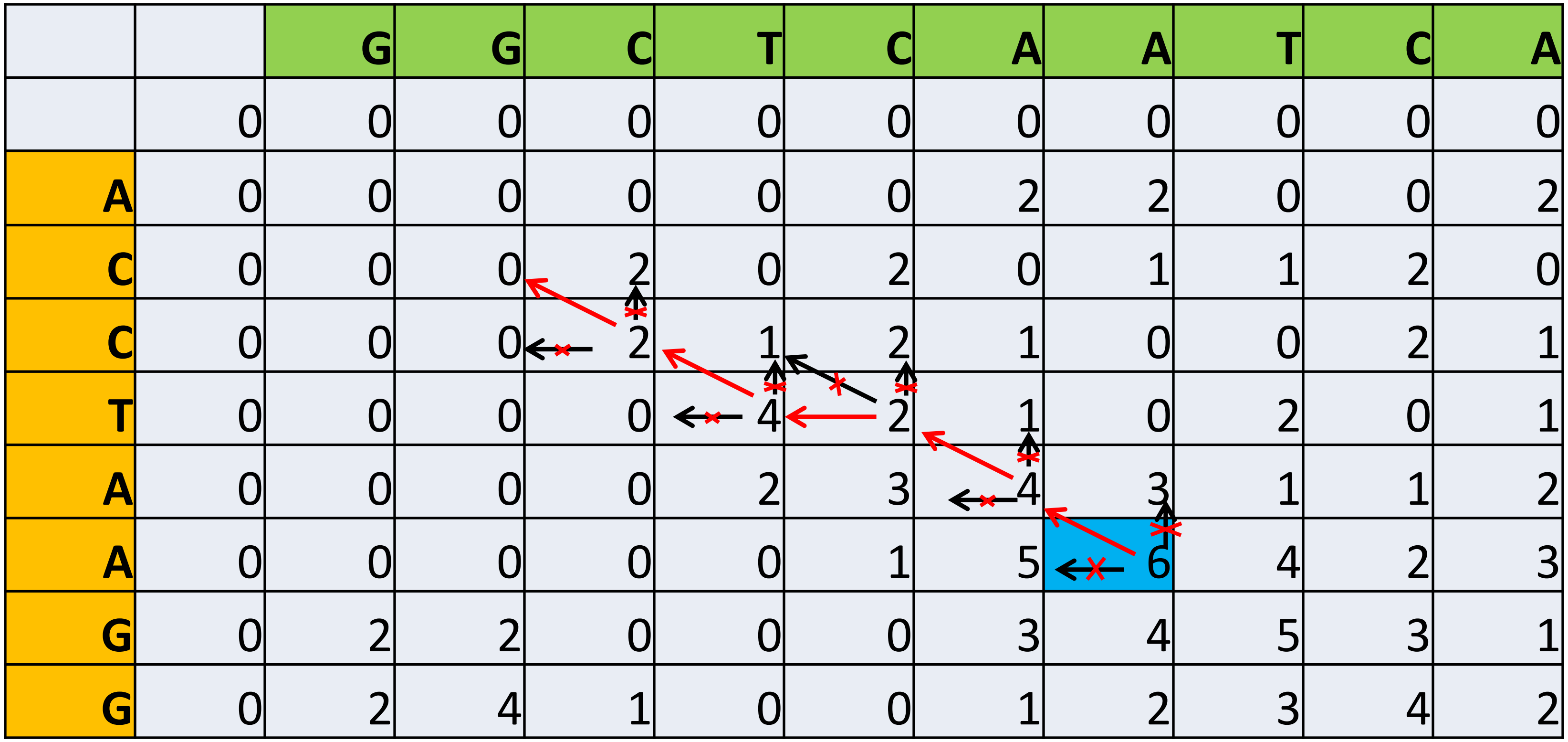
- Determine the best **local** alignment and the maximum alignment score for

- **Sequence A:** ACCTAAGG
- **Sequence B:** GGCTCAATCA
- Scoring conditions:
  - $s(a_i, b_j) = +2$ if $a_i = b_j$,
  - $s(a_i, b_j) = -1$ if $a_i \neq b_j$ and
  - $s(a_i, -) = s(-, b_j) = -2$

Local alignment illustration (2 of 2)

		G	G	C	T	C	A	A	T	C	A
	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	2	2	0	0	2
C	0	0	0	2	0	2	0	1	1	2	0
C	0	0	0	2	1	2	1	0	0	2	1
T	0	0	0	0	4	2	1	0	2	0	1
A	0	0	0	0	2	3	4	3	1	1	2
A	0	0	0	0	0	1	5	6	4	2	3
G	0	2	2	0	0	0	3	4	5	3	1
G	0	2	4	1	0	0	1	2	3	4	2

Local alignment illustration (3 of 3)



		G	G	C	T	C	A	A	T	C	A
	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	2	2	0	0	2
C	0	0	0	2	0	2	0	1	1	2	0
C	0	0	0	2	1	2	1	0	0	2	1
T	0	0	0	0	4	2	1	0	2	0	1
A	0	0	0	0	2	3	4	3	1	1	2
A	0	0	0	0	0	1	5	6	4	2	3
G	0	2	2	0	0	0	3	4	5	3	1
G	0	2	4	1	0	0	1	2	3	4	2

```
CTCAA      GGCTCAATCA
CT-AA      ACCT-AAGG
```

Best score: 6

in the whole seq. context

Aligning proteins Globally and Locally

Université de Liège

Protein Alignment

- Protein local and global alignment follows the same rules as we saw with DNA/RNA
- Differences
  - alphabet of proteins is 22 residues long
  - special scoring/substitution matrices used
  - conservation and protein proprieties are taken into account
    - E.g. residues that are totally different due to charge such as polar Lysine and apolar Glycine are given a low score

Substitution matrices

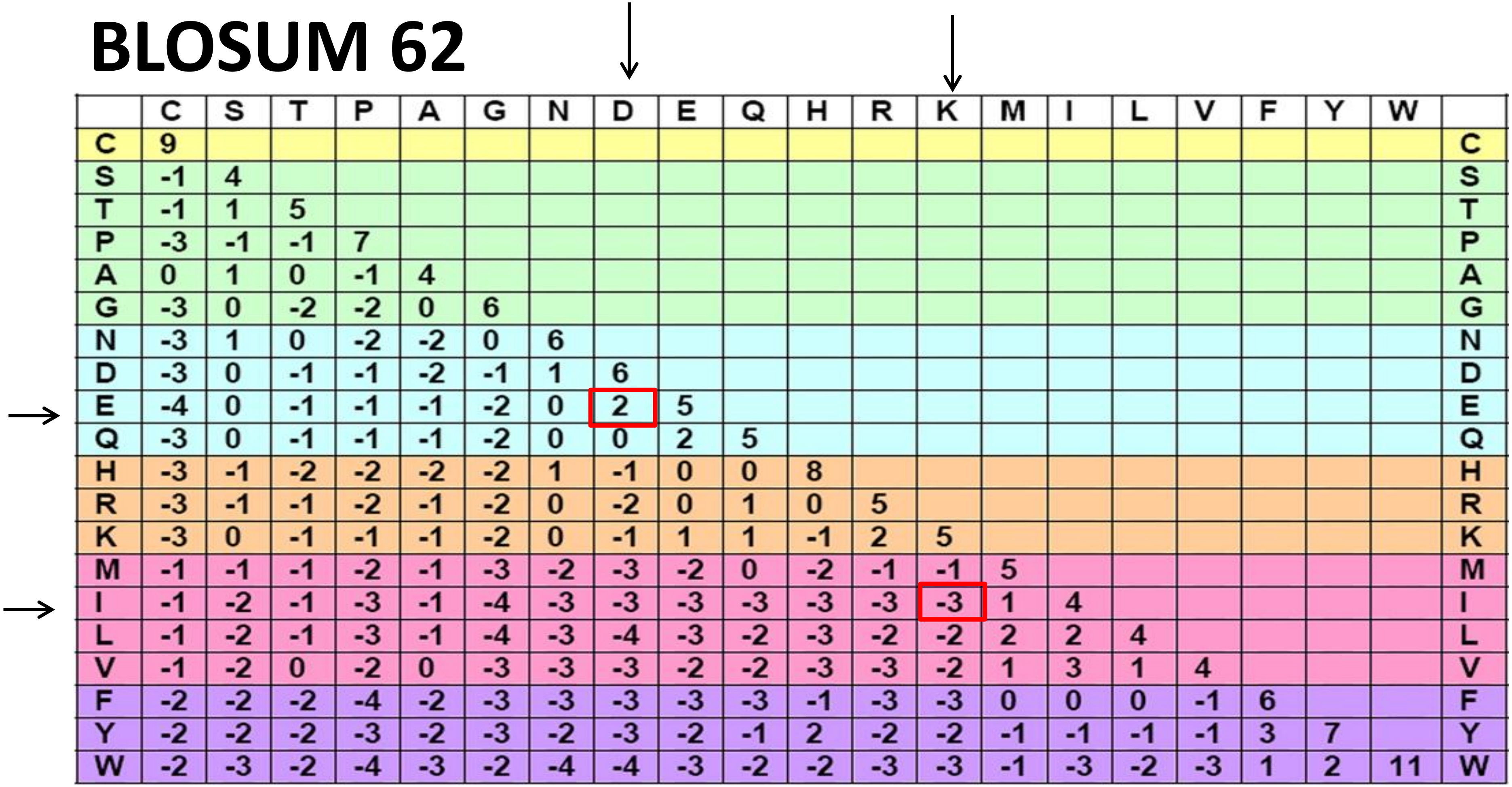
- Since protein sequences are more complex, matrices are collection of scoring rules
- These are 2D matrices reflecting comparison between sequence A and B
- Cover events such as
  - mismatch and perfect match
- Need to define gap penalty separately
- Popular **BLO**cks **SU**bstitution **M**atrix **(BLOSUM)**

BLOSUM-x matrices

- Constructed from aligned sequences with specific x% similarity
  - matrix built using sequences with no more then 50% similarity is called **BLOSUM-50**

- For highly mutating / dissimilar sequences use
  - BLOSUM-45 and lower
- For highly conserved / similar sequences use
  - BLOSUM -62 and higher

BLOSUM 62



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

- What diagonal represents? **perfect match between a.a.**
- What is the score for substitution E→D (acid a.a.)? **Score = 2**
- More drastic substitution K→I (basic to non-polar)? **Score = -3**

Practical problem:

Align following sequences both globally and locally using BLOSUM 62 matrix with gap penalty of -8

Sequence A: `AAEEKKLAAA`
Sequence B: `AARRIA`

Aligning globally using BLOSUM 62

		A	A	E	E	K	K	L	A	A	A
	0	-8	-16	-24	-32	-40	-48	-56	-64	-72	-80
A	-8	4	-4	-12	-20	-28	-36	-44	-52	-60	-68
A	-16	-4	8	0	-8	-16	-24	-32	-40	-48	-56
R	-24	-12	0	8	0	-6	-14	-22	-30	-38	-46
R	-32	-20	-8	0	8	2	-4	-12	-20	-28	-36
I	-40	-28	-16	-8	0	5	-1	-2	-10	-18	-26
A	-48	-36	-24	-16	-8	-1	4	-2	2	-6	-14

```
AAEEKKLAAA
AA--RRIA--
```

Score: -14

Other alignment options? Yes

Aligning locally using BLOSUM 62

		A	A	E	E	K	K	L	A	A	A
	0	0	0	0	0	0	0	0	0	0	0
A	0	4	4	0	0	0	0	0	4	4	4
A	0	4	8	3	0	0	0	0	4	8	8
R	0	0	3	8	3	2	2	0	0	3	7
R	0	0	0	3	8	5	4	0	0	0	2
I	0	0	0	0	0	5	2	6	0	0	0
A	0	4	4	0	0	0	4	1	10	4	4

```
KKLA
RRIA
```

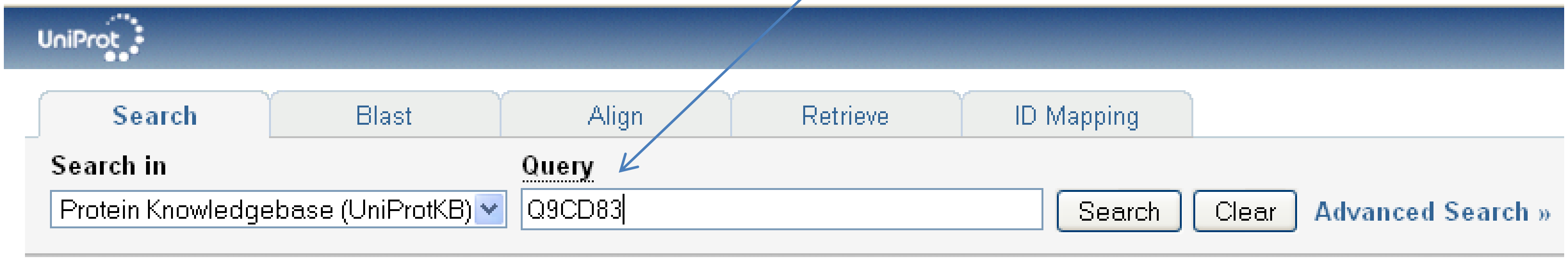
Score: 10

Using R for:

- Sequence Retrieval and Analysis

Protein database UniProt

- UniProt database (http://www.uniprot.org/) has high quality protein data **manually** curated
- It is manually curated
- Each protein is assigned **UniProt ID**

