

Using Genetic Sequencing to Identify Specimens Found in Tuna Stomachs

Abstract

This project uses genetics to help create a database of the eating habits of Albacore Tuna (*Thunnus alalunga*). Data is collected both through anatomical identification of species in the tuna stomachs, and through genetic identification of samples taken from a tuna stomach, which are then compared against a library of known fish DNA. The genetic results include four species found in the tuna stomach, *Sardinops caeruleus*, *Euphausia pacifica*, *Thunnus alalunga*, and *Scomber japonicus*. The results show how DNA sequencing can be utilized to identify otherwise unidentifiable species and also prove to contribute valuable information that broaden our understanding of the tuna population patters. This then increases our ability to understand how to sustain the economically valuable tuna populations.

Introduction

The Larger Impact.

Due to their economic value, tuna have been the subject and interest of many studies, with several studies being done on thee eating habits of tuna. (McHugh, 1952) There are still many mysteries revolving around tuna, such as why tuna move through only a limited area off our coast when there are broad expanses where temperatures and other conditions are favorable. (Pinkas, 1971) They are constantly threatened by the

fisheries industry because of their value on the market, and sustainability is becoming more of a challenge as demand for tuna increases. (Hoyle et. al, 2006) In a recent study looking specifically at population size of yellowfin tuna, the overall average weights of tuna that are caught “have consistently been much less than those that would maximize the average maximum sustainable yield, indicating that, from the yield-per-recruit standpoint, the yellowfin in the Easter Pacific Ocean are not harvested at the optimal size.” (Hoyle et. al, 2006) These trends in sustainability are becoming a threat to all tuna and it is because of this that understanding the ecological role that tuna play within our oceans has become a necessity for protecting them. By understanding the eating habits of tuna, we can better understand variations in their population patterns and in doing so can more accurately create efficient sustainability programs.

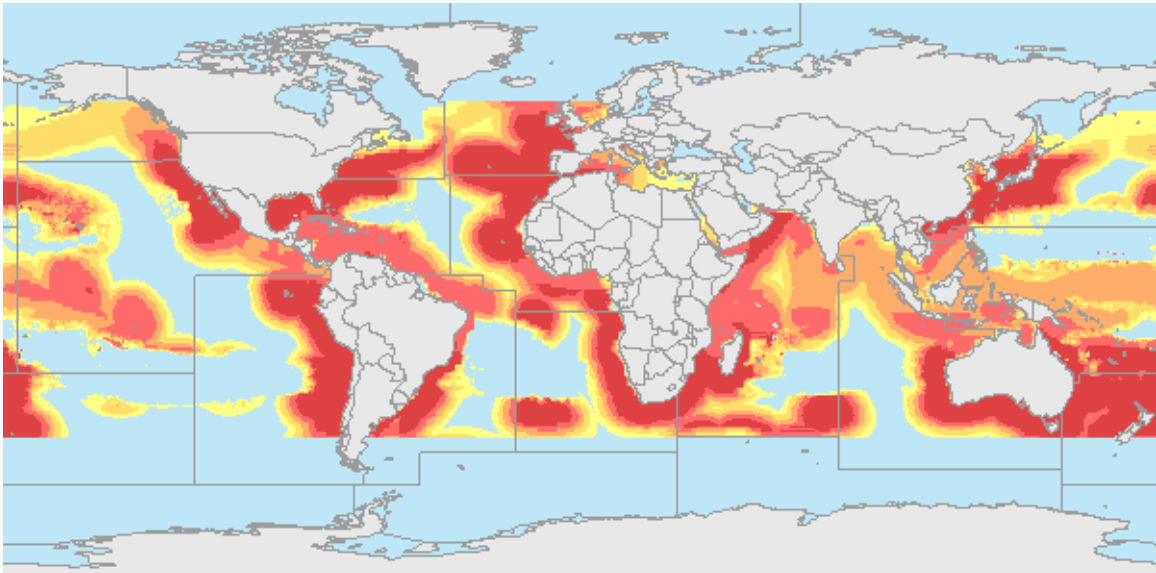
Project Background

The larger portion of this project was constructed and directed by a graduate student at the Scripps Institution of Oceanography, Sarah Glaser. Glaser’s primary objective was to create a database that consisted of what she found in the stomachs of Albacore Tuna (*Thunnus alalunga*). Since it is not possible to create a database of tuna eating habits by observation alone, Glaser collected samples of Albacore from the local waters off the coast of Southern California via vessel, and analyzed them in her laboratory at Scripps Institute of Oceanography.

