

Global Alignment using Needle

The screenshot displays the EMBL-EBI Job Dispatcher interface for a Needle global alignment job. The browser tabs show '3-DemoEMBOSS.mp4', 'EMBOSS Needle < EMBL-EBI', and 'EMBOSS Water < EMBL-EBI'. The address bar shows the URL: ebi.ac.uk/jdispatcher/psa/emboss_needle/summary?jobid=emboss_needle-I20240913-054055-0806-54719162-p1m&js=pass.

At the top, a yellow banner reads: 'Welcome to the new Job Dispatcher website. We'd love to hear your feedback about the new webpage!'. Below this, the 'Results for Job ID' section shows the job ID: `emboss_needle-I20240913-054055-0806-54719162-p1m`. There are 'Copy' and 'Resubmission' buttons.

The interface has three tabs: 'Tool Output' (selected), 'Result Files', and 'Submission Details'. Under the 'Tool Output' tab, there is a 'Download' button and a 'Tool output' section containing the following text:

```
#####
# Program: needle
# Runday: Fri 13 Sep 2024 09:41:00
# Command line: needle
#
# -auto
#
# -stdout
#
# -sequence emboss_needle-I20240913-054055-0806-54719162-p1m.sequence
# -sequence emboss_needle-I20240913-054055-0806-54719162-p1m.sequence
# -matrix: DYNAMIC2
# -gapopen 10.0
# -gapextend 0.5
# -endopen 10.0
# -endextend 0.5
# -format3 pair
# -spacedata
# -spacedata
# Align format: pair
# Report file: stdout
#####

#
#
# Aligned sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: DYNAMIC2
# Gap penalty: 10.0
# Extend penalty: 0.5
#
#
# Length: 32
# Identity: 30/32 (93.8%)
# Similarity: 30/32 (93.8%)
# Gap: 0/32 (0.0%)
# Score: 100.0
#
#
#####

EMBOS_001 1 ACGTACGTACGTACGTACGTACGTACGT 32
          1111111111111111111111111111111111
EMBOS_001 1 ACGTACGTACGTACGTACGTACGTACGT 32

#-----
#-----
```

At the bottom, there is a disclaimer: 'If you use this service, please consider citing the following publication: The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. More information about this bioinformatics application can be found in its [bio.tools](#) record. Please read the provided [Help & Privacy](#) before seeking help from our support staff. If you have any feedback or experienced any issues please let us know via [EMBL-EBI Support](#). Read our [Privacy Notice](#) if you are concerned with your privacy and how we handle personal information.'

The footer of the page states: 'EMBL-EBI is the home for big data in biology.'

The above alignment was generated using the EMBOSS Needle tool, which performs a global alignment on two DNA sequences. In this case, the sequences ACGTACGTACGTACGTACGTACGTACGTACGTACGT and ACGTACGTAGGTACGTACGTAGGTACGTACGTACGT were aligned. As shown in the results, the identity between the sequences is 30/32 (93.8%), indicating that most of the nucleotides match perfectly between the two sequences. The few differences are highlighted by the gaps introduced in the alignment, which is expected in global alignment as it tries to match the entire sequence length.

Local Alignment Using Water

3-DemoEMBOSS.mp4 x EMBOSS Needle < EMBL-EBI x EMBOSS Water < EMBL-EBI x +

ebi.ac.uk/jdispatcher/psa/embooss_water/summary?jobid=embooss_water-120240913-055909-0100-44168801-p1m&js=pass

EMBL-EBI home Services Research Training About us EMBL-EBI

EMBOSS Water

Pairwise Sequence Alignment (PSA)

Job Dispatcher Help & Privacy Your Jobs Input form [Feedback](#)

Welcome to the new **Job Dispatcher** website. We'd love to hear your [feedback](#) about the new webpage! [X](#)

Results for Job ID **embooss_water-120240913-055909-0100-44168801-p1m** [Copy](#) [Resubmission](#) [X](#)

