

UCHIME v4.0.87 Quick Reference

Reference database mode

```
usearch --uchime seqs.fasta --db ref.fasta --uchimeout output.uchime
[--uchimealns alnfile]
```

De novo mode

```
usearch --uchime seqs.fasta --uchimeout output.uchime
[--uchimealns alnfile]
```

In de novo mode, sequences must have the string /ab=xx/ somewhere in the label, where xx is a floating point number indicating its relative abundance. E.g.,

```
>FQ56RFF4A/ab=2.34/
```

Output file

The --uchimeout file is a tab-separated file with the following 17 fields.

Field	Name	Description
1	Score	Value >= 0.0, high score means more likely to be a chimera.
2	Query	Sequence label
3	Parent A	Sequence label
4	Parent B	Sequence label
5	IdQM	%id between query and model made from (A, crossover, B)
6	IdQA	%id between query and parent A.
7	IdQB	%id between query and parent B
8	IdAB	%id between parents (A and B).
9	IdQT	%id between query and closest reference sequence / candidate parent.
10	LY	Yes votes on left
11	LN	No votes on left
12	LA	Abstain votes on left
13	RY	Yes votes on right
14	RN	No votes on right
15	RA	Abstain votes on right
16	Div	Divergence ratio, i.e. IdQM - IdQT
17	YN	Y (yes) or N (no) classification as a chimera. Set to Y if score >= threshold set by the --minh option.

Command-line options

--input filename

--uchime filename

Query sequences in FASTA format.

If the --db option is not specified, uchime uses de novo detection. In de novo mode, relative abundance must be given by a string /ab=xxx/ somewhere in the label, where xxx is a floating-point number, e.g. >F00QGH67HG/ab=1.2/.

Both the --input and --uchime options may be used in the stand-alone UCHIME program. If you are using USEARCH, then you must use the --uchime option.

--db filename

Reference database of chimera-free sequences in FASTA format. Optional, if not specified uchime uses de novo mode.

WARNING The database is searched ONLY on the plus strand. You MUST include reverse-complemented sequences in the database if you want both strands to be searched.

--abskew x

Minimum abundance skew. Default 1.9. De novo mode only.

Abundance skew is:

$$\min [\text{abund}(\text{parent1}), \text{abund}(\text{parent2})] / \text{abund}(\text{query}).$$

--uchimeout filename

Output in tabbed format with one record per query sequence.

First field is score (h), second field is query label.

For details, see manual.

--uchimealns filename

Multiple alignments of query sequences to parents in human-readable format. Alignments show columns with differences that support or contradict a chimeric model.

--minh h

Minimum score to report chimera. Default 0.3. Values from 0.1 to 5 might be reasonable. Lower values increase sensitivity but may report more false positives. If you decrease --xn, you may need to increase --minh, and vice versa.

--mindiv div

Minimum divergence ratio, default 0.5. Div ratio is 100% - %identity between query sequence and the closest candidate for being a parent. If you don't care about very close chimeras, then you could increase --mindiv to, say, 1.0 or 2.0, and also decrease --min h, say to 0.1, to increase sensitivity. How well this works will depend on your data. Best is to tune parameters on a good benchmark.

--xn beta

Weight of a no vote, also called the beta parameter. Default 8.0. Decreasing this weight to around 3 or 4 may give better performance on denoised data.

--dn n
Pseudo-count prior on number of no votes. Default 1.4. Probably no good reason to change this unless you can retune to a good benchmark for your data. Reasonable values are probably in the range from 0.2 to 2.

--xa w
Weight of an abstain vote. Default 1. So far, results do not seem to be very sensitive to this parameter, but if you have a good training set might be worth trying. Reasonable values might range from 0.1 to 2.

--chunks n
Number of chunks to extract from the query sequence when searching for parents. Default 4.

--[no]ovchunks
[Do not] use overlapping chunks. Default do not.

--minchunk n
Minimum length of a chunk. Default 64.

--idsmoothwindow w
Length of id smoothing window. Default 32.

--minsmoothid f
Minimum fractional identity over smoothed window of candidate parent. Default 0.95.

--maxp n
Maximum number of candidate parents to consider. Default 2. In tests so far, increasing --maxp gives only a very small improvement in sensitivity but tends to increase the error rate quite a bit.

--[no]skipgaps
--[no]skipgaps2
These options control how gapped columns affect counting of diffs. If --skipgaps is specified, columns containing gaps do not count as diffs. If --skipgaps2 is specified, if column is immediately adjacent to a column containing a gap, it is not counted as a diff. Default is --skipgaps --skipgaps2.

--minlen L
--maxlen L
Minimum and maximum sequence length. Defaults 10, 10000.

--ucl
Use local-X alignments. Default is global-X. On tests so far, global-X is always better; this option is retained because it just might work well on some future type of data.

--queryfract f
Minimum fraction of the query sequence that must be covered by a local-X alignment. Default 0.5. Applies only when --ucl is specified.

--quiet
Do not display progress messages on stderr.

--log filename

Write miscellaneous information to the log file. Mostly of interest to me (the algorithm developer). Use --verbose to get more info.

--self

In reference database mode, exclude a reference sequence if it has the same label as the query. This is useful for benchmarking by using the ref db as a query to test for false positives.