

# Aligning Short Reads with BWA-SW

When you select the *Tools > Align to reference > Align short reads* item in the main menu, the *Align Sequencing Reads* dialog appears. Set value of the *Align short reads method* parameter to *BWA-SW*. The dialog looks as follows:

**Align Sequencing Reads**

Alignment method: BWA-SW

Reference sequence:

Result file name:

Library: Single-end ☐ Prebuilt index ☒ SAM output

Short reads

Path	Type	Order
------	------	-------

Add Remove

**Base Options**

Score for a match	1	Number of threads:	8
Mismatch penalty	3	Size of chunk of reads	1000000
Gap open penalty	5	Score threshold (divided by match score)	30
Gap extension penalty	2	Z-best	1
Band width	50	Number of seeds to start rev alignment	5
Mask level	0.50	<input type="checkbox"/> Prefer hard clipping in SAM output	

NOTE: bwa-sw performs alignment of long sequencing reads (Sanger or 454). It accepts reads only in FASTA or FASTQ format. Reads should be compiled into single file.

Start Cancel Help

There are the following parameters:

*Reference sequence* — DNA sequence to align short reads to. This parameter is required.

*Result file name* — file in SAM format to write the result of the alignment into. This parameter is required.

*Prebuilt index* — check this box to use an index file instead of a source reference sequence. Also you can [build it manually](#).

*SAM output* — always save the output file in the SAM format (the option is disabled for *BWA*).

*Short reads* — each added short read is a small DNA sequence file. At least one read should be added.

You can also configure other parameters.

*Score for a match (-a)* — score of a match.

*Mismatch penalty (-b)* — mismatch penalty.

*Gap open penalty (-q)* — gap open penalty.

*Gap extension penalty (-r)* — Gap extension penalty. The penalty for a contiguous gap of size  $k$  is  $q+k*r$ .

*Band width (-w)* - Band width in the banded alignment.

*Mask level (-c)* - Coefficient for threshold adjustment according to query length. Given an  $l$ -long query, the threshold for a hit to be retained is  $a*\max\{T, c*\log(l)\}$ .

*Number of threads (-t)* - Number of threads in the multi-threading mode.

*Size of chunk of reads (-s)* - Maximum SA interval size for initiating a seed. Higher  $-s$  increases accuracy at the cost of speed.

*Score threshold (divided by much score) (-T)* - minimum score threshold.

*Z-best (-z)* - Z-best heuristics. Higher  $-z$  increases accuracy at the cost of speed.

*Number of seeds to start rev alignment (-N)* - Minimum number of seeds supporting the resultant alignment to skip reverse alignment.

*Prefer hard clipping in SAM output (-H)* - use hard clipping in the SAM output. This option may dramatically reduce the redundancy of output when mapping long contig or BAC sequences.

Select the required parameters and press the *Start* button.