The Journal of Experimental Secondary Science

Wild Type

- Laboratory vs. Wild Flies
- Discovery and Analysis of Palindromes
- Expression of Novel Splice Variants in the MTAP Gene
- Thymoquinone as a Novel Antibiotic and Chemotherapeutic Agent
- Genetic Variation of the Trout Population in the Lake Champlain Basin
Staff and Reviewers - July 2012

Academic Advisory Board
Cornelis Murre, PhD
Professor of Molecular Biology, UC San Diego

A. Ramachandra Rao, PhD
Emeritus Professor of Civil Engineering, Purdue University

Vanessa Svihla, PhD
Assistant Professor, Teacher Education
University of New Mexico

Nagesh Ragavendra, M.D.
Professor of Radiological Sciences, UCLA Medical Center

Editorial Board
Ravi Dinakar, PhD, Biology IC - Molecular Biology
Editor in Chief

Alicia Bicknell, PhD, Biology IC - Biochemistry
Post Doctoral Fellow, UMASS Medical School

Bernadette Roche, PhD
Chair, Dept. of Biology, Loyola University

Dahlia McGregor, PhD
President, New York Biology Teachers Association

Peter J. Cook, PhD
Research Fellow, Sloan Kettering Institute

Jennifer Stubbs, PhD
Scientist, Pathway Genomics

Judith A. Scheppler, PhD
Coordinator, Student Inquiry and Research, IMSA

Nancy C. Norton, PhD
Associate Professor, Chemistry and Biochemistry
University of Arizona

Claire Standen, PhD
Instructor in Molecular Medicine, UMASS Medical School

Diana Laboy-Rush, PhD
STEM Solutions Manager, Learning.com

Barbara Nicholson, PhD
Professor, Central Connecticut State University

Staff
Technology Director:
Scott Fenton, BS, Physics IC - Mechanics
Senior Infrastructure Engineer, Mphasis
CEO, Gnosis Tutoring

Legal Counsel:
Vijay Dinakar, Esq., Dinakar and Associates

Consultant:
Daniel Antonelli, Esq., Antonelli & Antonelli

Proof Editor:
Charissa Sorrentino, MS, English IC - Exposition
Teacher, Immaculata High School

July 2012 Issue Reviewers

Adrian R. Krainer, PhD
St. Giles Foundation Professor of Molecular Genetics
Program Chair, Cancer & Molecular Biology
Cold Spring Harbor Laboratory

Jacqueline Novatt, PhD
Post-Doctoral Fellow
Cold Spring Harbor Laboratory

Olga A. Timofeeva, PhD
Assistant Professor
Georgetown University
Lombardi Comprehensive Cancer Center

John Cuozzo, PhD
Director, Protein Biochemistry
X-Chem, Inc.

Peter Faletra, PhD
Delegate to AAAS Council
Past President of the NAAS

Ashton Powell, PhD
Faculty Associate, UNC Center for Genomics and Society

Patrick C. Mathias, PhD
Illinois Medical Scholar, Department of Bioengineering
University of Illinois at Urbana-Champaign

Holly Soutter, PhD
Senior Research Scientist, X-Chem, Inc.

Kevin Knobbe, PhD IC - CERN
Professor of Physics, Estrella College

Jason Osborne, PhD
President and Executive Director
Paleo Quest

Reviewers from Editorial Board:

Peter J. Cook, PhD
Research Fellow, Sloan Kettering Institute

Jennifer Stubbs, PhD
Scientist, Pathway Genomics

Alicia Bicknell, PhD
Post Doctoral Fellow, UMASS Medical School
# Table of Contents

**Editorials:**


p. 1 - 5

**In School Articles:**

**A Comparison of Laboratory-Reared Stock and Captured Fruit Flies** *(Drosophila melanogaster)* using Upward Movement, Phototaxic, and Starvation Assays Reveals Significant Behavioral Differences

Trinity Russell* and Richard Kurtz

Student1, Teacher2: Commack High School, 1 Scholar Lane, Commack, NY 11725

p. 6 - 11

**Discovering the Distribution of Palindromic Sequences in the SMAD4 Gene using Large and Medium Deletions and the Resulting RNA Structure Predictions**

Andrew Ninh* and Lisa Battig

Student1, Teacher2: Fountain Valley High School, 17816 Bushard Street, Fountain Valley, CA 92708

p. 12 - 18

**Internship Articles:**

**Thymoquinone as a Novel Antibiotic and Chemotherapeutic Agent: a Natural Therapeutic Approach on Staphylococcus aureus, Bacillus anthracis, and Four NCI-60 Cancer Cell Lines**

Syed Hossain*, Erin Sikes-Thurston, Stephen H. Leplla, and Alexander N. Wein

Student1, Teacher2: Herndon High School, 700 Bennett Street, Herndon VA 20170

Mentor3: Microbial Pathogenesis Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 33 North Drive, Bethesda, MD 20892

p. 19 - 22

**Mutations in Diaphyseal Medullary Stenosis- Malignant Fibrous Histiocytoma Related Gene MTAP Affects Expression of Splice Variants SV2 and SV5**

Alla Uts, Estela Arras, Olga Camacho-Vanegas, Catalina Camacho-Vanegas, and John A. Martignetti

Student1, Teacher2: High School for Mathematics, Science and Engineering, NY, NY 10031

Intern1, Mentor4: Mount Sinai Medical Center, New York, NY

p. 23 - 27

**Genetic Analysis of the Brown, Brook, and Tiger Trout Populations in the Lake Champlain Basin**

Markie Palermo, Sandra Latourelle, and Nancy L. Elwess

Student1,2, Intern2, Mentor: State University of New York at Plattsburgh

p. 28 - 31
It is an exciting time in science education. We are at the start of a new vision for teaching science that promises to transform the experiences of students in all grades across the country. This new vision calls for science classrooms that bring science alive for students, emphasizing the satisfaction of pursuing compelling questions and the joy of discovery. The vision is articulated in a new report from the National Research Council: “A Framework for K-12 Science Education”, which provides a blueprint for new state standards in science education. Based on this framework a consortium of educators and scientists from 26 states are developing Next Generation Science Standards. These standards will be available for states to adopt to guide what students learn in science for the next decade or more.

