

- **Uploading your assignment as a PDF is mandatory. The system will only accept a single PDF file and won't accept anything else or more than one file.**
- Find and save sequences for NM_000558.3 and AF230076.1 from NCBI.
- Select one of the dot-matrix analysis programs provided below and perform analysis for the given sequences.
 - <http://myhits.isb-sib.ch/cgi-bin/dotlet>
 - <https://www.bioinformatics.nl/emboss-explorer/>
- 1) Change the tool parameters such as the threshold and window size to get different results (not all tools might have the same parameters - change whatever parameter the tool you use has). Explain the differences you observe.
- 2) Submit your dotplot image results, the name of the tool, and the parameters you've used. Pick one of the images as the best one. Shortly explain why do you think it is the best result.

Nucleotide

Nucleotide

Advanced

Search

Help



COVID-19 Information



[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

FASTA

Send to:

Change region shown

Customize view

Homo sapiens alpha-2-globin (HBA2) gene, complete cds

GenBank: AF230076.1

[GenBank](#) [Graphics](#)

>AF230076.1 Homo sapiens alpha-2-globin (HBA2) gene, complete cds
CGCCCCGGCCGGGCGTCCCCCGCGCCCCAAGCATAAACCTGGCGCGCTCGCGGCCCGGCACTCTTCTGG
TCCCCACAGACTCAGAGAGAACCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCCT
GGGGTAAGGTTCGGCGCGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAAAGTGAGGCTCCCTCCCCTG
CTCCGACCCGGGCTCCTCGCCCGCCCGGACCCACAGGCCACCCTCAACCGTCTGGCCCCGGACCCAAAC
CCCACCCCTCACTCTGCTTCTCCCCGAGGATGTTCTGTCTTCCCCACCACCAAGACCTACTTCCCGC
ACTTCGACCTGAGCCACGGCTCTGCCAGGTTAAGGGCCACGGCAAGAAGGTGGCCGACGCGCTGACCAA
CGCCGTGGCGCAGTGGACGACATGCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAGCTT
CGGGTGGACCCGGTCAACTTCAAGGTGAGCGGCGGGCCGGGAGCGATCTGGGTCGAGGGGCGAGATGGCG
CCTTCTCTCAGGGCAGAGGATCACGCGGGTTGCGGGAGGTGTAGCGCAGGCGGCGGCTGCGGGCCTGGG
CCGCACTGACCCTTCTCTGCACAGCTCCTAAGCCACTGCCTGCTGGTGACCTGGCCGCCACCTCCC
CGCCGAGTTTACCCTCGGGTGACGCGCTCCCTGGACAAAGTTCTGGCTTCTGTGAGCACCCTGCTGACC
TCCAAATACCGTTAAGCTGGAGCCTCGGTAGCCGTTCTCTGCCCGCTGGGCTCCCAACGGGCCCTCC
TCCCCTCCTTGACCGGCCCTTCTGGTCTTTGAATAAAGTCTGAGTGGGCGGACGCTGTGTGTCCTG
GGTCTCTGTGTCGGGAATGTGCCAACAA

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the HBA2 gene

Molecular and Hematological Analysis of Alpha- and Beta-Thal [Genet Test Mol Biomarkers. 2021]

Molecular Characterization and Hematological

Nucleotide

Nucleotide ▾

Search

Advanced

Help



COVID-19 Information



[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

FASTA ▾

Send to: ▾

Change region shown ▾

Customize view ▾

Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA

NCBI Reference Sequence: NM_000558.3

This sequence has been updated. [See current version.](#)

[GenBank](#) [Graphics](#)

>NM_000558.3 Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA
ACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTC
AAGGCCGCTGGGGTAAGGTCGGCGCGCAGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGGATGTTCC
TGTCCTTCCCCACCAAGACCTACTTCCCGCACTTCGACCTGAGCCACGGCTCTGCCAGGTTAAGGG
CCACGGCAAGAAGGTGGCCGACGCGCTGACCAACGCCGTGGCGCACGTGGACGACATGCCAACGCGCTG
TCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGACCCGGTCAACTTCAAGCTCCTAAGCCACT
GCCTGCTGGTGACCTGGCCGCCACCTCCCGCCGAGTTACCCCTGCGGTGCACGCCTCCCTGGACAA
GTTCTGGCTTCTGTGAGCACCGTGTGACCTCCTAAATACCGTTAAGCTGGAGCCTCGGTGGCCATGCTT
CTTGCCCTTGGGCTCCCCCAGCCCTCCTCCCTTCTGACCCGTACCCCGTGGTCTTTGAATAA
AGTCTGAGTGGGCGGC

Analyze this sequence

[Run BLAST](#)[Pick Primers](#)[Highlight Sequence Features](#)[Find in this Sequence](#)

Recent activity

[Turn Off](#) [Clear](#)

Homo sapiens hemoglobin, alpha 1 (HBA1),
mRNA

Nucleotide

<http://myhits.isb-sib.ch/cgi-bin/dotlet>

Tool: Dotlet JS (*beta*)

There For dotlet JS, there is image contrast adjustment **(1)**, and there are red lines **(2)** to indicate where the sequences coincide in the dot-plot simultaneously (as coordinates).

The parameters used in this study are **window size (3)** and **type of the matrix (4)**.

There is also a bar where the sequence positions can be adjusted manually **(5)**. From here you can replace red lines(2) to indicate where the sequences coincide in the dot-plot simultaneously.

<http://myhits.isb-sib.ch/cgi-bin/dotlet>

Tool: Dotlet JS

SEQUENCE 1

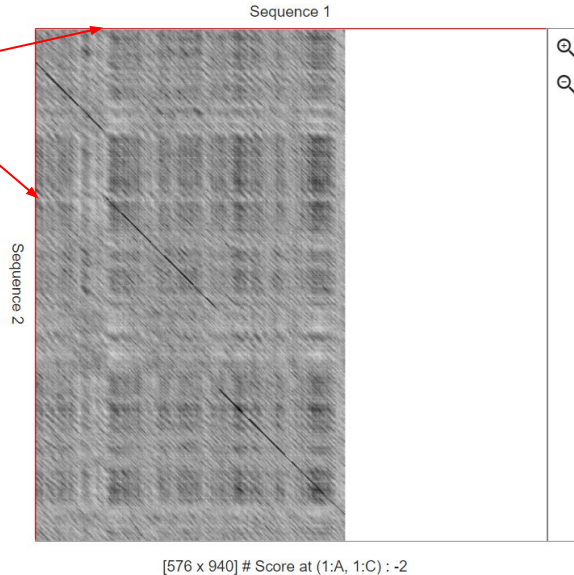
SEQUENCE 2

Window size
15

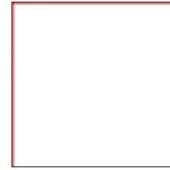
Scoring matrix
BLOSUM 62



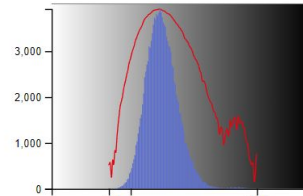
2



[576 x 940] # Score at (1:A, 1:C) : -2



These red lines indicate the position of the red lines of 2.



This graph represents the frequency distribution.

