

# Using sequins with metagenomics.

Metagenome sequins are a set of synthetic DNA controls that reflect the sequence complexity, GC content, phylogenetic diversity and abundance of a natural microbial community. The sequins are ‘spiked-in’ to your DNA sample, which together undergo to library preparation, sequencing and analysis. The sequins can then be distinguished from you sample DNA in the output library by their synthetic sequence, and analyzed as internal controls. r

Sequins are compatible with all standards library preparation and sequencing methods. This protocol describes the laboratory steps required to re-suspend and spike the sequins into your DNA sample, as well as the bioinformatic steps required to analyze sequins in your output library.

For further details on the design, validation and use of sequins, we refer users to ‘**Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis**’ by Hardwick et. al., (2018) *Nature Communications*.

## 1. Laboratory steps

### 1.1. Receiving sequins.

Your sequins should arrive in tubes within a sealed package. Once received, store the sequin tubes at -20C until you are ready to use them.

### 1.2. Preparing sequin stocks.

The sequins should be first resuspended to the correct dilution. Each tube typically contains at least 200ng of lyophilized sequin DNA (please note that the exact amount may vary). Please use the table below to guidethe amount and dilution of sequins that should be used according to the sample DNA amount:

| 1. Sample DNA Amount. | 2. Resuspension Volume (ddH <sub>2</sub> O) | 3. Resuspended Sequins Conc. | 4. Resuspended Sequin Volume to Add to Sample. | 5. Mass of Sequins Added. | 6. Final Sequin Concentration. |
|-----------------------|---|------------------------------|--|---------------------------|--------------------------------|
| 10 ng                 | 2000 ul                                     | 0.1 ng/ul                    | 2 ul   | 0.2 ng                    | 2 %                            |
| 100 ng                | 200 ul                                      | 1 ng/ul                      | 2 ul   | 2 ng                      | 2 %                            |
| 200 ng                | 100 ul                                      | 2 ng/ul                      | 2 ul   | 4 ng                      | 2 %                            |
| 1000 ng               | 20 ul                                       | 10 ng/ul                     | 2 ul   | 20 ng                     | 2 %                            |

*Table provides guidelines for amount of sequins to add to sample DNA.*

For example, we recommend adding sequins to your sample at a 2% fraction by mass, so that approximately 2% of the reads in your output library will be derived from sequins. Therefore, if the input requirement for your library is 100ng, then you will want to add 2ng of sequin DNA to 100ng of sample DNA to achieve ~2% fraction by mass. To achieve this, first resuspend the sequins in 200ul of sterile dH<sub>2</sub>O or TE Buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) to reach a ~1ng/ul concentration, and then add ~2ul of this sequin resuspension to your sample.

### 1.3. Spike sequins into sample.

Please note that we recommend you confirm the concentration of your sequin resuspension before addition to your sample. This can be achieved using a Qubit™ or a similar instrument (please note that we have experienced inaccurate quantification of sequins when using Nanodrop™).

Once you have added sequins to your DNA sample, the combined sample/sequins mixture is then used as input into your preferred library preparation protocol as per manufacturer’s instructions.

### 1.4. Store remaining sequins.

Once you have re-suspended your sequins, we recommend you store them as single-use aliquots at -20 C to prevent unnecessary future freeze-thaw cycles. Frozen DNA sequins aliquots are stable for at least 6 months. These individual sequin aliquots should then be thawed and added to DNA samples just prior to library preparation.

## 2. Bioinformatic steps.

To analyze sequins, we have developed a software toolkit, named *Anaquin*, that accepts either .FASTQ or .BAM formats, and can be integrate into your bioinformatic pipeline. When *anaquin meta* processes the .FASTQ or .BAM files, it performs three main functions:

**Partition.** *Anaquin meta* will partition the library into smaller sub-libraries comprising either sample or sequin reads/alignments.

**Calibrate.** The number or fraction of sequin reads in a library can be modulated. For example, *anaquin meta* can calibrate the number of sequins reads to comprise 1% of the library (using the `--calibrate` option). This tool is useful for matching dilutions between multiple replicates and samples.

