Here is your in-class homework:

The folder skink_dna contains DNA sequences for a number of lizard species. Each DNA sequence is in a unique file, where the filename is SPECIESNAME_ID.txt. Your task is to write a script that does the following:

[here I assume that your working directory contains the **skink_dna** folder, but you are NOT working within **skink_dna**]

1) Using ‘dir’, get the list of filenames from skink_dna. (`dir(‘skink_dna/’)`)

2) Using paste and/or any other functions necessary, modify each of the filenames to reflect the full path such that you can open any file from the current working directory. Thus, if you had a file ‘abracadabra.txt’ in the ‘skink_dna’ folder, the character string required to open the file from your current working directory would be

```
skink_dna/abracadabra.txt
```

Remember that the slash is required to move into different subdirectories (or folders, as you may think of them).

3) You should now have a vector of complete filenames, and you should be able to loop over the list of filenames and open each in turn from your current working directory.

4) When you have verified that you can do this, use the loop construct you’ve written above and modify it so that you loop over the list of filenames, opening each file in turn. As you open each file, you are to extract summary statistics and store them in a list.

Your list should have the following components:

> i) name: the name of the species/sample (e.g., GREEDLR0324). Note that this does not include the .txt extension. If your list is named `result`, then this component might be `result$name.`
>
> ii) The length of each sequence, e.g., `result$length`
>
> iii) The number of A nucleotides, e.g., `result$countA`
>
> vi) Number of G nucleotides
>
> vii) Number of T nucleotides
>
> viii) Number of missing/bad characters (denoted by N)
>
> ix) The number of possible start codons in the sequence. A start codon is the string ‘ATG’.

When you are done looping, convert the list to a dataframe. Column names should correspond to the list components you’ve assigned above.

Part II

Now we want to put all of those DNA sequences together in a single file. Loop over the list of filenames, open each in turn and scan the contents. You should then write both the sample name (e.g., GREEDLR0324) and the DNA sequence to a file separated by exactly 2 tab spaces. Write all the DNA sequences to the same file, with each species on a separate line. You will need to use an append=TRUE argument in whatever function you use to write the DNA sequences (cat? write? write.table?) to file. What happens if append != TRUE?

Suppose you write everything to a file `allDNA.txt`. This file would then contain the following:

```
GREEDLR0324         ACTGGGTACTGCCE etc etc.....
LEONDLR013          ACTGGGTACTGCCE etc etc.....
PANTDLR0025         ACTGGGTACTGCCE etc etc.....
HELEWAM131025       ACTGGGTACTGCCE etc etc.....
```