A major goal of the framework is to shift the emphasis in science education from teaching facts to immersing students in doing science. A central tenet of the framework is that throughout elementary, middle and high school, students should have opportunities to DO science. This might look different in second grade, eighth grade and tenth grade, but at all levels students have the capacity to think scientifically and engage in the practices of science.

The new framework is based on the research of scores of researchers in education and the professional wisdom of numerous experienced educators. The framework brings to life a major conclusion that has cut across many of the reports published by BOSE; that to learn science deeply and to appreciate the wonder and beauty of science students need to have opportunities to engage in scientific practices. That is, students need to do the things that scientists do: pose questions, plan and carry out investigations, gather and interpret data, develop arguments about their findings, and then communicate their arguments to peers.

The emphasis on practices – on doing science – represents a shift from what happens in typical science classrooms where there has too often been an emphasis solely on the facts of science. Emphasizing practices acknowledges that reading textbooks, listening to lectures and memorizing terms alone does not lead to deep understanding. Nor does focusing only on facts help students appreciate how scientists develop an understanding of the world and the rigorous process by which the scientific community comes to accept one explanation or theory as better than another. We now know from research that students come to understand scientific ideas best when they actively use their knowledge while engaging in the practices of science.

Unfortunately, at present many students in grades K-12 do not have access to opportunities that allow them to experience science as envisioned in the framework. For example, a previous report of BOSE, America’s Lab Report found that most laboratory experiences for high school students were not high quality. They were often disconnected from the ideas presented in lectures or the textbooks and were of the “cookbook” variety where students follow a scripted set of steps with little opportunity to express their own ideas and pursue their own curiosity.

The vision of the framework will help to address this problem of access. It emphasizes that ALL students can learn science and that they should have opportunities to engage in the full range of science practices. The overarching goal of the framework is to ensure by the end of 12th grade that all students have some appreciation of the beauty and wonder of science, possess sufficient knowledge of science and engineering to engage in public discussions on related issues, are careful consumers of scientific and technological information related to their everyday lives, are able to continue to learn about science outside school, and have the skills to enter careers of their choice, including (but not limited to) careers in science, engineering, and technology.

Now more than ever, it is essential for every American citizen to understand science. Science, engineering, and the technologies they
connecting high school students to opportunities with NASA

Rick Varner, MEd
Aerospace Education Specialist, AESP
NASA Goddard Space Flight Center, Greenbelt, MD
Richard.S.Varner@nasa.gov
http://www.aesp.psu.edu/

It is no secret that we learn through the combined experiences that make up our lives. This fact has been well established and studied. Experiential learning and “learn by doing” mantras have long been a foundation for the best practices of teaching; however, technological advances over the past decade and the pervasive use of the internet in our daily lives begs to renew old questions. Are simulated or digitally replicated experiences as efficacious as the “real thing”? When are the two types of experiences most productive and where do negotiations begin and end to measure the benefits and drawbacks of each?

During the dawn of computer simulations in science education, it was easy to focus on two scenarios for use of simulations in place of the hands-on lab experience: if the laboratory experience was too dangerous with the available safety precautions and if the cost of the laboratory experience exceeded its practical classroom use. As classroom and laboratory budgets have diminished, computer simulations may replace more activities that were once a common part of the secondary high school experience.

Today’s simulations are growingly sophisticated and there is an ever-astonishing blur between reality and CGI for the high-end productions. This is clearly evident in the film industry. For example, it doesn’t take long for the viewer to forget that half of the characters in the fantasy movie Avatar do not actually exist outside of the production.

Randy Bell and Lara Smetana have published research through NSTA that suggests four content and instructional considerations when using simulations: 1. Use computer simulations to supplement, but not replace instruction, 2. Keep the instruction student centered, 3. Clearly recognize the limitations of the simulations and 4. Make the content, not the technology, the focus of the simulation.

Bell and Smetana’s work reinforces that simulations should not stand alone in place of investigational experiences and instruction. They are just one other tool available to teachers and students on the quest for understanding and learning. Real world opportunities will always include computers crunching the data or creating the models to help us better understand the data, but studying the Earth sciences, for example, should involve getting wet or dirty in the process. It is for this reason that the Apollo astronauts spent a great deal of time in the desert southwest to learn the in situ skills required to don the hat of a geologist during their sample collecting journeys on the lunar surface.

For the past seven years, I have had the great privilege to work with NASA Education and many of its missions. Throughout this experience, I have learned volumes about what is available to students for practical and authentic research. Short of boarding a spacecraft and orbiting the Earth, many of these opportunities are available via Internet based data sets and imagery. Teachers and students can use this data to interpret a myriad of changes in our Earth’s environment. Most of this online data is not a simulation, rather reflective of the actual data collected remotely by Earth Observing Satellites and Sensors. As Bell and Smetana suggest students would need to have
a conceptual understanding of the questions they would pose for their research. Research conducted through NASA data would be very much student and content centered, as the data exists for interpretation by the researcher and, in most cases, is not in a prepackaged right or wrong answer scenario.

Need a place to start? Try the NASA Earth Observatory site. This site has articles authored at a variety of reading skill levels and connected to the discussions being held by the scientists who manage the research for the respective satellites and sensors. From the background information of the NASA Earth Observatory, NASA Earth Observations (NEO) offers a site where NASA Earth Science Data can be visually manipulated and data sets can be compared across environmental parameters.

NASA's Giovanni is also a web based application “that provides a simple and intuitive way to visualize, analyze, and access vast amounts of Earth science remote sensing data without having to download the data.” Giovanni is a bit more involved than NEO, but extremely useful for real time research that is applicable across the scientific community.

If this feels like jumping in at the deep end, the My NASA Data site has been used by teachers at the middle grades and offers an introductory approach to using NASA Data for research. My NASA Data offers You Tube tutorials and lessons, and a series of lessons and project suggestions. It is a more guided path for the secondary teacher or student, but provides many of the skills that lead to the paths less worn for use of NEO and Giovanni.

These resources represent just a handful of Earth Science opportunities for students interested in these fields of study. On the NASA Student Programs web site there are more than eighty opportunities described across a wide variety of disciplines and populations. These opportunities are not all data set sources. In deed, most offer the very real world connections to NASA personnel and resources that many student researchers may assume are not within their reach.