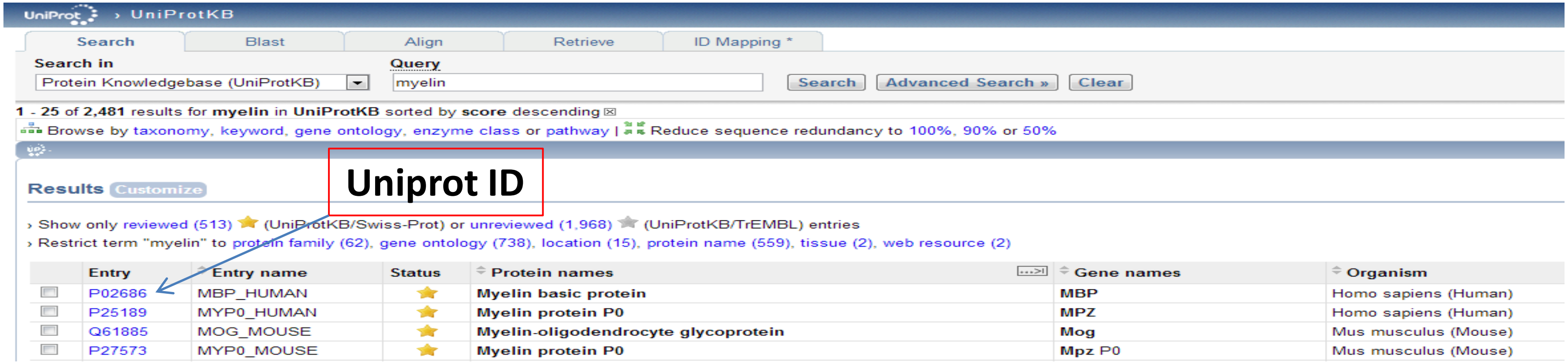


Search in: Protein Knowledgebase (UniProtKB)

Query: Q9CD83

Retrieving data from UniProt

- In search field one can enter either use UniProt ID or common protein name
  - **example:** myelin basic protein



Search in: Protein Knowledgebase (UniProtKB)

Query: myelin

1 - 25 of 2,481 results for **myelin** in **UniProtKB** sorted by **score** descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50%

Results Customize

Uniprot ID

› Show only reviewed (513) ★ (UniProtKB/Swiss-Prot) or unreviewed (1,968) ★ (UniProtKB/TrEMBL) entries

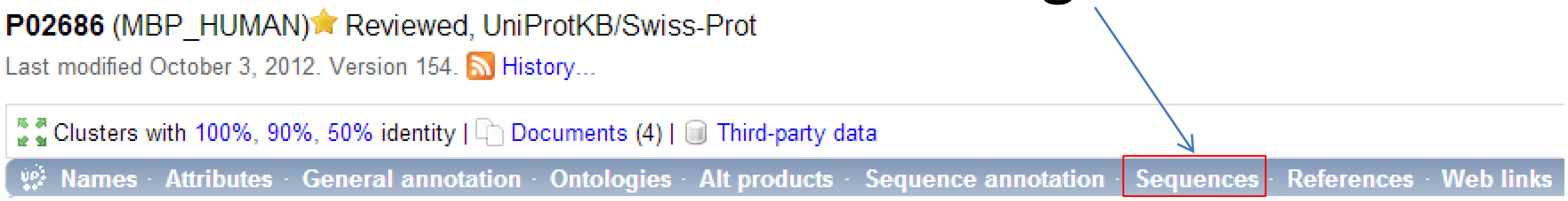
› Restrict term "myelin" to protein family (62), gene ontology (738), location (15), protein name (559), tissue (2), web resource (2)

	Entry	Entry name	Status	Protein names	Gene names	Organism
□	P02686	MBP_HUMAN	★	**Myelin basic protein**	**MBP**	Homo sapiens (Human)
□	P25189	MYP0_HUMAN	★	**Myelin protein P0**	**MPZ**	Homo sapiens (Human)
□	Q61885	MOG_MOUSE	★	**Myelin-oligodendrocyte glycoprotein**	**Mog**	Mus musculus (Mouse)
□	P27573	MYP0_MOUSE	★	**Myelin protein P0**	**Mpz** P0	Mus musculus (Mouse)

