Utilizing Homologous Genes to Distinguish Between Similar Species

Joshua M Black - Email: josh.black@byu.edu - NetID: jblack33
Dr. Perry Ridge – Bioinformatics

Project Purpose

The purpose of this project is to create software that analyzes the genomes of two similar animal species to determine small, similar, but not identical, regions that can be used to distinguish the species of animal specimen that are too similar to do so through visual inspection.

Project Importance

Biologists working in the field often are required to identify a species through visual inspection. If, within an ecosystem, more than one similar animal species exist that also have similar appearances, identification through physical inspection may be difficult or impossible. When visual identification fails biologists are required to analyze the animal's genome to determine, with certainty, its identity. For most species limited genetic mapping data exist that would aid in the identification, therefore the entire genome must be determined through expensive sequencing. The entire sequence must then be compared to the sequences of known species and determined to be the closest matching species. Sequencing itself can cost over $1000 and take days to complete. Polymerase Chain Reaction (PCR) is a process that that can be used to determine the identity of an incredibly small portion of a genome. Small portions of a genome can be used to determine identity (Trautner, 2013). This project's intended software aims to identify a small portion of the genome that can be used to determine a specimen's species. This project and software could potentially allow scientists to use PCR instead of full sequencing to identify similar species that are too similar to do so by visual inspection. Resulting in a fraction of the time and cost required to identify the species of an animal specimen.

Project Profile Body

Similar species share homologs. Homologs are genes that originated in common ancestor organisms but remain in organisms that have diversified through evolution over time. Homologs share genomic coding. PCR requires the use of primers—similar coding in the genomes—to identify the small region of the genome that the PCR will be run on to identify the genomic code in the region. Homologs are perfect regions of the genome to discover different genomic codes that exist after identical primers. If a similar species of fish shared a homologous gene that coded for a specific trait that had an identical primer in both species and had developed such that the 5th code after the primer in one species was Guanine (G) and Thymine (T) in another species, it would be possible to use a PCR reaction targeted to the primer to distinguish between the two species. The main goal of this project is to develop a faster, cheaper way to identify similar species that cannot be identified through physical observation.

The software proposed by this project will identify possible homologs based on similarity between the genomes of

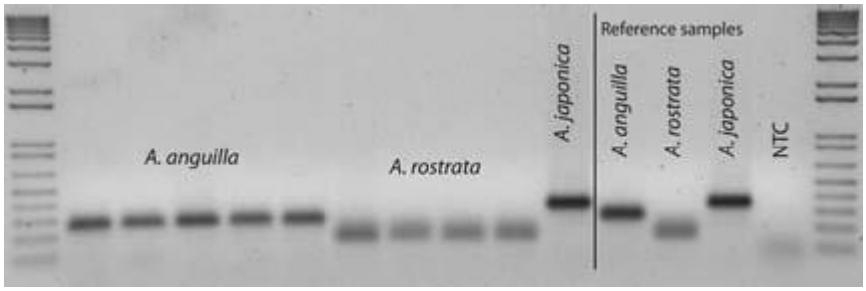

Figure 1 shows a PCR result identifying two distinct species.

two distinct species. This will utilize a BLAST algorithm that will identify the most similar regions of the genome code of both species (Altschul, 2009). A software will then identify possible regions for PCR by comparing similarities and identifying those areas that possess the required amount of variance between the two species.

After identifying possible regions for the PCR reaction the program will identify primers in the homologs that can be used to target the PCR reactions. This will require that each primer occur exactly once in each organism. If a primer exists in more than one location then the PCR reaction will have more than one target and will produce erroneous results.

If a match occurs the program will indicate to the user the region of the genome that will be tested and the primer target to be used in the PCR reaction. The user will then be able to use normal lab techniques to produce a definitive identification of the specimen in question.

Anticipated Academic Outcome

Dr. Ridge and I will present the outcome of this research at the International Society of Computational Biology annual meeting in July, 2017. We also plan on publishing the completed software package.

Qualifications

I am qualified for this project because of my strong academic career where I have earned a 3.94/4.00 GPA and my training and interest in genetics, biology, and software design. I have been able to learn the mechanics and processes of the biological tests used in this research through microbiology theory and laboratory instruction. I have been able to understand genetic variance through genetics classes that teach about evolution, genomic similarities, and differences in genomes. I have had instruction in manipulation and interpretation of the genomic code through technology in bioinformatics classes. I have also learned advanced programming concepts through computer science instruction and experience as a Lead Developer and Web Programmer at Brigham Young University.

My mentor, Dr. Ridge, was formally trained in both computer and the biological sciences. He has extensive experience developing algorithmic approaches to interpret biological data and applying these methods in large datasets. Dr. Ridge has been involved in bioinformatics research in a diversity of areas including genetics, biochemistry, immunology, molecular evolution, population genetics, medical diagnostics, and computer science. He has published more than 20 papers since starting at BYU three years ago, including multiple publications with undergraduate and graduate student co-authors. He has been a mentor to me through formal classroom instruction and personal tutelage. We meet regularly to discuss problems and progression of the project and he continues to help me understand difficult subject matter.

Project Timetable

This project will be completed by September, 2017 . Since the beginning of the semester we have been conducting research into the feasibility of this project and have begun to map and design the accompanying software.

Scholarly Sources

Altschul S, Gish W, Miller W, Myers E, Lipman D: Basic local alignment search tool. J Mol Biol 1990, 215(3):403–410.

Ferreira, Diana, et al. "Identification Of A Third Feline Demodex Species Through Partial Sequencing Of The 16 S R DNA And Frequency Of Demodex Species In 74 Cats Using A PCR Assay." Veterinary Dermatology 26.4 (2015): 239-e53. Academic Search Premier. Web. 24 Oct. 2016.

Camacho C. et al. (2009) "BLAST+: architecture and applications"

Trautner, J. H. "Stocking The Right Eel Species: A Fast PCR-Based Identification Assay To Discriminate European (Anguilla Anguilla (Linnaeus, 1758)), American (A. Rostrata (Lesueur, 1817)) And Japanese Eel (A. Japonica (Temminck & Schlegel, 1846))." Journal Of Applied Ichthyology 29.4 (2013): 912-915. Academic Search Premier. Web. 24 Oct. 2016

Yuskianti, V., et al. "Species-Specific PCR For Rapid Identification Of Ganoderma Philippii And Ganoderma Mastoporum From Acacia Mangium And Eucalyptus Pellita Plantations In Indonesia." Forest Pathology 44.6 (2014): 477-485. Academic Search Premier. Web. 24 Oct. 2016.