**Julie Garcia, L13593, Mus Musculus Prlr Gene**

In searching for homologs of my gene, the Prlr gene in Mus Musculus (house mouse), I used BLASTn, because my sequence is a nucleotide sequence (mRNA). I wanted to find homologs across species, so I searched the non-redundant Nucleotide database. I first searched for a human homolog. There were 38 matches. One was the “complete ids” of the human prolactin receptor Prlr (mRNA). It had an E-value of 0.0 and an Identity cover of 78% which was the best match. There were several transcript variants of the human Prlr gene in addition to several clones. I also found the human Prlr gene sequence of Chromosome 5 of the human genome. The top matches had one continuous stretch of alignment.

I then removed human from the search, and of course the Mus Musculus exact match came up first and four of the mRNA splice variants came up next. As I found in previous searches the Rat Prlr gene is the most highly conserved gene across a species. Other very similar rodent homologs were Peromyscus Maniculatus bairdii, Cricetulus griseus, Microtus ochrogaster, and Nannospallax galili. Other species with homologs of the Prlr gene were rabbit, elephant, leopard, domestic cat, killer whale, and polar bear.

When searching the Prlr gene using FASTA, I did notice the speed difference. FASTA was much slower than BLAST as stated in the lectures. The first result was again the exact match of the Prlr gene in Mus Musculus. Other nearly exact matches with E-value of 0.0 and Identity Scores above 95% were clones of the mRNA in mice and mutations of the Prlr gene in mice. Some FASTA results were different than anything that I’ve seen in BLAST like matches labeled “Novel therapeutic targets in cancer” and “novel compositions and methods for cancer”, which appear to come from a patent database as sequence targets for cancer therapeutics. Also high on the list was the rat prolactin receptor homolog, with an E-value of 0.0 and Identity score of 92.3%.

To perform the multiple sequence alignment, I retrieved the sequences of the top 10 matches of the Prlr gene and ran Clusta Omega and T-Coffee against my sequence. The accession numbers for the top 10 matches were X73372.1, L14811.1, BC006652.1, BC005555.1, NM\_011169.5, XM\_006520035.1, XM\_006520036.1, XM\_006520034.3, XM\_006520037.2, D10214.1. When I ran Clusta Omega I found that NM\_011169.5 had a long sequence at the start that did not match any of the other sequences, so I threw that one out and ran it again. This time I found a really long highly conserved region across all 10 RNA sequences (about~1800 bps long) which showed extremely high conservation between sequences. The sequences were made up of transcript variants within Mus Musculus, so high conservation was to be expected. I then noticed that four of the sequences were significantly longer, so I removed all of the XM\_ and one BC\_ sequence and ran it again. The four remaining sequences showed that they were nearly identical. I then ran the same 4 sequences in T-Coffee, and it gave nearly identical results.

I went back and excluded Mus Musculus when doing my BLAST search for homologs, in order to find conserved genes across other species. I found out that this gene is highly conserved across many species, so I picked out the most interesting ones including Rat (the most closely related sequence), leopard, rabbit, and human. When running these sequences through Clusta Omega, there were long stretches of highly conserved regions across these species (~1400 bp long).

I also ran a multiple sequence alignment on the same sequences, with two other tools, T-Coffee and Kalign. Kalign gave similar results to the Clustal Omega results, almost identical. T-Coffee found similar results in the middle of the genes, but found more matches towards the ends of the sequences as well, I suspect because it uses a combination of local and global alignment.

I ran T-Coffee Espresso, because I wanted to try a tool that included structural information in addition to just the sequences. This tool gave me an error when aa sequence was greater than 2500, so I only ran this on the mouse and rat Prlr gene. The results showed a score of 99 with the entire gene highly conserved across the species.