Evaluating Olfactory Genes in *Felis catus*

Abstract

By comparing human olfactory genes to a full *Felis catus* genome, as well as comparing other organisms' olfactory genes, creates a starting point for more complicated studies. By understanding which genes are shared or not shared, could be used in a plethora of different applications. *Felis catus* is an interesting species due to being a domesticated animal, while not having large gene changes over a long period of time after being domesticated.

Motivation

When it comes to domesticated animals, people generally associate dogs with a keen sense of smell. While dogs generally have more scent receptors in their nose than cats, it is postulated that cats can differentiate between smells than dogs. Also, cats still have a much greater sense of smell than their human companions [4]. Cats are also extremely interesting to study genetic changes from domestication. Varying sources give varying lengths for how long cats have been domesticated, but the consensus is somewhere around 10,000 years [5]; whereas we believe dogs have been domesticated for two to four times as long [6].

While many domesticated animals have been bred to selectively by humans, including cats, there is generally a utility that is trying to be bred into the animal. Selective breeding in cats is looking for creating different appearances in cats but does not have the goal to change a cat's senses like breeding for a hunting dog would. This is likely why it seems as though *Felis catus* genetics have not changed much at all over 10,000 years of domestication. Also, domesticated cats, still had to rely on their senses to stalk prey and evade predators up until about 100-200 years ago. Now that cats are being brought into human homes and being bred to be companions that do not need to hunt, will that effect their senses? Much like humans' sense of smell has waned over many generations of not needing to rely as heavily on our sense of smell to survive.

This project hopes to do some differential analysis of the G Protein-Coupled Receptors found in humans and mice and compare them against an entire *Felis catus* genome. The work done here would be useful in preliminary EDA to start a project analyzing cat olfactory genes. In humans and mice, these gene sequences have been located so they can be used as a benchmark for the cat genome.

Data

The entire nucleotide sequence of a specimen (named Cinnamon) of *Felis catus* comes from a study done that sought to analyze cat genes for comparative analysis. Part of their study did analyze olfaction, as part of an analysis of the senses [3].

Sequences of human olfactory genes was pulled from The Human Olfactory Data Explorer (HORDE), by pulling the sequences from a tab delimited file [2]

Mouse sequences were manually pulled from the Olfactory Receptor DataBase (ORDB) [1]. In the time to complete the project, a web mining program could not be built to pull all of the information automatically from the ORDB site. This site linked to ncbi where fasta files of each gene could be

downloaded. Horde also had mouse sequences in a file that could have also been used, but these sequences did not have as much information about each sequence as compared to downloading from NCBI. In the end, either probably could have worked, but to get a program to mine the data on ORDB, one would only need to mine the ascension numbers and could use esearch and efetch to pull all the sequences.

The mouse data only has around 20 genes due to having to manually download them, but the analysis will be done on mouse data as if all of it was present.

Process

Building the *Felis catus* BLAST DB

The first thing to do is to download all the necessary data, and then unzipping if necessary. The cat genome came as different fasta files for each chromosome. The idea is to build a blast database with this data, and in order to do that, the best way is to have all the chromosomes in one file. The easiest way to do this is running the following command in the directory that the fasta files are in:

```
cat *.fna > all_seq.fasta
```

Now that all of the sequences are in a single file (size of around 2.5 Gb) a blast database could be built so that a fast search could be done. This can be done with the following command in a linux shell.

```
makeblastdb -in all_seq.fasta -parse_seqids -blastdb_version 5 -title
"Felis_catus" -dbtype nucl -max_file_sz '4GB'
```

Interestingly, this database that has around 2.5 billion nucleotides, was created in about 24 seconds.

```
Building a new DB, current time: 12/16/2020 16:58:05
New DB name:   /mnt/d/INFO519/PROJECT/DATA/all_seq.fasta
New DB title:  Felis_catus
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 4000000000B
Adding sequences from FASTA; added 4508 sequences in 23.8673 seconds.
```

Extracting Human Olfaction Genes from .tsv

To run a blast search on the database that was just created, the human sequences are needed. To make things as easy as possible, it is best to have each sequence in it's own fasta file. Also, it is important to not lose data when making this conversion, so as much information about each sequence is stored in the fasta file's header. Lastly, the .tsv includes pseudo genes, so those genes are not converted to fasta files. All of this is done via python script that is found in /TOOLS/extract_seqs.py. The fasta files will be put into a the folder /DATA/human_seqs/genes.csv.

The end result is about 400 olfactory genes.