With the new NASA Education Portfolio currently being developed, the following areas have been identified as Key Achievements for fiscal year 2013:

• Providing experiential opportunities, internships, and scholarships for high school and undergraduate students
• Using NASA's unique missions, discoveries, and assets to inspire student achievement and educator teaching ability in STEM fields

Each of these outcomes are based on experiential learning that connects teachers and students to NASA personnel and its resources. At NASA's Goddard Space Flight Center, 200-300 high school and college interns regularly participate in real world experiences throughout the summer months and for many their research continues into the following semesters, often with technological career connections to NASA or one of its contractors upon graduation.

While I am partial to seeing students pursue studies in Science, Technology, Engineering or Mathematics (STEM), whatever career path a student pursues should be one for which they have developed a passion for learning and discovery. The next generation of careers may be based upon the foundations of today's learning, so building experiential foundations is essential to the life long learning process required to compete for future STEM careers. The “real world” of the future will consist of a whole new genre of experiences gloriously innovative and immersive. Take your steps now to start this exciting journey: become a participant, instead of a spectator, and collect experiences that will separate your portfolio from the rest.

References:


5. Giovanni, NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MS. http://disc.sci.gsfc.nasa.gov/giovanni/overview/what-is-giovanni.


The American Junior Academy of Sciences

Peter Faletra, PhD
Delegate to AAAS Council
Past President of the NAAS
peter.faletra@gmail.com
http://www.amjas.org/index.php

The American Junior Academy of Sciences (AJAS) is America’s only research honor society for high school students aspiring to become scientists. The AJAS believes that the most effective way for students to learn science is to perform authentic research, and to follow the scientific process to its logical end by writing a research paper and either publishing or presenting it to a group of scientists. If students are not given this opportunity, it is somewhat like a musician who composes music but never performs for an audience. The Journal of Experimental Secondary Science serves a pivotal role in the scientific process by providing a venue for students who have done outstanding research and been through a scientific peer review. The journal's rigorous review of student research by accomplished scientists allows young scientists to experience one of the most important and challenging elements of the scientific process.

Membership in the AJAS is by invitation only. Each state Academy of Science chooses premier high school students from their state to nominate as AJAS delegates. Eighth grade students are also invited in some cases. Each state Academy of Science also determines the guidelines by which their state AJAS delegates are chosen. Usually, scientists in each state's senior academy organize their respective junior academy of science. Under the direction of the Junior Academy of Science, a statewide scientific research program is organized whereby students who have done outstanding research are chosen to attend the national AJAS convention.

The national AJAS convention meets in conjunction with the American Association for the Advancement of Science (AAAS) annual symposium. The AAAS is the world's largest general science organization and publishes the journal "Science". With more than 138,000 members and 275 affiliated societies, the AAAS serves as an authoritative source for information on the latest developments in science and bridges gaps among scientists, policy-makers and the public to advance science and science education. An important purpose of the AAAS is to support young scientists in middle school and high school. It recognizes that these young people are its future and provides a venue for these young scientists to present their research along side the world's leading researchers at their annual symposium.

Early Matters

Desmond Murray, PhD
Associate Professor of Chemistry
Andrews University & Berrien RESA Math Science Center
Founding Director
Building Excellence in Science and Technology (BEST Early)
murrayd@andrews.edu
http://www.bestearly.com/

Why has it become accepted practice for our youth to play musical instruments, sports and video games early but not to conduct independent lab experiments and research until much later? This science education convention persists in the face of our own personal experiences and objective findings from neuroscience and pedagogy, for example, which declare affirmatively that - early matters.

As we all know, while there are some exceptions, by and large, our educational system does not provide real authentic independent research opportunities for the majority of students until graduate school. This means the majority of students complete four years of high school and four years of college without any research experience. This in itself means that they are not generally exposed to and competent in such marketable skills as careful observation, documenting data/experience accurately, communicating results effectively, and critically thinking through what their observations mean. These are skills urgently and broadly needed in our 21st century information-innovation economy, regardless of job, career or field.

In my opinion this is an unfortunate waste. A waste of time and a waste of talent. It is wasted time and opportunity since we all know that human curiosity is very high in our early years. Our current science education system does not maximize fully, leverage effectively, or
nurture efficiently this window of opportunity in the human life cycle. In fact, many have argued convincingly, and I agree completely, that we have traditionally done the opposite: discouraging curiosity and creativity, valuing conformity, rote-memorization and standardization rather than independence and critical thought. It is as if we all seem to forget that experiments made us modern.

Our current system also wastes the most valuable resource in the 21st century economy – our human resource. By not fully utilizing and harnessing our human resources early in our pursuit of knowledge and for the purpose of innovation, we are losing out. Students in this content-obsessed system become disinterested in science early and consequently the science talent pool and STEM workforce is diminishing. This means we are losing insights, discovery, creativity, and productivity in a world where more is needed to address global problems such as health, energy, rapid urbanization, and scarcity of essential resources from water to minerals.

For the last 17 years, I have been engaging both high school and college students, curricularly and non-curricularly, in early research. I can assure you that our students are not too young to research but are in fact very eager to do so. These years of experience has embolden my efforts to advocate and provide our youth with opportunities to conduct research early regardless of whether their career goals are in science, technology engineering, and mathematics.

There is no valid reason why high school students who drive cars, use computers, excel at video games, and navigate a host of 21st century technologies cannot also recrystallize solids, isolate DNA, rotovap solvents, reflux reactions, separate mixtures, analyze protein gels, or operate infrared, Raman, UV-Vis, NMR and AFM instruments.

Early research participation facilitates a host of desirable and needed outcomes, such as: (a) building the STEM workforce, (b) capturing the most innovative and productive years of the human life span, (c) increasing investments in young ‘homegrown’ researchers rather than foreign post-docs, (d) providing an avenue for seamless transitions and interactions between secondary and tertiary science education, and (e) facilitating a sustainable culture of innovation, discovery and development.

Our democratic western societies are built, in part, on the values of open and free. I believe that with respect to research, discovery and innovation, we need to institutionalize another value – early. I believe that universal adoption of early research participation programs is fully consistent with former United States President Franklin D. Roosevelt’s statement: “We cannot always build the future for our youth, but we can build our youth for the future.” Furthermore, it is in the spirit of our best pedagogical practice as eloquently, succinctly and powerfully expressed by the American poet, Theodore Roethke, “I learn by going where I have to go.”
A Comparison of Laboratory-Reared Stock and Captured Fruit Flies (*Drosophila melanogaster*) using Upward Movement, Phototaxic, and Starvation Assays Reveals Significant Behavioral Differences

Trinity Russell¹* and Richard Kurtz²

Student¹, Teacher²: Commack High School, 1 Scholar Lane, Commack, NY 11725

*Corresponding author: trinity.russell13@gmail.com

Abstract

Fruit flies (*Drosophila melanogaster*) are model research organisms and are frequently reared in research institutions. Research specimens domesticated in a laboratory-reared setting may have different behavioral phenotypes as compared to their wild counterpart. Additionally, it has been determined that the absence of “key” stimuli in the physical environment of captive animals may result in altered behavioral patterns¹. The purpose of this investigation was to test for differences in the behavioral phenotype of outdoor captured fruit flies as compared to laboratory-reared fruit flies with the use of three tests: an Upward Movement Assay, a Starvation Assay, and a Phototaxic Assay. Results from the Upward Movement Assay demonstrated statistically significant differences in the vertical moving speed of laboratory-reared flies and outdoor-caught flies. Results from the Phototaxic Assay revealed outdoor captured fruit flies exhibited a natural phototactic behavior while laboratory-reared flies exhibited an inverse phototactic behavior. The Starvation Assay proved that flies recently descended from outdoor caught fruit flies were able to withstand starvation twice as long as laboratory-reared flies. These results indicate a strong behavioral difference between flies that are descended from laboratory stock and flies that are caught from the outdoors. Research regarding the differences in domesticated organisms is an imperative topic for study because domestication and genetic drift have the potential to alter the behavioral phenotype. Changes in the behavioral phenotype may jeopardize the results of research experiments. Thus it is crucial to have a thorough comprehension of the behavior of outdoor caught fruit flies as compared to the behavior of their wild counterpart. It is suggested scientists change their fruit fly stock every few hundred generations in an effort to protect the natural gene pools of organisms which are bred in captivity for extended periods of time.

Introduction

Domestication is the evolutionary genetic change arising from the transition of a population from nature to deliberate human cultivation². Animals have been domesticated both unconsciously and methodically since the end of the Pleistocene Era. Due to cultivation and routine human interactions with species, selection pressures are created and the affected species is forced to adapt to a new environment. The transition of a free-living culture to captive status is often accompanied by changes in availability and/or accessibility of shelter, space, food, water, predation, and the social environment³. It is implied that the phenotype of the domesticated species will differ from its counterpart once it has undergone domestication. As natural selection has the ability to change the frequency of traits within a population, it has been discovered that the brain of an organism has the tendency to shrink when bred in captivity for extended periods of time³. Two approaches have been developed to gain a better understanding of the domestication process: a comparison of wild and domestic stocks (of a species), and the study of wild and domestic hybrids⁴.

This experiment utilized the first approach, a comparison of the wild and domestic stocks of a species, to understanding the domestication process. The wild stock (outdoor captured fruit flies) is used as a representative and ancestor of the domesticated population (laboratory-reared fruit flies). A comparative approach between specific populations of wild versus domesticated animals at a single point in time is applied to this study. This experiment and other similar research tests suggest altered behaviors are a result of altered genes. Specific alleles have a relatively large impact on the development of behavioral characteristics specific to domesticated animals. For example, a study was conducted to identify the genetic variation between cultivated rice and its wild progenitor. This study assessed the genetic basis of the changes associated with the process of rice domesticaion. A total of 19 traits related to domestication in cultivated rice were discovered⁵.

Another study demonstrated that laboratory-reared flies and wild fruit flies exhibit differences in an ovipositor choice test⁶. In the choice tests, using white and black artificial ovipositor domes, the wild flies’ selected black domes almost exclusively, but the laboratory-reared flies failed to display any preference⁶. Additionally, a study compared the behavior of wild and domestic stocks of Brook Trout. Results proved the domesticated group was much more vulnerable to trap-netting than the wild groups. After a week of trapping fish, 84% of the captured fish were from the domesticated group⁷.

*Drosophila*’s favorable characteristics make it an ideal research specimen. They require minimal care, space, and equipment. In addition, fruit flies are easily cultivated in the laboratory; they have a high fecundity, and a short generation time. Thus, they are highly susceptible to genetic drift and domestication. *Drosophila* is presently one of the most commonly used model organisms in biological research. They serve as a genetic model for numerous diseases. They are also used to study aging, oxidative stress, immunity, diabetes, cancer, obesity and drug abuse⁸.

Laboratory-reared fruit flies are essential in biological research and it is crucial to have a thorough understanding of their...
behavior compared to the behavior of a fly originating from a natural environment. Laboratory-reared fruit flies which are domesticated have the ability to jeopardize the validity of significant experimental results. Although this is known, domestication of fruit flies within a laboratory is hardly taken into consideration.

The laboratory-reared fruit flies were reared in the laboratory for approximately 1,500 generations which made them an ideal specimen for a study of domestication. Outdoor captured and laboratory reared fruit flies were both genetically variable. However, low genetic variation due to a consistent environment, laboratory conditions such as small culture vials, and genetic drift, accounts for predictably less genetic variation among the laboratory reared population. Neither of the fly populations were isogenic. Descendants of flies caught outdoors (wild) and descendants of laboratory-reared fruit flies (domesticated) were tested for differences in behavioral phenotypes with the use of three assays: an Upward Movement Assay, a Phototaxic Assay, and a Starvation Assay.

The Upward Movement Assay was used to test the flies’ vertical movement. The Phototaxic Assay was used to test fruit flies locomotive movement in response to the stimulus of light (phototaxic behavior). The Starvation Assay was constructed to test the flies’ abilities to withstand starvation.

**Materials and Methods**

Outdoor captured flies were caught from locations on Long Island, New York: Commack, East Northport, and Mount Sinai. Strain Oregon R laboratory reared fruit flies were obtained from Carolina Biological. These flies are known to have descended from flies cultivated in the laboratory for 54 years. Carolina Biological’s fruit fly culture has remained unchanged without the addition of fruit flies from external sources. Both outdoor captured and laboratory-reared fruit flies were cultivated in the school laboratory for no more than 20 generations. Fruit flies used for this experiment were descendants of either outdoor captured or laboratory-reared locations, and they were all cultivated at the same ambient conditions. The fruit fly life cycle consist of four stages: egg, larva, pupa, and adult. At a standard classroom temperature of 21°C, the fruit fly life cycle lasts two weeks. Flies were re-cultivated every two weeks. The Upward Movement Assay was used to determine the flies’ upward velocity toward a light source (figure 1). This assay consisted of a laboratory clamp, a 35cm glass tube with a 0.5 cm diameter, a bright light, a fly aspirator, syringe, and a stopwatch. This experiment was conducted in a dark room. The glass tube was held vertically by the laboratory clamp and the bright light was attached to the top of the glass tube. Flies were individually aspirated and syringed into the glass tube. The 25cm flight time was recorded. Data was averaged and recorded as time per centimeter. Each fly location consisted of 5 blocks and every block had 10 trials. In total, the Upward Movement Assay was comprised of 200 trials (Supplementary tables 1-4). The Phototaxic Assay was used to determine the strength of the flies’ phototaxic behavior. This assay consisted of Plexiglas, a syringe, a fruit fly aspirator, a bright light, and two flight arenas (light and dark). Plexiglas was cut into 3 fragments and glued together to create narrow corridors (2mm diameter) for the flies to travel (figure 2). One narrow canal branched off into two canals. At the end of each canal was either a light or dark flight arena. The bright flight arena was illuminated with an incandescent light and the dark flight arena was fully covered in black construction paper. The remainder of the apparatus was enclosed in red translucent paper. Flies are unable to distinguish red light and this prevented the flies from getting distracted by excess light. Fruit flies were injected into the device and they showed an attraction for either the light or dark flight arena. The data from this assay was recorded in 5 blocks: each block contained 5 trials totaling 100 trials (Supplementary tables 5-8). The Starvation Assay was used to test the flies’ ability to withstand starvation. This assay consisted of vials with moist cotton. Ten flies from each location were anesthetized and placed into vials with moist cotton. At the end of each day the number of flies alive were recorded. Each fly location had 5 blocks of 10 flies, totaling 200 trials (Supplementary tables 9-12).
Results

Laboratory-reared fruit flies exhibit behavior which was different from its wild counterpart. They exhibited a different locomotive behavior, a stronger phototaxic behavior, and they were able to survive longer when starved.

Results from the Upward Movement Assay demonstrate a statistical difference in vertical flight of lab reared and wild caught flies. Flies which were reared in laboratories took approximately six times longer to travel the same distance as flies which were captured from the outdoors. In addition, the behaviors of laboratory-reared flies were observed to be sporadic and inconsistent (figure 3). Results were assessed with an Analysis of Variance (ANOVA Microsoft Excel). A single factor 1-Way Analysis of Variance Test (ANOVA) was applied to the data. The upward movement probability was 1.12E-06 indicating that there was a statistical difference in the vertical flight speed of the flies (due mostly to the difference in lab-reared flies).

Results from the Phototaxic Assay demonstrate a statistical difference in phototaxic behavior of lab reared and wild caught flies ($p = 0.004$). Flies which were caught from the outdoors all exhibited a natural phototaxic behavior. Flies which were reared in laboratories exhibited an inverse phototaxic behavior (figure 4). A non-parametric test was applied to the ordinal data. The chi square probability was 0.004. This means that there is evidence to support the alternative hypothesis that the behavior/movement is dependent on the origin of the fly. The probability of accepting the alternative hypothesis when the null hypothesis should have been accepted is very small.

Results from the Starvation Assay show that fruit flies which are reared in laboratories survived a maximum of three days. Fruit flies which are captured from the outdoors survived a maximum of four to five days. Fruit flies which were reared in laboratories died more rapidly than flies which were captured from the outdoors (figure 5).

Figure 4. Phototaxic Assay Data. Wild caught flies were mostly attracted to the light flight arena and laboratory reared flies were mostly attracted to the dark flight arena. Wild caught fruit flies exhibited a natural phototaxic behavior. Laboratory reared flies (Carolina biological) exhibited an inverse phototaxic behavior. (Error bars = Standard Error)

Figure 5. Starvation Assay Data. Flies from outdoor captured locations (Mount Sinai, East Northport, and Commack) survived a maximum of four to five days when starved. Flies from Carolina Biological were all dead by the third day of starvation. (Error bars = Standard Error)
Discussion

Humans can play a role as a buffer between an animal and its environment. Fruit flies that are reared in laboratories have reduced sensitivity to their environment\(^2\). It is speculated that the laboratory-reared fruit flies' impaired vertical movement may be caused by a lack of predation. Laboratory-reared fruit flies may have not honed mechanisms for escaping predators. Without the need to hone flight skills, laboratory-reared fruit flies may have developed flight mechanisms that are different from their wild counterpart. Fruit flies which are reared in laboratories may not be exposed to a natural 12 hour light 12 hour dark photoperiod. It is speculated that an inconsistent photoperiod may cause a lab reared fruit fly to develop an inverse attraction to light. It has also been discovered that wild caught flies had evolved better mechanisms for storing sugars and fats as compared to laboratory-reared flies\(^3\). Higher extrinsic adult mortality rates leads to an early reproductive effort, eclosion at an earlier age and a smaller sized fruit fly\(^4\). Therefore, the inability to properly store fats and sugars, ultimately affect laboratory-reared offspring. A recent study which tests starvation in Oregon R fruit flies suggests there is a correlation between the sex of a fruit fly and its ability to withstand starvation. Results suggest that wild female fruit flies survive longer than wild male flies. In addition, wild female fruit flies have been discovered to survive longer than both male and female laboratory reared flies. The sex of a fruit fly is a possible source of variability within the results of the three behavioral assays\(^5\).

Laboratory reared, and outdoor captured fruit flies were both genetically variable. The impact of genetic diversity and its effect on the results is unknown due to the inability to determine the level of variability for each fly location. While it may be expected for the wild population to display a greater behavioral variance, compared to the domesticated stock, their behavior was surprisingly similar. The results of the three assays depicted the laboratory reared population to have a greater variation in behavior. The results allude to the assumption that laboratory reared fruit flies have a greater level of variability. A comparative study of genetic variation between outdoor captured and laboratory reared fruit flies would be a fruitful avenue for additional research. Furthermore, domestication may possibly be a result of chromosomal and clustered blocks of genes. This clustering of genes may provide explanations for the genetic basis of domestication\(^6\).

