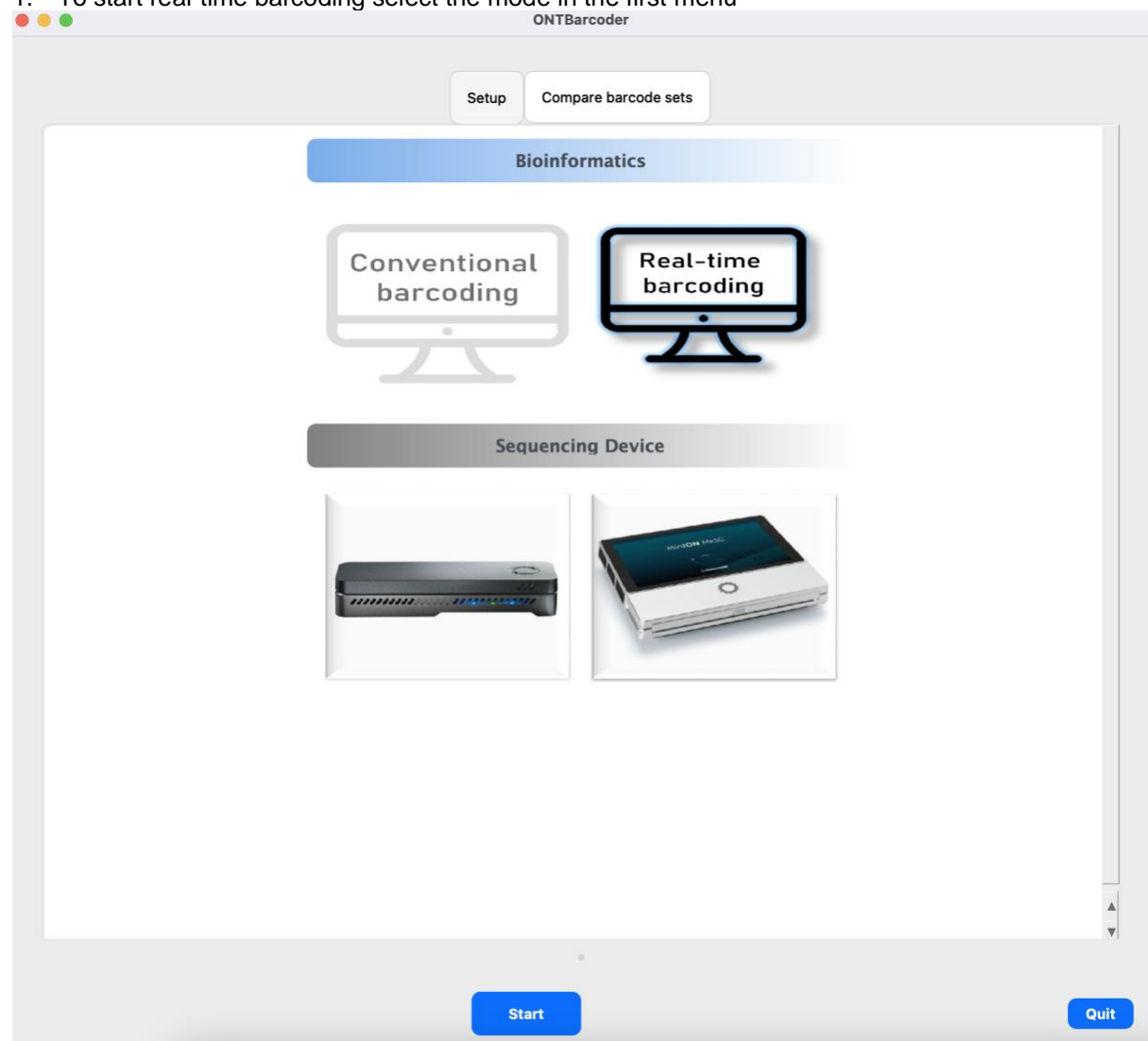


REAL-TIME BARCODING

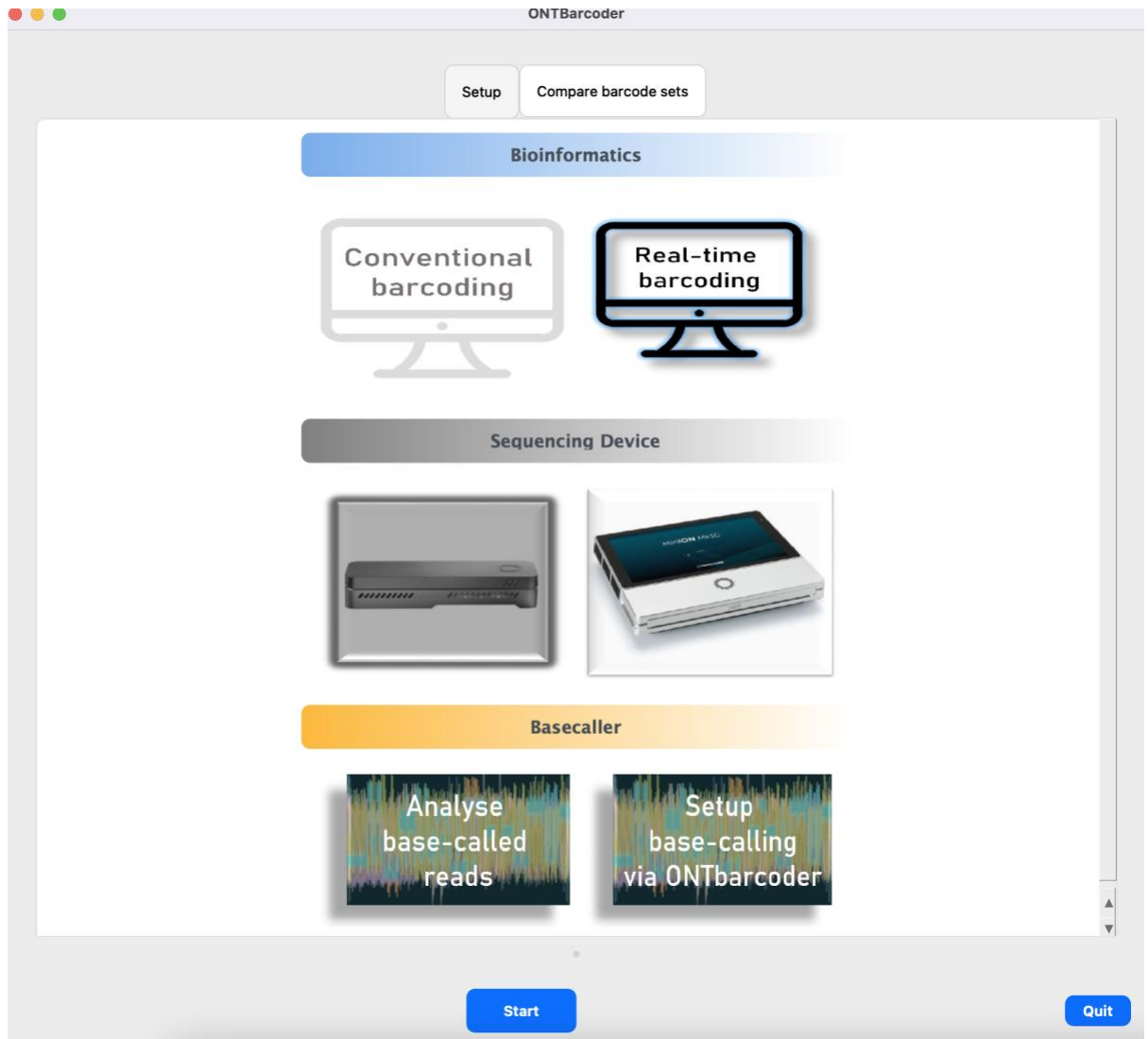
ONTbarcode (available from <https://github.com/asrivathsan/ONTbarcode>) is installed by unzipping the folder with the version of the program that supports the operating system on your computer. The folders are available from releases in Github. For MacOS, an .app bundle has been created.