Build the Human BLAST DB

Similarly, to how a cat BLAST DB was built, a human DB containing only human olfaction sequences will be useful to compare mouse olfaction to humans. This is done in a similar fashion using the following command:

```
makeblastdb -in all_human_seq.fasta -parse_seqids -blastdb_version 5 -
title "Homo_sapien" -dbtype nucl
```

Blast Searches

For all of the blast searches, there are many different fasta files that needed to be searched. From all of these searches, the information that is most important at this point is just the score of the alignment and some metrics on the alignment, or lack of alignment. Due to this, the following bash scrpt finds the top alignment and saves some of the information about the alignment:

```
#!/bin/bash
FILES="../DATA/human_seqs/*.fasta"
for f in $FILES
do
  echo "Processing $f"
  fn=${f##*/}
  fnn="${fn%.fasta}"
  blastn -query $f -db ../DATA/cat_blast/all_seq.fasta | egrep ">" -A
7 -m 1 > "../DATA/cat_blast_results/${fnn}.summary"
done
```

For all other BLAST searches, the above search script was slightly modified. The actual scripts can be found in the /TOOLS directory.

This script either outputs nothing, or it will output a file for each search that looks like the following:

```
>NC_018736.3 Felis catus isolate Cinnamon breed Abyssinian chromosome
E1,
Felis_catus_9.0, whole genome shotgun sequence
Length=63494689

 Score = 1088 bits (589),  Expect = 0.0
 Identities = 807/916 (88%), Gaps = 0/916 (0%)
 Strand=Plus/Plus
```

With a file like this for every matched gene, there are quite a few files that need to be parsed and relevant information extracted. This is done with another python script that can be found at /TOOLS/extract_blast_data.py. This script is used on all of the blast search data in this project to output a csv file with summarized data.

The script also lists the genes that are missing in the search, as well as the percent of missing genes.

Analysis

The majority of the analysis will be done on the human olfactory genes compared to a cat genome. This is due to not having many mouse olfactory genes.

Most interestingly are the genes that are present in humans, but not present in cats. About 9.5% of the genes tested were not found in the cat genome. The missing genes are the following:

```
['OR10A2', 'OR10A3', 'OR10A4', 'OR10A5', 'OR10A6', 'OR10A7',
'OR10AD1', 'OR10AG1', 'OR10C1', 'OR10D3', 'OR10G2', 'OR10G3',
'OR10G4', 'OR10G6', 'OR10G7', 'OR10G8', 'OR10G9', 'OR10H1', 'OR10H2',
'OR10R2', 'OR10X1', 'OR10Z1', 'OR12D3', 'OR14A16', 'OR14C36',
'OR2AK2', 'OR2G2', 'OR2H1', 'OR2H2', 'OR2L13', 'OR2Y1', 'OR4A5',
'OR4C12', 'OR4K17', 'OR5V1', 'OR6K6', 'OR6X1']
```

From here, an analysis of these genes could be done to find any patterns. After searching these genes in the HORDE database, some of them provided interesting insights, whereas others were not as interesting. The first missing gene 'OR10A2', when searched in the HORDE database, shows that it has strongly matched orthologs to other mammal species, so it is interesting that it is not present in the *Felis catus* genomes (Figure 1). On the other end, it seems that some of the missing genes, 'OR10R2', are those that are only a strong ortholog with an orangutan gene (Figure 2).

Most Similar ORs in other species (Orthologs)

Specie	Symbol	Percent	Length
Dog	OR10A5F	94%	303
Dog	OR10A5	94%	302
Dog	OR10A5D	93%	303
Dog	OR10A5E	93%	303
Mouse	Or10a5	92%	302
Mouse	Or10a2	90%	302
Rat	Or10a5	91%	302
Rat	Or10a5b	91%	302
Rat	Or10a2	88%	302
Cow	OR10A5	92%	303
Cow	OR10A5I	91%	302
Cow	OR10A5G	90%	301
Cow	OR10A5H	91%	299
Orang	OR10A5J	98%	303
Orang	OR10A5	95%	303
Horse	OR10A5	94%	303
Horse	OR10A5K	93%	303
Chimp	OR10A5GP	98%	303
Chimp	OR10A5	96%	303

Figure 1

Most Similar ORs in other species (Orthologs)

Specie	Symbol	Percent	Length
Dog	OR10R3B	77%	309
Mouse	Or10k2	56%	310
Rat	Or10j7	55%	305
Cow	OR10R3	76%	311
Orang	OR10R2B	96%	308
Horse	OR10R2C	87%	308
Chimp	OR10R3P	76%	309

Figure 2

When it comes to analyzing the matches, the first thing that is of interest is where these olfactory genes are located in the cat genome. Figure 3 shows that many of the matches were found in the 'D1' chromosome of the cat. Outside of the D1 chromosome, there is a decent spread across the chromosomes where matches occur, and then there are chromosomes where there are no matches.

