Biology Assignment Rashmeen Jaggi 31st May 2010

An Investigation into Molecular Genetics
Differentiating between Various Molecular Databases

This unit we looked at molecular genetics and their importance in human life. The experiment was mostly to find out how closely related two species are and how we can identify the differences. Using various protein sequences, we were able to compare and distinguish among various species. This report mainly focuses on relating four different species from the rodent family: Chinchilla, Guinea Pig, Crested porcupine and Hamster.

Research question

How are the four species; Chinchillas, Guinea Pigs, Crested porcupines, Hamsters of the rodent family related to each other? What are the similarities and differences among the four amino acid sequences and how can they be measured? To what extent are the four species different from each other?

The independent variables of this experiment were the four species: Chinchilla (Chinchilla Brevicaudata), Guinea Pigs (Cavia Porcellus), crested Porcupine (Hystrix Cristata) and Hamsters (Cricetidae). The dependant variables of this experiment were the amino acid sequences of the four sequences. Using the database search capacities on internet, the entire amino acid sequences were found and later using the Fasta programming on internet, the four species were compared with each other. All the data was then collected in a table. The following are the control variables of the experiment:

- The protein of the amino acid sequences was a control variable. For this experiment, the amino acid sequences for the protein Insulin were used, mainly because these are short and easier to compare.
- The type of family to which the four species belong to is also a control variable. For this experiment, the four chosen species belong to the rodent family. This will help us investigate the differences among the species of the same family.
- The characteristics of the species that were compared during the experiment were a control variable. For the experiment the following characteristics were compared only: Number of AA in the beta chain, Number of similar AA pairs (:), Number of conservative AA pairs (.), Number of different AA ( ).
- The database from which the amino acid sequences were obtained is also a control variable. All the sequences were obtained from www.uniprot.org
- The program used to compare the four amino acids is also a control variable. The variables were compared on the website: http://fasta.bioch.virginia.edu/fasta_www2/fasta_list2.shtml
- Another control variable of the experiment would be the number of species used for the experiment. Throughout the experiment, amino acid sequences of four different species are used.

Hypothesis

My prediction for this experiment is that the four species will be closely related to each other as they all belong to the same family; the rodent family. Statistics suggest that all species of the rodent family share the same traits such as having a single pair of incisors in each jaw, short and compressed tongues, the bones of the lower arm, the radius and ulna, are distinct, and the elbow joint permits free motion of the forearm. The number of toes and fingers range 3-5, though the thumb may be absent or vestigial. These similarities suggest that the species will share a few nucleotide basses that code the similar protein. The following pictures of my chosen species also suggest that they are linked together. The three species except for the crested porcupine appear to be similar in appearance and share the similar features.

Biology Assignment Rashmeen Jaggi 31st May 2010

Guinea Pig Porcupine Hamster Chinchilla

Table 1: Comparing Chinchilla and Guinea Pig

Species	**Chinchilla (Chinchilla Brevicaudata)**	**Guinea Pig (Cavia Porcellus)**
Number of AA in Beta chain	86	110
Amino Acid Sequence	FVNKHLCGSHLVDALYLVCGDRGFFYTPM AXXELEDPQVGQADPGVVPEAGRLQPLAL EMTLQXXGIVDQCCTSICTLYQLENYCN	MALWMHLLTVLALLALWGPNTGQAFVSRHLCGS NLVETLYSVCQDDGFFYIPKDRRELEDPQVEQT ELGMGLGAGGLQPLALEMALQKRGIVDQCCTGT CTRHQLQSYCN
Aligned Sequence	10 20 30 40 50 FVNKHLCGSHLVDALYLVCGDRGFFYTPMAXXELEDPQVGQADPGVVPEAGRLQPL ::..::::::.::..:: :: : :::: : ::::::: :.. :. :: :::: MALWMHLLTVLALLALWGPNTGQAFVSRHLCGSNLVETLYSVCQDDGFFYIPKDRRELEDPQVEQTELGMGLGAGGLQPL 10 20 30 40 50 60 70 80 60 70 80 ALEMTLQXXGIVDQCCTSICTLYQLENYCN ::::.:: :::::::: :: .::..::: ALEMALQKRGIVDQCCTGTCTRHQLQSYCN 90 100 110	
% Identity	64.0 % identity (79.1 % similar) in 86 aa overlap	
Number of similar AA pairs (:)	55	% of total 55/110 = 50%
Number of conservative AA pairs (.)	13	% of total 13/110 = 11.8%
Number of different AA ( )	42	% of total 42/110 = 38.2%