- We will use retrieve data for **P02686**

Understanding UniProt fields

- Information is divided into categories



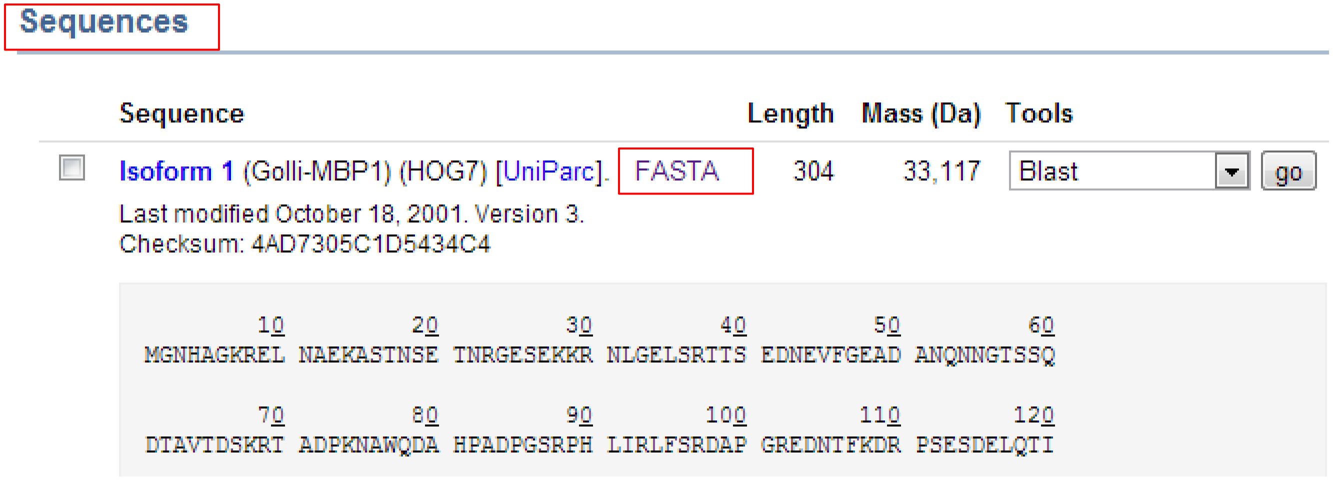
P02686 (MBP_HUMAN)⭐ Reviewed, UniProtKB/Swiss-Prot

Last modified October 3, 2012. Version 154. History...

Clusters with 100%, 90%, 50% identity | Documents (4) | Third-party data

Names · Attributes · General annotation · Ontologies · Alt products · Sequence annotation · Sequences · References · Web links

- Click on ‘**Sequences**’ category and then **FASTA**



Sequences

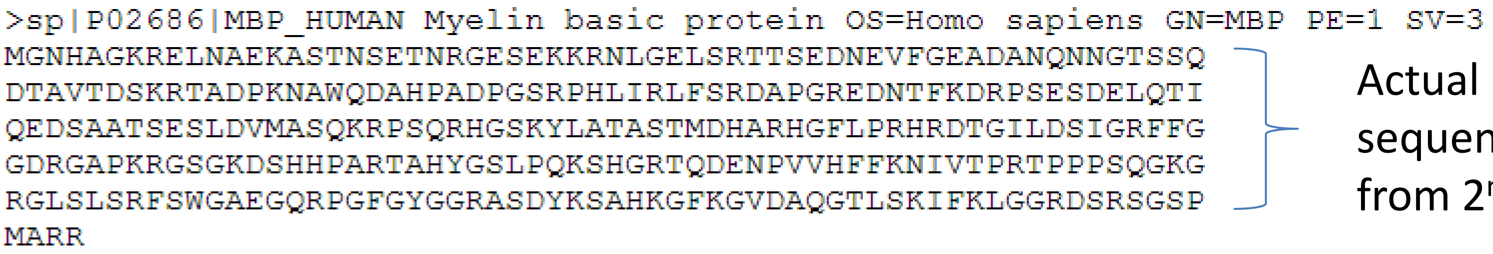
	Sequence	Length	Mass (Da)	Tools
□	**Isoform 1** (Golli-MBP1) (HOG7) [UniParc]. FASTA Last modified October 18, 2001. Version 3. Checksum: 4AD7305C1D5434C4	304	33,117	Blast ▾ go

```
        10         20         30         40         50         60
MGNHAGKREL NAEKASTNSE TNRGESEKKR NLGELSRTTS EDNEVFGEAD ANQNNGTSSQ

        70         80         90        100        110        120
DTAVTDSKRT ADPKNAWQDA HPADPGSRPH LIRLFSRDAP GREDNTFKDR PSESDELQTI
```

FASTA format

- FASTA format is widely used and has the following parameters
  - – Sequence name start with **>** sign
  - – The fist line corresponds to protein name



```
>sp|P02686|MBP_HUMAN Myelin basic protein OS=Homo sapiens GN=MBP PE=1 SV=3
MGNHAGKRELNAEKASTNSETNRGESEKKRNLGELSRTTSEDNEVFGEADANQNNGTSSQ
DTAVTDSKRTADPKNAWQDAHPADPGSRPHLIRLFSRDAPGREDNTFKDRPSESDELQTI
QEDSAATSESLDVMASQKRPSQRHGSKYLATASTMDHARHGFLPRHRDTGILDSIGRFFG
GDRGAPKRGSGKDSHHPARTAHYGSLPQKSHGRTQDENPVVHFFKNIVTPRTPPPSQGKG
RGLSLSRFSWGAEGQRPGFGYGGRASDYKSAHKGFKGVDAQGTLSKIFKLGGRDSRSGSP
MARR
```