1

Seq1: 326
AFCGACCTGCACGGCACAAGCTTCGGTGGACCCGGTCAACTTCAAGCTCCTAAGCCAATGCCTGCTGTGA
CAGCGGCTGCAGGGCTGGGCCGCACTGACCTTTCTCTGCACA GCTCTAAGCCAATGCTGCTGCTGCTGACCC
Seq2: 648

5

BLOSUM 62: Changes in the window size

BLOSUM 62 matrix, with window size = 15 is the default option of the Dotlet JS.

Here, window size is a parameter that indicates to compare a short sequence segment with all possible segments of the same size in the reference sequence [1]. If the result obtained based on this evaluation exceeds a certain threshold, this relationship can be seen in the dot plot [2].

In order to see the effect of the window size parameter, four different window sizes have been tried in addition to the default window size(15) in the BLOSUM 62 matrix.

The resolution is changed whenever the window size changes, because each window size is different from the sequence comparison range. Every time the window size is changed, the frequency distributions are also changed (red circles at the bottom in the next images). Since comparison could not be made in areas where the window size is 30 or greater than 30, although dot plot images were created (left), frequency graphics could not be generated.

SEQUENCE 1

SEQUENCE 2

Window size

3

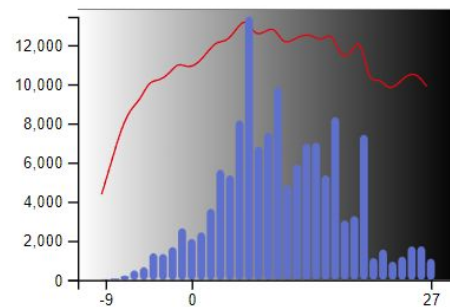
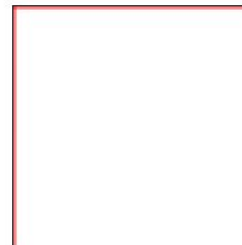
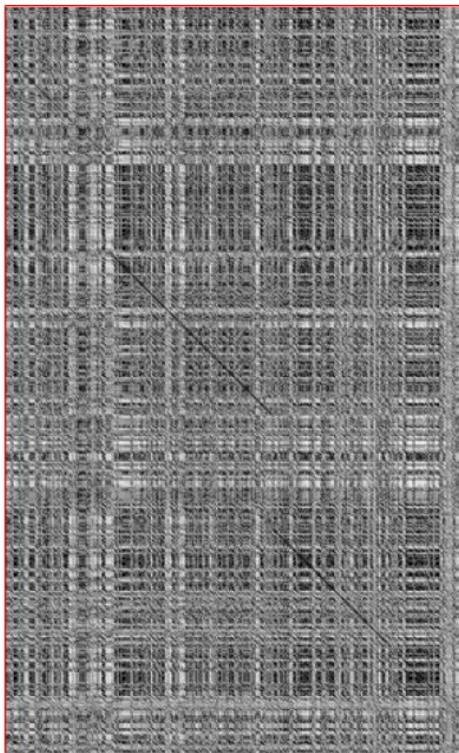
Scoring matrix

BLOSUM 62



Sequence 1

Sequence 2



[576 x 940] # Score at (1:A, 1:C) : -3

Default:

SEQUENCE 1

SEQUENCE 2

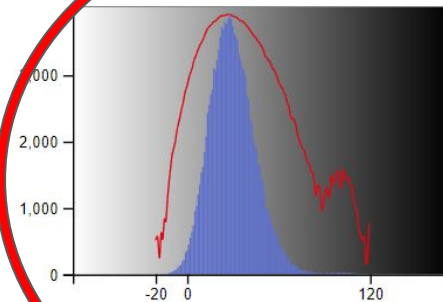
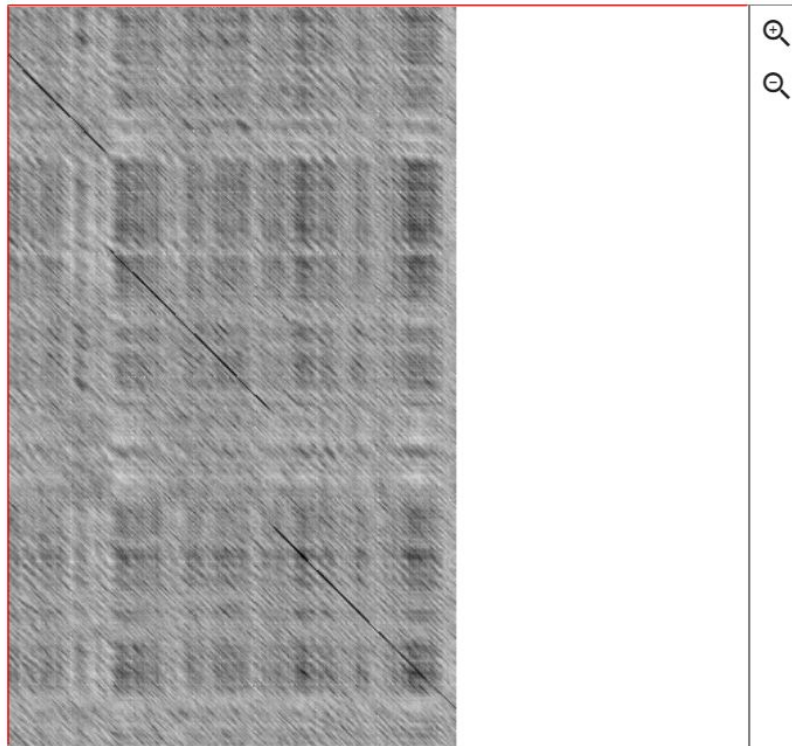
Window size
15

Scoring matrix
BLOSUM 62



Sequence 1

Sequence 2



[576 x 940] # Score at (1:A, 1:C) : -2

SEQUENCE 1

SEQUENCE 2

Window size

20

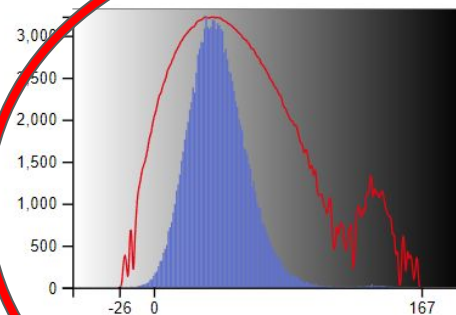
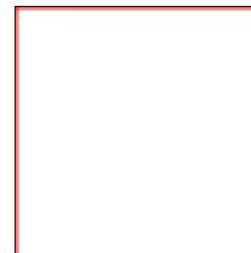
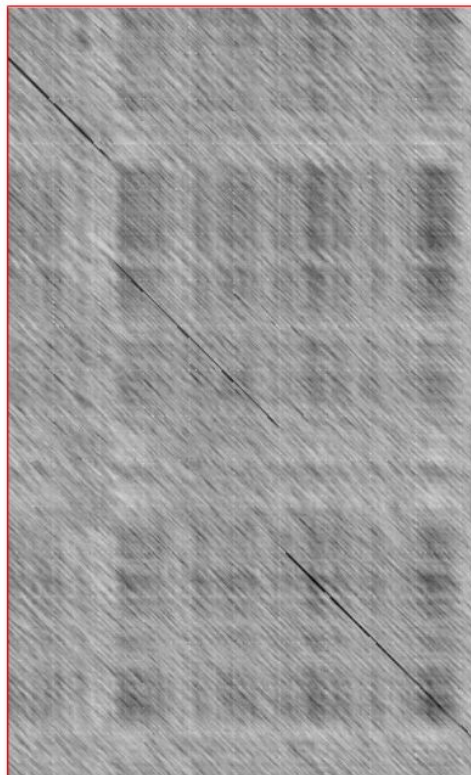
Scoring matrix