It has been shown that domestication has an important effect on the development of the domestic phenotype. Scientists who use fruit flies for experimental purposes are transferring free living insects to captive status. This domestic phenotype results in genetic drift, artificial selection, and relaxed selection\(^7\). Many organisms are domesticated in laboratory-reared settings and these organisms may have a different behavioral phenotype from their wild counterpart. Changes in behavioral phenotype can alter the outcome of experiments and invalidate scientific results. Scientists frequently use laboratory-reared fruit flies as experimental specimens and the effects domestication has on the behavior of a fly should be considered with respect to the results. It is suggested scientists change their fruit fly stock every few hundred generations. “Natural” gene pools should be protected when breeding animals in captivity for extended periods of time. In the future, additional outdoor captured fruit flies from Long Island will be compared to laboratory-reared fruit flies from Long Island. Flies from various states and various parts of the world will be tested to further validate the results. A test will be conducted to determine the rate of evolution for a fruit fly. With this data, scientists will be able to create a more precise time period for changing and restoring their fruit fly stock. A study will also be conducted to determine the specific genes responsible for the domestication of a fruit fly.

References


Supplementary Data

Table 1. Mount Sinai Upward Movement Data (wild caught, time in sec. to travel 1 cm).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Grouped Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8</td>
<td>1.2</td>
<td>0.6</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>0.5</td>
<td>0.8</td>
<td>0.7</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>0.6</td>
<td>0.5</td>
<td>1.0</td>
<td>1.1</td>
<td>0.9</td>
</tr>
<tr>
<td>4</td>
<td>0.5</td>
<td>1.3</td>
<td>1.5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>5</td>
<td>0.7</td>
<td>1.2</td>
<td>0.7</td>
<td>1.2</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>1.2</td>
<td>1.2</td>
<td>0.5</td>
<td>1.5</td>
<td>1.0</td>
</tr>
<tr>
<td>7</td>
<td>0.8</td>
<td>1.0</td>
<td>1.1</td>
<td>0.7</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>8</td>
<td>0.6</td>
<td>1.0</td>
<td>0.9</td>
<td>1.3</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>9</td>
<td>0.8</td>
<td>1.1</td>
<td>0.0</td>
<td>0.7</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>1.1</td>
<td>1.4</td>
<td>1.2</td>
<td>0.5</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Average 0.8 ± 0.1

Table 2. East Northport Upward Movement Data (wild caught, time in sec. to travel 1 cm).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Grouped Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>1.2</td>
<td>0.7</td>
<td>1.2</td>
<td>0.6</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>1.2</td>
<td>1.3</td>
<td>1.1</td>
<td>1.2</td>
<td>0.6</td>
<td>1.4</td>
</tr>
<tr>
<td>3</td>
<td>0.7</td>
<td>1.2</td>
<td>1.6</td>
<td>1.3</td>
<td>0.9</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>1.7</td>
<td>0.8</td>
<td>1.5</td>
<td>1.0</td>
<td>1.2</td>
<td>1.8</td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>0.7</td>
<td>0.5</td>
<td>1.4</td>
<td>0.6</td>
<td>1.7</td>
</tr>
<tr>
<td>6</td>
<td>1.2</td>
<td>0.6</td>
<td>0.7</td>
<td>1.2</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.9</td>
<td>1.6</td>
<td>0.6</td>
<td>1.2</td>
<td>0.6</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>1.1</td>
<td>0.6</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
<td>2.4</td>
</tr>
<tr>
<td>9</td>
<td>1.0</td>
<td>1.2</td>
<td>1.0</td>
<td>0.4</td>
<td>1.2</td>
<td>2.5</td>
</tr>
<tr>
<td>10</td>
<td>0.7</td>
<td>0.6</td>
<td>1.1</td>
<td>0.4</td>
<td>1.3</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Average 1.0 ± 0.1

Table 3. Commack Upward Movement Data (wild caught, time in sec. to travel 1 cm).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Grouped Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8</td>
<td>0.7</td>
<td>1.3</td>
<td>0.7</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>1.4</td>
<td>0.8</td>
<td>1.2</td>
<td>1.0</td>
<td>0.6</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>0.5</td>
<td>1.3</td>
<td>0.8</td>
<td>1.2</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>5</td>
<td>0.6</td>
<td>1.2</td>
<td>0.6</td>
<td>0.4</td>
<td>1.4</td>
<td>0.9</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.8</td>
<td>0.6</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>7</td>
<td>1.5</td>
<td>1.0</td>
<td>1.2</td>
<td>1.1</td>
<td>0.6</td>
<td>1.1</td>
</tr>
<tr>
<td>8</td>
<td>1.2</td>
<td>1.2</td>
<td>1.3</td>
<td>0.6</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>9</td>
<td>0.6</td>
<td>0.8</td>
<td>0.5</td>
<td>0.8</td>
<td>1.3</td>
<td>0.8</td>
</tr>
<tr>
<td>10</td>
<td>1.3</td>
<td>0.6</td>
<td>0.6</td>
<td>1.3</td>
<td>0.7</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Average 0.9 ± 0.1

Table 4. Carolina Biological Upward Movement Data (laboratory-reared, time in sec. to travel 1 cm).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Grouped Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7</td>
<td>0.6</td>
<td>2.1</td>
<td>1.3</td>
<td>1.7</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>1.7</td>
<td>2.6</td>
<td>2.3</td>
<td>13.3</td>
<td>4.5</td>
</tr>
<tr>
<td>3</td>
<td>13.3</td>
<td>13.3</td>
<td>12.1</td>
<td>1.6</td>
<td>0.5</td>
<td>6.0</td>
</tr>
<tr>
<td>4</td>
<td>0.9</td>
<td>8.0</td>
<td>1.0</td>
<td>40.0</td>
<td>0.6</td>
<td>10.1</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7</td>
<td>4.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>6</td>
<td>0.9</td>
<td>0.9</td>
<td>40.0</td>
<td>1.6</td>
<td>2.3</td>
<td>9.1</td>
</tr>
<tr>
<td>7</td>
<td>1.6</td>
<td>1.7</td>
<td>3.3</td>
<td>1.0</td>
<td>40.0</td>
<td>9.5</td>
</tr>
<tr>
<td>8</td>
<td>40.0</td>
<td>40.0</td>
<td>0.8</td>
<td>2.5</td>
<td>0.7</td>
<td>16.8</td>
</tr>
<tr>
<td>9</td>
<td>1.1</td>
<td>1.4</td>
<td>3.0</td>
<td>13.3</td>
<td>0.9</td>
<td>3.9</td>
</tr>
<tr>
<td>10</td>
<td>3.2</td>
<td>0.7</td>
<td>0.8</td>
<td>1.3</td>
<td>3.3</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Average 6.7 ± 0.9

