MEFGL Bioinformatics Surgery – Presentation of Problems Sign-Up sheet

|  |  |  |
| --- | --- | --- |
| Date | Name | Brief description of issue |
| 22/03/17 | Georgina Brennan | How to concatenate sequences by ID?  I have three fasta files containing sequences for three barcode markers (rbcl, matk, ITS2), with the species ID of the barcoded plants in the header. I want to concatenate the sequences in the three files based on species ID. The output should be a single fasta file with a header for each species and the concatenated sequences from the three markers. Note, not all species are found in all the barcode fasta files, but the output file should keep all species even if the species ID does not match with any other header in another fasta file. If possible, I would like to keep all the ID information in the header for all sequences – but this is not essential for now.    *Example*:    **Fasta file 1:**  >Poa\_annua\_3|NMW4597|NBGW1050|NMW4598  aagttt    **Fasta file 2:**  >Poa\_annua\_2|NMW4590|NBGW1060  aggttc    **Fasta file 3:**  >Poa\_annua\_3|NMW4555|NMW4582|NBGW1076  gtgtgc    **output fasta file:**  >Poa\_annua\_3|NMW4597|NBGW1050|NMW4598….  aagtttaggttcgtgtgc |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |