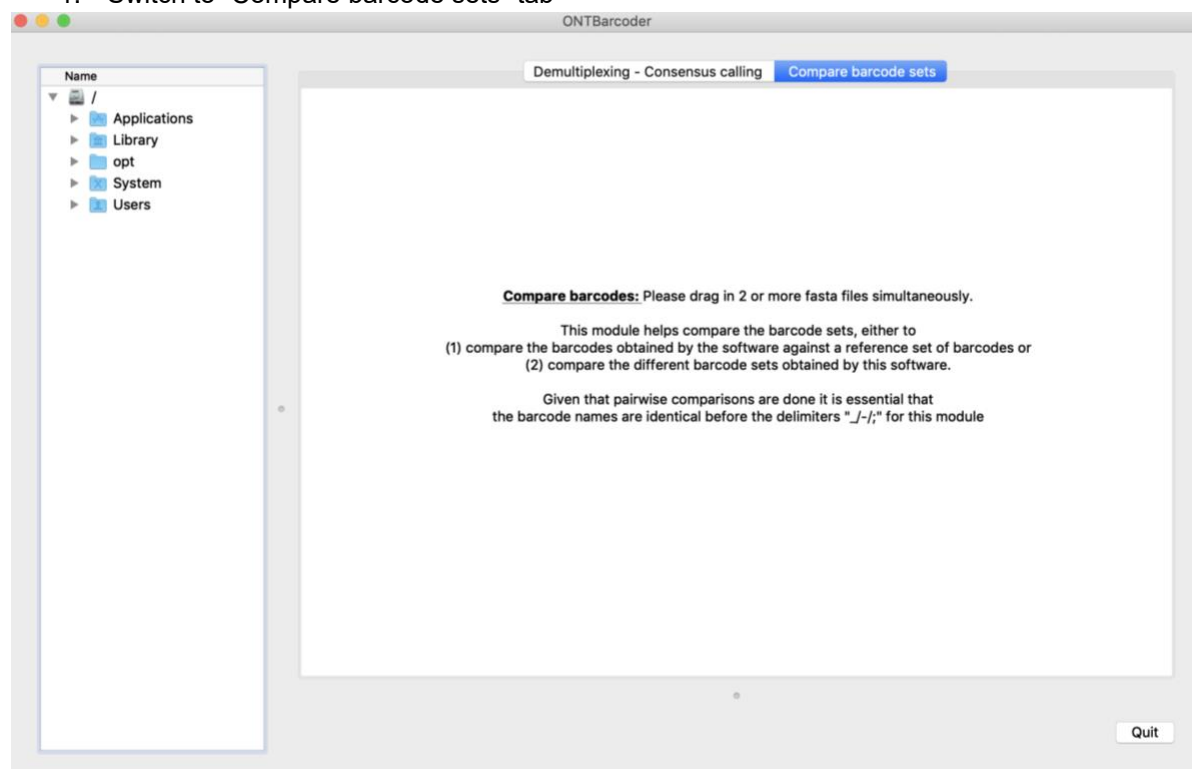


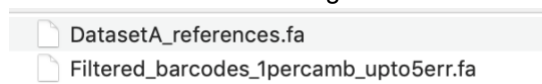
## Compare barcodes in the “Compare barcodes sets tab”

This can be used in different modes (4 modes: single barcode file against single reference, multiple barcode files against single reference, pairwise comparisons of two barcode files, and all-vs-all comparisons of multiple barcode files), which depend on how many input files are dragged in. Here we show how to compare the output for Dataset A with reference sequences generated with Sanger sequencing. The guide here uses images of ONTbarcoder1 but the steps are same for ONTbarcoder2