Actual protein sequence starts from 2nd line

Retrieving protein data with R and SeqinR

- Can "talk" programmatically to UniProt database using R and *seqinR* library
  - – *seqinR* library is suitable for
    - "Biological Sequences Retrieval and Analysis"
    - Detailed manual could be found here
  - – Install this library in your R environment

```
install.packages("seqinr")
library("seqinr")
```

  - – Choose database to retrieve data from

```
choosebank("swissprot")
```

  - – Download data object for target protein **(P02686)**

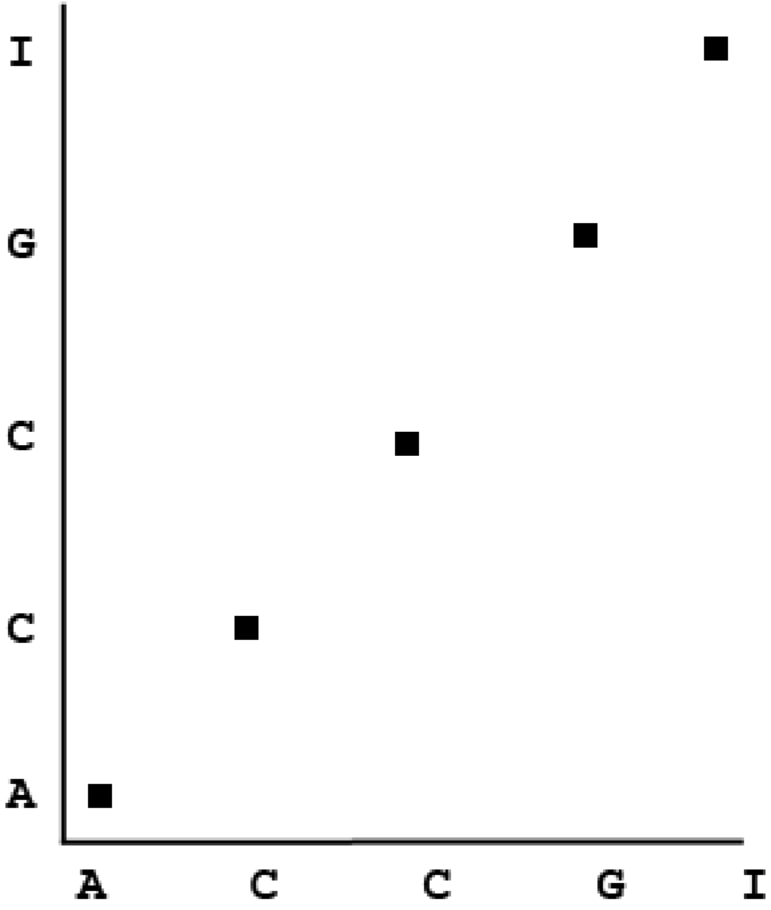
```
query("MBP_HUMAN", "AC=P02686")
```

  - – **See sequence of the object** MBP_HUMAN

```
MBP_HUMAN_seq = getSequence(MBP_HUMAN); MBP_HUMAN_seq
```

Dot Plot (comparison of 2 sequences) (1of2)

- 2D way to find regions of similarity between two sequences
  - Each sequence plotted on either vertical or horizontal dimension
  - If **two a.a.** from two sequnces at given positions are **identical** the **dot** is plotted
  - **matching** sequence segments appear as **diagonal lines** (that could be parallel to the absolute diagonal line if insertion or gap is present)