Real-time barcoding module of ONTbarcode accepts is compatible with both Mk1B and Mk1C and accepts input in the form of either a folder where base-called fastq files are stored or a folder containing fast5 are stored

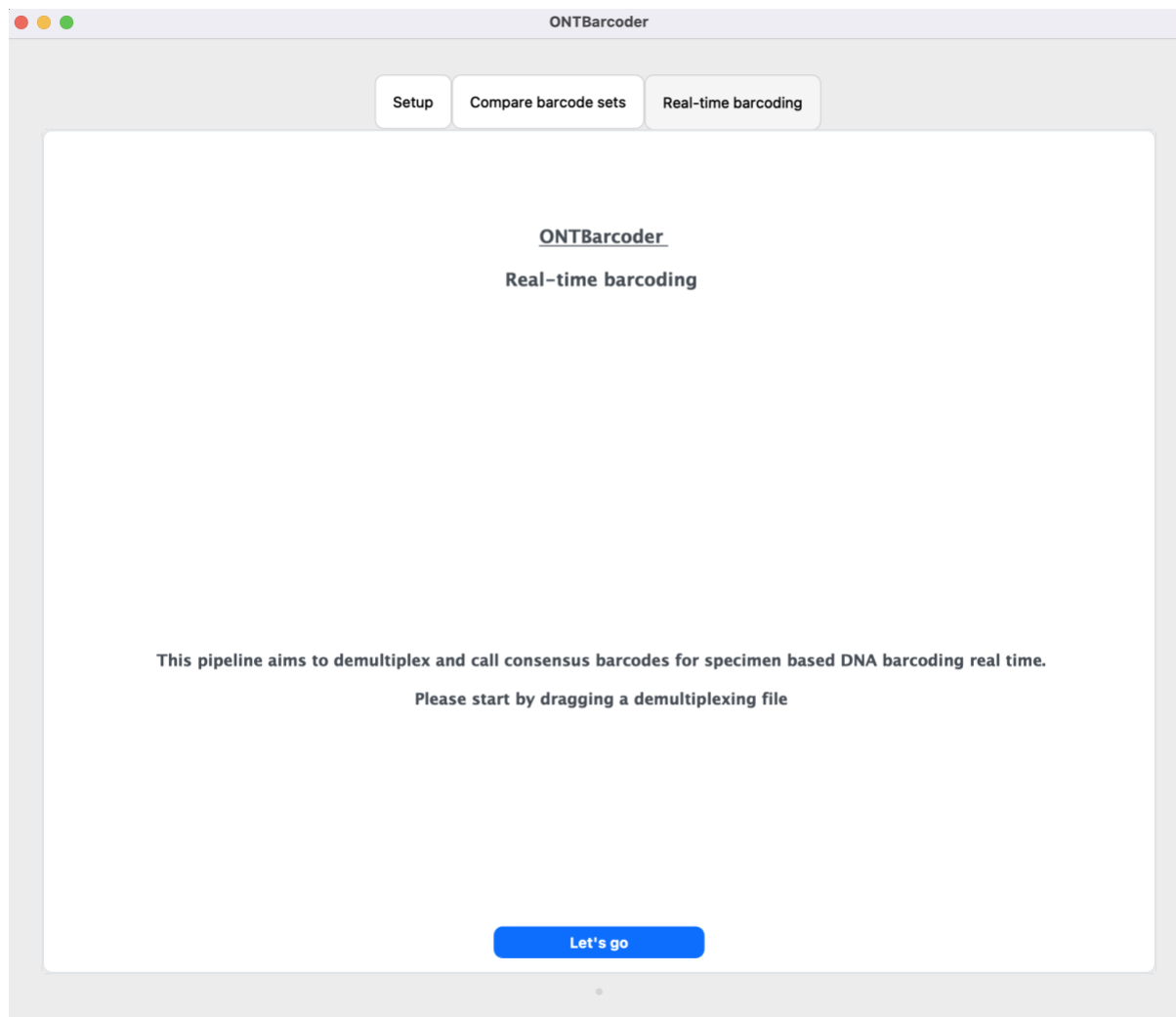
1. To start real-time barcoding select the mode in the first menu



2. Next select the sequencing device of interest



3. Finally select the type of reads you would be inputting, folder containing base-called fastq files, or folder containing fast5 files for basecalling and press "Start".
4. The next screen will ask you to drag in a demultiplexing file. Drag your demultiplexing file and press the "Let's go" button that appears



5. The four modes (Mk1b+analysis of basecalled reads, Mk1b+analysis of fast5, Mk1c+analysis of basecalled reads, Mk1c+analysis of fast5), will give you different option menus. Common to all is the basic set up for the barcode calling.