Table 2: Comparing Crested Porcupine and Hamster

Species	**Crested Porcupine (Hystrix Cristata)**	**Hamsters (Cricetidae)**
Number of AA in Beta chain	51	51
Amino Acid Sequence	FVNQHLCGSHLVEALYLVCGNDGFFYRPK AGIVDQCCTGVCSLYQLQNYCN	FVNQHLCGSHLVEALYLVCGERGFFYTPKSGIV DQCCTSICSLYQLENYCN
Aligned Sequence	10 20 30 40 50 FVNQHLCGSHLVEALYLVCGNDGFFYRPKAGIVDQCCTGVCSLYQLQNYCN ::::::::::::::::::::. :::: ::.::::::::..::::::.:::: FVNQHLCGSHLVEALYLVCGERGFFYTPKSGIVDQCCTSICSLYQLENYCN 10 20 30 40 50	

% Identity	86.3% identity (96.1% similar) in 51 aa overlap		
Number of similar AA pairs (:)	46	% of total	46/51 = 90.2%
Number of conservative AA pairs (.)	5	% of total	5/51 = 9.8%
Number of different AA ( )	2	% of total	2/51 = 3.9%

Table 3: Comparing Crested Porcupine and Chinchilla

Species	**Crested Porcupine (Hystrix Cristata)**	**Chinchilla (Chinchilla Brevicaudata)**
Number of AA in Beta chain	51	86
Amino Acid Sequence	FVNQHLCGSHLVEALYLVCGNDGFFYRPK AGIVDQCCTGVCSLYQLQNYCN	FVNKHLCGSHLVDALYLVCGDRGFFYTPMAXXE LEDPQVGQADPGVVPEAGRLQPLALEM TLQXXGIVDQCCTSICTLYQLENYCN
Aligned Sequence	10 20 30 FVNQHLCGSHLVEALYLVCGNDGFFYRPKA---------- :::.:::::::::.:::::::. :::: : : FVNKHLCGSHLVDALYLVCGDRGFFYTPMAXXELEDPQVGQADPGVVPEAGRLQPLALEMTLQXX 10 20 30 40 50 60 40 50 GIVDQCCTGVCSLYQLQNYCN ::::::::..:.::::.:::: GIVDQCCTSICTLYQLENYCN 70 80	
% Identity	45.8% identity (54.2% similar) in 83 aa overlap	
Number of similar AA pairs (:)	41	% of total	41/86 = 47.7%
Number of conservative AA pairs (.)	6	% of total	6/86 = 6.9%
Number of different AA ( )	39	% of total	39/86 = 45.3%

Table 4: Comparing Guinea Pig and Hamster

Species	**Guinea Pig (Cavia Porcellus)**	**Hamsters (Cricetidae)**
Number of AA in Beta chain	110	51
Amino Acid Sequence	MALWMHLLTVLALLALWGPNTGQAFVSRH LCGSNLVETLYSVCQDDGFFYIPKDRREL EDPQVEQTELGMGLGAGGLQPLALEMALQ KRGIVDQCCTGTCTRHQLQSYCN	FVNQHLCGSHLVEALYLVCGERGFFYTPKSGIV DQCCTSICSLYQLENYCN
Aligned Sequence	10 20 30 40 50 60	