BLOSUM 62



Sequence 1

Sequence 2



[576 x 940] # Score at (1:A, 1:C) : -1

SEQUENCE 1

SEQUENCE 2

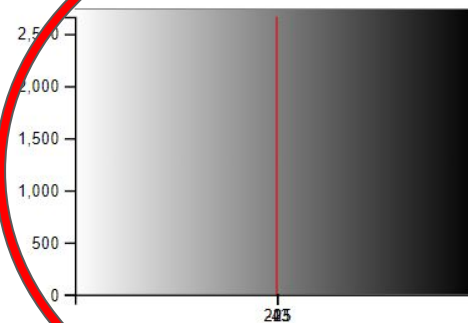
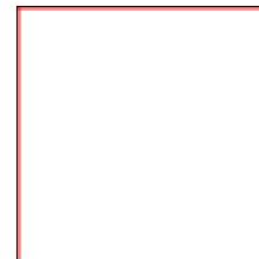
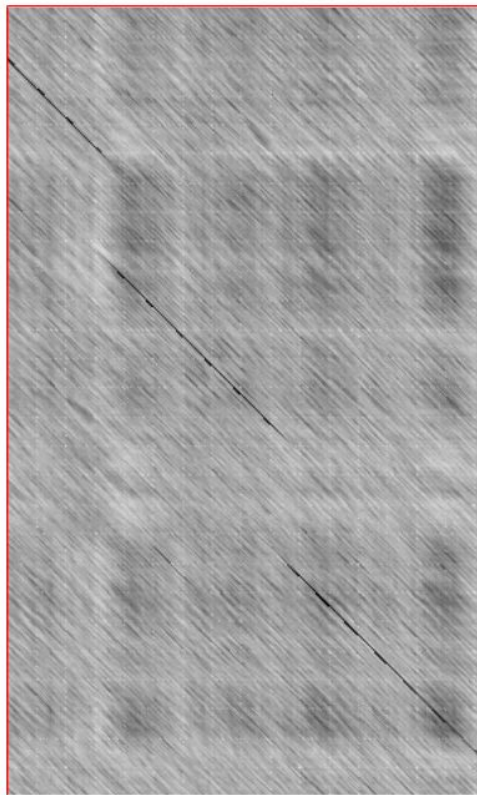
Window size

30

Scoring matrix
BLOSUM 62



Sequence 1



[576 x 940] # Score at (1:A, 1:C) : 1

SEQUENCE 1

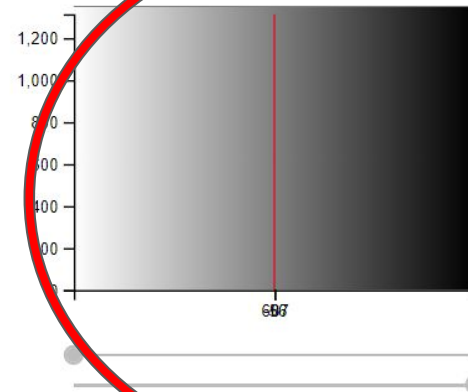
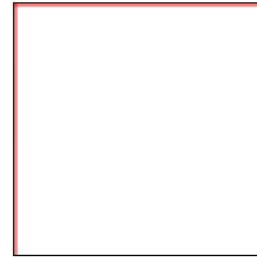
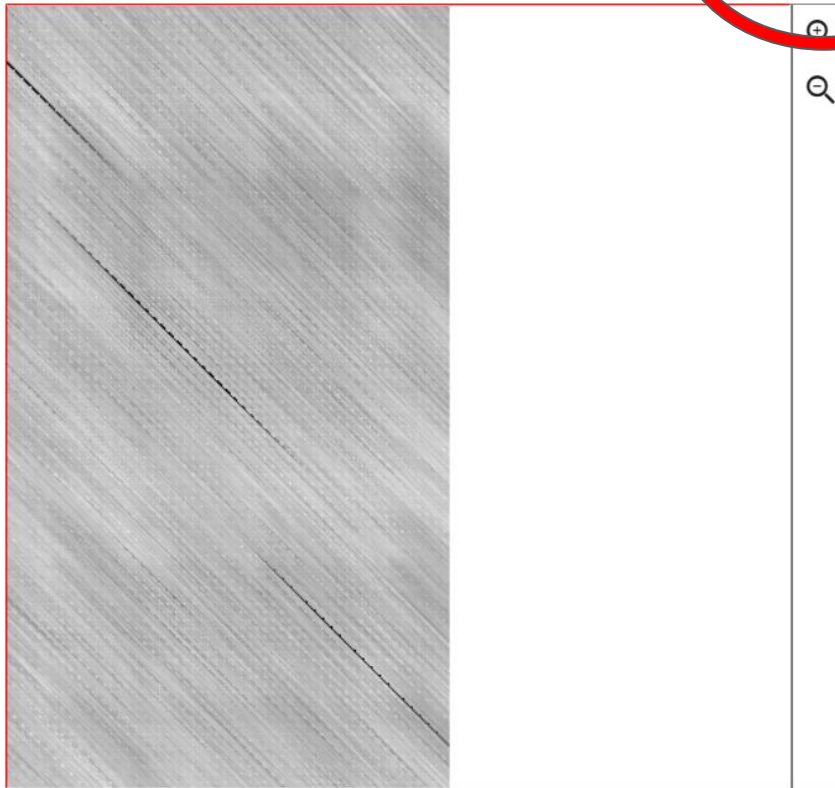
SEQUENCE 2

Window size
100

Scoring matrix
BLOSUM 62



Sequence 1



[576 x 940] # Score at (1:A, 1:C) : 53

Different Matrices with same window size (= 15)

In addition to the BLOSUM 62 matrix-which is the default matrix option-, Dotlet JS offers five more matrix options. The way that each matrix handles data is different, but the results are similar. For example, while identity matrix represents the pairwise sequence identity as 0-1 [3], PAM matrices are calculated as log-odds ratios [4], and BLOSUM matrices are a deep-scoring matrices like PAM but can align more closely related sequences [4] .

The main difference between the results obtained by the matrices is the difference between the distribution ranges in the matches of the sequences. The main reason for this is that the ranges handled by the matrices are different from each other, as mentioned above, in the calculation dimension. The frequency graphs for which this is represented are indicated by the red circles:

Identity Matrix:

SEQUENCE 1

SEQUENCE 2

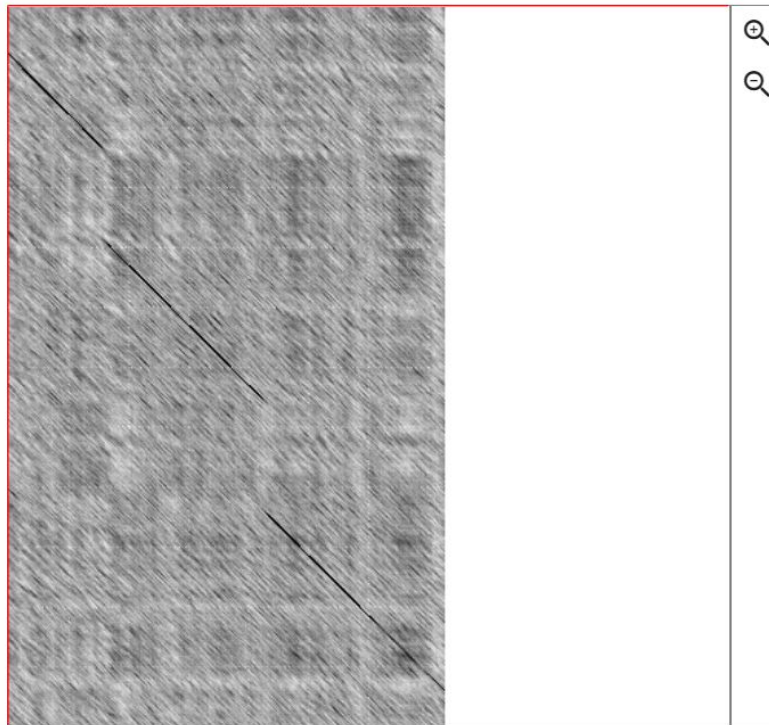
Window size
15

Scoring matrix
Identity

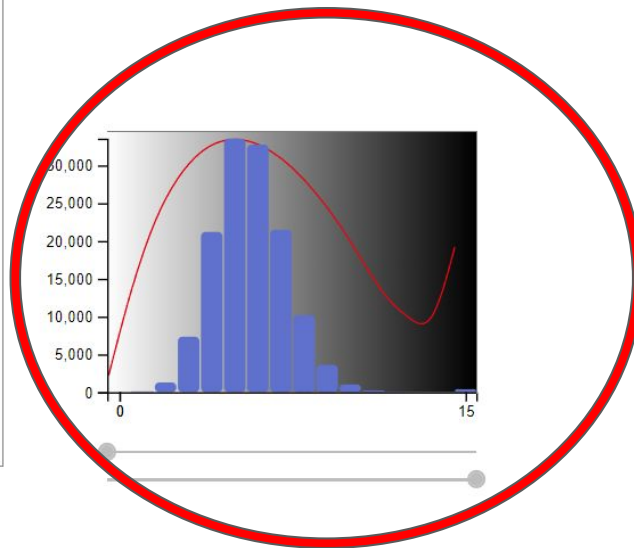
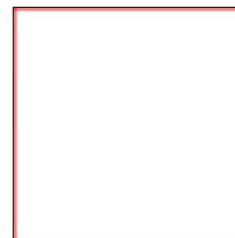


Sequence 1

Sequence 2



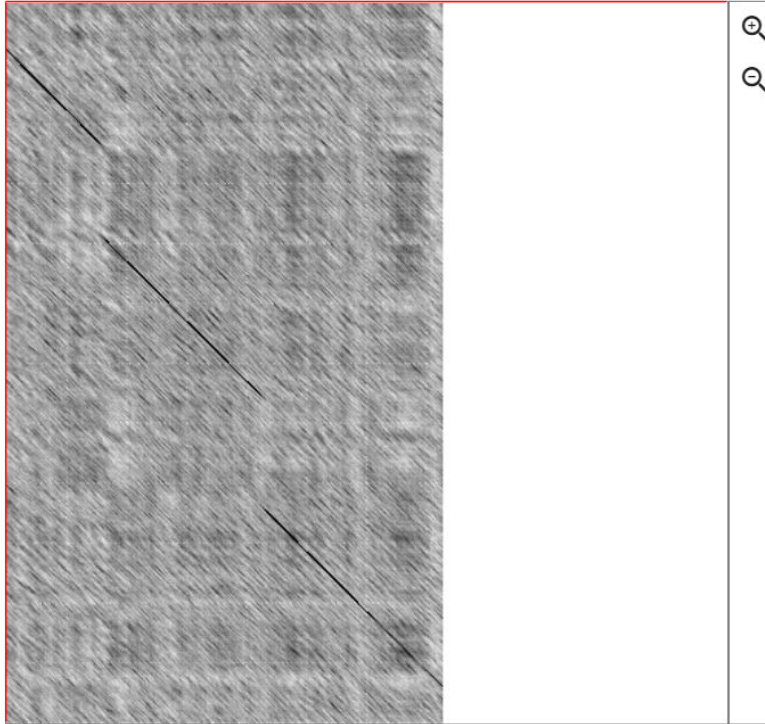
[576 x 940] # Score at (1:A, 1:C) : 1



BLOSUM 45:

Sequence 2

Sequence 1



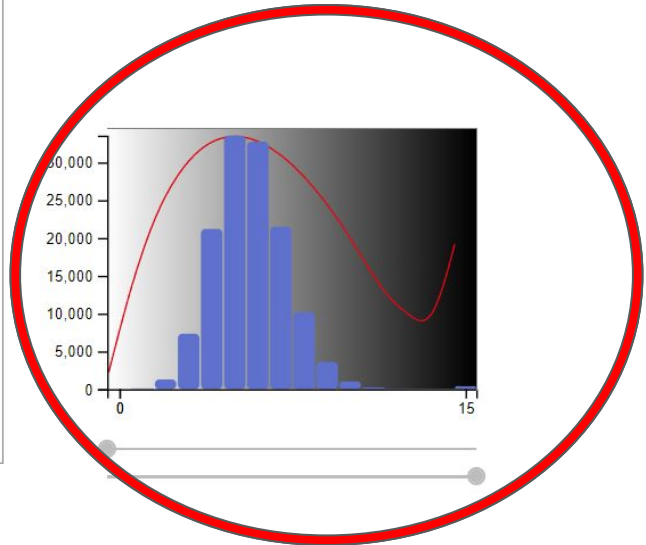
[576 x 940] # Score at (1:A, 1:C) : 1

SEQUENCE 1

SEQUENCE 2

Window size
15

Scoring matrix
BLOSUM 45



Default:

SEQUENCE 1

SEQUENCE 2

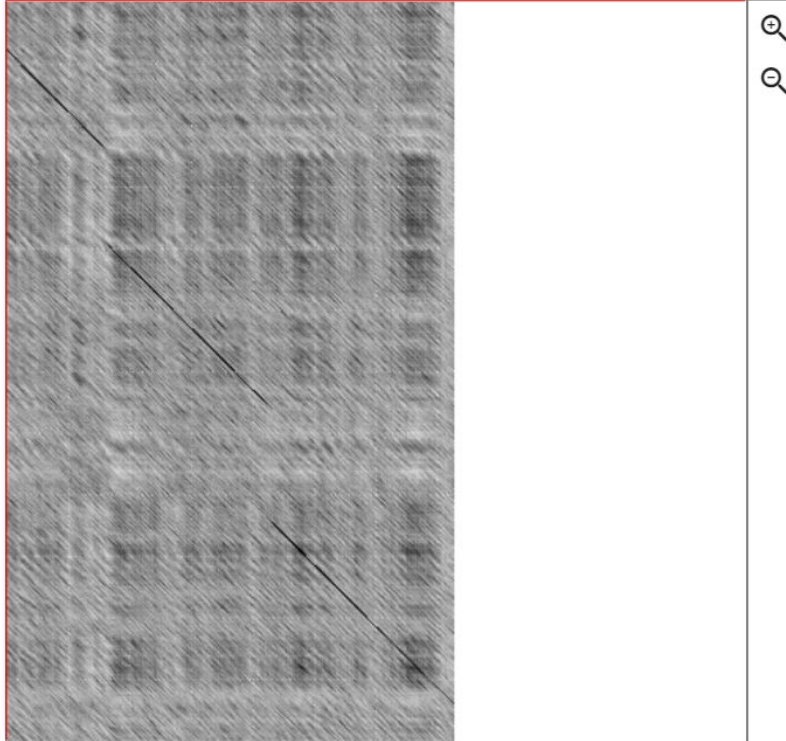
Window size
15

Scoring matrix
BLOSUM 62

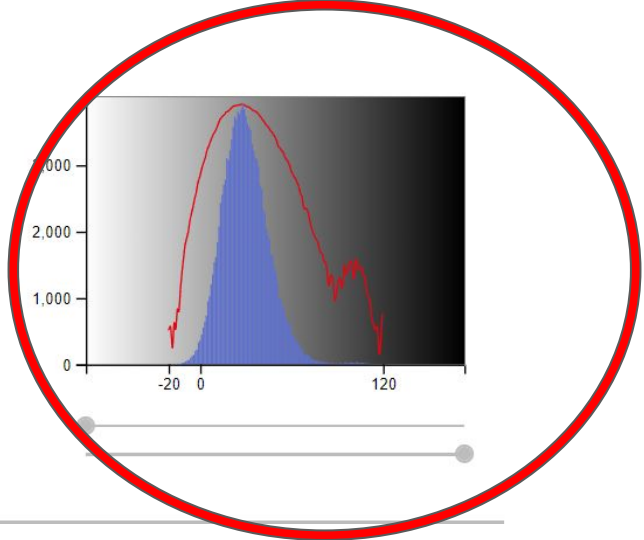
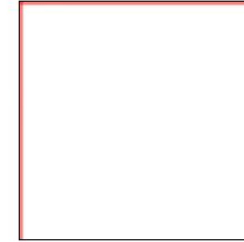


