

nguyenet_Assignment4

June 10, 2020

1 Assignment #4

1.1 Part 1. This part does not involve computer programming.

In humans there are two sex chromosomes: X and Y. Males are XY and females are XX. In chickens there are also two sex chromosomes: Z and W. Males are ZZ and females are ZW.

The question for this part is whether or not the human X chromosome is homologous to the chicken Z chromosome. We will answer this by using BLAST to align some human genes to the chicken genome and some chicken genes to the human genome (a more advanced study would include more genes than we are considering). If the genes on the human X chromosome map to the chicken Z chromosome (and vice versa), this supports the idea that the two species' sex determination system had a common origin. If instead the genes on the human X chromosome map to chromosomes other than the chicken Z chromosome (and vice versa), this supports the idea that the two sex determination systems evolved independently.

Complete the following table:

Inputed corresponding chicken chromosome (last column) for human gene.

Inputed corresponding human chromosome (second to last column) for chicken gene.

Gene Chromosome in Humans Chromosome in Chicken

Human Gene 1 5 Z

Human Gene 2 5 Z

Human Gene 3 9 Z

Human Gene 4 9 Z

Human Gene 5 X 1

Human Gene 6 X 1

Human Gene 7 X 4

Chicken Gene 1 22 1

Chicken Gene 2 X 1

Chicken Gene 3 X 1

Chicken Gene 4 X 4

Chicken Gene 5 5 Z

Chicken Gene 6 9 Z

Chicken Gene 7 5 Z

Based on this information, do you think the human chromosome X and the chicken chromosome Z are homologous?

The information in the table do not suggest that human chromosome X and the chicken chromosome Z are homologous. Genes on the human X chromosome do not map to the chicken Z chromosome, nor vice versa. Instead, the data supports the theory that the two sex determination systems evolved independently of one another.

1.2 Part 2. This part does involve computer programming.

Consider the multiple sequence alignment file “part-ace2-multi.txt” in the Lecture 7 folder. Write a function (or functions) to go through all the alignment blocks in this file and count the number of positions where all 4 alleles appear (A, C, G, T). We are only going to count an allele if it appears as a CAPITAL letter.

```
[6]: # import regular expressions
import re
```

Function 1

```
[7]: def ctPosACGT(listSeq):
    '''counts positions where all four alleles are present across sequences'''
    count = 0
    for j in range(len(listSeq[0])): # walks through each position in the
    ↪specie sequence; all sequences same length
        ctACTG = 0 # keeps count of how many bases found at said position
        for base in 'ACGT': # walks through each base alphabetically
            for i in range(len(listSeq)): # walks through each specie sequence
                if listSeq[i][j] == base:
                    ctACTG += 1 # increase 'ctACTG' by 1 if base is found
                    break # break out of loop for looking through each specie;
    ↪moves onto next base
            if ctACTG == 4:
                count += 1
    return count
```

Main Function

```
[8]: def allPosACGT(file):
    '''looks across all blocks in multiple sequence alignment file,
    and counts all positions where all four alleles are present across
    ↪sequences'''
    totPosACGT = 0 # keeps track of total count
    infile = open(file)
    for line in infile:
        if re.match("a score", line): # new block found
```

```

speciesSeq = [] # keeps track of seq for each specie in block
for subLine in infile:
    if re.match("s ", subLine): # species found
        temp = re.split("\s+", subLine)
        speciesSeq.append(temp[6])
    elif re.match("\s+", subLine): # end of block found
        break # break out of this loop
    blockPosACGT = ctPosACGT(speciesSeq) # for block, count # of
    ↪positions with all bases ACGT
    totPosACGT += blockPosACGT # add to 'totPosACGT' count
infile.close()
return totPosACGT

```

Run “part-ace2-multi.txt” through main function, which will go through all the alignment blocks in this file and count the number of positions where all 4 alleles appear (A, C, G, T).

```

[9]: ace2 = allPosACGT("part-ace2-multi.txt")
    print(ace2)

```

22