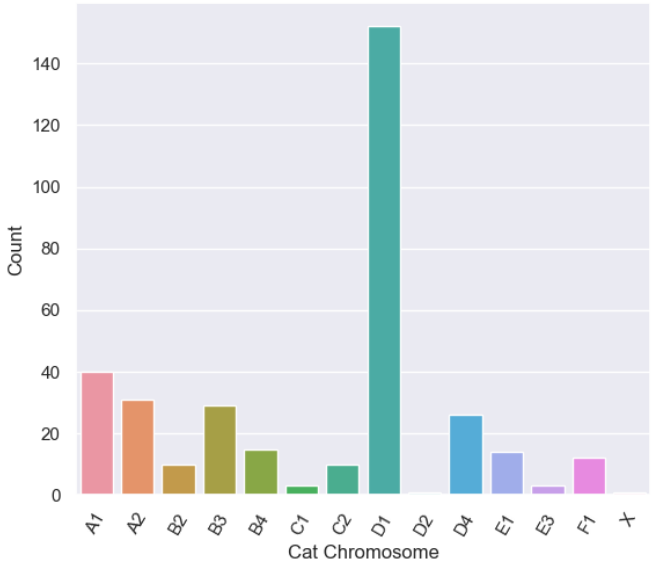

Figure 3

Figure 4 and Figure 5 show the distribution of Identity and Gaps percentage. The identity percentage is most frequently in the high 80s, and the gap percentage is almost always <3%. These values give confidence that the genes that are listed as matched and the genes that are missing are likely correctly labeled. For example, if the gap percentage was high and identity was even higher, this would likely be due to serendipity as opposed to actually showing a match. Also, with the gaps being so low, and even 0% in many matches, we know that these gene alignments are almost all substitutions.

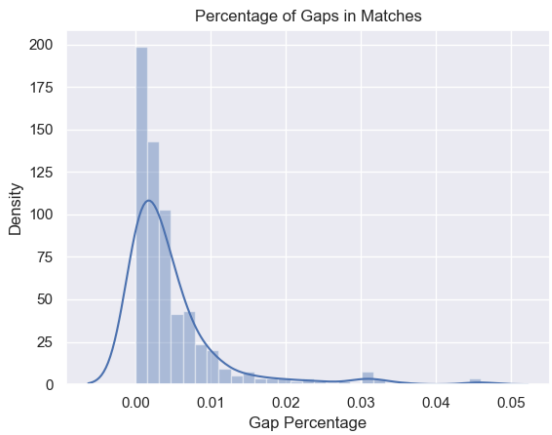

Figure 4

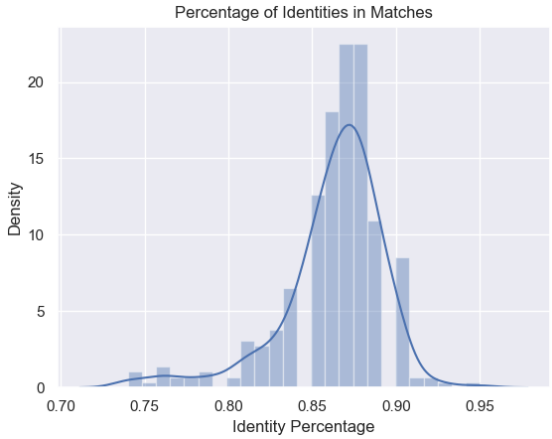

Figure 5

Similar analysis could be done when comparing mice to cats, and mice olfactory genes to human olfactory genes. The output that these graphs have been built from could serve as a reference when analyzing specific human olfactory genes. Or this process could be done to compare the human olfaction genes to other organisms. Although it is not the express idea of this project, building a database of the comparison of human genes to find which occur in other organisms is something that is partially done in the HORDE database, but only is done with a handful of different species. Having a more extensive database of comparisons could aid not only in studying olfaction, but even evolutionary biology.

References

[1] Crasto C..,Marenco L., Miller P.L., and Shepherd G.S. (2002) Olfactory Receptor Database: a metadata-driven automated population from sources of gene and protein sequences. *Nucleic Acids Research*1:354-360

[2] Olender T, Nativ N, Lancet D. HORDE: comprehensive resource for olfactory receptor genomics.Methods Mol Biol. 2013;1003:23-38.

[3] M. J. Montague, G. Li, B. Gandolfi, R. Khan, B. L. Aken, S. M. J. Searle, P. Minx, L. W. Hillier, D. C. Koboldt, B. W. Davis, C. A. Driscoll, C. S. Barr, K. Blackistone, J. Quilez, B. Lorente-Galdos, T. Marques-Bonet, C. Alkan, G. W. C. Thomas, M. W. Hahn, M. Menotti-Raymond, S. J. O'Brien, R. K. Wilson, L. A. Lyons, W. J. Murphy, and W. C. Warren, "Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication," Proceedings of the National Academy of Sciences, vol. 111, no. 48, pp. 17230–17235, 2014.

[4] "A Cat's Five Senses," CVMBS News, 12-Aug-2020. [Online]. Available: https://vetmed.tamu.edu/news/pet-talk/a-cats-five-senses/. [Accessed: 16-Dec-2020].

[5] I. C. Care, International Cat Care, 23-Nov-2020. [Online]. Available: https://icatcare.org/advice/the-origins-of-cats/. [Accessed: 16-Dec-2020].

[6] "Ancient genomes heat up dog domestication debate," Nature News. [Online]. Available: https://www.nature.com/news/ancient-genomes-heat-up-dog-domestication-debate-1.22320. [Accessed: 16-Dec-2020].