	MALWMHLLTVLALLALWGPNTGQAFVSRHLCGSNLVETLYSVCQDDGFFYIPKDRRELEDPQVEQ ::..::::::::::.:: :: . :::: ::. FVNQHLCGSHLVEALYLVCGERGFFYTPKS---------- 10 20 30 70 80 90 100 110 TELGMGLGAGGLQPlALEMALQKRGIVDQCCTGTCTRHQLQSYCN ::::::::. :. .::..::: ----------GIVDQCCTSICSLYQLENYCN 40 50		
% Identity	39.5% identity (52.3% similarity) in 86 aa overlap		
Number of similar AA pairs (:)	34	% of total	34/110 = 30.9%
Number of conservative AA pairs (.)	11	% of total	11/110 = 10%
Number of different AA ( )	65	% of total	65/110 = 59.1%

Table 5: Comparing Chinchilla and Hamster

Species	**Chinchilla (Chinchilla Brevicaudata)**	**Hamsters (Cricetidae)**	
Number of AA in Beta chain	86	51	
Amino Acid Sequence	FVNKHLCGSHLVDALYLVCGDRGFFYTPM AXXELEDPQVGQADPGVVPEAGRLQPLAL EM TLQXXGIVDQCCTSICTLYQLENYCN	FVNQHLCGSHLVEALYLVCGERGFFYTPKSGIVD QCCTSICSLYQLENYCN	
Aligned Sequence	10 20 30 40 50 60 FVNKHLCGSHLVDALYLVCGDRGFFYTPMAXXELEDPQVGQADPGVVPEAGRLQPLALEMTLQXX :::.::::::::::::::::::::::: . FVNQHLCGSHLVEALYLVCGERGFFYTPKS---------- 10 20 30 70 80 GIVDQCCTSICTLYQLENYCN :::::::::::::::::: GIVDQCCTSICSLYQLENYCN 40 50		
% Identity	52.3% identity (58.1% similar) in 86 aa overlap		
Number of similar AA pairs (:)	45	% of total	45/86 = 52.3%
Number of conservative AA pairs (.)	5	% of total	5/86 = 5.9%
Number of different AA ( )	36	% of total	36/86 = 41.9%

Table 6: Comparing Guinea Pig and Crested Porcupine

Species	**Guinea Pig (Cavia Porcellus)**	**Crested Porcupine (Hystrix Cristata)**
Number of AA in Beta chain	110	51
Amino Acid Sequence	MALWMHLLTVLALLALWGPNTGQAFVSRH LCGSNLVETLYSVCQDDGFFYIPKDRREL EDPQVEQTELGMGLGAGGLQPLALEMALQ KRGIVDQCCTGTCTRHQLQSYCN	FVNQHLCGSHLVEALYLVCGNDGFFYRPKAGIV DQCCTGVCSLYQLQNYCN
Aligned Sequence	10 20 30 40 50 60 MALWMHLLTVLALLALWGPNTGQAFVSRHLCGSNLVETLYSVCQDDGFFYIPKDRRELEDPQVEQ ::..::::::.:::.:: :: .::::: :: FVNQHLCGSHLVEALYLVCGNDGFFYRPKA---------- 70 80 90 100 110 TELGMGLGAGGLQPLALEMALQKRGIVDQCCTGTCTRHQLQSYCN ::::::::::.:. .:::.::: ----------GIVDQCCTGVCSLYQLQNYCN 40 50	
% Identity	43.0% identity (53.5% similar) in 86 aa overlap	
Number of similar AA pairs (:)	37	% of total 37/110 = 33.6%
Number of conservative AA pairs (.)	9	% of total 9/110 = 8.2%
Number of different AA ( )	64	% of total 64/110 = 58.2%

