

2020

Biological Pathways Associated with Wild and Domestic Animals

Cayla Steemer
Prairie View A&M University

Laura Carson
Prairie View A&M University

Jim Hu
Texas A&M University

Gloria Regisford
Prairie View A&M University, gcregisford@pvamu.edu

Follow this and additional works at: <https://digitalcommons.pvamu.edu/pursue>



Part of the [Bioinformatics Commons](#), [Biology Commons](#), and the [Genetics and Genomics Commons](#)

Recommended Citation

Steemer, Cayla; Carson, Laura; Hu, Jim; and Regisford, Gloria (2020) "Biological Pathways Associated with Wild and Domestic Animals," *Pursue: Undergraduate Research Journal*: Vol. 3 : Iss. 1 , Article 4.
Available at: <https://digitalcommons.pvamu.edu/pursue/vol3/iss1/4>

This Article is brought to you for free and open access by Digital Commons @PVAMU. It has been accepted for inclusion in Pursue: Undergraduate Research Journal by an authorized editor of Digital Commons @PVAMU. For more information, please contact hvkoshy@pvamu.edu.

Biological Pathways Associated with Wild and Domestic Animals

Cover Page Footnote

A special thank you to: • CBG Grant awarded to EGCR • Department of Biology, Prairie View A&M University • Cooperative Agricultural Research Center, Prairie View A&M University • College of Agriculture and Human Sciences, Prairie View A&M University



Biological Pathways Associated with Wild and Domestic Animals

Cayla Steemer,¹ Laura Carson, PhD,¹ Jim Hu, PhD,² and Gloria Regisford, PhD³

¹Department of Agriculture, Nutrition and Human Ecology, Prairie View A&M University

²Department of Biochemistry and Biophysics, Texas A&M University

³Department of Biology, Prairie View A&M University

Corresponding Author

Dr. Gloria Regisford

Department of Biology

Prairie View A&M University

PO Box 519: MS 2008

Prairie View, TX 77446

Phone: (936) 261 -3165

Email: gcregisford@pvamu.edu

Abstract

Background: Zoonotic diseases are problematic, in that, they impact both wild and domestic animals alike. Thus, there is a need to investigate the genomes of wild and domestic. Gene ontology (GO) is a major bioinformatics initiative, whereby descriptions of gene products across the database are developed and unified to describe all species. This process is performed by biocurators, who gather, annotate, and validate information on the databases, consequently affording easy access to accurate and updated data. In this study, we investigated the biocuration of two biological processes, DNA integration, which is used for DNA coding, and the tricarboxylic acid (TCA) cycle that occurs in all aerobic organisms. The objective of this study was to compare the genomes of 271 mammals, birds, reptiles and some aquatic animals to determine the number of wild versus domesticated animals, where DNA integration and the TCA cycle have annotations. We hypothesized that there would be more annotations on domesticated animals than wild animals because of easier access to domesticated animal genomes. **Methods:** To test this hypothesis, we first accessed the National Center of Biotechnology Information (NCBI) to retrieve the taxonomy ID for 271 species of animals found in this study. Then the European Bioinformatics Institute (EBI) database, QuickGO, was accessed to retrieve all annotations associated with the taxonomy ID of the species. Data was assembled into a wiki-database that is now publicly available online. **Results and Conclusion:** Data indicated that more annotations for DNA integration and the TCA cycle were 22% higher in domestic animals than in wild animals. Therefore, we propose that more biocuration needs to be done for wild animals. The number of wild animals' sequences available are growing, but if they are not annotated, detailed investigations are not possible. If more resources are dedicated to the investigation of the genomes of wild animals, more work can be done to study the genetic factors affecting zoonotic diseases.

Keywords: bioinformatics, zoonotic diseases, wildlife, domesticated animals

Introduction

Information is constantly being acquired in the life sciences, and with an abundance of informational databases, it is a necessity to keep collected works up to date and accessible to the rest of the world. Genomic databases are used by many people, including students and researchers (Hayamizu, 2015). Furthermore, databases have become an essential part of the biological sciences, because they often serve as gateways to biological data (Binns, 2009). Biocurators take information from scientific literature and describe the data using annotation protocols (Jenkinson, 2008). Additionally, biocurators assign Gene Ontology terms (GO terms) to a specific gene product, referred to as GO annotation. These GO annotations represent a biological process, molecular function or cellular components (Blake, 2014). The European Bioinformatics Institute (EBI) database collects and stores data from life science experiments and provides many services and tools such as Quick Gene Ontologies (QuickGO) (Squizzato, 2015). Quick GO is a GO annotation database that allows scientists to share data, perform complex queries and analyze results (Ashburner, 2000).

