

VAT Manual

Versatile Alignment Tool

Hao Xuan
xuan13hao@ku.edu

January 25, 2026

Contents

1 Overview	3
2 Syntax and Command Overview	3
2.1 Syntax	3
2.2 Commands	3
2.3 General Options	3
2.4 Database Construction Options	3
2.5 Alignment Options	3
2.6 Alignment Modes	4
2.7 Advanced Options	4
2.8 View Options	4
3 Getting Started	4
3.1 Basic Workflow	4
4 Generating VAT Databases	4
4.1 Command	4
5 DNA Alignment Modes	5
5.1 Whole Genome Sequencing (WGS)	5
5.2 Splice-aware RNA-seq Alignment	5
5.3 Circular RNA Detection	6
5.4 DNA Homology Search	6
5.5 Whole Genome Alignment (WGA)	6
6 Protein Alignment	7
7 Translated Alignment (blastx)	7
8 Output Files	7
9 Description of All Parameters	7
9.1 General Parameters	7
9.2 Scoring Parameters	8
9.3 Advanced Parameters	8

1 Overview

VAT (Versatile Alignment Tool) is a high-performance sequence alignment framework for DNA and protein sequences. VAT supports multiple alignment paradigms including whole-genome mapping, split reads, such as splice-aware RNA-seq alignment, homology search, metagenomic classification, and translated alignment.

2 Syntax and Command Overview

2.1 Syntax

```
VAT COMMAND [OPTIONS]
```

2.2 Commands

- **makevatdb**: Build VAT database from a FASTA file.
- **dna**: Align DNA query sequences against a DNA reference database.
- **protein**: Align protein query sequences against a protein reference database.
- **blastx**: Align translated DNA queries against a protein reference database.
- **view**: View or convert VAT alignment archive (.vatr) files.
- **dna2pro**: Convert DNA sequences into protein sequences.

2.3 General Options

- **-h, --help**: Display help message.
- **-p, --threads**: Number of CPU threads (default: 4).
- **-d, --db**: Reference database file.
- **-a, --vaa**: VAT alignment archive (.vatr) file (optional).
- **--dbtype**: Database type: nucl or prot.

2.4 Database Construction Options

- **-i, --in**: Input reference FASTA file.

2.5 Alignment Options

- **-q, --query**: Input query file.
- **-k, --maxtarget_seqs**: Maximum number of reported target sequences (default: 25).
- **--top**: Report alignments within this percentage of the top score.
- **-e, --evalue**: Maximum E-value threshold (default: 0.001).
- **--report_id**: Minimum identity percentage to report.
- **-S, --seed_len**: Seed length (default: 16 for DNA, 8 for protein).
- **--gapopen**: Gap open penalty.
- **--gapextend**: Gap extension penalty.

- **--match:** Match score.
- **--mismatch:** Mismatch penalty.

2.6 Alignment Modes

- **--wgs:** Whole-genome sequencing alignment.
- **--wga:** Whole-genome alignment.
- **--splice:** Splice-aware RNA-seq alignment.
- **--circ:** Circular RNA detection.
- **--chimera:** Chimera alignment detection.
- **--metagenomic:** Metagenomic alignment mode.
- **--dnah:** DNA homology search.
- **--long:** Long-read alignment mode.
- **--SEN:** Sensitive protein alignment mode.

2.7 Advanced Options

- **--xdrop:** Ungapped X-drop threshold.
- **-X, --gapped_xdrop:** Gapped X-drop threshold.
- **--band:** Dynamic programming band width.
- **-N, --num_shapes:** Number of seed shapes (0 = all).

2.8 View Options

- **-o, --out:** Output file.
- **-f, --outfmt:** Output format (tab/sam/paf/pairwise).

3 Getting Started

3.1 Basic Workflow

The standard VAT workflow consists of two stages:

1. **Database Generation:** Convert FASTA references into binary VAT indexes.
2. **Alignment / Analysis:** Map reads or sequences against pre-built databases.

4 Generating VAT Databases

4.1 Command

```
VAT makevatdb --in reference.fa --dbtype nucl|prot -p 8
```

- **--in**
 - **Default:** –

- **Type:** FASTA file
- **Description:** Input reference FASTA file.
- **--dbtype**
 - **Default:** nucl
 - **Type:** nucl | prot
 - **Description:** Database type: nucleotide or protein.
- **--threads**
 - **Default:** 4
 - **Type:** int
 - **Description:** Number of CPU threads used for index construction.