Table 7: Concluding Table

Species	**Chinchilla**	**Guinea Pig**	**Crested Porcupine**	**Hamster**
Chinchilla	N/A	64.0 % identity (79.1 % similar)	45.8% identity (54.2% similar)	52.3% identity (58.1% similar)
Guinea Pig	64.0 % identity (79.1 % similar)	N/A	43.0% identity (53.5% similar)	39.5% identity (52.3% similarity)
Crested Porcupine	45.8% identity (54.2% similar)	43.0% identity (53.5% similar)	N/A	86.3% identity (96.1% similar)
Hamster	52.3% identity (58.1% similar)	39.5% identity (52.3% similarity)	86.3% identity (96.1% similar)	N/A

The first six tables above compare the four species and as we can see, the last table summarizes the overall data in tables 1-6. Table 7 suggests:

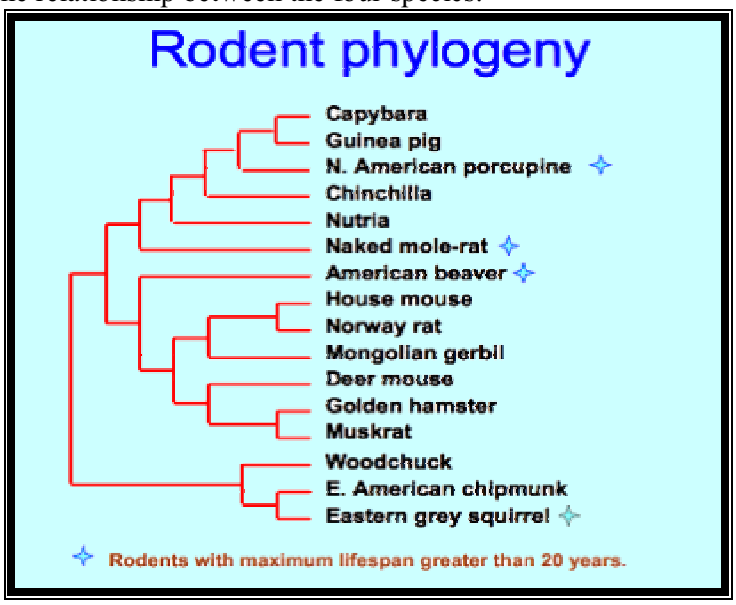
- The order of the three species closer to the Chinchilla is: Guinea Pig, Hamster and Porcupine.
- The order of the three species closer to the Guinea Pig is: Chinchilla, Porcupine and Hamster.
- The order of the three species closer to the Porcupine is: Hamster, Chinchilla and Guinea Pig.
- The order of the three species closer to the hamster is: Porcupine, Chinchilla and Guinea Pig.

The above tables suggest that the four species are not highly related to each other which could be due to the reason that the rodent family is the largest order of placental mammals, with a approximately 2, 050 species. The diversity of species in the rodent family could cause the relatedness of two species to reduce. And in order to investigate the reasoning behind this theory, geneticists around the world have been working on the Human Genome Project which is to sequence our entire DNA and locate all of the functionally important sequences within the DNA, such as genes. The project aims to investigate and discover the genes responsible for certain genetic disorders and to improve diagnoses, treatments and therapies.

The Human Genome Project intends to investigate certain scientific ideas by relating bioinformatics, proteomics and microarrays to the Human Genome Project. Bioinformatics combines biological science, computer science and information technology and amalgamates the data from the genomic research to facilitate the discovery of new biological insights. A database known as the BLAST allows one to compare gene or protein sequence rapidly. It helps find corresponding genes in different organisms. Proteomics is the study of the identities, structures, interactions and abundances of all organisms. The process involves separating the desired proteins into an individual ones and then with the use of bioinformatics, comparing them with to a unique protein. These two processes could lead to the development of new indicators for diagnosing diseases and identifying new targets for therapeutic drugs. Microarrays are an important tool of genomic revolution as it can identify the active genes in a cell. With the use of tiny sequences from thousands of genes, it can differ among various cells.