There are several annotations that could be used to compare the genomes of wild and domesticated animals. For this experiment, the annotations for the Tricarboxylic Acid (TCA) cycle and DNA integration were chosen and the rationale for this choice follows. Both the TCA cycle and DNA integration processes are known to be functional in animals. Although these two processes are individually being studied by scientists, the number of annotations for these processes in domestic and wild animals is unknown. Because many species are the subject of current genomic investigations, the sequencing data of wild and domestic animals are now publicly available and will allow for a comparison.

The TCA cycle is a very important biological process that produces energy and is known to occur in all animals (Alisdair, 2004). The energy production cycle is also found in most plants, animals, fungi, and bacteria. In this experiment, we chose the TCA cycle as our control, as it is found in almost all genomes. Thus, the number of TCA cycle annotations should be consistent regardless of whether the species is wild or domesticated.

DNA integration is the biological process that controls the integration of foreign material into the DNA of a host cell. Mammalian genomes are susceptible to foreign DNA insertions both naturally and experimentally. Meaning that viruses that are DNA- or RNA-based viruses can integrate into the mammalian genomes with ease

(Müller, 2001). Mechanisms involved in the illegitimate integration of DNA could also be involved in viral DNA integration (Würtele, 2003). Viral DNA integration and retroviral DNA integration share similar characteristics as DNA integration in mammalian cells, but they are only found within prokaryote cells. DNA integration is represented in the host cell of the mammalian genomes. Therefore, we chose to use DNA integration annotations, as these annotations mark areas of a genome that have been changed or impacted by a virus within the host cells. We want to isolate this biological process for its annotations within the wild and domesticated animal genomes, but since DNA integration is not as prominent as TCA cycle, there might not be a consistency of annotations.

The objective of this study is to compare the genomes of 271 variety of mammals, birds, reptiles and aquatic animals to determine the number of wild versus domesticated animals, in which DNA integration and the TCA cycle are annotated. Since the TCA cycle and DNA integration occurs in all organisms and are both important biological processes for the organisms studied, we hypothesized that there would be more annotations on domesticated animals than wild animals, because of easier access to domesticated animals for genetic investigations.

Methods

The methods used involved utilizing three main databases. National Center of Biotechnology Information (NCBI), European Bioinformatics Institute (EBI) and a wiki-database. The wiki-database was created to house the information collected in this study. For this experiment, we focused on mammals, reptiles, fish, and birds. All data was transferred into the wiki-database on the “Species Table Page.” (Figure 1). A separate table was created for each animal that was investigated. Each species table shows the following genomic information: the common name, category, genomic status (genome sequencing), the research lab (in which the genome was sequenced), link sources among other (Figure 1).

The wiki-database houses all the information collected on animals, plants and insects. We separated the species into different categories that determined the importance of those species. The wiki-database housed the genomic information needed to create a summary of each “Species Table” for the 271 species investigated.

The screenshot shows a Wiki-database page titled "User:Shabnam/PVAMU Undergraduate Research". At the top, there is a navigation bar with links for "Read", "Edit", "View history", and "More". A search box is located on the right. Below the navigation bar, there is a note: "Please do not use <h1> style headings (a heading wrapped in a single = sign) - it is visually indistinguishable from the page title." followed by "Next Group Meeting TBA".

The main content area includes:

- Objectives** [edit]: A list of four objectives:
 1. Find species that are agriculturally important.
 2. Assess GO annotations
 3. Pick biological pathways.
 4. Provide a survey of functional annotation for selected pathways for agriculturally important species of your interest.
- List of Agriculturally Important Species** [edit]: A section titled "Some possibly useful resources:" with a link to "USDA Agricultural Research Service".
- Species Table**: A table with the following data:

Category	Species Name	Common Name	NCBI Taxon ID	Genomic Status	Research Lab(s)	Link(s)	Additional Notes
Insect (Pest)	Icerya purchasi	Cottony Cushion Scale	249532	No Data		<ul style="list-style-type: none"> Wikipedia Wikipedia Texas A&M Agrilife Extension Biocontrol University of California Riverside Journal of Economic Entomology International Organization for Biological Control 	Icerya Purchasi is a pest to citrus and pittosporum.

Figure 1: Wiki-database for information collected on the species investigated.

To use NCBI, we first had to find the binomial nomenclature for each animal using their common names. Next, we identified whether the species was wild or domesticated by using information stored in the NCBI Taxonomy database. A new table was created for each of the 271 species investigated

We assessed Gene Ontology (GO) annotations in the EBI, QuickGO database for each species. Lastly, we inserted GO annotation data found into our "GO Annotation Survey". The "GO Annotation Survey" is linked to each species in the "Species Table" under their species name. The "GO Annotation Survey" shows display the NCBI taxonomy ID at the top. Then goes into further detail of the breakdown of different categories that are shown in the far-left column in the survey. The middle column houses the proteins/annotations found for that taxonomy ID, then the next two columns display the individual biological pathways and their proteins/annotations found.

Results

The graphs in Figures 2 & 3 break down the biological pathways for species with annotations versus the total number of species investigated. Below the graph shows the overall comparison of the domesticated species and the wild species and how they vary within the evidence found.

Of the 71 domesticated mammal, fish and reptilian species examined only 1% (1/71) had DNA integration markers annotated, 8% (6/71) had the TCA cycle and 28% (20/71) had both DNA integration and TCA cycle markers annotated in their genomes. The number of annotations found for each biological pathway was not in abundance. Yet, enough to prove that there are annotations for the domesticated species that were found in this study.

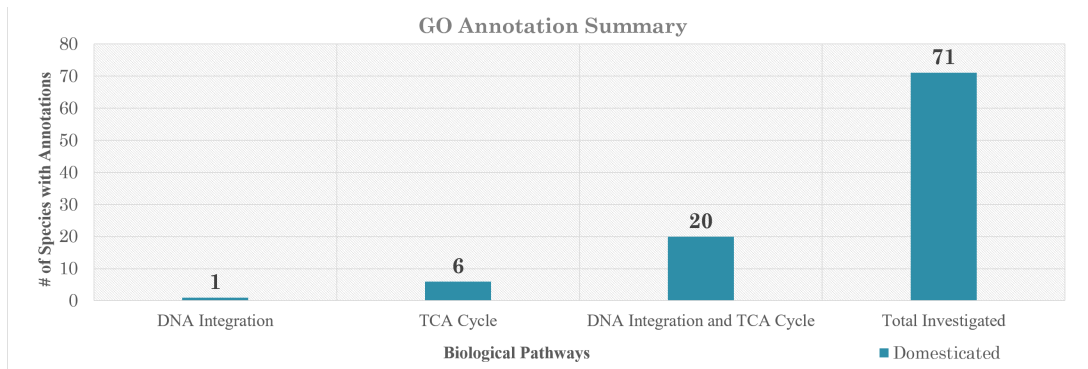


Figure 2: GO Annotation Summary

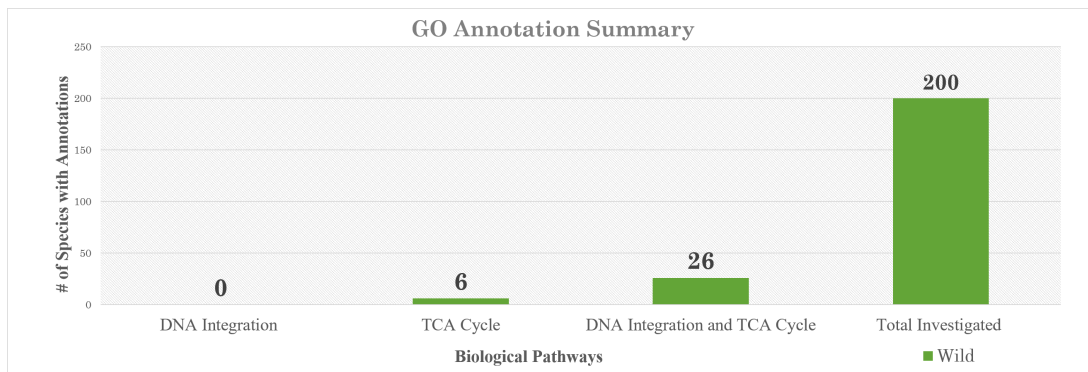


Figure 3: Wild Species Go Annotation Summary

Of the 200 wild mammal, fish, and reptilian species investigated, 0% (0/200) had DNA integration alone, 3% (6/200) had TCA cycle and 13% (26/200) had both DNA integration and TCA cycle. The number of annotations found for each biological pathway was not as expected but results indicate that some wild species did have annotations associated with their species.

Overview of GO Annotation Survey

After identifying the genomes with annotations using the GO Annotation surveys. We formulated Wiki species pages. Table 1 shows the GO Annotation survey for *C. hircus*. It shows the important components addressed in the methods.

As shown in Table 1 below, you will see that it displays the NCBI Taxonomy ID at the top. The left column below the Tax ID shows all the proteins and GO Annotations for the species. The right side shows the GO term and ID for the TCA cycle and DNA integration. The Parent GO ID and term name are for the biological processes investigated. The survey also shows the total number of GO Annotations which are broken down into the evidence and the aspects. For each species, they have annotations that display electronic and manual evidence and then they are categories in either molecular function, biological processes, or cellular components.

Table 1: Survey of Goat Species Wiki Page and GO Annotation

NCBI Taxonomy ID	9925		
Parent GO ID	GO:0006099	GO:0015074	
Parent GO Term Name	TCA	DNA Integration	
Number of Proteins	22,479	2	0
Total Number of GO Annotations	21,565	2	0
Electronic Evidence	20,850	2	0
Manual Evidence	21	0	0
Molecular Function Aspect	6,555	0	0
Biological Process Aspect	8,080	2	0
Cellular Component Aspect	7,844	0	0

There are two main types of evidence used to separate the annotations, electronic and manual. Electronic evidence is a simple algorithm that biocurator uses to search for the desired gene products such as the TCA cycle pathway. These are automatic annotations imputed by biocurators, but they are essentially hypothetical annotations. This causes the total number of GO annotations to fluctuate because electronic evidence will be deleted or added at any time. Thus, the importance of the need for manual evidence as well. In QuickGO, there are multiple different types of manual evidence used but we chose to look for manual evidence that is found from experimentation. This type of manual evidence is important because it provides accurate annotations for each species that we investigated.

There are three different aspects used to categorize each annotation. The molecular function is how the gene product performs, the biological process is that gene products involved in, and which cellular components the gene products are found. We input data at each aspect for the total 271 species. Both the TCA cycle and DNA integration both fall into the aspect of the biological process aspect.

We compare the differences between both the wild and domesticated species to test our hypothesis that there would be more annotations on domesticated animals than wild animals. Shown in Tables 2 through 4, the domesticated species have a higher amount of annotations when compared to our wild species in Tables 5 through 8. The *B. taurus* (cattle) gene products have many protein products and some annotations for the biological processes. The *O. cuniculus* (rabbit) and the *F. catus* (cat) being that they are both domesticated species are like the *B. taurus* in terms of their GO Annotation surveys. Shown in Tables 5 and 6, we have two of the wild species, *P. Alecto* (black flying fox) and *P. troglodytes* (chimpanzee) used for research that possess a high amount of annotations, yet shown in Tables 7 and 8, the other wild species, *S. tatarica* (Saiga antelope) and *P. lotor* (raccoon), a low amount of gene products are found along with any for the biological processes of the TCA cycle and DNA integration.

Table 2: Survey of Cattle Species Wiki Page and GO Annotation

NCBI Taxonomy ID	9913		
Parent GO ID	GO:0006099	GO:0015074	
Parent GO Term Name	TCA	DNA Integration	
Number of Proteins	27,886	54	6
Total Number of GO Annotations	353,986	119	7
Electronic Evidence	297,431	107	6
Manual Evidence	3,345	0	0
Molecular Function Aspect	92,295	0	0
Biological Process Aspect	146,821	119	7
Cellular Component Aspect	114,870	0	0

Table 3: Survey of Rabbit Species Wiki Page and GO Annotation

NCBI Taxonomy ID		9986	
Parent GO ID		GO:0006099	GO:0015074
Parent GO Term Name		TCA	DNA Integration
Number of Proteins	19,673	33	10
Total Number of GO Annotations	223,372	57	10
Electronic Evidence	218,338	57	10
Manual Evidence	739	0	0
Molecular Function Aspect	60,217	0	0
Biological Process Aspect	94,671	57	10
Cellular Component Aspect	68,484	0	0

Table 4: Survey of *Cat* Species Wiki Page and GO Annotation

NCBI Taxonomy ID		9685	
Parent GO ID		GO:0006099	GO:0015074
Parent GO Term Name		TCA	DNA Integration
Number of Proteins	25,671	38	12
Total Number of GO Annotations	264,365	76	12
Electronic Evidence	234,771	67	12
Manual Evidence	6	0	0
Molecular Function Aspect	76,197	0	0
Biological Process Aspect	19,542	76	12
Cellular Component Aspect	79,626	0	0

Table 5: Survey of Black Flying Fox Species Wiki Page and GO Annotation

NCBI Taxonomy ID		9402	
Parent GO ID		GO:0006099	GO:0015074
Parent GO Term Name		TCA	DNA Integration
Number of Proteins	14,382	21	2
Total Number of GO Annotations	78,874	36	2
Electronic Evidence	78,799	36	2
Manual Evidence	0	0	0
Molecular Function Aspect	30,814	0	0
Biological Process Aspect	24,093	36	2
Cellular Component Aspect	23,967	0	0

Table 6: Survey of Chimpanzee Species Wiki Page and GO Annotation

NCBI Taxonomy ID		9598	
Parent GO ID		GO:0006099	GO:0015074
Parent GO Term Name		TCA	DNA Integration
Number of Proteins	78,081	121	19
Total Number of GO Annotations	533,813	224	19
Electronic Evidence	506,230	216	19
Manual Evidence	5	0	0
Molecular Function Aspect	177,010	0	0
Biological Process Aspect	197,850	224	19
Cellular Component Aspect	158,953	0	0

Table 7: Survey of Saiga antelope Species Wiki Page and GO Annotation

NCBI Taxonomy ID		34875	
Parent GO ID		GO:0006099	GO:0015074
Parent GO Term Name		TCA	DNA Integration
Number of Proteins	38	0	0
Total Number of GO Annotations	467	0	0
Electronic Evidence	467	0	0
Manual Evidence	0	0	0
Molecular Function Aspect	131	0	0
Biological Process Aspect	143	0	0
Cellular Component Aspect	193	0	0

Table 8: Survey of *Raccoon* Species Wiki Page and GO Annotation

NCBI Taxonomy ID		9654	
Parent GO ID		GO:0006099	GO:0015074
Parent GO Term Name		TCA	DNA Integration
Number of Proteins	266	0	0
Total Number of GO Annotations	1,913	0	0
Electronic Evidence	1,913	0	0
Manual Evidence	0	0	0
Molecular Function Aspect	440	0	0
Biological Process Aspect	642	0	0
Cellular Component Aspect	831	0	0

When comparing the 71 domesticated species to the 200 wild species, majority of the wild species annotated did not possess a high amount of proteins and GO annotations like the Saiga antelope and raccoon, leading one to speculate that there is more focus

on domesticated species and species widely used for laboratory research. Such works enable the identifications and detail of their genomes annotated.

Conclusion

During our experiment, we can establish that domesticated and wild species had annotations associated with the biological processes, TCA cycle and DNA integration. We were able to validate the hypothesis because domesticated animals possessed more annotations than wild animals. Though evidence of biological processes was minimal and mostly electronic, it still validates that these two processes found. Additionally, through observation, the findings in this experiment determine that wild species annotated did not possess a high number of gene annotations. Therefore, a lack of GO annotations reflects the poor annotation of genes that are known, or genes being missed in the genome records. The problem of a low quantity of identified genes would suggest is with the structural annotation missing genes, or not finding homologs of the genes present. For example, *S. tatarica*, the lack of DNA sequencing and genome record for this species, justifying the low number of genes. The *P. lotor* also shows there only being a record of the mitochondrial DNA being sequenced, explaining the low count of genes present. We were able to find that more biocuration with wild animals because little is still known about their genomes.

