### Part 1: Find Genes in Candidate Regions [4 pts total]

Find all of the **protein-coding genes** in the *Calypte anna* (Anna's hummingbird) genome that fall **within the candidate regions** under selection. Make sure that you are using the latest genome build: bCalAnn1\_v1.p. When you have gotten your final list of genes, be sure to save or download your results to a separate file, as you will be referring to them again later.

#### 1A. What is the **total number** of protein-coding genes that fall within a candidate region *(1 pt)*?

There are 17 protein-coding genes.

#### 1B. How did you find these genes *(3 pts)*?

We used NCBI’s tools and filters to find them using the *Calypte anna* organism, protein-coding genes, and falling within the candidate regions:

"Calypte anna"[Organism] AND ("genetype protein coding"[Properties] AND alive[prop]) AND ((NC\_044245[nucl\_accn] AND 000056181726[CHRPOS] : 000056876451[CHRPOS]) OR (NC\_044248[nucl\_accn] AND 000032923716[CHRPOS] : 000033354545[CHRPOS]) OR (NC\_044274[nucl\_accn] AND 000043020578[CHRPOS] : 000043645624[CHRPOS])) AND alive[prop]

### Part 2: Find Known Genes Related to Pigmentation/Melanin [4 pts total]

Find all of the **protein-coding genes** in **humans** that have known **phenotypes and/or diseases** related to pigmentation and melanin. We want to cast a wide net to make sure we get all of the genes, so you should search for several variations on the terms "pigment" and "melanin": pigmentation, pigmented, hyperpigmented, hyperpigmentation, hypopigmented, hypopigmentation, melanocyte, melanosome, and melanin.

Once you have the results of your search(es), again be sure to save them to a separate file so that you can refer back to them for parts 3. and 4.

#### 2A. What is the **total number** of protein-coding genes from human with known phenotypes related to melanin and/or pigmentation *(1 pt*)?

There are 35 genes.

#### 2B. How did you find these genes *(3 pts)*?

This was found in NCBI, looking for phenotypes or diseases associated with pigmentation/melanin, and protein-coding genes through the human organism:

(pigment[Disease/Phenotype] OR pigmentation[Disease/Phenotype] OR pigmented[Disease/Phenotype] OR hyperpigmentation[Disease/Phenotype] OR hypopigmentation[Disease/Phenotype] OR melanocyte[Disease/Phenotype]) AND "Homo sapiens"[porgn] AND "genetype protein coding"[Properties]

### Part 3: Find the Overlap Between Gene Sets [3 pts total]

Figure out which, if any, of the genes in your candidate gene list are also found in the list of known pigmentation/melanin genes. Note that comparing 2 lists of genes is *not* something that would be very straightforward to do within NCBI, so you will need to come up with your own method to do this. It could involve using Python or R or some of the shell scripting techniques we have talked about, or could even be done in Excel.

*The hummingbird genes use the same nomenclature as the human genes, so you do not need to worry about doing any kind of conversion to the names-you can compare the gene lists directly.*

#### 3A. Which gene(s) from the candidate gene regions overlap with the known pigmentation genes *(1 pt)*?

**TYRP1** was the gene from the overlap with candidate gene regions overlap with the known pigmentation genes.

N300\_15523

tyrosinase related protein 1

5,6-dihydroxyindole-2-carboxylic acid oxidase

#### 3B. How did you find the overlap *(2 pts)*?

We used R:

a <- read.table("./lab1pt1gene\_result.txt",sep='\t',header=T)

b <- read.table("./lab1pt2gene\_result.txt",sep="\t",header=T)

humans <- vector()

for (i in 1:nrow(b))

{

#print(b[i,"GeneID"])

humans<-c(humans,b[i,"Symbol"])

}

#humans #Works, gives 35 IDs

bird <- vector()

for (j in 1:nrow(a)) {

bird<-c(bird,a[j,"Symbol"])

}

#bird

intersect(humans, bird)

### Part 4: Find More Details on the Top Candidate Gene [4 pts total]

If you did not find any overlap in the gene lists in Part 3, then just select the gene of your choice from the list of known human pigmentation genes and use that to answer the following questions. If you found multiple genes overlapping between the 2 lists, then just select *one* and use it to answer the questions.

<https://www.ncbi.nlm.nih.gov/gene/7306>

#### 4A. Where (i.e., in which tissue) is this gene most highly expressed in humans? *(1 pt)*

#### 

#### *It is most highly expressed in the skin for humans.*

#### 4B. What biological processes is this gene involved in? *(1 pt)*

#### 

#### *It is involved in the melanin biosynthetic pathway (positive regulation), acetoacetic acid metabolic process, melanocyte differentiation, and melanosome organization.*

#### 4C. What family of proteins does the product of this gene belong to, and how many other members of this family are found in humans? *(1 pt)*

#### 

#### *It belongs to the tyrosinase family. There are 15 members of this family found in humans. (*[*https://www.uniprot.org/uniprotkb?facets=model\_organism%3A9606&query=%28family%3A%22tyrosinase%20family%22%29*](https://www.uniprot.org/uniprotkb?facets=model_organism%3A9606&query=%28family%3A%22tyrosinase%20family%22%29)*)*

#### 4D. What (if any) human diseases are associated with mutations in this gene/protein? *(1 pt)*

#### *Defects in this gene are the cause of rufous oculocutaneous albinism and oculocutaneous albinism type III.*