

How to extract fasta sequences from a multi-fasta file based on matching headers in a separate file?

Author : Tariq Abdullah

Categories : [Bioinformatics Programming](#)

Date : April 5, 2019

This is a simple Perl script to extract FASTA sequences from a large fasta file depending on the matching fasta headers present in another file.

For example, your fasta sequences are present in a file named, "input.fa" and the headers are in another file called "headers.txt".

```
#!/usr/bin/perl use warnings; use strict; my $headerfile = 'headers.txt'; my $input = 'input.fa'; open( HEADERFILE, '<', $headerfile ) or die $!; chomp ( my @headers = map { split } <$headerfile> ); # splitting lines on whitespaces. close HEADERFILE; my %seqs; open( INPUTFILE, '<', $input ) or die $!; { local $/ = ''; #Reading until blank line while ( <$input> ) { my ( $header, $sequence ) = m/>\s*(\S+)\n(.*)/ms; $sequences{$header} = $sequence; } open( my $seqsfile, ">", "input.fa" ); foreach my $header (@headers) { if ( $sequences{$header} ) { print $header, "\n"; print $sequences{$header}, "\n"; } } close( $seqsfile ); } close INPUTFILE; exit;
```

Sharing is caring. Spread the love!

- [Print](#)
- [Email](#)
- [LinkedIn](#)
- [Twitter](#)
- [Facebook](#)
- [Google](#)
-