Species Studied

Thunnus alalunga belong to the family Scombridae, shared with mackerels and bonitos as well as other tuna species. (Agbayani, 1999) Albacore can grow to 140 cm in

length and up to 60 kg in weight, with a population doubling time range of 1.4 to 4.4 years. (Agbayani, 1999) This highly migratory species of tuna can be located in tropical and temperate waters of all oceans including the Mediterranean Sea but not at the surface between 10°N and 10°S. (Agbayani, 1999) Basic behaviors include population concentration along thermal discontinuities, while forming mixed schools with skipjack tuna (*Katsuwonus pelamis*), yellowfin tuna (*Thunnus albacares*) and bluefin tuna (*T. maccoyii*). (Agbayani, 1999) Below is a map charting out the distribution of albacore tuna throughout the worlds oceans.



(Agbayani, 1999)

Project Background Continued

Glaser separated the subjects initially by storing the stomach contents of the Albacore in individual jars and labeled them accordingly. When classifying the stomach contents by species, Glaser looked at physical features of the stomach contents to determine how to categorize the sample to the lowest possible taxon. For approximately

95% of the samples studied, the anatomical features of the food specimen in the albacore stomach were enough to classify into species-specific categories. For the remaining samples taken, the digestion process had degraded the physical state of the species to the point where it was not identifiable to the species-specific extent. Since discarding samples that were unidentifiable via anatomical features would skew the data, it was decided to use alternate identification methods in order to classify the degraded specimens into the proper species.

The only other option of classification in the case of this project was genetic identification. Since genetic sequencing, a process that generates specific base pair sequences of a selected gene, only needs a small amount of tissue, this project was a perfect candidate since although the unidentifiable specimens had been shredded, there was plenty of tissue left over to sequence. My responsibility and contribution to this project included collecting samples of unidentifiable species and using genetic sequencing to determine the species-specific taxonomy of these samples.

Genetic Sequencing

Genetic sequencing is a fairly recent technology that is used throughout the field of biology, including marine biology, but its origins have a deep-rooted history. In 1865, Gregor Mendel, known today as the “father of genetics,” traced inheritance patterns of traits in pea plants and showed that they followed a pattern. (Wikipedia, 2006) Combined with Charles Darwin’s release of the Origin of Species in 1859, this knowledge laid the foundation for genetic studies. (Wikipedia, 2006) For years to come new knowledge about the structure and function of Deoxyribonucleic Acid (DNA) would evolve and in 1977 DNA would be sequenced for the first time by Fred Sanger, Walter Gilbert, and

Allan Maxam working in Sanger's lab to complete the entire genome sequence of Bacteriophage Φ -X174. (Wikipedia, 2006) A DNA sequence is a series of letters representing the nucleotide order of a specific gene in DNA, breaking it down into its most basic components. The four possible nucleotide subunits in a DNA strand are adenine, cytosine, guanine and thymine which are represented by the letters A, C, G and T respectively. (Wikipedia, 2006) This sequence is acquired from raw tissue through the genetic sequencing process which will be described in more detail in the methods section.