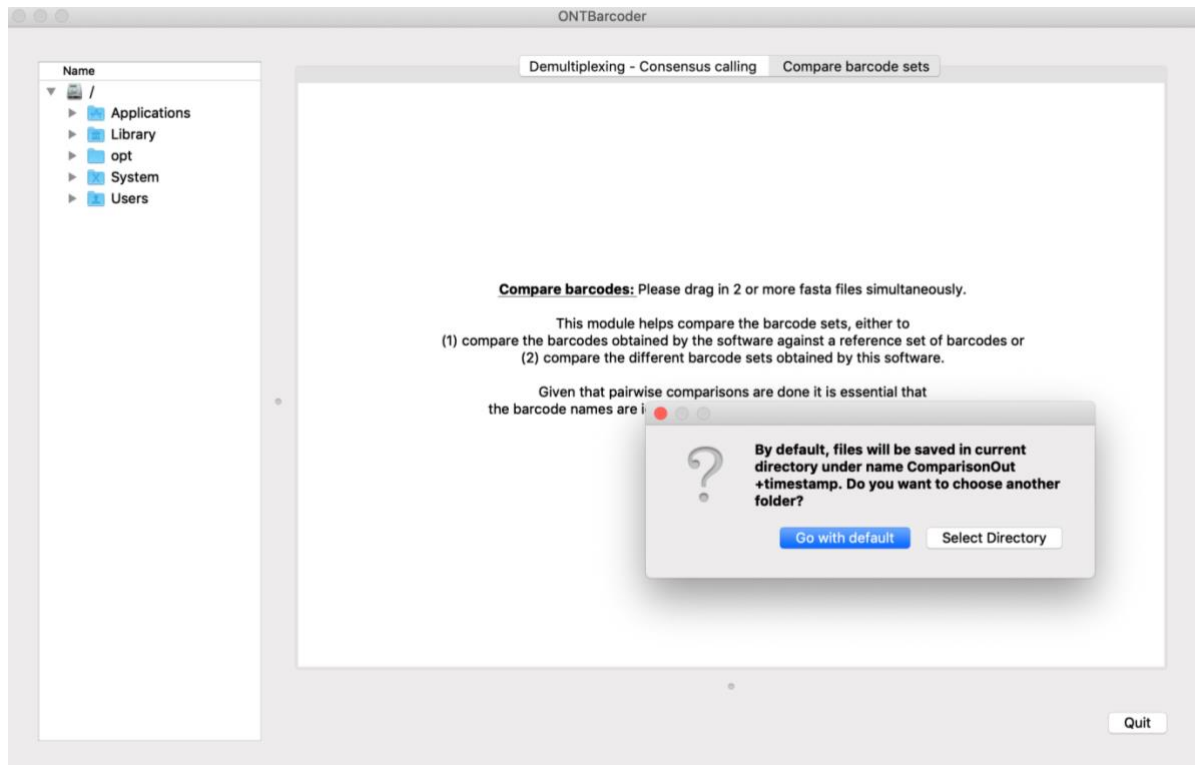
1. Switch to “Compare barcode sets” tab



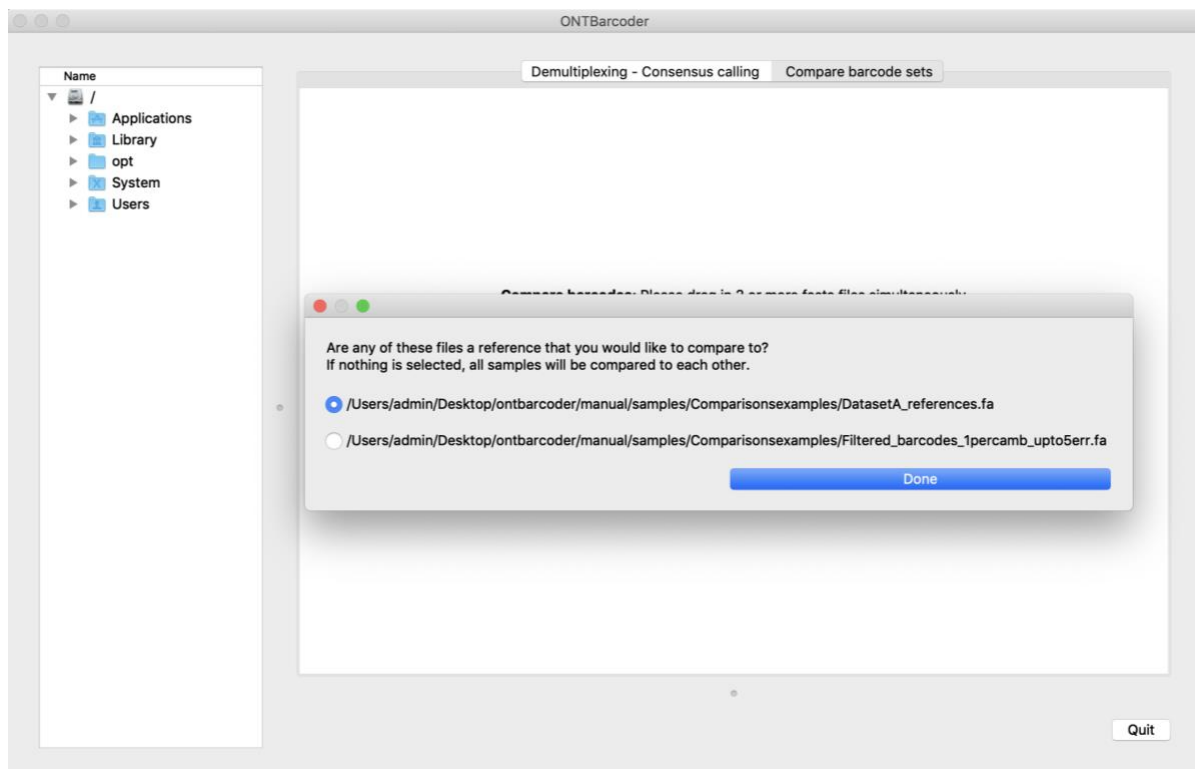
2. Drag at least two fasta files into the window. For the example, go to the “[Main\\_barcode\\_results](#)” folder and use the following two files:



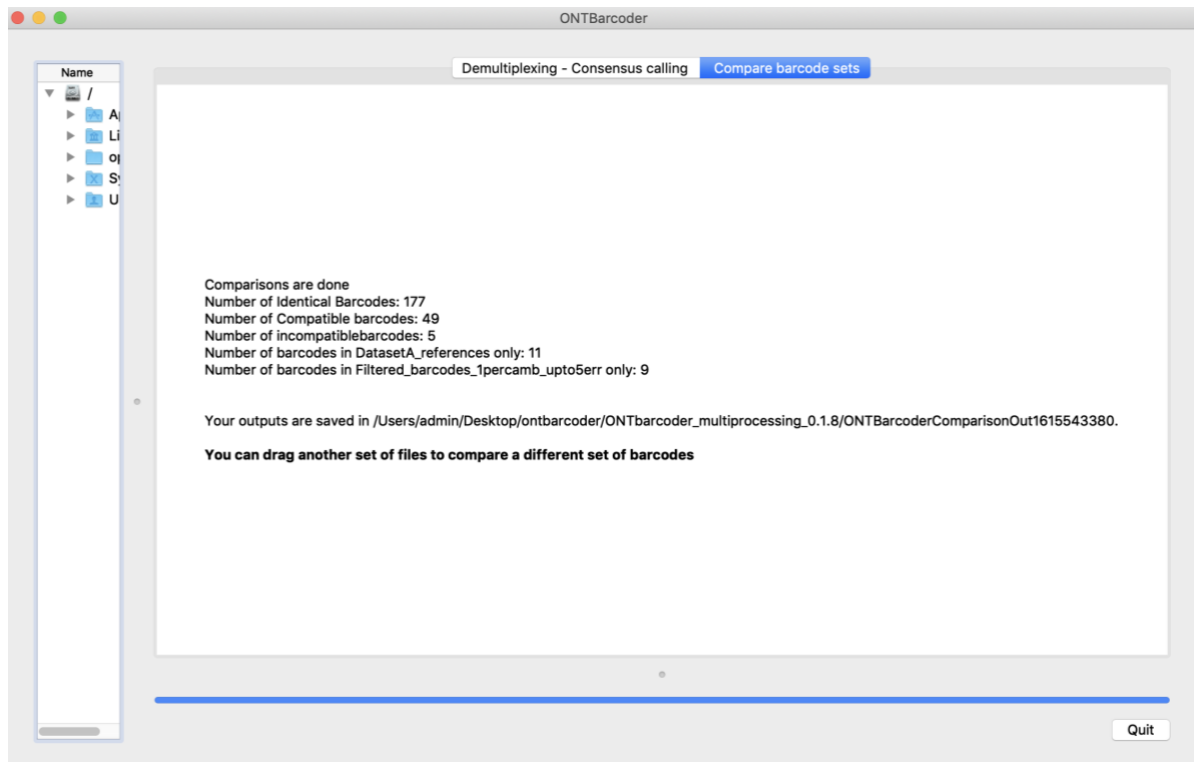
3. Select an output mode, if you select default, it will create a folder within the directory containing the software starting with ONTBarcoderComparisonOut. If you choose “Select Directory” ensure that the directory name has no empty space and is empty. Generally avoid complex characters in your names ( space, brackets, slashes, colons)



