Generating Hidden Markov Models Protocol

Main Objective: Develop and automate a way to expediate RNA / DNA analysis using HMM.

I am going with the assumption this is from the HMM_SEQ repository that we are using

1. First do the alignment with the families using the following command : `bash Generate_Alignment_Clustalo.sh family_list.txt`
   a. For the ones that did not get aligned due to 1 sequence, you would have to do a diff on the family_list.txt without the extension (refer to the hmm_create.fs shell script to remove extensions off of files) vs family_list.txt
2. After the alignment is complete… Note this will take awhile
3. Start building your hidden markov models with the following command : `bash hmm_create.fs family_list.txt`
4. After building the HMM we would have to create our own local database on our machine. Use the following command to open all the hmm files into a file with no extension : `cat family_HMM/* > FamilyDB`
5. After that is done, we need the index the HMMS so we can perform queries. Use the following command : `hmmpress FamilyDB`
6. Now we can start performing queries from using the following hmm_scanner.fs to make queries: i.e :
   a. `bash hmm_scanner.fs 01_kaplan_sequences_individual_sequences/2016/kaplan_1/`
7. There are two ways to extract the top hits from either the individual output files created from each query or open all the queries into one text and grepping the top hits using unique identifies.
8. Now we need to extract the top hits from each output file generated. Here is the code : `bash top_hits.fs output`