*R*Pubs *by RStudio*

Making quick maps of multigene families using blast, genome assemblies and gggenes

Introduction

This module will help you go from a set of long-read genome contigs to a simple representation of a complex multigene family like the major histocompatibility complex. This is not a substitute for full-on annotation, but it may be useful if you want a quick overview of a multigene family across several haplotypes of the same species. We will use the scrub jay MHC as an example and we'll assume that you have already generated formatted tables of blast hits using the command-line blast. For this you will need to setup a blast database of your genome to be blasted, as well as find the appropriate query sequences of genes you are interested in representing. I have found these resources to be helpful in doing this: https://www.ncbi.nlm.nih.gov/books /NBK569856/ (https://www.ncbi.nlm.nih.gov/books/NBK569856/) and https://www.metagenomics.wiki/tools /blast/blastn-output-format-6 (https://www.metagenomics.wiki/tools/blast/blastn-output-format-6). For example, let's imagine I have several genomes (fasta files) ending with the suffice "ren". On the Cannon cluster, you can run the command:

```
module load blast/2.6.0+-fasrc01
for file in *ren*;
  do makeblastdb -in "$file" -dbtype nucl -out ${file%.*};done
```

You can then make an 'uber' database of all your genomes (in this case, the "MCZ" files) with this command:

```
blastdb_aliastool -dblist "MCZ_orn_365326.hap1.fa_ren MCZ_orn_365326.hap2.fa_ren MCZ_
orn_365338.hap1.fa_ren MCZ_orn_365338.hap2.fa_ren \
MCZ_orn_366490.hap1.fa_ren MCZ_orn_366490.hap2.fa_ren MCZ_orn_366493.hap1.fa_ren MCZ_
orn_366493.hap2.fa_ren MCZ_orn_366494.hap1.fa_ren MCZ_orn_366494.hap2.fa_ren" \
-dbtype nucl -out scrubjay_9_haps -title "10 scrub jay renamed genomes combined"
```

Getting your blast output together

Now by blasting to this library (called "scrubjay_9_haps_ren") you can blast to all the individual libraries simultaneously. You will want to set your output format when you blast so that it generates a table that is easily manipulated. I did this by using the -outfmt "6" command, further specifying the precise columns to generate using the codes available here (https://www.metagenomics.wiki/tools/blast/blastn-output-format-6). So the commmand to generate a table for mhc class II exon 3 using part of this scrub jay sequence as a probe (https://www.ncbi.nlm.nih.gov/nuccore/U23958.1) might look like this:

```
blastn -db scrubjay_26_haps -query Mhc_U23958.1_exon3.fa -out sj_mhc2_exon3_blast.out
-task megablast -outfmt "6 qseqid sseqid pident sstart send sstrand"
```