When doing background research on this project, I noticed that all of the previous studies involving eating habits of tuna involved the traditional way of classifying a sample taxonomically, but noticed that none had incorporated genetic sequencing into the study. This is mainly because the other studies were performed prior to the discovery and utilization of the genetic sequencing method. Genetic sequencing has been used in similar projects in marine biology. For example, a recent study was done to determine the drifting patterns of plankton which could not be identified anatomically but only by genetic sequencing, which demonstrates how genetics expands the range of what marine biologists can study. (Bucklin et. al, 2000)

Building Genetic Libraries

While proving the usefulness of genetic sequencing, this project also demonstrates the importance of building genetic libraries. A genetic library is a collection of gene sequences from known species. The largest and most useful library in existence today belongs to the National Center for Biotechnology Information (NCBI) in the form of GenBank which houses the genome sequencing data. GenBank can be searched with a program called Blast, invented by David Lipman, which can align the sequence of an

unknown sample to a sequence of a known species. My genetic portion of this project showed how imperative it is to have complete and thorough genetic libraries of known species. Without resources like NCBI's GenBank, a genetic sequence of an unknown origin would be useless because without known samples to compare it to, the sequence would be worthless.

Another segment of this project involved creating a genetic database at Scripps Institute of Oceanography for known fish species. These specimens were sequenced for the 16s gene that was to be compared with the tuna stomach specimens. I, along with SIO graduate students, successfully sequenced over 250 different fish for the 16s gene. The purpose of this was to create a control group so that the unknown sequences could have known sequences to match up to, and to also contribute the sequences to GenBank's library. GenBank builds its library through private institutions submitting genetic data, which is a cooperative effort to continually expand the library.

Materials and Methods

The Species of tuna studied in this project was *Thunnus alalunga*. The tuna specimens were collected roughly 15 miles off the coast of San Diego, California. Tuna were then brought back to a lab at Scripps Institute of Oceanography and dissected for their stomach content, which was then stored in 100% Ethanol.

Both the known sample library, which currently consists of 264 known species, and the unknown tuna stomach samples were sequenced for the 16s gene. The sequencing protocol used consisted of a 7-step process as follows:

1. Extraction

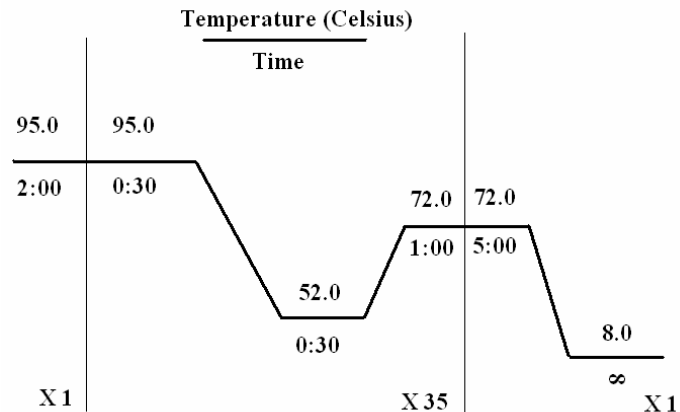
Using a small piece of tissue from the sample, this process extracts and purifies the DNA to prepare it for amplification. The Dneasy Handbook for Animal Tissues (2004) was used as a basis for the extraction protocol.

2. Polymerase Chain Reaction (PCR)

PCR enzymatically replicates DNA without using a living organism, such as *E. coli* or yeast. The technique allows a small amount of the DNA molecule to be amplified many times, in an exponential manner, to allow for an easier analysis.

Concentration table and temperature profile below.

<i>Additive</i>	<i>Volume</i>
DNA grade H2O	19.0 ul
10x Buffer	2.5 ul
dNTP's	0.5 ul
16s Ar (primer)	0.5 ul
16s Br (primer)	0.5 ul
Sigma Red Taq	1.0 ul
DNA Template	1.0 ul
Total	25.0 ul



3. Gel Electrophoresis

Gel electrophoresis is a technique used to separate molecules based on physical characteristics such as size, shape, or isoelectric point. Gel electrophoresis was used in this project to verify the competence of the extraction and PCR before proceeding with the rest of the sequencing, to make sure there was adequate DNA

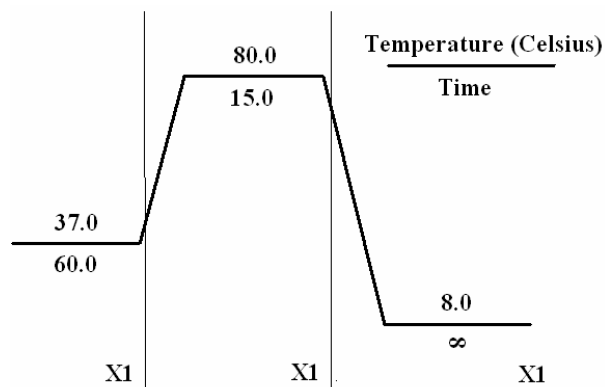
to sequence without contamination or impurities. Protocol used involved using a 0X174 ladder (control) and running each gel for 30 minutes on 80 volts.

4. ExoSAP

The ExoSAP stage uses two enzymes, Exo-I and SAP, which eliminate the primers left over from the PCR done in step 2. This step is critical because any left over enzymes from the PCR could hinder the following sequencing reaction.

Concentration table and temperature profile below.

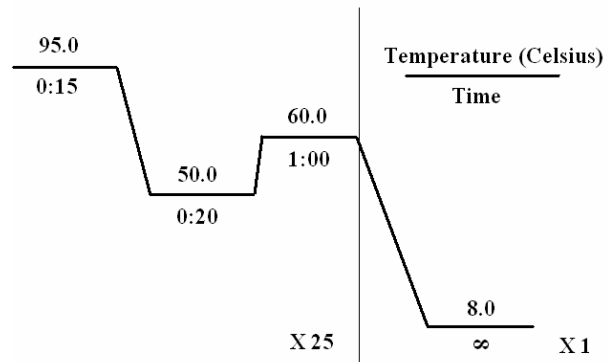
<i>Additive</i>	<i>Volume</i>
PCR product	5.0 ul
SAP	0.5 ul
Exo-I	0.5 ul
10x SAP buffer	1.0 ul
DNA grade H2O	3.0 ul
Total	10.0 ul



5. Sequencing reaction

In the sequencing reaction, the primer works as it did in the PCR process by multiplying the bands of DNA but the ET-dye premix acts as a terminator which terminates the replication in different areas and labels each of the dideoxynucleotide chain-terminators with a separate fluorescent dye, which fluoresces at a different wavelength. This type of sequencing is known as dye terminator sequencing. Concentration table and temperature profile below.

<i>Additive</i>	<i>Volume</i>
Cleaned PCR product	2.0 ul
ET-dye premix	4.0 ul
Primer	0.5 ul
DNA grade H2O	3.5 ul
Total	10.0 ul



6. Sephadex

This is the final cleaning stage before the sequenced DNA can be put through a machine. Sephadex is a trademark for cross-linked dextran gel. It is normally manufactured in a bead form and most commonly used for gel filtration columns. Here it removes any impurities that could hinder the final stage of sequencing. The sequencing reaction product was spun through the sephadex in a centrifuge for three minutes at 3000 rpm.

7. Sequence

The final stage consists of running the Sephadexed sample through a machine which magnetically separates the DNA stand in relation to its size and reads the wavelengths emitted by the fluorescent dye. The samples are automatically downloaded into another computer where analysis and comparison of the base pair order can be made.

Results

In total, 31 samples total were taken from 5 different stomachs of tuna and 4 different species variations were found among the different samples. The table below summarizes the results.

Stomach #	Sample #	Description of contents	Species Match
PU5001	1	<i>Body tissue of known Sardine</i>	<i>Sardinops caeruleus</i>
PU5001	2	<i>Eyeball from shrimp-like animal, resistant to digestion</i>	<i>Euphausia pacifica</i>
PU5001	3	<i>Eyeball from shrimp-like animal with small amount of tissue</i>	N/A
PU5001	4	<i>Larger piece of tissue, brown and rough in texture</i>	<i>Thunnus alalunga</i>
PU5001	5	<i>White fish tissue</i>	N/A
PU5001	6	<i>Fish scale, reflective, very small</i>	N/A
PU5001	7	<i>White fish tissue</i>	N/A
PU5001	8	<i>White fish tissue</i>	<i>Sardinops caeruleus</i>
PU5001	9	<i>Shiny black coarse material</i>	N/A
PU5001	10	<i>Shiny black coarse material</i>	N/A
PU5002	11	<i>Vertebral column with no tissue attached</i>	N/A
PU5002	12	<i>Forelimb from a crab, species unknown</i>	N/A

