



## BWA.indexer Documentation

**Description:** Builds a BWT index from a set of DNA sequences.  
**Author:** Heng Li, Broad Institute  
**BWA Version:** 0.5.9  
**Contact:** Marc-Danie Nazaire, [gp-help@broadinstitute.org](mailto:gp-help@broadinstitute.org)

### Summary

The BWA.indexer builds a BWT index from a set of DNA sequences. This module takes a sequence files in FASTA format, and outputs a set of 6 files in a ZIP archive. These files together constitute the index. For more information on the FASTA format, see the NIH description here at <http://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml>.

This document is adapted from the BWA documentation for release 0.5.9. For more information about BWA.indexer, see the [BWA project site](#). BWA.indexer was developed at the Wellcome Trust Sanger Institute and the Broad Institute.

### Memory Requirements

Depending on the options specified, BWA.indexer requires between 2.5GB and 3.5GB of memory to run.

### Speed

Indexing the human genome takes approximately 3 hours. Indexing smaller genomes is significantly faster, but requires more memory.

### References

BWA manual page: <http://bio-bwa.sourceforge.net/bwa.shtml>.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*. 2009;25:1754-1760. [PMID: 19451168] (<http://www.ncbi.nlm.nih.gov/pubmed/19451168>)

### Parameters

Name	Description
fasta.file (required)	A single file containing sequences in FASTA format.

# GenePattern

algorithm (required)	<p>The algorithm to use to construct the BWT index. Options include:</p> <ul style="list-style-type: none"><li>• is: The IS linear-time algorithm for constructing a suffix array. It requires <math>5.27 \cdot N</math> memory, where <math>N</math> = database size. IS is moderately fast, but does not work with databases larger than 2GB.</li><li>• bwtsw: The algorithm implemented in BWT-SW. This method works with the whole human genome, but it does not work with databases smaller than 10MB and it is usually slower than IS.</li></ul> <p>Default: is</p>
color.space.index (required)	<p>Whether to build a color-space index. The input FASTA should be in nucleotide space. Default: no</p>
output.prefix (required)	<p>A prefix for the output file name.</p>

## Output Files

1. Eight files comprise the index, and are output in a ZIP archive (<output prefix>.zip). The file names are in the following formats:
  - <output prefix>.amb
  - <output prefix>.ann
  - <output prefix>.bwt
  - <output prefix>.pac
  - <output prefix>.rbwt
  - <output prefix>.rpac
  - <output prefix>.rsa
  - <output prefix>.sa

## Platform Dependencies

<b>Module type:</b>	RNA-seq
<b>CPU type:</b>	any
<b>OS:</b>	Macintosh, Linux
<b>Language:</b>	C++, Perl