Tool Output Result Files Submission Details

Tool output [Download](#)

```
#####
# Program: water
# RunDate: Fri 13 Sep 2024 09:59:12
# Command line: water
# -auto
# -stdout
# -sequence embooss_water-120240913-055909-0100-44168801-p1m.a.sequence
# -sequence embooss_water-120240913-055909-0100-44168801-p1m.b.sequence
# -detailfile ERL050m02
# -gapopen 1.0
# -gapextend 0.5
# -format3 pair
# -spstorein0
# -spstorein2
# Align_format: pair
# Report_file: stdout
#####

#
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: ERL050m02
# Gap_penalty: 1.0
# Extend_penalty: 0.5
#
# Length: 34
# Identity: 30/34 (88.2%)
# Similarity: 30/34 (88.2%)
# Gaps: 4/34 (11.8%)
# Score: 170.0
#
#
#####
EMBOSS_001 1 ACCTACTAC-CTACTACTAC-CTACTACTACT 32
            ||||||||| ||||||||| |||||||||
EMBOSS_001 1 ACCTACTACTA-GCTACTACTACTA-GCTACTACTACT 32

#
#
#
#####
```

3-DemoEMBOSS.mp4 x EMBOSS Needle < EMBL-EBI x EMBOSS Water < EMBL-EBI x +

ebi.ac.uk/jdispatcher/psa/embooss_water/summary?jobid=embooss_water-120240913-060144-0580-22567665-p1m&js=pass

EMBL-EBI home Services Research Training About us EMBL-EBI

EMBOSS Water

Pairwise Sequence Alignment (PSA)

Job Dispatcher Help & Privacy Your Jobs Input form [Feedback](#)

Welcome to the new **Job Dispatcher** website. We'd love to hear your [feedback](#) about the new webpage! [X](#)

Results for Job ID **embooss_water-120240913-060144-0580-22567665-p1m** [Copy](#) [Resubmission](#) [X](#)

Tool Output Result Files Submission Details

Tool output [Download](#)

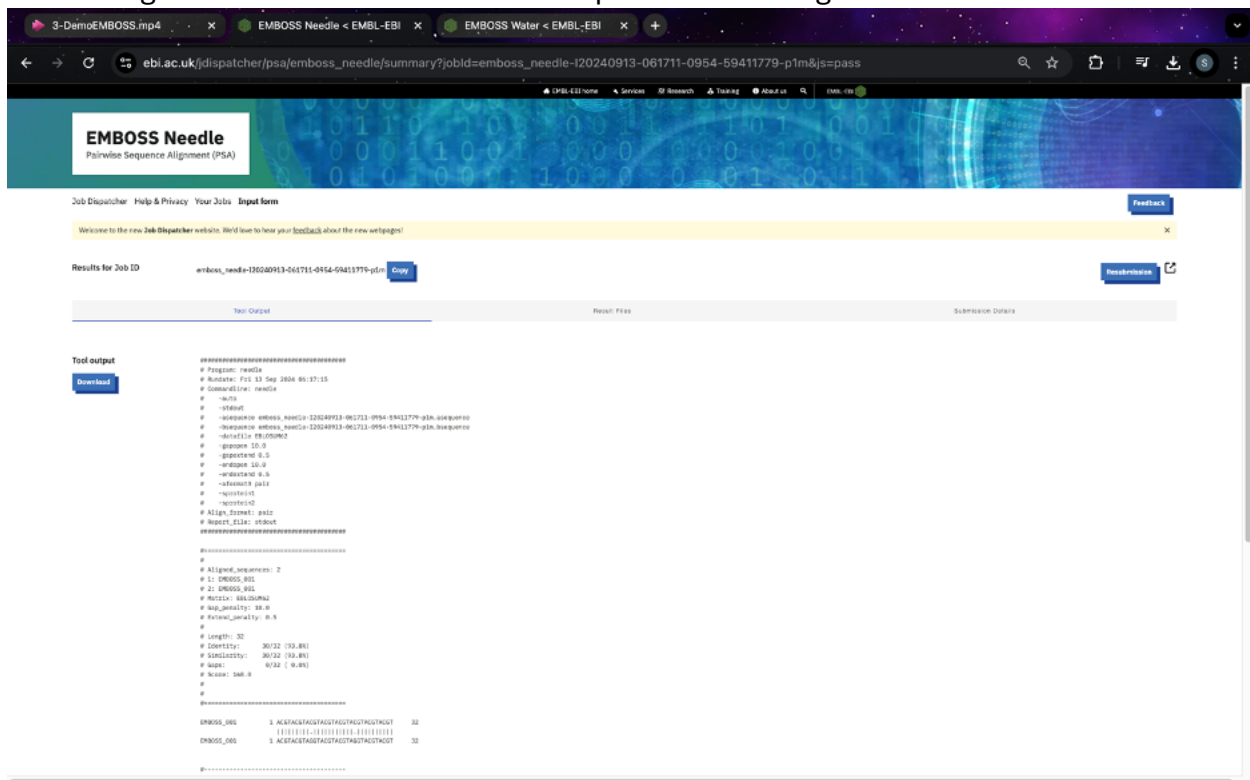
```
#####
# Program: water
# RunDate: Fri 13 Sep 2024 09:01:48
# Command line: water
# -auto
# -stdout
# -sequence embooss_water-120240913-060144-0580-22567665-p1m.a.sequence
# -sequence embooss_water-120240913-060144-0580-22567665-p1m.b.sequence
# -detailfile ERL050m02
# -gapopen 10.0
# -gapextend 0.5
# -format3 pair
# -spstorein0
# -spstorein2
# Align_format: pair
# Report_file: stdout
#####

#
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: ERL050m02
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 32
# Identity: 30/32 (93.8%)
# Similarity: 30/32 (93.8%)
# Gaps: 0/32 (0.0%)
# Score: 168.0
#
#
#####
EMBOSS_001 1 ACCTACTACTACTACTACTACTACTACTACTACT 32
            ||||||||| ||||||||| |||||||||
EMBOSS_001 1 ACCTACTACTACTACTACTACTACTACTACTACT 32

#
#
#
#####
```