**Report.** *Anaquin meta* generates several useful reports, including an analysis of library performance (`meta_summary.stats`), quantitative accuracy (`meta_sequins_table.tsv`) as well as individual sequin performance (`meta_sequins.tsv`).

Anaquin can be downloaded from [www.sequinstandards.com/software](http://www.sequinstandards.com/software). Then run:

```
unzip anaquin_3.14.zip
cd anaquin_3.14.0
make
```

To demonstrate the following steps, we have used the following example RNAseq libraries that can be downloaded from [www.sequinstandards.com/resources/](http://www.sequinstandards.com/resources/)

```
communityA_metaquin_MixA.R1.fq.gz
communityA_metaquin_MixA.R2.fq.gz
```

### 2.1. Analysis of .FASTQ libraries (option 1).

To directly analyze the sequin from your library .FASTQ files, run the following command:

```
$ anaquin meta -t 24 -o results --calibrate 0.005 -1 communityA_metaquin_MixA.R1.fq.gz \
-2 communityA_metaquin_MixA.R2.fq.gz
```

In this example command, we used `--calibrate 0.005` to calibrate sequin reads to 0.5% of the NGS library. Results are provided in output directory (**see Section 3**)

### 2.2. Analysis of .BAM alignments (option 2).

#### 2.2.1. Build index with decoy chromosome.

Alternatively, users can align their library to a combined index comprising the reference microbial genomes of interest, as well as the sequin decoy chromosomes (chrQ\*). The sample reads will thereby align to the microbial genomes, whilst the sequin reads will align to the decoy chromosome.

In this example, we have aligned the library using BWA (Li et. al., 2009) to reference actinomycete genomes (these genome sequences (*actinomycete\_genomes.fa*) can be downloaded from: [www.sequinstandards.com/resources/](http://www.sequinstandards.com/resources/) )

1. The user must first build the combined index. We first concatenate the sequin decoy chromosomes to the microbe genome sequences::

```
$ gunzip actinomycete_genomes.fa.gz
$ cat ../actinomycete_genomes.fa metasequin_decoy_2.2.fa >actinomycete_decoy.fa
$ bwa index -p actinomycete_decoy.bwa actinomycete_decoy.fa
```

2. We then align the NGS library to the combined index:

```
$ bwa mem -t 24 actinomycete_decoy.bwa communityA_metaquin_MixA.R1.fq.gz \
communityA_metaquin_MixA.R2.fq.gz | samtools view -S -b | samtools sort \
>CommunityA_MixA.align.sort.bam
```

3. We finally use *anaquin meta* to analyse the .BAM alignment files.

```
$ anaquin meta -t 24 -o results --calibrate 0.005 --combined CommunityA_MixA.align.sort.bam
```

In this example command, we used `--calibrate 0.05` to calibrate sequin reads to 0.5% of the NGS library. Results are provided in output directory (specified using `-ouput`; see Section 3)

### 2.3. Output results.

When anaquin meta is complete, the following files are generated in the output directory:

*anaquin.log* – Log files recording the usage and execution of sequin processes.

*meta\_sequin\_table.tsv* – Quantification of sequins.

*meta\_sequin.tsv* - Detailed report on each individual sequins.

*meta\_ladder\_table.tsv* – Quantification of synthetic DNA ladder.

*meta\_ladder.tsv* - Sequin alignments/reads derived from the synthetic DNA ladder.

*meta\_sample\_\** - Sample alignments/reads (excludes sequins).

*meta\_sequin\_\** - Alignments/reads derived from sequins.

*meta\_ladder\_\** - Sequin alignments/reads derived from the synthetic DNA ladder.

*meta\_vector\_\*.fq.gz* - Residual alignments/reads derived from sequin plasmid sequences (used during sequin manufacture).

*For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.*

*Information in this document is subject to change without notice.*

*Material safety data sheets (MSDSs) are available at [www.sequistadnards.com/resources/](http://www.sequistadnards.com/resources/)*

*Revision History | Publication Number 5.1 | Revision Date March 2020*

*© Garvan Institute of Medical Research, 2016-2020*

*The trademarks mentioned herein are the property of their respective owners.*

*ioanalyzer are registered trademarks of Agilent Technologies, Inc.*