4. You can now select a set of reference barcodes if you want all other files to be compared to the barcodes in the reference set. Alternatively, you can just press “Done” to do all pairwise comparisons between the fasta files. Here, since we want the reference mode, we select the reference Sanger file and click “Done”

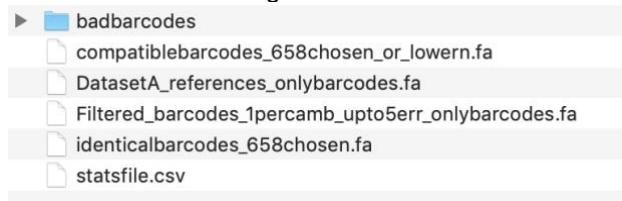


5. The output will be as shown below and the text can be copied. It is also in the output folder

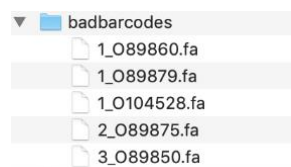


The barcodes in the sets are compared and classified into three categories: “identical” where sequences are a 100% match and lack ambiguities, “compatible” where the sequences only differ by ambiguities, and “incompatible” where the sequences differ by at least one base pair.

6. The output folder contains the following:



- “[statsfile.csv](#)” is the key summary folder having summary of the comparisons and the various incompatible barcodes with the edit distances
- Identicalbarcodes\* and compatible\* barcode files gives one sequence per dataset, in case the user wants to dereplicate barcodes
- The \*onlybarcodes.fa files contain barcodes that cant be compared
- “badbarcodes” folder contains incompatible barcodes and the value before “\_” represents edit distance.



Note that if more than 2 input files were provided, each pairwise comparison is given in a separate folder and a summary of all comparisons is then given the outermost folder

## Summary

Barcode module	comparison	Input required	Results
	Comparison of barcodes to references	<ul style="list-style-type: none"><li>One or more barcode fasta file(s)</li><li>Reference fasta file <i>See format specifications</i></li></ul>	<ul style="list-style-type: none"><li><a href="#">“statsfile.csv”</a> that describes the overall summary and erroneous barcodes</li></ul>
	Comparison of barcodes to each other	<ul style="list-style-type: none"><li>Two or more barcode fasta file(s) <i>See format specifications</i></li></ul>	<ul style="list-style-type: none"><li><a href="#">“statsfile.csv”</a> that describes the overall summary and erroneous barcodes</li></ul>

## Format Specifications

<a href="#">Files for comparison module</a>	<ol style="list-style-type: none"><li>Specimen barcodes in different files should have identical name before the following delimiter characters (“_/-/;”)</li><li>The barcodes for at least some of the specimens should be present in multiple input files</li></ol>
---------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------