Reads longer than this are accepted for demultiplexing

Expected barcode length, used for QC and determining expected product length

Genetic code

Start the run

Demultiplexing Settings

Minimum Length: 658

Length of barcode: 658

Window to define product length (length of barcode +/-): 100

Window for primer and tag search: 100

Tag Mismatch: 2

Consensus settings

Minimum Coverage: 20

Large Coverage, or coverage beyond which you don't examine every increment: 100

Step size for larger coverages: 10

Maximum Coverage, set 0 for no limit: 0

☒ Subset higher coverage to Large coverage based on read lengths?

Main consensus calling frequency: 0.3

Range of frequencies to assess: 0.2,0.5

to be examined at step size of: 0.05

Genetic Code: 5. The Invertebrate Mitochondrial Code

Proceed

Advanced features can be edited if you want to test if they improve your yield:

Primers & tags are searched within the first/last specified number of bp of a read

Number of mismatches allowed in the tag region

Barcodes are called only if they meet this coverage

Other consensus calling frequencies to test

Specifies the steps for the range. In current settings 0.2,0.25,0.3,0.35,0.4,0.45,0.5 are tested

Window size for determining product length (explan+primer/tag length)+window size determine product length

After reaching the large coverage, barcodes refining is attempted only after obtaining enough number of reads defined by the step size

Cap on coverage after which refinement of barcodes stops. 0 disables the cap

Main consensus calling frequency, i.e. at least 30% of the bases should be occupied by a single nucleotide for a consensus to be called

Demultiplexing Settings

Minimum Length: 658

Length of barcode: 658

Window to define product length (length of barcode +/-): 100

Window for primer and tag search: 100

Tag Mismatch: 2

Consensus settings

Minimum Coverage: 20

Large Coverage, or coverage beyond which you don't examine every increment: 100

Step size for larger coverages: 10

Maximum Coverage, set 0 for no limit: 0

☒ Subset higher coverage to Large coverage based on read lengths?

