|  |  |
| --- | --- |
| **Louse Case Study:**  **Part 3 – Genetics, Genomics, and Gene Function** | Wikimedia Commons |

## Objectives

Upon completion of this 3-part module, each student should be able to:

* Evaluate how an organism’s form and function is influenced by their genes, evolutionary history, interspecific interactions, and environmental selection pressures.

Upon completion of this week’s submodule, each student should be able to:

* Predict gene function based on exploration of related protein domains.
* Evaluate a gene for evidence of selection.

**Introduction**

An organism's complete set of genetic materials is called its **genome**. This includes the DNA found in the nucleus, mitochondria, and plastids (in the case of plants and algae). The DNA that gets translated into proteins makes up the coding regions. In humans, only 2% of the genome actually codes for proteins, the rest of the genome is made up of non-coding regions. **Genetics** and **genomics** are two fields of molecular biology that study genes and often get confused. Genetics is the study of single genes and their inheritance. Genomics is the study of the sequence, organization, structure, expression, interaction, and regulation of whole genomes.

Genomes are shaped by evolutionary history, inherited from ancestors and modified by mutation and natural selection. At the most proximal level, the form and function of an organism is determined by its genes. Today you will be investigating a louse gene with an unknown function to determine if it might be important in the evolution of the lice ecomorphs.

**Activity 1: Predicting Gene Function**

Since genes code for proteins and proteins determine form and function, lice species with different ecomorphs must have specific genetic differences that account for their different shapes.In this activity, you will investigate the genes of feather lice ecomorphs to look for patterns associated with their differences in form and function. The genomics work we will be doing draws from louse genomes Dr. Julie Allen (University of Nevada, Reno) has sequenced as part of her research. We do not know what the genes you will investigate do, where they are expressed in the cell, or how they evolved. We don’t know if these genes are related to the ecomorph divergence or how natural selection might be affecting their evolution. Your work will be helping Dr. Allen to better understand the genomes of the organisms she studies.

You will be assigned homologous protein sequences from one gene for 8 bird lice species – 3 from head ecomorphs, 3 from body ecomorphs, and 2 from wing ecomorphs. By **homologous**, we mean that the genes or proteins share a common ancestry – they must be derived from some ancestral gene or protein. Most of the time, homologs in different organisms are still doing the same job as the ancestral gene or protein. In this case, the homologs are called **orthologs**. Occasionally, a gene gets duplicated and one copy may end up mutating enough that it takes on a different function. Such homologs would be called **paralogs** – same ancestry, but now different functions. A homologous outgroup sequence from human head lice will also be provided.

You will be generating hypotheses about the function(s) of your gene and using several web-based genomic analysis tools to look for evidence that your gene might be involved in the development of the different ecomorphs.

The lice genera you will be working with are listed in the table below.

Table 1. Lice genera, ecomorphs, and hosts.

|  |  |  |
| --- | --- | --- |
| **Genus** | **Ecomorph** | **Host** |
| *Chelopistes* | body | Chacalacs |
| *Saemundssonia* | head | Gull |
| *Quadraceps* | wing | Gull |
| *Osculotes* | body | Hoazin |
| *Pessoaiella* | wing | Hoazin |
| *Ibidoecus* | head | Ibis |
| *Stirnidoecus* | head | Purple Starling |
| *Discocorpus* | body | Tinamou |

**BLAST (Basic Alignment Search Tool)**

The National Center for Biotechnology Information (**NCBI**), which is part of the National Institutes of Health, provides access to biomedical and genomic information for scientific and medical researchers. NCBI includes a wide variety of databases, as well as tools for analyzing the data they contain. Researchers submit known DNA and protein sequences from organisms they study into these data bases. **BLAST** is a program that allows researchers to input a nucleotide or amino acid sequence and search the databases at NCBI to try to identify the source of the sequence or find closely related sequences. In this activity, you will BLAST your protein sequence to investigate the possible function of the gene you have been assigned.