Sequence 1

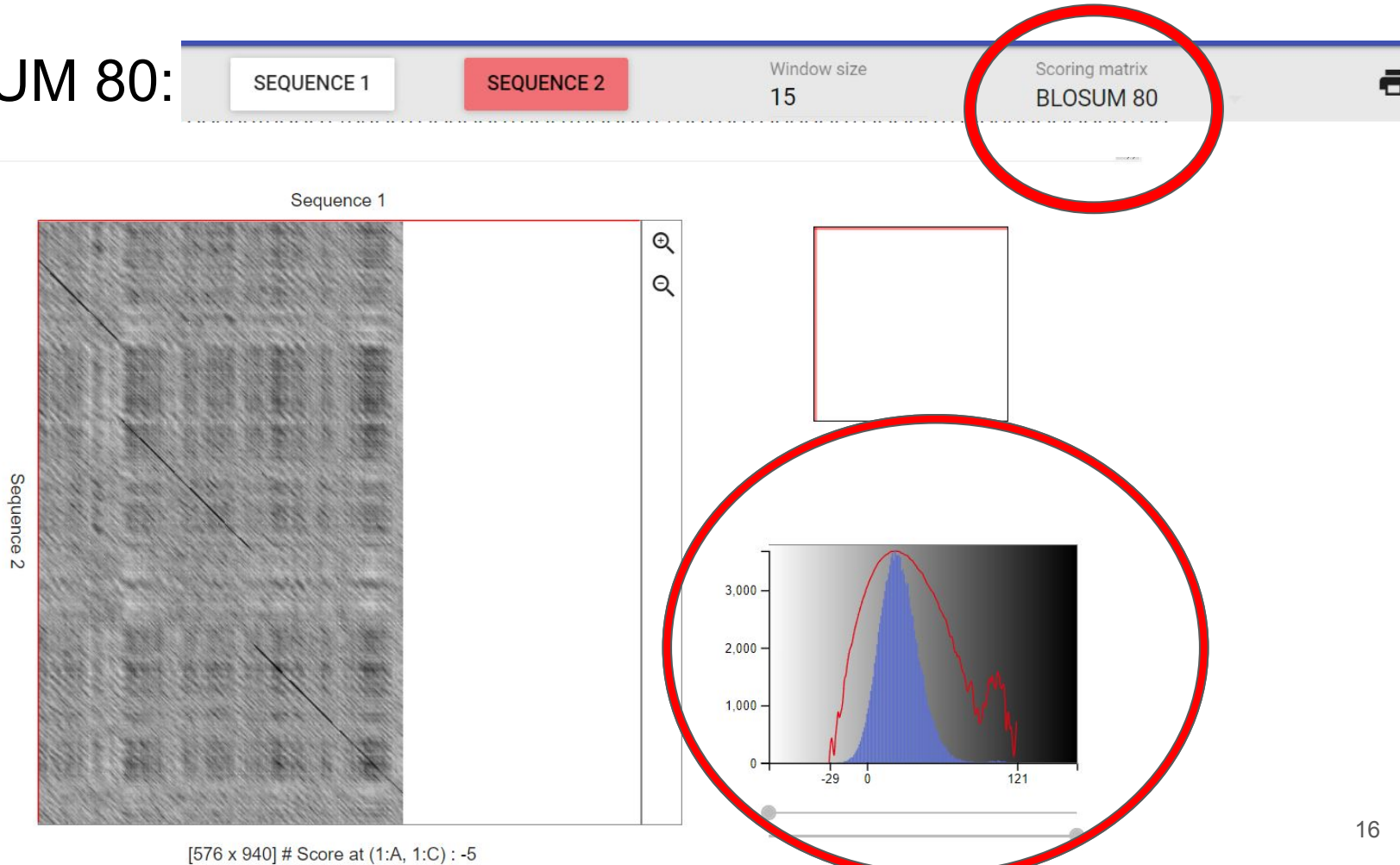
Sequence 2




[576 x 940] # Score at (1:A, 1:C) : -2

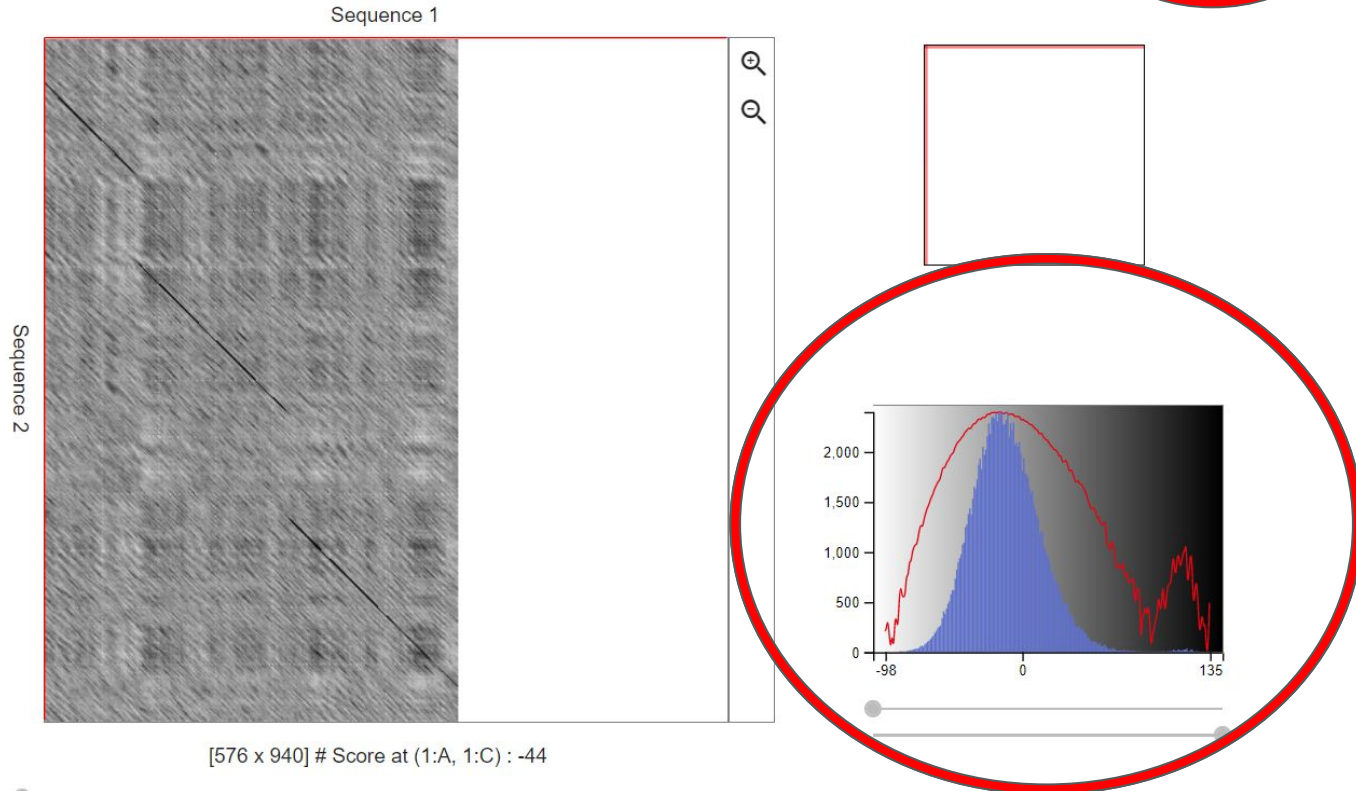


BLOSUM 80:



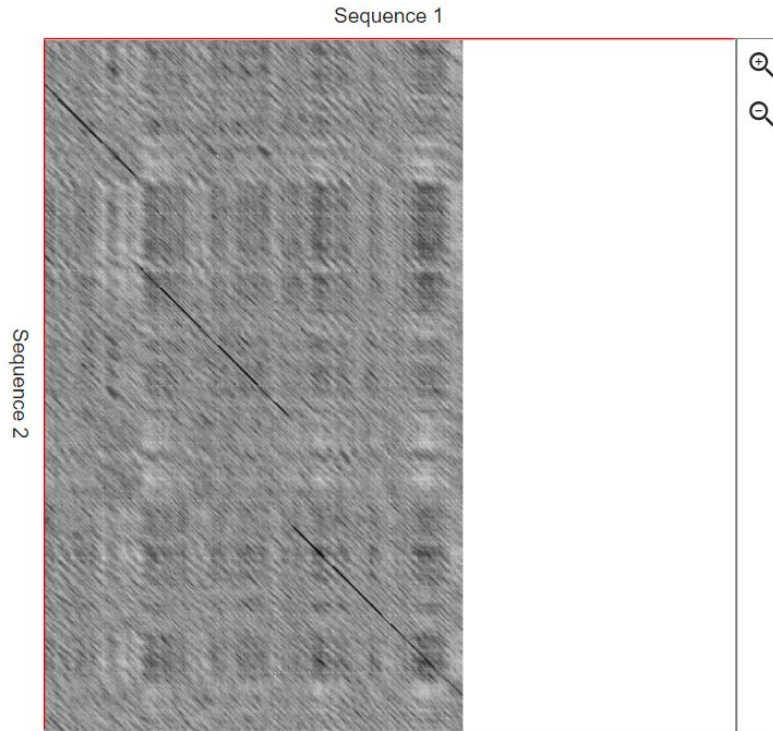
PAM 30:

SEQUENCE 1 **SEQUENCE 2** Window size 15 Scoring matrix **PAM 30** 

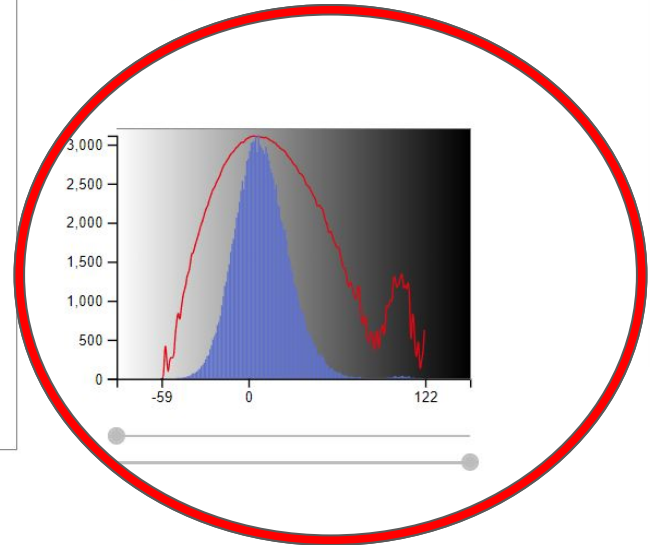


PAM 70:

SEQUENCE 1 SEQUENCE 2 Window size 15 Scoring matrix PAM 70



[576 x 940] # Score at (1:A, 1:C) : -25



The Best Image : BLOSUM 62 (default option)

The default option BLOSUM 45 gave the best image. This is likely due to BLOSUM's success in evaluating evolutionary related or divergent sequences. Because the two given sequences are evolutionarily close to each other, Homo sapiens alpha-2-globin (HBA2) (gene) and Homo sapiens hemoglobin, alpha 1 (HBA1), (mRNA), the coding sequences of these two genes are identical and they differ from one another with very minor mutations [5].

