Name:

Comparative Genomics

RepeatMasker Lab

Genome Setup

1. After downloading, uncompressing, and creating the symbolic links, use ls to view the contents of the project directory and paste a screen shot into the lab worksheet.
2. Is the human Y chromosome more AT-rich than the chicken W chromosome? Base your answers on the hints in the lab text file and show your work. DO NOT INCLUDE Ns IN YOUR ESTIMATES!

Analysis

1. After running RepeatMasker for both the human and chicken datasets, ls the contents of your project directory and paste a screen shot here.
2. less human.chrY.log

What is the name of the search engine?

What is the title of the database that was used?

View the species/taxa search. What does the ‘lineage’ represent?

1. less chicken.chrW.log

What is the name of the search engine?

What is the title of the database that was used?

View the species/taxa search. What does the ‘lineage’ represent?

1. less human.chrY.fa.tbl

Does the GC level listed match your answer from the grep results above?

How many bases were masked in the human Y chromosome?

What is the most abundant type and clade of TE on the human Y chromosome? Support your answer.

1. less chicken.chrW.fa.tbl

Does the GC level listed match your answer from the grep results above?

How many bases were masked in the chicken W chromosome?

What is the most abundant type and clade of TE on the chicken W chromosome? Support your answer.

Let's compare our results to the complete genome repeat analyses on the RepeatMasker website.

First, human:

1. According to the repeat landscape for the human genome, what is the most abundant *recently active* type of transposable element in the human genome?
2. Are the number of ancestral and lineage-specific repeat families in your analysis compatible with what is listed for RepeatMasker Library db20140131? Provide an explanation for any discrepancy.
3. Are the percentage of interspersed repeats for the whole genome comparable to your RepeatMasker search of the Y chromosome? What about if you just consider the regions excluding Ns?

Now, chicken:

1. According to the repeat landscape for the chicken genome, what is the most abundant *recently active* type of transposable element in the chicken genome?
2. Are the number of ancestral and lineage-specific repeat families in your analysis compatible with what is listed for RepeatMasker Library db20140131? Provide an explanation for any discrepancy.
3. Are the percentage of interspersed repeats for the whole genome comparable to your RepeatMasker search of the W chromosome? What about if you just consider the regions excluding Ns?