Main consensus calling frequency: 0.3

Range of frequencies to assess: 0.2,0.5

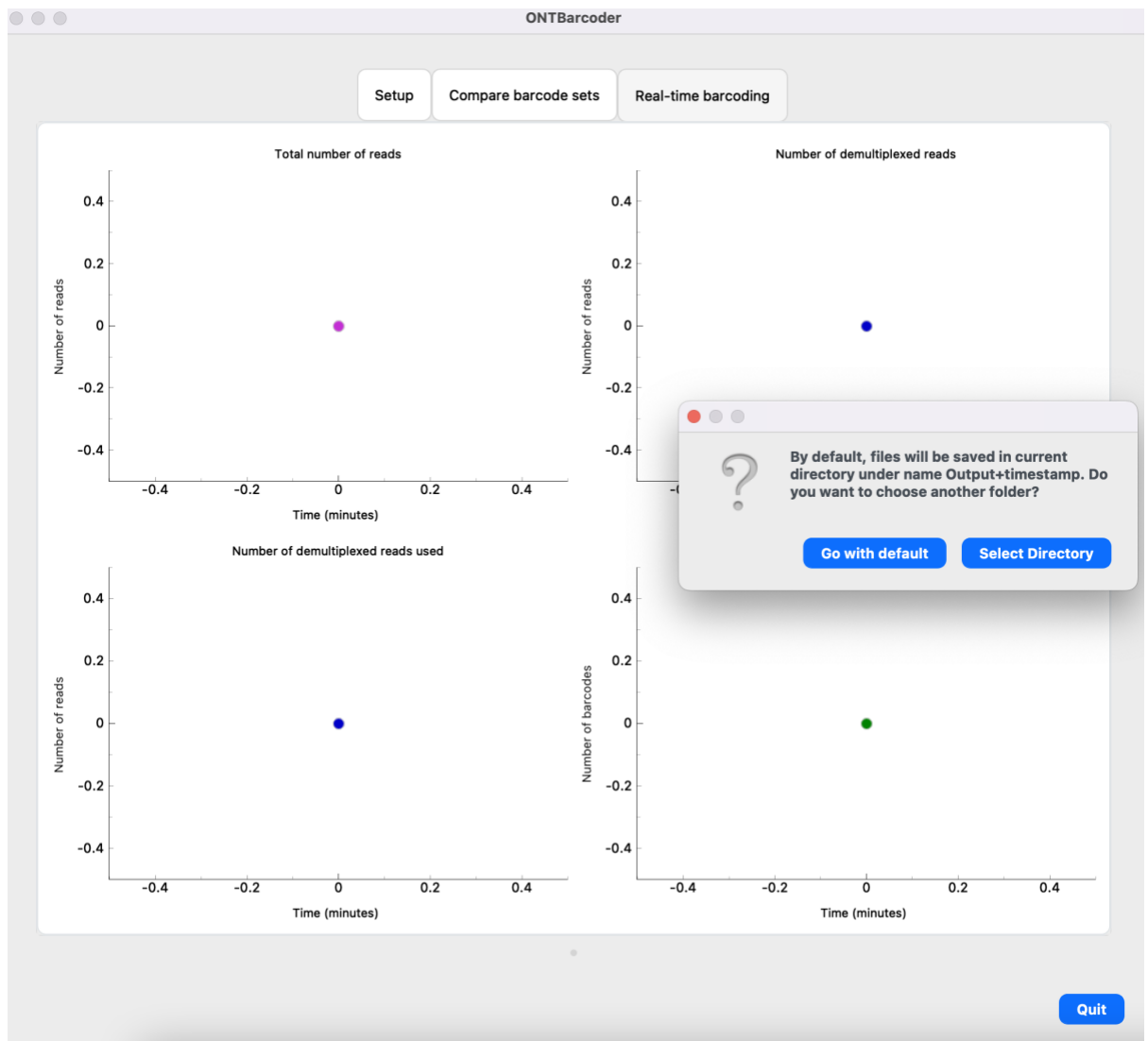
to be examined at step size of: 0.05

Genetic Code: 5. The Invertebrate Mitochondrial Code

Proceed

6. At this point we recommend you start your sequencing run.

7. Select where you want the output to be stored.



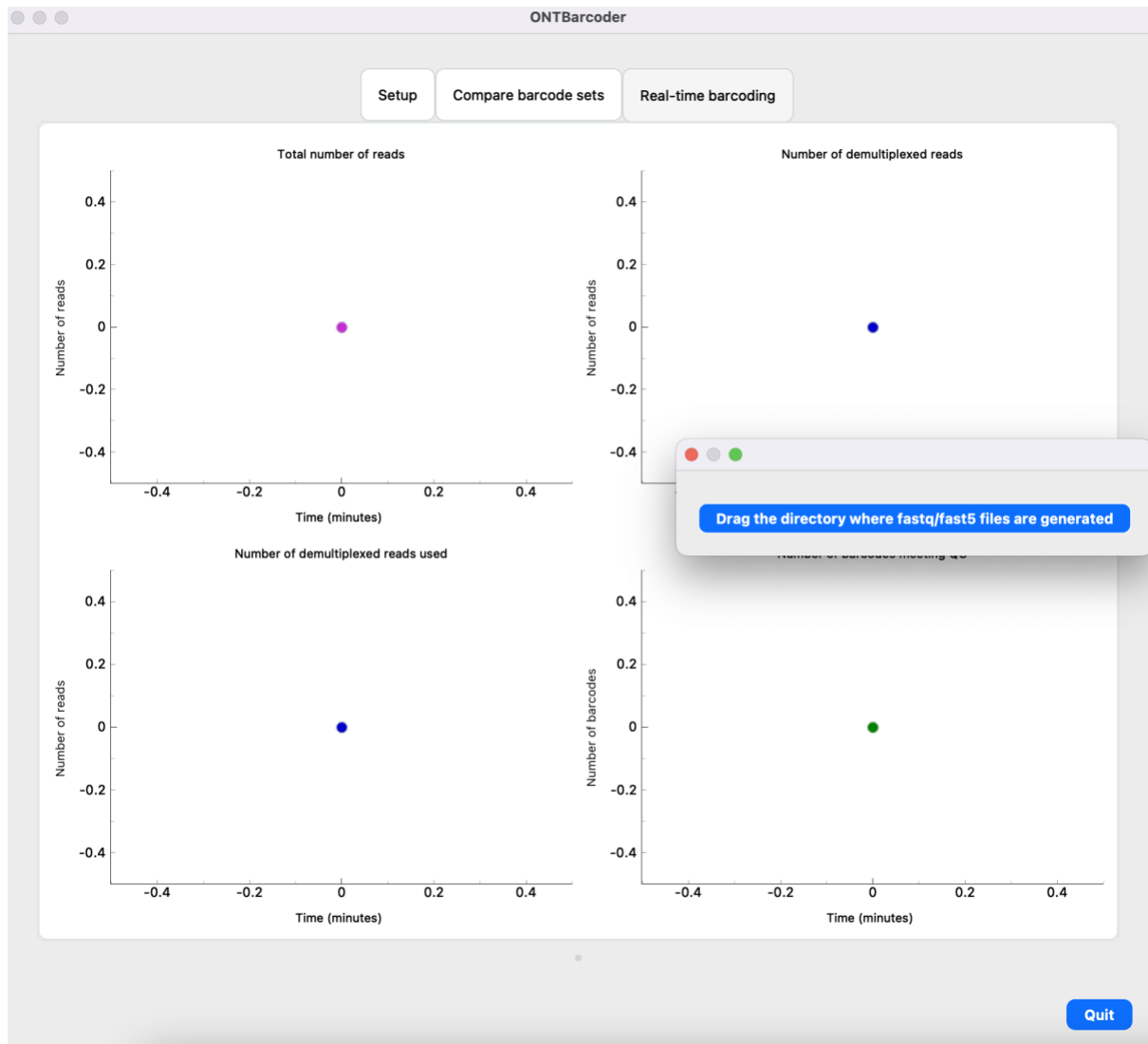
Next we detail the four modes that can be run in this software

Mode 1: Mk1B + Analysis of base-called reads

Here ONTbarcoder is run in the same computer as the one sequencing. Basecalling is set up by MinKnow.

What is needed:

1. Demultiplexing file (already used in set up)
2. Path to folder where fastq files are stored. It is important that the path contains the fastq files themselves. MinKNOW if set up for filtering based on qscore- saves reads in folders called **pass** or **fail**. In this case use the pass folder. However, overall we recommend that MinKNOW is not set up filtering.



At this point a pop up menu will appear asking the User to drag in the directory where the fastq files are being generated. Drag in this folder, and press “Click to proceed”