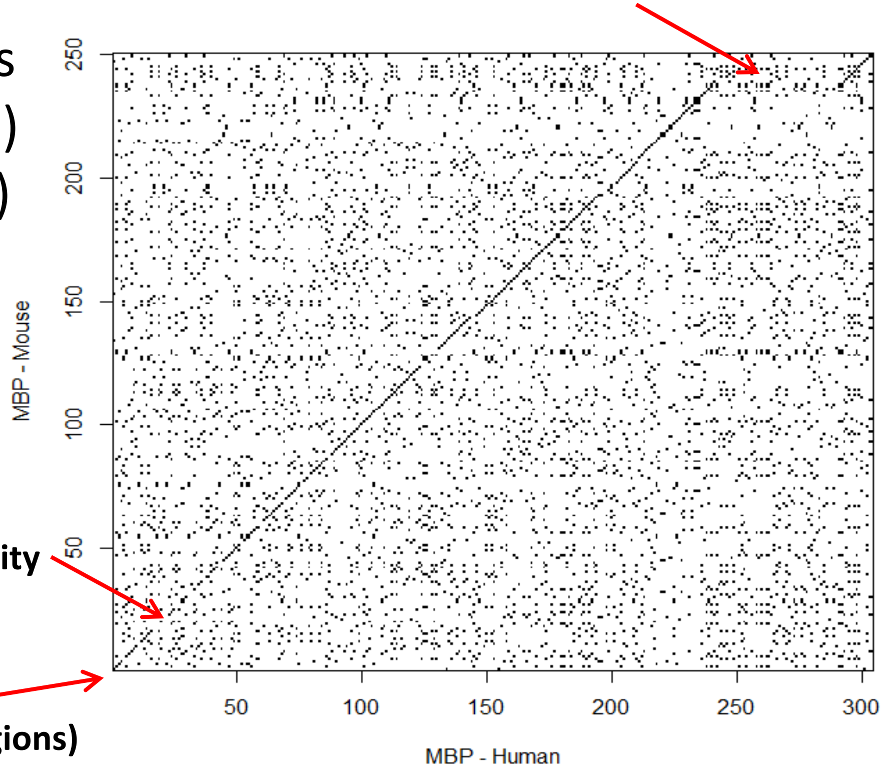
Dot Plot (comparison of 2 sequences) (2of2)

- Let’s compare two protein sequences
  - Human MBP (Uniprot ID: **P02686**)
  - Mouse MBP (Uniprot ID: **P04370**)

- Download 2^{nd} mouse sequence

```
query("MBP_MOUSE", "AC=P04370");
MBP_MOUSE_seq = getSequence(MBP_MOUSE);
```

INSERTION in MBP-Human or GAP in MBP-Mous



Breaks in diagonal line = regions of dissimilarity

Shift in diagonal line (identical regions)

- Visualize dot plot

```
dotPlot(MBP_HUMAN_seq[[1]], MBP_MOUSE_seq[[1]],xlab="MBP - Human", ylab = "MBP - Mouse")
```

- Is there similarity between human and mouse form of MBP protein?
- Where is the difference in the sequence between the two isoforms?

Using R and `Biostrings` library for:

- Pairwise **global** and **local** alignments

Université de Liège

Installing `Biostrings` library

- Install library from Bioconductor

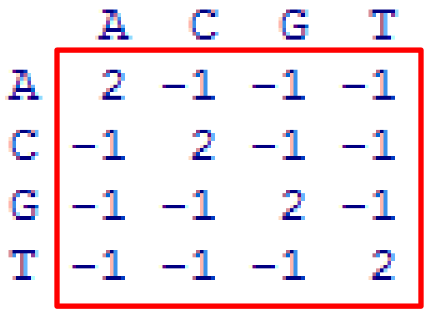
```
source("http://bioconductor.org/biocLite.R")
biocLite("Biostrings")
library(Biostrings)
```

- Define substitution martix (e.g. for DNA)

```
DNA_subst_matrix = nucleotideSubstitutionMatrix(match = 2,
                                  mismatch = -1, baseOnly = TRUE)
```

- The scoring rules
  - Match: $s(a_i, b_j)$ = 2 if $a_i = b_j$
  - Mismatch : $s(a_i, b_j)$= -1 if $a_i \neq b_j$
  - Gap: $s(a_i, -)$ = -2 or $s(-, b_j)$ = -2

`DNA_subst_matrix`



```
   A  C  G  T
A  2 -1 -1 -1
C -1  2 -1 -1
G -1 -1  2 -1
T -1 -1 -1  2
```

Global alignment using R and Biostrings

- Create two sting vectors (i.e. sequences)

```
seqA = "GATTA"
seqB = "GTTA"
```

- Use pairwiseAlignment() and the defined rules

```
globalAlignAB = pairwiseAlignment(seqA, seqB,
    substitutionMatrix = DNA_subst_matrix, gapOpening = -2,
               scoreOnly = FALSE, type="global")
```

- Visualize best paths (i.e. alignments)

```
globalAlignAB

Global PairwiseAlignedFixedSubject (1 of 1)
pattern: [1] GATTA
subject: [1] G-TTA
score: 2
```

Local alignment using R and `Biostrings`

- Input two sequences

```
seqA = "AGGATTTTAAAA"
seqB = "TTTT"
```

- The scoring rules will be the same as we used for global alignment

```
globalAlignAB = pairwiseAlignment(seqA, seqB,
     substitutionMatrix = DNA_subst_matrix, gapOpening = -2,
              scoreOnly = FALSE, type="local")
```