Local alignment is a sequence comparison method that identifies the best matching region between two sequences without attempting to align the entire sequence length. This approach is particularly useful when comparing sequences that may share regions of similarity but differ significantly elsewhere. Unlike global alignment, which attempts to

For this task, the EMBOSS Water tool was used to perform a local alignment between two sequences: ACGTACGTACGTACGTACGTACGTACGTACGT and ACGTACGTAGGTACGTACGTAGGTACGTACGT. The Water tool identified and aligned only the most similar regions of these sequences, focusing on the highest scoring match. The results showed an identity of 30/32 (93.8%) with a similarity of 32/32 (100%), demonstrating a strong alignment within the region of interest. This highlights the advantage of local alignment when analyzing sequences that do not align across their entire lengths. In contrast to global alignment (performed earlier with Needle), which aligns all parts of both sequences even if gaps are needed, the Water tool only aligns the regions that share the most homology, ignoring non-matching ends. The local alignment in this case successfully captured the highly similar core of the two sequences while disregarding the less relevant mismatches at the edges.



In this task, we performed global alignments using the EMBOSS Needle tool with two different gap penalty schemes: a linear-like gap penalty and an affine gap penalty. The linear-like penalty was simulated by setting the gap opening penalty to 1, while the affine gap penalty was achieved by setting the gap opening penalty to 10. In both cases, the gap extension penalty was kept at 0.5. Additionally, the end gaps were not penalized, as the "End Gap" setting was set to "False" to avoid penalizing gaps at the ends of the sequences. The linear-like penalty (gapopen = 1) encourages more frequent gaps since the cost of opening a gap is relatively low. As a result, this alignment introduced a greater number of gaps throughout the sequences to achieve a better overall match. This allowed for more flexibility in matching the sequences, as evidenced by the alignment having a similarity score of 93.8%, but with more interruptions caused by the frequent gaps.

On the other hand, the affine gap penalty (gapopen = 10) discourages the introduction of new gaps due to the high cost of opening them. This resulted in fewer gaps in the alignment, but the gaps tended to be longer. In this case, the similarity score remained the same at 93.8%, but the pattern of gaps was different, reflecting the preference for fewer but more extended gaps. The identity score in both alignments was the same, but the gap structure was different, reflecting how the choice of gap penalties affects the alignment. By comparing the two alignments, we can see the impact that different gap penalties have on sequence alignment. The linear-like penalty results in more gaps but offers more flexibility, while the affine penalty favors longer, uninterrupted stretches of matching sequences by discouraging the frequent introduction of new gaps. Both approaches are valuable depending on the biological context of the sequences being aligned.