The run should start

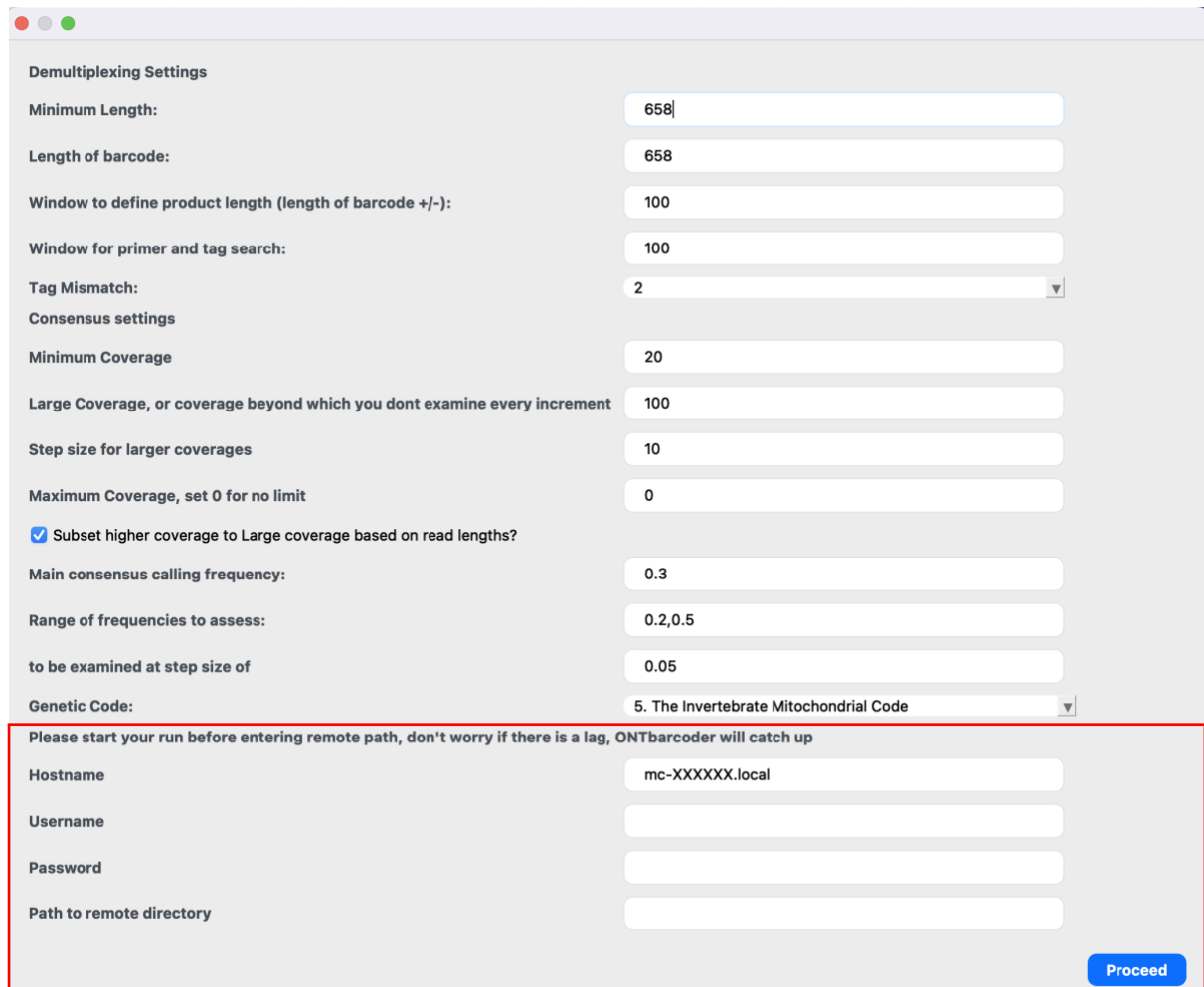
Mode 2: Mk1C + Analysis of base-called reads

In this mode sequencing is in a remote device like Mk1C, and you can connect to it either over ethernet (best practice) or hotspot.

What is needed:

1. Demultiplexing file (already used in set up)
2. Path to folder where fastq files are stored in the remote device. You can connect over ssh to the remote device and copy the text of the path , either via WinSCP/Putty in Windows or terminal in Mac/Linux. It is important that the path contains the fastq files themselves. MinKNOW if set up for filtering based on qscore- saves reads in folders called **pass** or **fail**. In this case use the pass folder. However, overall we recommend that MinKNOW is not set up filtering.
3. Details to login to mk1c (host name, username, password)

The big initial menu that pops up contains a section in the end



The screenshot shows a configuration window for MinKNOW. The top section, 'Demultiplexing Settings', includes fields for 'Minimum Length' (658), 'Length of barcode' (658), 'Window to define product length (length of barcode +/-)' (100), 'Window for primer and tag search' (100), 'Tag Mismatch' (2), 'Consensus settings' (Minimum Coverage: 20, Large Coverage: 100, Step size for larger coverages: 10, Maximum Coverage: 0), and a checked checkbox for 'Subset higher coverage to Large coverage based on read lengths?'. Other fields include 'Main consensus calling frequency' (0.3), 'Range of frequencies to assess' (0.2,0.5), 'to be examined at step size of' (0.05), and 'Genetic Code' (5. The Invertebrate Mitochondrial Code). A red box highlights the bottom section, which contains the instruction 'Please start your run before entering remote path, don't worry if there is a lag, ONTbarcoder will catch up' and fields for 'Hostname' (mc-XXXXXX.local), 'Username', 'Password', and 'Path to remote directory'. A 'Proceed' button is located at the bottom right of this section.