PU5002	13	<i>White tissue, large and tooth-shaped.</i>	<i>Thunnus alalunga</i>
PU5003	14	<i>Tissue from Euphausid, species unknown</i>	N/A
PU5003	15	<i>Tissue from Euphausid, species unknown</i>	N/A
PU5005	16	<i>White tissue, floating</i>	<i>Thunnus alalunga</i>
PU5005	17	<i>White tissue, floating</i>	<i>Thunnus alalunga</i>
PU5005	18	<i>Pink tissue, on bottom of container</i>	N/A
PU5005	19	<i>Pink tissue, on bottom of container</i>	N/A
PU5005	20	<i>Eyeball from Euphausid, species unknown</i>	N/A
PU5005	21	<i>Eyeball from Euphausid, species unknown</i>	N/A
PU5005	22	<i>2 Eyeballs from Euphausid, species unknown</i>	<i>Euphausia pacifica</i>
PU5005	23	<i>2 Eyeballs from Euphausid with tissue attached, species unknown</i>	N/A
PU5005	24	<i>Pink tissue, on bottom of container</i>	N/A
PU5005	25	<i>Pink tissue, on bottom of container</i>	N/A
SD5052	26	<i>White tissue from midsection of unknown</i>	<i>Scomber japonicus</i>
SD5052	27	<i>White tissue from midsection of unknown</i>	N/A
SD5052	28	<i>Black/Brown tissue from Sardine mouth</i>	N/A
SD5052	29	<i>White Tissue from behind gill of unknown</i>	N/A

SD5052	30	<i>White tissue from midsection of unknown</i>	<i>Scomber japonicus</i>
SD5052	31	<i>White tissue from midsection of unknown</i>	<i>Scomber japonicus</i>

The N/A in the species match column means that the sequence was not sufficient enough to produce a statistically significant match of a species. 11 of the 31 samples sequenced with a statistically significant match of a species. The frequency of occurrence of each species was, *Sardinops caeruleus* (2), *Euphausia pacifica* (2), *Thunnus alalunga* (4), *Scomber japonicus* (3). The table below shows the statistics for the species sampled with the lowest error value.

<i>Species</i>	Common Name	Base Pairs	Error Value
<i>Sardinops caeruleus</i>	Pacific sardine	560	0
<i>Euphausia pacifica</i>	Krill	220	5e-85
<i>Thunnus alalunga</i>	Albacore	211	2e-95
<i>Scomber japonicus</i>	Chub mackerel	535	0

Common names (Kozloff, 1996)

Many samples failed to produce readable sequences either because the initial sample was non-biological or because contamination or a lack of ample DNA amplification occurred during the sequencing process. Regardless, the sequenced DNA samples that yielded results produced data that contributed to the tuna diet study.

Discussion

Found Species

The first found species, *Sardinops caeruleus* was not a surprise result. When the tuna used for the study were caught, this species of Sardine was used as bait to attract the

tuna. The DNA match is important however, because it serves as a control to prove the validity of the DNA sequenced results. Since the Sardines were relatively intact when the tuna stomachs were dissected, it is assumed that the Sardines are from the bait boat and not from the natural environment.

The second found species; *Euphausia pacifica* was a good example of how DNA sequencing can help in identifying the species when digestion had done considerable damage to the specimen to the extent where it was impossible to establish a species through anatomical features alone. From the specimen alone it was suggested that the small shrimp like creature was a Euphausid, but it wasn't until DNA sequencing verified the species that it was found to be *Euphausia pacifica*. This find was worthy contribution to the study that would not have otherwise included *Euphausia pacifica* into the database of tuna prey.