*Procedure*

1. Navigate to the NCBI website: <https://www.ncbi.nlm.nih.gov>
2. Click on the BLAST link in the menu on the right.
3. Click on the box labeled “Protein BLAST”.
4. Select one of your protein sequences (include the genus name) and copy it.
5. Paste the sequence into the text box in BLAST.
6. Click the button labeled “BLAST”. This processing step may take a while (especially if a large group of students are all using the site at the same time).
7. Once the results are returned, take a look at the matches presented in the region labeled “Sequences Producing Significant Alignments”. **Write the description of the top hit, including species name if present, in the space below. Do you recognize this species? If not, do a web search. What is the common name of the species?**
8. Select the tab labeled “Graphic Summary” from just above your results. Click on the first image (below the statement “click on the image below for detailed results”). This will take you to the NCBI conserved domain database. This site has information about proteins similar to the one you are investigating. Read any information provided about your protein. There may be some information at the top and additional information will pop up if you move your mouse over the regions of the figure. **Summarize the information about the protein function below.**

**Pfam**

The Pfam database is a large collection of protein families. Proteins can be characterized by functional regions, or domains, that have been determined by researchers to have specific functions. Proteins that have similar domains are said to be in the same “protein family”. Even though the function of the protein we are investigating is not known, if we can identify some of these known domains in the structure of our protein, we can narrow down which protein family it may belong to and therefore, what the protein’s function might be in lice.

*Procedure*

1. Navigate to the PFam (Protein Family) website (<http://pfam.xfam.org>). Click on sequence search, enter your protein sequence in the text box, and click “Go”. **What protein families are identified as significant matches?**
2. Click on at least the top two matches and read the information about the protein families (you may need to try different tabs to find the information). Some of the language in these descriptions will be complex. Use web resources to try to help you understand what the functions of these protein families are. **Summarize this information below in your own words**.

**PSORT (Protein Subcellular Localization Prediction)**

One piece of information that would help you determine the function of your gene would be where the protein expressed by the gene is located in the lice cells. For example, if you could determine that the protein was only found inside the mitochondrion, this could indicate that the protein is involved in cellular respiration.

PSORT is an algorithm that predicts the subcellular localization sites of proteins based on their amino acid sequences. The output will be presented with these abbreviations below. Dual localization is indicated with a dash between two abbreviations, e.g. cyto\_nucl indicates a protein would be found in both the nucleus and the cytoplasm. Higher numbers represent a stronger match to that cellular region.

cyto: cytoplasm

nucl: nucleus

cysk: cytoskeleton

extr: extracellular

mito: mitochondria

pero: peroxisome

plas: plasma membrane

chlo: chloroplast

E.R.: Endoplasmic reticulum

Golg: Golgi body

vacu: vacuole

lyso: lysosome

Procedure

1. Navigate to the PSORT (Protein Subcellular Localization Prediction) website (<https://wolfpsort.hgc.jp/>). This tool will make a prediction about where in the cell your protein might be located. Paste your protein sequence into the text box. Select “animal” as organism type and hit “submit”. **Once the results are shown, record the output below. Based on these results, where is it most likely that your protein is localized within the lice cells?**

**Activity 2: Evidence of Selection on your Gene [Sequence Comparison]**

In this activity, you will be comparing the amino acid sequences of the eight lice species from three different ecomorphs. If your gene were important in the expression of the different ecomorphs, we would expect that there would be differences in the DNA sequences for your gene between the different ecomorphs. If these differences in the DNA sequences lead to differences in the amino acid sequences, the folding of your protein might be altered in ways that affect the morphology of the lice.

Your instructor will provide you with aligned protein sequences.

*Procedure*:

1. View the aligned protein sequences.
2. If an asterisk (\*) is shown beneath a column of letters, this indicates that every genus in your tree has the same amino acid in this spot. Ignore these: they are not informative.
3. If two dots are shown beneath a column, there is variation in this amino acid among the genera, however the mutation is not expected to result in a change in protein structure or function. If one dot is shown beneath the column, the amino acid substitution is also expected to be minor.
4. Focus on the columns that do not have a mark under them. These contain amino acid differences that would be predicted to change protein form and function. Exam these columns.
5. If there are any sites where all genera from one ecomorph share the same amino acid substitution (different from the other ecomorphs), this would be a gene to examine further for an influence on ecomorph development.
6. Are there any sites showing this pattern? If so, how many sites show this pattern? [**Note**: It is not expected that you will find many, if any, of these sites]

**Summary**

Summarize what you know about your gene, the possible function, where it is localized in the cell, and whether there is any evidence that this gene might be related to the determination of lice ecomorphs. Google additional resources to go beyond what you learned in lab. Use your own words and deductions. You will not be evaluated on whether you are correct (in fact, no one knows yet if you are), but on the clarity of your discussion and evidence of diligent effort.