Demultiplexing Settings	
Minimum Length:	658
Length of barcode:	658
Window to define product length (length of barcode +/-):	100
Window for primer and tag search:	100
Tag Mismatch:	2
Consensus settings	
Minimum Coverage	20
Large Coverage, or coverage beyond which you dont examine every increment	100
Step size for larger coverages	10
Maximum Coverage, set 0 for no limit	0
<input checked="" type="checkbox"/> Subset higher coverage to Large coverage based on read lengths?	
Main consensus calling frequency:	0.3
Range of frequencies to assess:	0.2,0.5
to be examined at step size of	0.05
Genetic Code:	5. The Invertebrate Mitochondrial Code

Please start your run before entering remote path, don't worry if there is a lag, ONTbarcoder will catch up

Hostname	mc-XXXXXX.local
Username	
Password	
Path to remote directory	

Proceed

Host name to connect to Mk1C. Typically it can be accessed by mc-xxxxx.local

Please start your run before entering remote path, don't worry if there is a lag, ONTbarcoder will catch up

Username to access mk1c (should be provided with mk1c)

Path to directory where fastq files are stored

Password to access mk1c (should be provided with mk1c)

mc-XXXXXX.local

Username

Password

Path to remote directory

Success

Mode 3 and 4: Mk1B/Mk1C + Basecalling is set up via ONTbarcoder

In these modes, you can configure ONTbarcoder to do basecalling via guppy. Depending on whether you use Mk1B and Mk1C, Mode 1 and Mode 2 menus will appear. However an additional menu will appear for configuring Guppy

What is needed:

1. Demultiplexing file (already used in set up)
2. Path to folder containing fast5 (either locally, same as Mode1) or remotely (same as Mode 2)
3. Guppy (GPU mode) installed in the computer and its path.
4. Path to guppy basecalling model cfg file

Outputs

- > Aligned
- > Consensus
 - consensus_all.fa
 - consensus_good.fa
- > Demultiplexed
 - dmpfile
 - dmpfile_tagfr
- > livefastq
 - log.txt
- > Processing
- > Subsets

“**consensus_all.fa**” stores the final iteration consensus barcodes for all dataset irrespective of quality, while “**consensus_good.fa**” stores the QC filtered barcodes and is the recommended barcode set

Summary

Expected input and output

1. Conventional barcoding	Input required	Results
Mode 1: Mk1B + Basecalled reads	<ul style="list-style-type: none"> • Demultiplexing file • Local directory where fastq files are stored, can be dragged in 	consensus_good.fa (QC filtered) consensus_all.fa (all barcodes)
Mode 2: Mk1C+ Basecalled reads	<ul style="list-style-type: none"> • Demultiplexing file • Path to remote directory where fastq files are stored, has to be pasted in text menu 	
Mode 3: Mk1B + Fast5	<ul style="list-style-type: none"> • Demultiplexing file • Local directory where fastq files are stored, can be dragged in • Path to guppy executable • Path to guppy config file 	
Mode 3: Mk1C + Fast5	<ul style="list-style-type: none"> • Demultiplexing file • Path to remote directory where fastq files are stored, has to be pasted in text menu • Path to guppy executable • Path to guppy config file 	

Format Specifications

File	
1. FASTQ file	Standard fastq, generated after basecalling with ONT software
2. Demultiplexing File	<p>A 5-column csv file with the following headers: SpecimenID, TagFsequence, TagRsequence, PrimerF, PrimerR</p> <p>You can only demultiplex one primer pair at a time. FASTQ files with data for multiple pairs, have to be processed sequentially.</p> <p>Please avoid unusual characters in Specimen ID (e.g. characters like “(){}[]V.,;*\$” will lead to crashes)</p>

3. Directory containing fastq file	Folder containing fastq files generated by ONT can be compressed or not compressed
4 Directory containing fast5 files	Folder containing fast5 files generated by ONT