

The screenshots below show the analysis of *Brassica napus* sequence data

Step 4.1.1 – demultiplexing with bclfastq2

Command used:

```
$ nohup /opt/bcl2fastq2-2.18.12/bin/bin/bcl2fastq --runfolder-dir
/runs/190301_NS500446_0582_AHLWC3BGX9 --ignore-missing-bcl --output-dir
Demultiplex_Output_190301_NS500446_0582_AHLWC3BGX9 --barcode-mismatches
1 --aggregated-tiles AUTO -r 16 -d 16 -p 16 -w 16
```

Output (screenshot of a read in fastq format):

[illegible]

NOTE: the sequences indicated in red and blue are the adapter sequences that will be removed during steps 4.1.2 and 4.1.3. The sequence in light purple is the miRNA sequence

Step 4.1.2 – adapter trimming

Command used:

```
$ /opt/cutadapt-1.15/bin/cutadapt -a TGGATTCTCGGGTGCCAAGG -n 5 -O 4 -m 10 -j 0 --nextseq-trim 10 -o  
Sample1_Brassica_R1_cutadapt.fastq.gz Sample1_Brassica_R1.fastq.gz
```

Output (screenshot of a read in fastq format after trimming):

```
@NS500446:582:HLWC3BGX9:1:11101:10309:2245  
1:N:0:GGTAGC  
TCGCTTCAATGTCGAGAAAGAGGGATGTAAAA  
+  
AAAAAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
```

Step 4.1.3 – trimming of random nucleotides from the HD adapters

Command used:

```
$ seqtk trimfq -b 4 -e 4 Sample1_Brassica_R1_cutadapt.fastq > Sample1_Brassica_R1_4bp_trimmed.fastq
```

Output (screenshot of read in fastq format after trimming):

```
@NS500446:582:HLWC3BGX9:1:11101:10309:2245  
1:N:0:GGTAGC  
TTCAATGTCGAGAAAGAGGGATGT  
+  
AEEEEAEAAAAAAAAAAAAAAAAAAAA
```

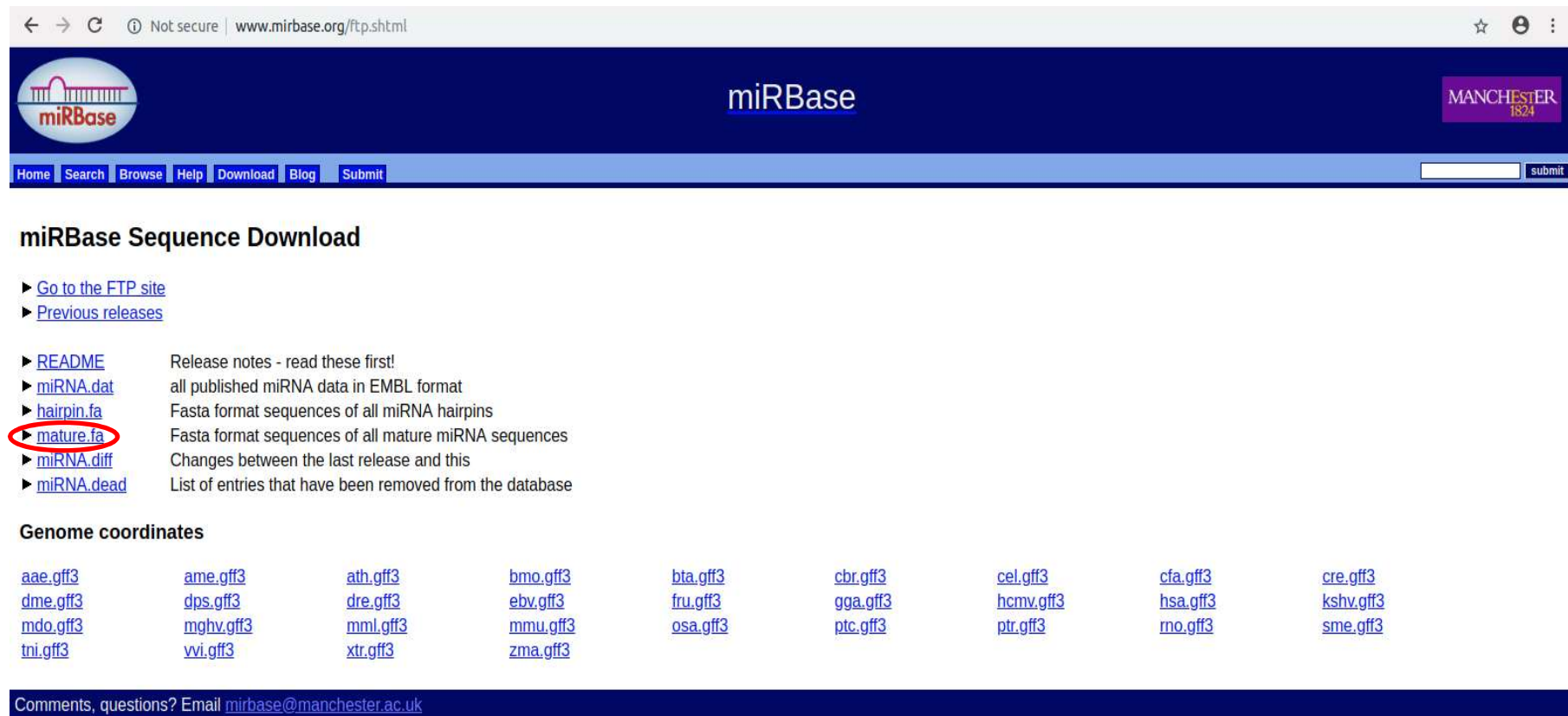
Step 4.1.4 – Discard sequences shorter than 10 nucleotides