The similarities of these genes can be seen by moving the red lines:

Default:

SEQUENCE 1

SEQUENCE 2

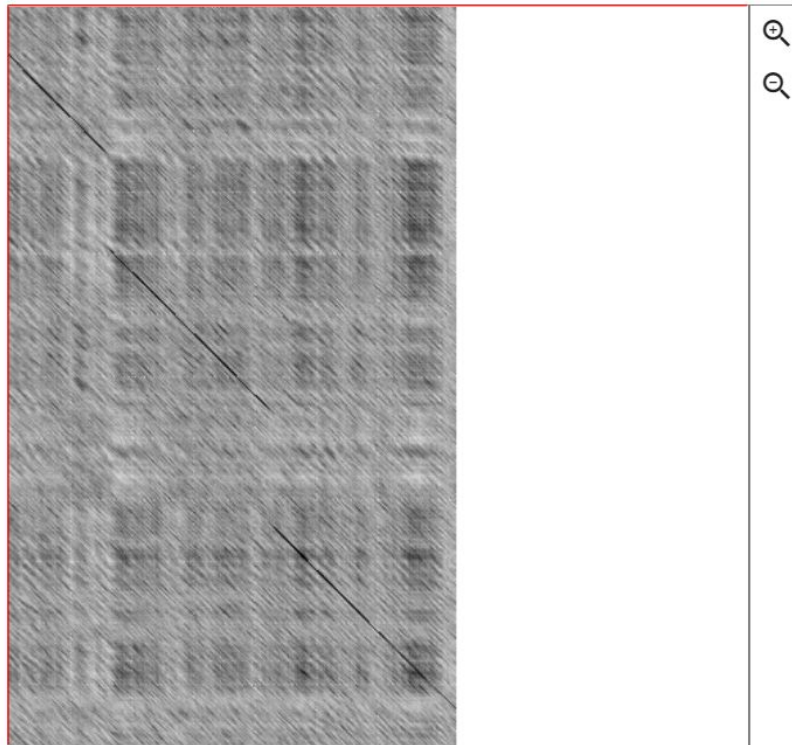
Window size
15

Scoring matrix
BLOSUM 62

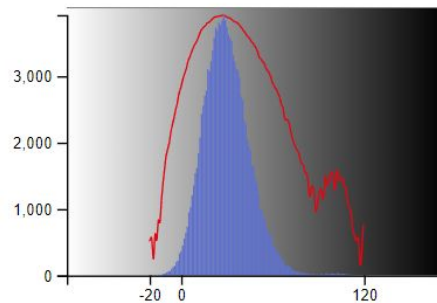
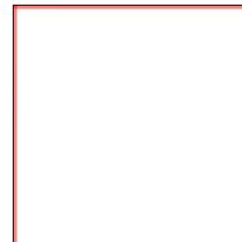


Sequence 1

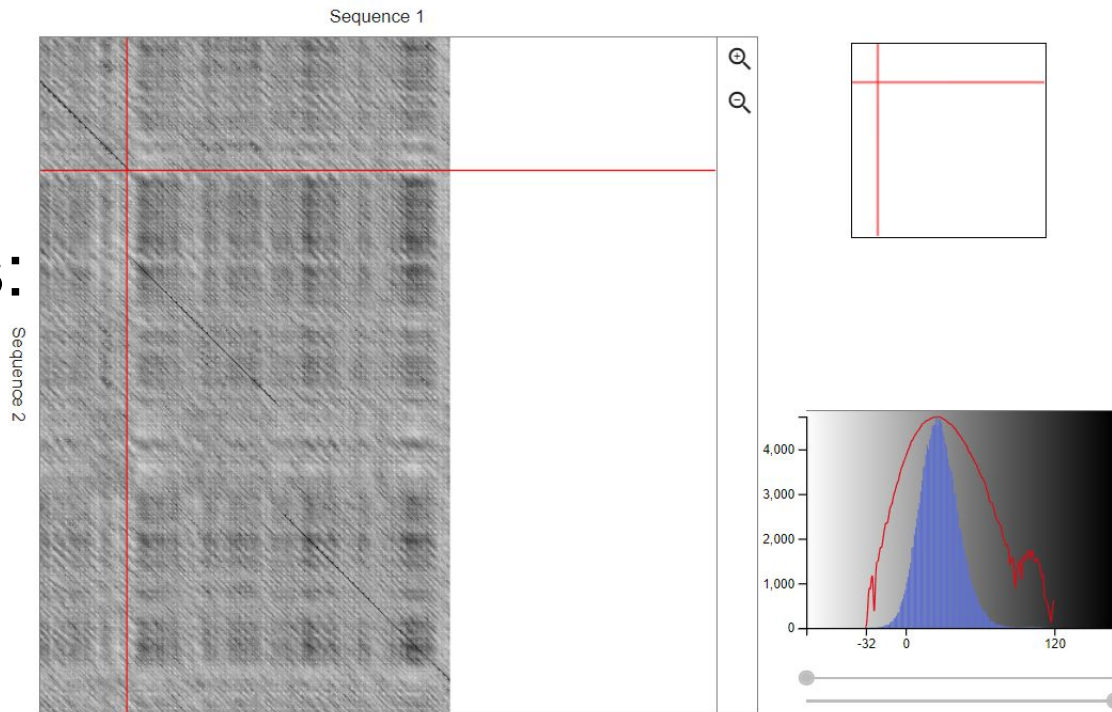
Sequence 2



[576 x 940] # Score at (1:A, 1:C) : -2

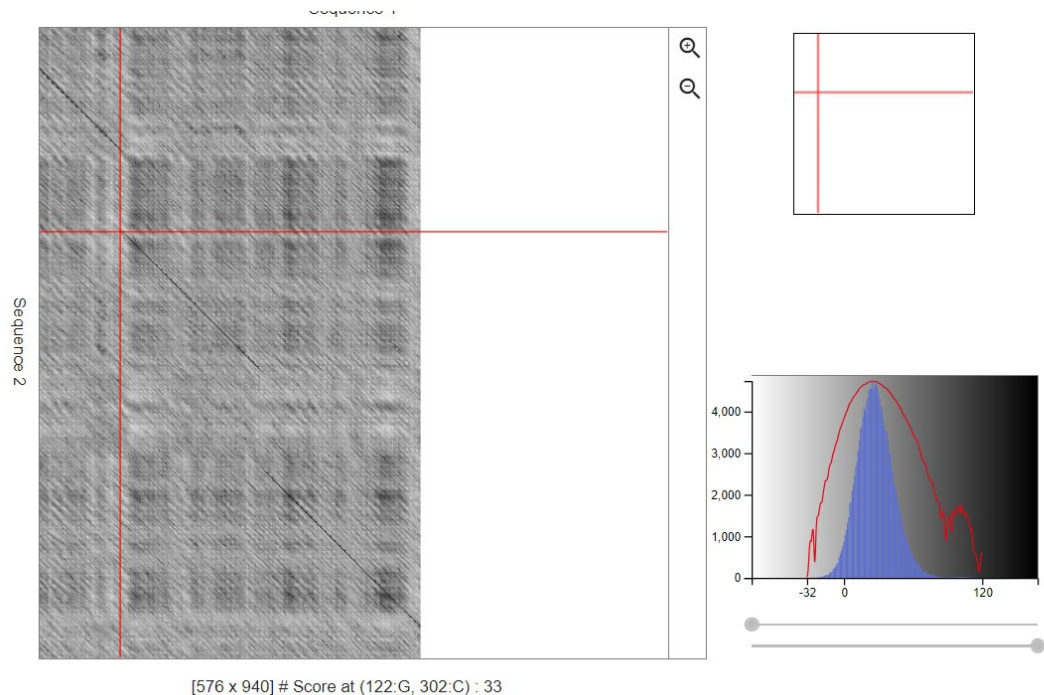


Changes of the sequence intersection lines:



AAGGTCG **GCGC**GCACGCTGGCAGTATGGT **GCGG**AGGCCCTGGAAGAG **GATGTTCC**TGTCC **TCCCC**ACCACCA
 GGTCGGC **GCGC**ACGCTGGCAGTATGGT **GCGG**AGGCCCTGGAGAAGT **GAGCTCC**CTCGCTCCGACCCG

Changes of the sequence intersection lines:

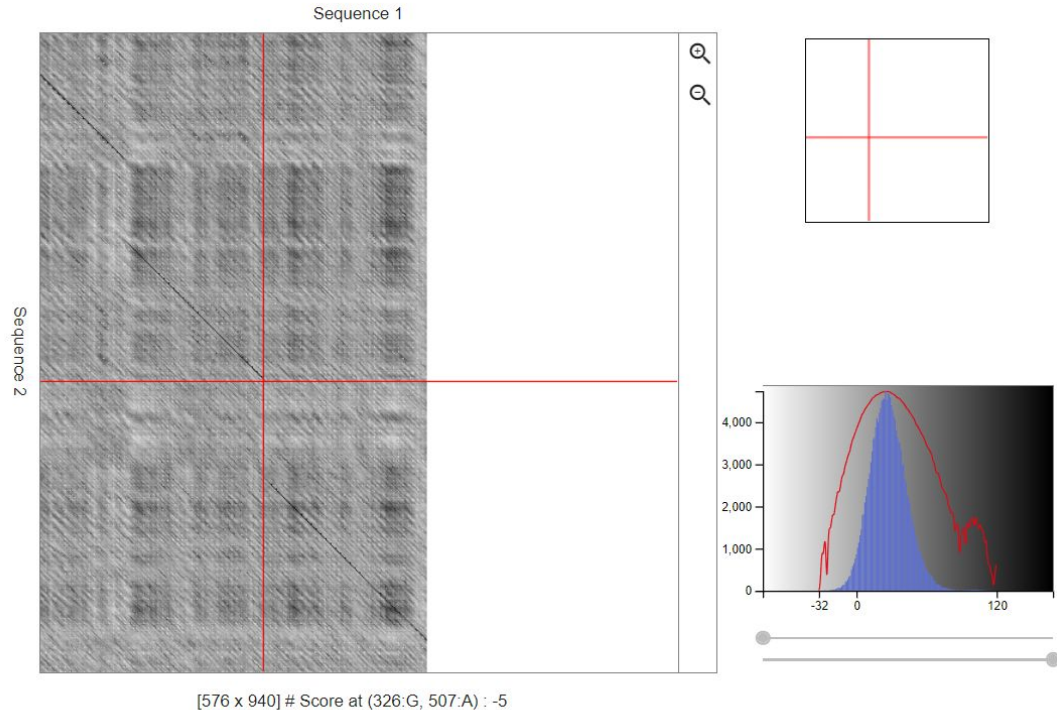


Seq1:122

AAGGTCGGCGCGCAGCTGGCGAGTATGGTGC GGAGGCCCTGGAGAGGATGTCCTGTCCTTCCCACCACCA
 GCCCGGACCCAAACCCACCCCTCACTCTGCTTCTCCCCGCAGGATGTCCTGTCCTTCCCACCACCAAGA

Seq2:302

Changes of the sequence intersection lines:

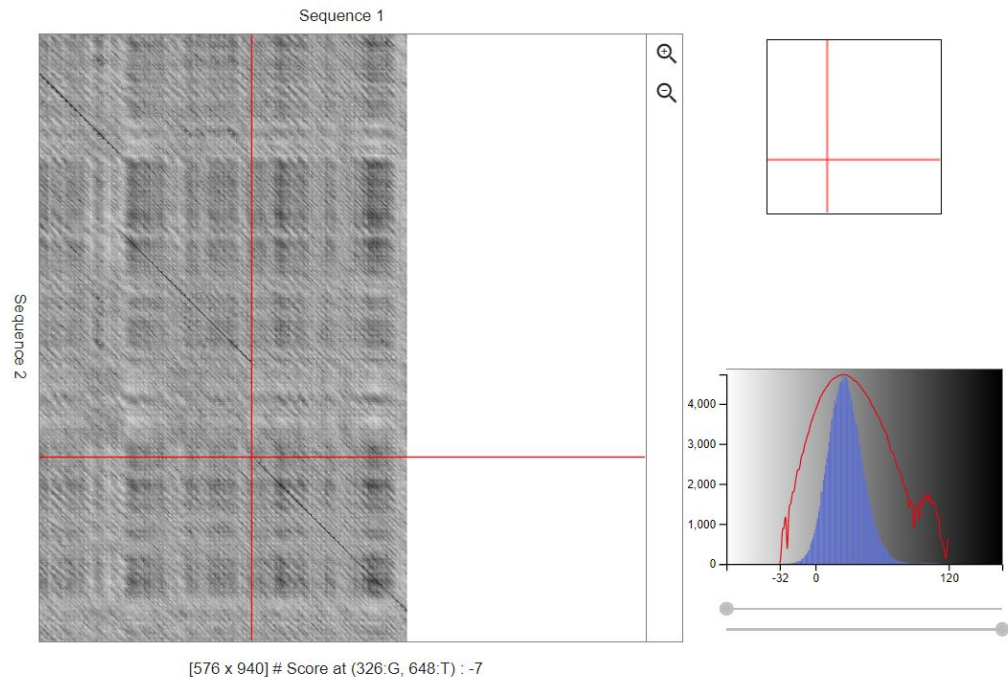


Seq1: 326

AGCGACCTGCACGCGCACAAAGCTTCGGGTGGACCCGGTCAACTTCAAGCTCCTAAGCCACTGCCTGCTGGTGA
 ACCTGCACGCGCACAAAGCTTCGGGTGGACCCGGTCAACTTCAAGGTGAGCGGCGGGCCGGGAGCGATCTGGT

Seq2: 507

Changes of the sequence intersection lines:



Seq1: 326

AGCGACCTGCACGGCACAAGCTTCGGGTGGAGCCGGTCAACTTCAAGCTCCTAAGCACTGCCTGCTGGTGA

CGCGGGCTGCGGGCTGGGCCGCACTGACCTCTTCTCTGCACAGCTCTCAAGCCACTGCCTGCTGGTGACCC

Seq2: 648

References

- [1]: <https://genomeevolution.org/wiki/index.php/Dotplot>
- [2]: <http://petang.cgu.edu.tw/Bioinfomatics/MANUALS/Genehelp/compareparam.html>
- [3]: Campanella, J.J., Bitincka, L. & Smalley, J. MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics* 4, 29 (2003). <https://doi.org/10.1186/1471-2105-4-29>
- [4]: Pearson W. R. (2013). Selecting the Right Similarity-Scoring Matrix. *Current protocols in bioinformatics*, 43, 3.5.1–3.5.9. <https://doi.org/10.1002/0471250953.bi0305s43>
- [5] <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=3039>