The discovery of *Thunnus alalunga*, or Albacore, was an extremely interesting result, being the same species stomach from which it was taken from. At first analysis it appears as if the Albacore may be eating each other, however, upon closer inspection it is found that this is not the case. When looking at the descriptions of the samples that sequenced for Albacore, it comes to attention that they were all from very small pieces of tissue that were not attached to a larger body. This suggests that the tissue collected initially was not from a prey of the albacore but rather tissue from the albacore itself. During the dissection of the albacore for its stomach contents, it is very possible for pieces of albacore tissue to accidentally incorporate into the stomach sample container.

Finally, the found species, *Scomber japonicus*, or chub mackerel, was another surprise result to find in the tuna stomach. Tuna are known to compete with mackerel for

the same food sources, but it was not known that mackerel could fall prey to a tuna. This discovery alters the way we can chart out food sources for tuna, and adds a new dynamic to explain their population patterns and migrations.

The four found results are not equally enlightening to the overall struggle for tuna sustainability, but they all contribute to the understanding of Tuna and therefore can assist in effectively sustaining Tuna populations, which are in constant threat from over fishing and environmental hazards. The original purpose of this project was to demonstrate how DNA libraries combined with sequencing technology could contribute to studies that involve unidentifiable species. The project was not only successful in that regard, but also went above its expectations by making vital discoveries to the tuna diet study, which include never before discovered information that can contribute to the fight for sustainability in our oceans.

Failed Samples

As for the samples that failed to produce significant results, there are many reasons why a sample would fail. For the samples with descriptions of dark material instead of lighter tissue, it is possible that the sample taken was non-biological and therefore had no DNA to sequence. Contamination is another possibility for failure of a sample. If the DNA is not properly purified or DNA from another source is mixed in with the sample, a clean sequence would not be found. If good tissue was selected, and contamination did not occur, it is also possible for the sequencing machine to malfunction. Since genetic sequencing is a fairly novel technology, the machinery involved has a tendency to make errors. Many samples showed strong bands of DNA when put through PCR, however, clean sequences would not result. A combination of

operator and mechanical error caused approximately 2/3 of the samples to fail, which was a learning experience for myself, and hopefully this percentage will decrease in any future work.

Future Work

My portion of this project has been completed, however, the larger project of compiling a database of tuna prey is an ongoing development. Sarah Glaser is still actively collecting samples of albacore from different areas of the Eastern Pacific Ocean. The genetic portion of this project has been proved successful once, and so it will be utilized in the future if needed.

Acknowledgements

I would like to thank Dr. Ron Burton for providing me with the facilities and supervision to perform all of the genetic sequencing done in this project. I would also like to thank Sarah Glaser for allowing me to participate in the project she created. Lastly I would like to thank the Environmental Systems department for giving me this wonderful learning opportunity.

References

- A. Bucklin, S. Kaartvedt, M. Guarnieri and U. Goswami. "Population genetics of drifting (*Calanus* spp.) and resident (*Acartia clausi*) plankton in Norwegian fjords." *Journal of Plankton Research*, Oxford University Press 2000. Vol.22 no.7 pp.1237-1251.

- Agbayani, Eli. "Fish base." WorldFish Center, FishBase Project. 1999
<http://filaman.ifm-geomar.de/Summary/SpeciesSummary.php?id=142>
- "History of genetics." Wikipedia, The Free Encyclopedia. 14 May 2006, 20:47 UTC. Wikimedia Foundation, Inc. 8 Jun 2006
http://en.wikipedia.org/w/index.php?title=History_of_genetics&oldid=53201459
- Hoyle, Simon D.; Maunder, Mark N. "Status of yellowfin tuna in the eastern Pacific Ocean in 2004 and outlook for 2005." Inter-American Tropical Tuna Commission Stock Assessment Report 6 2006: 3-102.
- Kozloff, E. N. "Marine Invertebrates of the Pacific Northwest." University of Washington Press, Washington, United States of America. 1996.
- McHugh, J. L. "The food of albacore (*Germo alalunga*) off California and Baja California. Scripps Inst. Oceanography." Bull, 6 (4) 1952: 161–172.
- Pinkas, Leo. "Food habits of albacore, bluefin tuna, and bonito in California waters." State of California, Dept. of Fish and Game, 1971.