Generated files use the internal .vatf format and must not be modified.

5 DNA Alignment Modes

5.1 Whole Genome Sequencing (WGS)

Optimized for short- and long-read genome resequencing.

```
VAT dna -d ref.fa -q reads.fa --wgs -o out.sam -f sam -p 8
```

- **--wgs**
 - **Default:** false
 - **Type:** flag
 - **Description:** Enable WGS alignment mode.
- **--long**
 - **Default:** false
 - **Type:** flag
 - **Description:** Enable long-read alignment.
- **--seed_len**
 - **Default:** 15
 - **Type:** int
 - **Description:** Seed length for genome mapping.

5.2 Splice-aware RNA-seq Alignment

Detects exon–intron junctions and alternative splicing events.

```
VAT dna -d ref.fa -q rna.fa --splice -o out.sam -f sam
```

- **--splice**
 - **Default:** false
 - **Type:** flag

- **Description:** Enable splice-aware alignment.
- **--long**
 - **Default:** `false`
 - **Type:** `flag`
 - **Description:** Enable long-read alignment.
- **--seed_len**
 - **Default:** `15`
 - **Type:** `int`
 - **Description:** Seed length for genome mapping.

5.3 Circular RNA Detection

Detects back-spliced junctions characteristic of circular RNAs.

- **--circ**
 - **Default:** `false`
 - **Type:** `flag`
 - **Description:** Enable circular RNA detection mode.

5.4 DNA Homology Search

Sensitive nucleotide homology alignment for evolutionary analysis.

- **--dnah**
 - **Default:** `false`
 - **Type:** `flag`
 - **Description:** Enable DNA homology alignment mode.

5.5 Whole Genome Alignment (WGA)

Aligns entire genomes and outputs alignment archives.

```
VAT dna -d human.fa -q chim.fa --wga -a wga_out -p 32
VAT view -a wga_out.vatr -o hg_chim.tab -f tab -p 32
```

- **--wga**
 - **Default:** `false`
 - **Type:** `flag`
 - **Description:** Enable whole genome alignment mode.

6 Protein Alignment

```
VAT protein -d prot.fa -q query.fa -o out.tab -f tab
```

- **--seed_len**
 - **Default:** 8
 - **Type:** int
 - **Description:** Protein seed length.
- **--num_shapes**
 - **Default:** 0
 - **Type:** int
 - **Description:** Number of seed shapes (0 = all).

7 Translated Alignment (blastx)

```
VAT blastx -d prot.fa -q dna.fa -o out.tab -f tab -p 8
```

8 Output Files

VAT produces alignment results in standard formats.

- **--outfmt**
 - **Default:** tab
 - **Type:** tab|sam|pafl|pairwise
 - **Description:** Output format.
- **--vaa**
 - **Default:** -
 - **Type:** file
 - **Description:** Binary alignment archive (.vatr).

9 Description of All Parameters

Default values are shown in blue, allowed types in magenta.

9.1 General Parameters

- **--threads**
 - **Default:** 4
 - **Type:** int
 - **Description:** Number of CPU threads.
- **--evalue**

- **Default:** 0.001
- **Type:** float
- **Description:** Maximum allowed E-value.
- **--maxtarget_seqs**
 - **Default:** 25
 - **Type:** int
 - **Description:** Maximum number of reported targets.

9.2 Scoring Parameters

- **--match**
 - **Default:** 5
 - **Type:** int
 - **Description:** Match score.
- **--mismatch**
 - **Default:** -4
 - **Type:** int
 - **Description:** Mismatch penalty.
- **--gapopen**
 - **Default:** -6
 - **Type:** int
 - **Description:** Gap opening penalty.
- **--gapextend**
 - **Default:** -1
 - **Type:** int
 - **Description:** Gap extension penalty.

9.3 Advanced Parameters

- **--xdrop**
 - **Default:** 18
 - **Type:** int
 - **Description:** Ungapped X-drop threshold.
- **--gapped_xdrop**
 - **Default:** 18
 - **Type:** int
 - **Description:** Gapped X-drop threshold.