Table 5. Mount Sinai Phototaxic Data (wild caught, Light versus Dark).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Avg. Light</th>
<th>Avg. Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L</td>
<td>D</td>
<td>L</td>
<td>L</td>
<td>D</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>D</td>
<td>L</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>l</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>D</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>L</td>
<td>L</td>
<td>D</td>
<td>L</td>
<td>L</td>
<td>80</td>
<td>20</td>
</tr>
</tbody>
</table>

Attracted to light 75%
Attracted to dark 25%

Table 6. East Northport Phototaxic Data (wild caught, Light versus Dark).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Avg. Light</th>
<th>Avg. Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>L</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>D</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>L</td>
<td>D</td>
<td>L</td>
<td>D</td>
<td>L</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>L</td>
<td>D</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>40</td>
<td>60</td>
</tr>
</tbody>
</table>

Attraction toward light 60%
Attraction toward dark 40%
Table 7. Commack Phototaxic Data  
(wild caught, Light versus Dark).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Trial</th>
<th>Avg. Light</th>
<th>Avg. Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>1</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>2</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>L</td>
<td>D</td>
<td>L</td>
<td>L</td>
<td>3</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>L</td>
<td>4</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>5</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

Attracted to light: 72%  
Attracted to dark: 28%  
Average: 72  
Std. Dev.: 26.8  
Std. Error: 12

Table 8. Laboratory-reared Phototaxic Data  
(wild caught, Light versus Dark).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Trial</th>
<th>Avg. Light</th>
<th>Avg. Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>L</td>
<td>1</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>D</td>
<td>D</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>2</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>D</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>3</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>4</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>L</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>5</td>
<td>20</td>
<td>80</td>
</tr>
</tbody>
</table>

Attracted to light: 39%  
Attracted to dark: 61%  
Average: 32  
Std. Dev.: 26.8  
Std. Error: 12

Table 9. Mount Sinai Starvation Data  
(wild caught, number of flies alive at the end of each day).

<table>
<thead>
<tr>
<th>Day</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>9.2</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7.2</td>
<td>1.3</td>
<td>0.6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3.1</td>
<td>2.1</td>
<td>0.9</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 10. East Northport Starvation Data  
(wild caught, number of flies alive at the end of each day).

<table>
<thead>
<tr>
<th>Day</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8.2</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>6.2</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 11. Commack Starvation Data  
(wild caught, number of flies alive at the end of each day).

<table>
<thead>
<tr>
<th>Day</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8.4</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5.8</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 12. Mount Sinai Starvation Data  
(laboratory-reared, number of flies alive at the end of each day).

<table>
<thead>
<tr>
<th>Day</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5.4</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Discovering the Distribution of Palindromic Sequences in the SMAD4 Gene using Large and Medium Deletions and the Resulting RNA Structure Predictions

Andrew Ninh\textsuperscript{1*} and Lisa Battig\textsuperscript{2}

\textsuperscript{1}Student\textsuperscript{1}, \textsuperscript{2}Teacher\textsuperscript{2}: Fountain Valley High School, 17816 Bushard Street, Fountain Valley, CA 92708

*Corresponding author: andrewninh@ieee.org

Abstract

The SMAD4 gene codes for cell-signaling proteins that prevent abnormal vascular growths. DNA palindromes are inversely proportional sequences that play roles in gene expression through the formation of stem-loops and disease/tumor detection. Previous research approximated that there were 100 palindromes in every 1000 base pairs of a randomly generated sequence. A Java program was written to mutate fasta sequences based on ARUP’s SMAD4 database, information from NCBI, and random locations and another was written to find palindromes in the DNA sequences and output their lengths. These lengths were plotted sequentially using the Mathematica software. By measuring shifts in each mutated plot superimposed on the wild type plot, the number of pixels shifted between peaks was recorded and, using a scale which was measured to be 12 pixels per 100 palindromes, converted into the number of palindromes deleted. The amount of base pairs (bp) deleted was proportional to the amount of palindromes deleted. The relationship between bp and palindromes is described by the equation \( p = \text{round}(-0.242996 + 0.425309\times l) \) such that \( p \) represents the number of palindromes and \( l \) represents the length of a sequence. This linear regression shows that palindromes are evenly distributed throughout the SMAD4 gene assuming the distribution follows a Poisson distribution. Out of every 1000 bp, there are approximately 420 palindromes in the SMAD4 mRNA which is approximately 475 palindromes in SMAD4’s Primary Assembly genomic region. The SMAD4 gene exhibits 275 (137.5\%) more palindromes than the randomly generated palindromic distribution projected by previous research. Finding the distribution of palindromes in RNA molecules can lead to future research and classification of key regions that determine the shape of secondary, tertiary, and quaternary structures.

Introduction

Palindromic sequences in DNA are sequences of nucleotides that are identical when read from the 5’ to 3’ end on one strand and 3’ to 5’ end of the complementary strand (and vice versa). Genetic palindromic sequences, whose latter half is the reverse complement of the former half, differ from lexical palindromes which read exactly the same forwards and backwards. Palindromic sequences have a tendency of creating hairpin loops which are secondary RNA structures. These hairpin loops (also known as stem-loops) have many purposes which include protein synthesis repression\textsuperscript{1}. Nonrandom distribution patterns of DNA palindromes have also been found in cancer cells and are attributed to mycogene amplification\textsuperscript{2,3}. Secondary, tertiary, and quaternary structures of RNA are important to its function and expression; if an RNA molecule was mutated by a deletion or other mutation (particularly large mutations), RNA expression will be compromised as it may be too susceptible to enzymes or it may be unable to unwind during expression.

Small mutations involve the deletion of a few base pairs and can cause frameshifts (unless a multiple of 3 base pairs are deleted). Large mutations can range from the deletion of entire genes to exons of base pairs and tend to be more detrimental to the functioning of a gene.

