DNA Storage Decoding Instructions

This document will describe the steps needed in order to read the following archive:

Archive Order ID	DS-14
Codec Version	0.9
Archive Size	16.6 KB
Oligo Count	N/A

Data Recovery Methods

If you wish to recover the digital data stored in this DNA archive, there are three possible approaches:

OPTION 1: SEND THE POOL TO TWIST

You may contact Twist via **https://www.twistbioscience.com/contact** to request data recovery.

You'll need to ship the capsule containing the archived data to:

Twist Bioscience
Attention: DNA Data Storage
681 Gateway Blvd
South San Francisco, CA, 94080, USA

Please specify the archive order ID in the shipment.

Twist will extract the data from the capsule, sequence and decode the DNA, and provide the decoded digital files back to you.

OPTION 2: USE OF A THIRD-PARTY DNA SEQUENCING PROVIDER

You may also choose to use a third-party DNA sequencing company to read back the DNA archive. You'll need to send the capsule and opener along with the instructions printed below (steps 1 and 2) to the sequencing provider. They will then send back the sequenced DNA as a series of 'reads' - you'll then need to carry out step 3 below to decode the digital files from these sequence reads.

OPTION 3: SEQUENCE THE DNA YOURSELF

If you have access to a device capable of Illumina-compatible short-read sequencing, you can carry out all of the steps below yourself.

General Instructions for Recovery

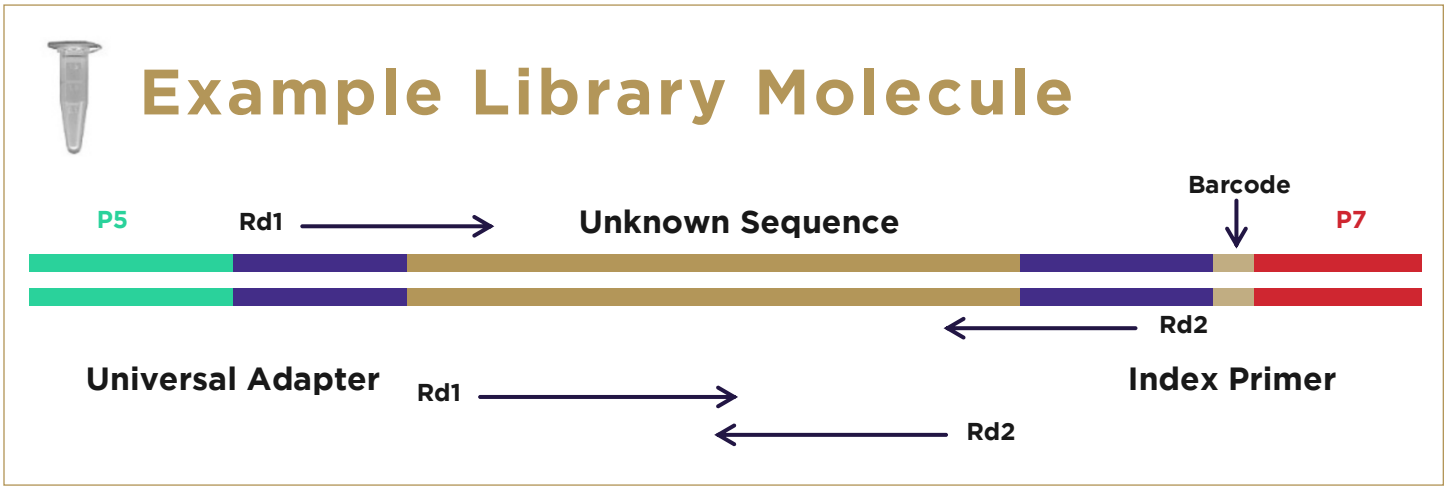
The recovery of the digital data stored in this pool requires three steps:

STEP 1: RECOVER AND RESUSPEND THE DNA

Your encoded data is stored inside an airtight stainless-steel capsule. Use the provided opener device to open the capsule. Once the capsule is open, add 20 µL of water or TE buffer (10 mM Tris, 0.1 mM EDTA) and resuspend the DNA (stored dry at the bottom of the capsule). Once resuspended, draw all liquid from the capsule and transfer to a tube for storage. Once extracted, the suspended DNA should be kept refrigerated at -20C to preserve DNA quality. If multiple uses are expected (e.g. sequencing the pool multiple times to test recovery), make individual aliquots in separate single-use tubes to prevent potential damage to DNA caused by freeze/thaw cycles.

STEP 2: SEQUENCE THE DNA

The oligonucleotides in this pool are intended to be sequenced using Illumina TruSeq 1x150 single-end sequencing. Each oligo in this pool is 192 nucleotides long and all oligos were synthesized with flanking read primers. For sequencing, this means that they only require the addition of P5/P7 adaptors and indexing barcodes during PCR prior to sequencing. That is, according to the following schematic:



The Rd1 and Rd2 sequences are already present on molecules in this pool. To amplify the DNA sequences in the pool, order and use the following primer sequences:

5' read primer: GTCAGATGTGTATAAGAGACAG

3' read primer: CGTGTGCTCTTCCGATCT

These primers (with attached P5/P7 sequences and barcodes) can be used to amplify the pool via PCR to ensure sufficient mass for sequencing.

Sequencing depth for this pool must be sufficient to generate at least 10 million single-end, high-quality (i.e. Phred quality score exceeding 20) reads to guarantee accurate data recovery.

Once sequencing is complete, reads are generally stored in a FASTQ-formatted file which can often be compressed via gzip, resulting in a file with a '.fastq.gz' extension.

STEP 3: DECODE THE SEQUENCED DNA

The provided decoding software expects as input an uncompressed text file containing NGS read sequences, one per line. The sequencing process removes the 5' read primer above and often includes additional bases after the 3' read primer - the decoding code will treat these sequences appropriately.

Preparing sequencing reads

If your sequencing data is stored in a FASTQ-formatted file, you'll need to convert this file to a plain text file. The easiest way to do this is via use of the UNIX sed tool via:

```
gunzip -c myreads.fastq.gz | awk '/^@/{getline;print}' > myreads.txt
```

Running the decoder Docker image

The decoder is provided as a Docker image. Download the docker image file and then import the image into a local Docker repository via:

```
docker import twist_bioscience_codec_v2_external.tar.gz twist_decoder:v2
```

Once loaded, you'll need to run the image to create a container via:

```
docker run -it twist_decoder:v2 /bin/bash
```

This will start the container and provide a bash shell prompt. In a second shell, determine the ID of the running container via:

```
docker ps
```

You will see something like:

```
CONTAINER ID   IMAGE              COMMAND       CREATED         STATUS         PORTS     NAMES
18ab14601b4d   twist_decoder:v2   "/bin/bash"   6 seconds ago   Up 5 seconds             confident_sutherland
```

You can then copy the read data into the running container via (where `18ab14601b4d` is the container ID of the running container):

```
docker cp myreads.txt 18ab14601b4d:/tmp/data/
```

Decoding the sequencing reads

To decode data, start by importing the reads:

```
/usr/twist/bin/twist_import /tmp/data /tmp/data/imported
```

Then cluster the reads by data index:

```
/usr/twist/bin/twist_cluster /tmp/data/imported /tmp/data/cluster
```

Then merge the clustered reads:

```
/usr/twist/bin/twist_cluster_merge /tmp/data/imported /tmp/data/cluster /tmp/data/merged
```

And finally decode the encoded data:

```
/usr/twist/bin/twist_decode /tmp/data/merged/ /tmp/data/decoded/
```

And copy the decoded file back out of the container:

```
docker cp 18ab14601b4d:/tmp/data/decoded .
```

The decoded file(s) will be present in the decoded folder

You can then exit out of the container shell and stop the running container via:

```
exit
```