Command used:

```
$ awk 'BEGIN {FS = "\t" ; OFS = "\n"} {header = $0 ; getline seq ; getline qheader ; getline qseq ; if  
(length(seq) >= 10) {print header, seq, qheader, qseq}}' Sample1_Brassica_R1_4bp_trimmed.fastq >  
Sample1_Brassica_R1_4bp_trimmed_more_than_10bp_long.fastq
```

Step 4.2.1 – Mapping to miRBase

Download mature miRNA sequences (mature.fr database):



The screenshot shows the miRBase website interface. The browser address bar displays 'www.mirbase.org/ftp.shtml'. The page has a dark blue header with the miRBase logo and a navigation bar with links: Home, Search, Browse, Help, Download, Blog, and Submit. Below the header, the section 'miRBase Sequence Download' is visible. It contains a list of links with descriptions: 'Go to the FTP site', 'Previous releases', 'README' (Release notes - read these first!), 'miRNA.dat' (all published miRNA data in EMBL format), 'hairpin.fa' (Fasta format sequences of all miRNA hairpins), 'mature.fa' (Fasta format sequences of all mature miRNA sequences), 'miRNA.diff' (Changes between the last release and this), and 'miRNA.dead' (List of entries that have been removed from the database). The 'mature.fa' link is circled in red. Below this, the 'Genome coordinates' section lists various species and their corresponding GFF3 files in a grid format. At the bottom, a footer bar contains the text 'Comments, questions? Email mirbase@manchester.ac.uk'.

miRBase

Home Search Browse Help Download Blog Submit

miRBase Sequence Download

- ▶ [Go to the FTP site](#)
- ▶ [Previous releases](#)
- ▶ [README](#) Release notes - read these first!
- ▶ [miRNA.dat](#) all published miRNA data in EMBL format
- ▶ [hairpin.fa](#) Fasta format sequences of all miRNA hairpins
- ▶ [mature.fa](#) Fasta format sequences of all mature miRNA sequences
- ▶ [miRNA.diff](#) Changes between the last release and this
- ▶ [miRNA.dead](#) List of entries that have been removed from the database

Genome coordinates

aae.gff3	ame.gff3	ath.gff3	bmo.gff3	bta.gff3	cbr.gff3	cel.gff3	cfa.gff3	cre.gff3
dme.gff3	dps.gff3	dre.gff3	ebv.gff3	fru.gff3	gga.gff3	hcmv.gff3	hsa.gff3	kshv.gff3
mdo.gff3	mghv.gff3	mmi.gff3	mmu.gff3	osa.gff3	ptc.gff3	ptr.gff3	rno.gff3	sme.gff3
tni.gff3	vvi.gff3	xtr.gff3	zma.gff3					

Comments, questions? Email mirbase@manchester.ac.uk

Replace U residues by T using the following command:

```
$ sed -i '/^>/! s/U/T/g' mature.fa
```

Step 4.2.2 – select your organisms of interest

Command used:

```
$ awk '/Brassica napus/{{print; nr[NR+1]; next}; NR in nr' mature.fa > mature_Brassica_92mirs.fa
```

Output (screenshot of *B. napus* sequences):

```
>bna-miR156a MIMAT0004445 Brassica napTs miR156a  
TGACAGAAGAGAGTGAGCACA  
>bna-miR171g MIMAT0004446 Brassica napTs miR171g  
TGATTGAGCCGCGCCAATATCT  
>bna-miR393 MIMAT0004447 Brassica napTs miR393  
TCCAAAGGGATCGCATTGATC
```

Step 4.2.3 – build an index for the file created in step 4.2.2

Command used:

```
$ bowtie2-build mature_Brassica_92mir.famature_Brassica_92mir
```

Step 4.2.4 – Align the sequencing reads to the database, requiring that a read maps entirely to a miRNA of the database, without any mismatches

Command used:

```
$ bowtie2 -N 0 -L 10 --score-min C,0,0 --end-to-end --time -x mature_Brassica_92mir -U  
Sample1_Brassica_R1_4bp_trimmed_more_than_10bp_long.fastq -S  
Sample1_Brassica_EXACT_end_to_end.sam
```

Output (screenshot of alignment results):

NS500446:582:HLWC3BGX9:1:11103:17410:17691	4	*	0	0	*	*	0	0	GCCACAGCGTCATACGCGCTGAGCGTTCTTT
NS500446:582:HLWC3BGX9:1:11103:3794:17785	0	bna-miR159	1	42	21M	*	0	0	TTTGGATTGAAGGGAGCTCTA
NS500446:582:HLWC3BGX9:1:11103:10467:17921	4	*	0	0	*	*	0	0	CCCTTCGGGCGTGTCGTAGATAATCATGACA
NS500446:582:HLWC3BGX9:1:11103:19321:17935	4	*	0	0	*	*	0	0	AAAAGATCAAGTCAATTTTAATAGTAAAAAA
NS500446:582:HLWC3BGX9:1:11103:3383:18136	0	bna-miR156a	1	39	20M	*	0	0	TGACAGAAGAGAGTGAGCAC
NS500446:582:HLWC3BGX9:1:11103:15304:18169	4	*	0	0	*	*	0	0	AGTATAATAATGGGATGCGTGAGA
NS500446:582:HLWC3BGX9:1:11103:15600:18209	4	*	0	0	*	*	0	0	GGTCTGGGCGGATATGCAATTGTTATGATA
NS500446:582:HLWC3BGX9:1:11103:15175:18254	4	*	0	0	*	*	0	0	AAGCGGTGGTTTTGGGGAAAAGA A/6E//E
NS500446:582:HLWC3BGX9:1:11103:14518:18296	4	*	0	0	*	*	0	0	CCCATCTCGGCAAGGGGCTCGGACACCAGTT
NS500446:582:HLWC3BGX9:1:11103:19529:18313	4	*	0	0	*	*	0	0	AAAGAAAGCACCCAGCGCGAGCAGTTCACCA
NS500446:582:HLWC3BGX9:1:11103:20179:18411	4	*	0	0	*	*	0	0	CTCTGTTGCCTGAACATCTCGACAGGGAATA
NS500446:582:HLWC3BGX9:1:11103:8783:18582	4	*	0	0	*	*	0	0	AGCAGCTTGTAGGACCGGGCCAGGGCGGTGA
NS500446:582:HLWC3BGX9:1:11103:2424:18728	4	*	0	0	*	*	0	0	AAGGTTGCGGCTTGGCATTGTGTACTATGAT
NS500446:582:HLWC3BGX9:1:11103:6262:18818	4	*	0	0	*	*	0	0	TTTTTATATGCCTACCATAACATGCAATCT
NS500446:582:HLWC3BGX9:1:11103:7995:18820	4	*	0	0	*	*	0	0	AGGAAGCTAAACCGATGGTTAGGA
NS500446:582:HLWC3BGX9:1:11103:22095:18826	0	bna-miR172a	1	42	21M	*	0	0	AGAATCTTGATGATGCTGCAT
NS500446:582:HLWC3BGX9:1:11103:12535:18974	4	*	0	0	*	*	0	0	CTAGCTGCGGACATTAGACTGG AEA6/AE
NS500446:582:HLWC3BGX9:1:11103:20618:19219	4	*	0	0	*	*	0	0	ATTCAATCGGGTTTGATTAAGGGT
NS500446:582:HLWC3BGX9:1:11103:5864:19398	4	*	0	0	*	*	0	0	TCCGGACCGGTTGGCATCGCCGCGGTCTGCG
NS500446:582:HLWC3BGX9:1:11103:8704:19460	4	*	0	0	*	*	0	0	AACGCCATCCCCCGGCCGACAACCAACCCA
NS500446:582:HLWC3BGX9:1:11103:26628:19507	4	*	0	0	*	*	0	0	CGTACAACCAAGAAAGCTCGGAGTGTTGCGC
NS500446:582:HLWC3BGX9:1:11103:17160:19578	4	*	0	0	*	*	0	0	CCAAGTACCTTATGTAAAAA AEEEA/E
NS500446:582:HLWC3BGX9:1:11103:10356:19622	0	bna-miR167b	1	39	20M	*	0	0	TGAAGCTGCCAGCATGATCT