Conclusion

The above data does not support my hypothesis as my hypothesis stated that the four species will be closely related to each other as they belong to the same family. However, the results above suggest that the pair of the Chinchilla and the Guinea Pig is not highly related to the pair of the Porcupine and the Hamster. Looking at the low identity percentages it seems that even though the four species belong to the same family, they are not completely related to each other. The diagram below is the Rodent phylogeny that explains the relationship between the four species:



The phylogeny explains how closely related the four species are. As we can see, the species: the chinchilla, the guinea pig and the porcupine fall in the same category, however, the hamsters fall in a

different one. The above explains that the genes of a hamster do not follow the same pattern as the genes of the other species; it is different to their proteins. This suggests that as the diversity of species in the rodent family increases, the relatedness among the species reduces. The wider the range, lower the relatedness among the species that belong to the family. It is not necessary that the protein sequences in the genes of the species of the same family are similar; they might follow the same pattern till some extent but not completely.

Also, the results above could be different due to the reason that the length of the sequences differed. The length of the sequences above ranged 50-110 which also caused the identity percentage to reduce as the number of nucleotide bases was different. When two sequences with different lengths are compared then eventually the identity percentage reduces because one does not have enough nucleotides in order to be compared to the nucleotides of the other sequence. Therefore, the length of the sequence could also affect the identity percentage.

Research suggests that guinea pigs produce insulin that deviates more strongly than some other mammals belonging to the rodent family. The amino acid sequences of the chinchillas differ from those of the nonhystricomorph-rodents whereas other insulins of species such as Crested Porcupines hold an intermediate position. The insulin sequence of the guinea pig differs from other species in a perspective such as having a lower anabolic activity and an improved growth promoting activity. This specie has lost the capability of forming zinc-binding hexamers and so have porcupines and chinchillas. A molecular evolution of pancreatic ribonuclease allows chinchillas to group with other species such as coypu and casiragua of the rodent family and the guinea pig with cuis and capybara, and both groups together with porcupines. This supports the above diagram and states that hamsters are different compared to most other rodent family species.

Research also suggests that the common insulin gene was duplicated in the ancestors of the rodents, with expression of only one of the two gens in most rodent species. Chinchillas share methionine 29 of the B-chain with the rat and mouse insulin sequence. However, the evidence of two insulin genes in other rodent species other than rat and mouse is negative or week. A hamster has only one insulin gene even though the presence of two insulins in hamster pancreas is reported. Also, only one gene has been found in the guinea pig genome. This suggests that the guinea pig and hamster are somewhat related but not completely which is another reason why they have at least 39.5% of resemblance. Many scientists also suggest the presence of pig/rat-type insulin in the cells of a guinea pig caused by very ancient gene duplication. This could be another reason why the four species are so different from each other.

Most of the results above that I achieve do not support my hypothesis. In future, I would use a larger variety of species for my project, so that could learn more about the differences among different species. I would like to also use a different amino acid protein for my experiment such as hemoglobin so that I could learn more about the differences among the different species caused by different types of proteins. This will help me discover and understand key factors affecting the differences among species; no matter what family they belong to. Next, I would also try to use sequences of equal lengths so that it is a fair test. To conclude, I don’t think that the method requires any changes and cold be redone by the changes made above because the websites I used for my experiment gave me accurate results.

Bibliography

1. "Insulin." *UniprotKB*. N.p., n.d. Web. 6 June 2010. <http://www.uniprot.org/uniprot/?query=name%3Ainsulin+AND+hamster&sort=score>.
2. "Rodent Insulin." *mbe.ox*. N.p., n.d. Web. 6 June 2010. <http://mbe.oxfordjournals.org/cgi/reprint/4/1/10.pdf>.
3. "Sequence Comparing." *FASTA*. N.p., n.d. Web. 6 June 2010. <http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi>.