- Visualize alignment

```
globalAlignAB
Local PairwiseAlignedFixedSubject (1 of 1)
pattern: [5] TTTT
subject: [1] TTTT
score: 8
```

Aligning protein sequences

- Protein sequences alignments are very similar except the substitution matrix is specified

```
data(BLOSUM62)
BLOSUM62
```

- Will align sequences

```
seqA = "PAWHEAE"
seqB = "HEAGAWGHEE"
```

- Execute the global alignment

```
globalAlignAB <- pairwiseAlignment(seqA, seqB,
    substitutionMatrix = "BLOSUM62", gapOpening = -2,
            gapExtension = -8, scoreOnly = FALSE)
```

Summary

- We had touched on practical aspects of
  - Global and local alignments
- Thoroughly understood both algorithms
- Applied them both on DNA and protein seq.
- Learned on how to retrieve sequence data
- Learned on how to retrieve sequences both with R and using UniProt
- Learned how to align sequences using R

Resources

- Online Tutorial on Sequence Alignment
  - http://a-little-book-of-r-for-bioinformatics.readthedocs.org/en/latest/src/chapter4.html
- Graphical alignment of proteins
  - http://www.itu.dk/~sestoft/bsa/graphalign.html
- Pairwise alignment of DNA and proteins using your rules:
  - http://www.bioinformatics.org/sms2/pairwise_align_dna.html
- Documentation on libraries
  - Biostings: http://www.bioconductor.org/packages/2.10/bioc/manuals/Biostrings/man/Biostrings.pdf
  - SeqinR: http://seqinr.r-forge.r-project.org/seqinr_2_0-7.pdf

Homework – HW2

Homework 2 – literature style (type 1)

You are asked to **analyze critically** by writing a report and **present** one of the following papers in a group:

1. *Day-Williams AG, Zeggini E **The effect of next-generation sequencing technology on complex trait research**. Eur J Clin Invest. 2011 May;41(5):561-7*
   - **A review paper on popular NGS under the context of genetics of complex diseases**
2. *Do R, **Exome sequencing and complex disease: practical aspects of rare variant association studies**. Hum Mol Genet. 2012 Oct 15;21(R1):R1-9*
   - **A more technical paper on how deep sequencing can help in association studies of rare variants to disease phenotypes under context of statistical genetics**
3. *Hurd PJ, Nelson CJ. **Advantages of next-generation sequencing versus the microarray in epigenetic research**. Brief Funct Genomic Proteomic. 2009 May;8(3):174-83*
   - **An overview paper describing on how NGS technology can be used in the context of epigenetic research. NGS technology described in detail**
4. *Goldstein DB. **Sequencing studies in human genetics: design and interpretation. Nat Rev Genet**. 2013 Jul;14(7):460-70* **(password protected)**
   - **This paper describes on how NGS could be interpreted and contrasted to GWAS. The paper focuses on functional interpretation of genetic variants found in the data**

Homework 2 – computer style (type 2)

- You would implement the Needleman–Wunsch global alignment algorithm in R
  - Follow the pseudo-code provided
  - Will translate it into R
  - Will understand alignment in-depth
  - Provide copy of your code and write a short report
    - Report should contain information on scoring matrix and rules used
    - Example sequences used for alignment
    - In code use comments (# comment)

Homework 2 – Q&A style (type 3)

- Here you would need to answer questions
  - Complete the local and global alignment of DNA and protein sequences graphically
  - Use seqinR library to retrieve protein sequences
  - Use Biostrings library to do alignment of sequences
  - Complete missing R code
  - Copy output from R as a proof
  - Calculate alignment scores

Feedback on HW1

HW 1a feedback

- Some almost confused the name of **the disease abbreviation** with the **disease associated genes** (e.g. HDL syndromes has no HDL1 gene but PRNP gene is associated with HDL1)
- Some printed the whole genome sequence **around** the disease gene, but your were asked to print only the **protein coding region (CDS)**
- Would be nice to get more screen snapshots and see the search query used to find articles
  - From HW1a: “Provide below the search **key words** used to obtain the results”

HW 2b feedback

- Computer style (type 2):
  - Good analysis on gene level with literature searches
  - Could of addressed results variation before and after cleaning data. What is overlap in results before and after QC?
  - Would be nice to have top 10 SNPs and corresponding p-values **before** and **after** cleaning
  - Overall, well done
- Q&A style (type 2)
  - The issue of loading *.phe and *.raw files
    - Set working directory in R where these files are located via
      - **setwd()**
    - Check current location by **getwd()**