We used bioinformatics to bring awareness of information that is still unknown about most animal genomes. The majority of the domesticated and wild species that were found during this study are vectors for zoonotic and infectious diseases. It is apparent that there is an increasing need for genomics in disease risk and susceptibility for endangered species (Irizarry, 2016). Alonso Aguirre argued that “habitat destruction, globalization, and species loss have led to ecosystem disruptions altering infectious disease transmission patterns” (Aguirre, 2017). That is why we wanted to examine a pathway that deals with foreign DNA transmission. We also wanted to search for gene annotations associated with DNA integration because “some viral DNA integration events are [similar] in nature to any type of foreign DNA integration” (Würtele, 2003). But due to the lack of annotations, we found that more exploration for wild animals for DNA integration is needed. In using tools like bioinformatics for genomic research, we can make connections that could provide knowledge on zoonotic disease etiology.

Acknowledgements

A special thank you to:

- CBG Grant awarded to EGCR
- Department of Biology, Prairie View A&M University
- Cooperative Agricultural Research Center,
Prairie View A&M University
- College of Agriculture and Human Sciences,
Prairie View A&M University

References

- Alisdair, F., Carrari, F., and Sweetlove, L. (2004). *Current Opinion in Plant Biology* 7.3: 254-61.
- Aguirre, & Alonso. (2017, December 15). Changing Patterns of Emerging Zoonotic Diseases in Wildlife, Domestic Animals, and Humans Linked to Biodiversity Loss and Globalization. Retrieved from <https://academic.oup.com/ilarjournal/article/58/3/315/4745734>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J.M., Davis, A. P., Dolinskyi, K., Dwight, S.S., Eppig, J. T., Harris, M. A., Hill, D.P., Issel-Tarver, L., Kassarski, A., Lewis, S., Matese, J.S., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. (2000). *Nature Genetics*, 25(1), 25–29. Doi:10.1038/75556
- Berg, J. M. (1970, January 01). *The Citric Acid Cycle*. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK21163/>
- Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., & Apweiler, R. (2009). *Bioinformatics*, 25(22), 3045-3046. doi:10.1093/bioinformatics/btp536
- The Gene Ontology Consortium, (2014). *Gene Ontology Consortium: going forward*, *Nucleic Acids Research*, 43(D1), D1049 - D1056.

<https://doi.org/10.1093/nar/gku1179>

Hayamizu, T., Baldock, R., & Ringwald, M. (2015). Mammalian Genome, 9584(1432-1777), 1-9. Doi:10.1007/s00335-015-9584-9

Irizarry, K. J., Bryant, D., Kalish, J., Eng, C., Schmidt, P. L., Barrett, G., & Barr, M. C. (2016). Integrating Genomic Data Sets for Knowledge Discovery: An Informed Approach to Management of Captive Endangered Species. *International journal of genomics*, 2016, 2374610.

Jenkinson, A., Albrecht, M., Birney, E., Blankenburg, H., Down, T., Finn, R., & Hermjakob, H., et al. (2008). BMC Bioinformatics 9.S8: 7. BMC Bioinformatics. BioMed Central. Web. 22 July 2008.

Müller, K., Heller, H., & Doerfler, W. (2001). Foreign DNA Integration. *Journal of Biological Chemistry*, 276(17), 14271-14278. doi:10.1074/jbc.m009380200

Squizzato, S., Park, Y., Buso, N., Cowley, A., Li, W., Uludag, M., Pundir, S. Cham, J., McWilliam, H., & Lopez, R. (2015). *Nucleic Acids Research*, 43(W1), W585-W588. Doi:10.1093/nar/gkv316

Würtele, H., Little, K. C., & Chartrand, P. (2003). Illegitimate DNA integration in mammalian cells. *Gene Therapy*, 10(21), 1791-1799.

doi:10.1038/sj.gt.3302074