Step 4.2.5 – Discard the reads that did not align

Command used:

```
$ samtools view -F 4 Sample1_Brassica_EXACT_end_to_end.sam > Reads_aligned_to_Mirs.sam
```

Output (screenshot of aligned reads):

```
NS500446:582:HLWC3BGX9:1:11102:14519:4559    0    bna-miR172a    1    42    21M    *    0    0    AGAATCTTGATGATGCTGCAT    AEEEEEEEE/AEEEEEEEE    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11102:9989:9496    0    bna-miR159    1    42    21M    *    0    0    TTTGGATTGAAGGGAGCTCTA    AEEEEEEEEEEEEEEEEAE    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11102:14281:12506    0    bna-miR159    1    42    20M    *    0    0    TTTGGATTGAAGGGAGCTCT    AEAEAE/EEE6/AEAEAE    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11103:13785:1821    0    bna-miR159    1    42    21M    *    0    0    TTTGGATTGAAGGGAGCTCTA    AAEEEEEEEEEEEEEEEE6    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11103:2084:4738    0    bna-miR159    1    42    21M    *    0    0    TTTGGATTGAAGGGAGCTCTA    AEEEEEEEEEEEEEEEE    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11103:13346:5936    0    bna-miR172d    1    42    21M    *    0    0    AGAATCTTGATGATGCTGCAG    AEE6EEEEEEEE6EEEEEEEE    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11103:7229:13663    0    bna-miR159    1    42    21M    *    0    0    TTTGGATTGAAGGGAGCTCTA    AEEE/EE/EEEE6AEEAEEEE    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11103:14686:13855    0    bna-miR167b    1    39    21M    *    0    0    TGAAGCTGCCAGCATGATCTA    A/AE/E/EAEEE/EE/EEEE    AS:i:0    XS:i:0
NS500446:582:HLWC3BGX9:1:11103:4811:16181    0    bna-miR159    1    42    21M    *    0    0    TTTGGATTGAAGGGAGCTCTA    AEEEEEAEEEEEEEEEEEE    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11103:2594:17444    0    bna-miR159    1    42    21M    *    0    0    TTTGGATTGAAGGGAGCTCTA    AEEEEEAEEEEEEE/<AEEE    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11103:3794:17785    0    bna-miR159    1    42    21M    *    0    0    TTTGGATTGAAGGGAGCTCTA    AEEEEEEEEEEEEEEEEEA    AS:i:0    XN:i:0
NS500446:582:HLWC3BGX9:1:11103:3383:18136    0    bna-miR156a    1    39    20M    *    0    0    TGACAGAAGAGAGTGAGCAC    AEEEEEEEEEEEEEEEE    AS:i:0    XS:i:0
```