The SMAD4 gene (MADH4, Mothers against decapentaplegic homolog 4) codes for the SMAD4 protein. The SMAD4 protein suppresses tumors through cell signaling by regulating the TGF\( \beta \) protein\textsuperscript{4}. The protein also recognizes and binds to sites with the 8 bp palindromic sequences (GTCTAGAC). Mutations in this gene cause many cancers (50\% of pancreatic cancers involve a mutated SMAD4 gene), particularly colorectal cancers\textsuperscript{5}.

In addition to the gene’s connection with cancers, SMAD4 mutations are also common in autosomal dominant syndromes such as Juvenile Polyposis Syndrome (JPS), Hereditary Hemorrhagic Telangiectasia (HHT), and a hybrid of the two diseases (JPS-HHT)\textsuperscript{6,7}. JPS is characterized by small benign polyps which can increase a patient’s risk of gastrointestinal cancers\textsuperscript{8}. HHT is characterized by arteriovenous malformations (AVMs). AVMs are abnormal connections of veins and arteries that clump together. AVMs can occur in various parts of the body but usually occur in the nervous system. Patients with AVMs are at an increased risk of bleeding. This is especially fatal if the AVM is located in the brain or spinal cord.

The purpose of this study is to find a possible correlation and proportionality between the amount of deleted base pairs of a large deletion mutation and the resultant number of palindromes removed and stating this correlation in an equation. By finding how many palindromic sequences were deleted in each large mutation and comparing this ratio with palindromic deletions by medium deletions, an accurate number of palindromes in however many base pairs were found. The results only account for perfect palindromic sequences, not approximate palindromes. It is important to find DNA palindromes in RNA because of tendencies of forming hairpin loops and pseudoknots which are key components of secondary, tertiary, and quaternary structures; by finding concentrated spots of palindromes, further research can be done into classifying specific regions of RNA molecules that are key in RNA functioning.
By using large and medium deletion mutations an accurate rate at which palindromes are distributed throughout the gene were found. This is based upon the assumption that palindromes are uniformly distributed throughout the gene through a Poisson distribution\(^9\). If the amount of palindromes per base pair is proportional, then palindromes are distributed similarly to palindromic distributions in randomly generated and short genomes (such as that of bacteria)\(^11\).

Deletions in the SMAD4 gene, which cause frameshift mutations, cause a shift (by deletion of palindromes) in the distribution of palindromic sequences that is proportional to the number of base pairs deleted. The shift in palindromic sequence distribution is similar to the frameshift in a nucleotide sequence following deletions and other frameshift mutations. As deletions occur and palindromic sequences are removed, the remaining palindromic sequences are shifted leftwards.

Materials and Methods
The appropriate .fasta files from GenBank were downloaded: Homo sapiens Chromosome 18, GRCh37.p5 Primary Assembly (region: 48,556,583 to 48,611,412) and Homo sapiens SMAD family member 4 (SMAD4)\(^{12}\) (accession number NM_005359). The large deletion mutations on the SMAD4 database were converted into a readable .csv file (6 columns: (1) Segment (exon, etc.) (2) Type of mutation (3) Start location of the mutation (integer) (4) End location of the mutation (integer) (5) Nucleotide change (6) Type of protein change) titled “mutations.csv”. Three Java programs were written: (Supplementary Data Index 1) reads mutation information from the mutations.csv file and accordingly mutates a .fasta file (Supplementary Data Index 2) finds palindromes and prints the length of each palindrome (length) (Supplementary Data Index 3) reads and writes .fasta files which are necessary file types for genetic sequences. Parameters were set on lengths of palindromes because 2 base pair (bp) palindromes are not important and a single nucleotide would not be a DNA palindrome. Due to the lack of large deletion mutations from the ARUP database, randomly chosen coding sequences (such as exons) were chosen to be deleted and added to the mutations.csv file. The information is divided up by columns and stored into 6 arrays (this way, the element [0] of each array would pertain to the first mutation and so on). The ListLinePlot function on Mathematica was used to plot the sequential lengths (sequence x length) of both the Primary Assembly and mRNA palindromes (see figures 1-2). Each respective plot's number of palindromes was counted. The SMAD4 mRNA .fasta file and the mutation-list on the .csv file were run through a Java program which created and saved 15 mutated .fasta files. The 15 mutated mRNA sequences were then run through the palindrome-finding program. Each sequence resulted in a string of numbers and commas (in the format “x,”). The outputs were saved into respective .csv files to be graphed. The .csv files are in a Cartesian system: the x-coordinates are the sequential appearance of each palindrome in the gene and the y-coordinates are the lengths of each palindrome. Lengths are chosen as the chosen y-coordinate because they are unchanging (unlike frequency) and would thereby make comparison and change measurement more accurate. Afterwards, a Mathematica program was written and employed the ListLinePlot function to plot the sequential lengths and then superimposed the mutated lengths graphs on the non-mutated lengths graphs (Supplementary Data Index 4). The lengths and frequencies superimpositions were saved in .png files. The .png image files were analyzed using the ImageJ program by using the straight-line tool to measure the “pixel shift” (the frame shift is measured by differences in pixels). The two superimposed graphs should be identical except one is shifted leftwards due to the deletion mutation. This process was repeated for different conspicuous peaks (multiple trials whose average value were recorded). Using the x-axis (increments of 10 occurrences) as a scale, pixel differences were measured which were used to compare and measure each pixel shift between peaks. By using accurate pixel-palindrome scales, the number of palindromes deleted per pixel shift was determined. These proportions were recorded and plotted on Mathematica to find a graph of best fit to determine a mathematical model of the distribution of palindromes.

Results
Figures 1-2 are plots of the number of palindromes in the SMAD4 mRNA and the Primary Assembly region. The Primary Assembly has 24,653 palindromes while mRNA has 3,738. The Primary Assembly was measured to contain on average, using random regions, 6.6 times as many palindromes as mRNA. It is important to note the amount of large palindromes found in the Primary Assembly compared to those found in the mRNA. Figures 3 and 4 show the overlap of wild type SMAD4 and deletions found in JPS and pancreatic cancer.

Tables 1-2 demonstrate that the A mutations are large mutations from the SMAD4 database\(^{13}\), B mutations are medium deletions generated from the list of coding sequences provided by GenBank, and C mutations are randomly generated deletions based on the location of large coding sequences/exons.

By measuring each plot using the ImageJ program, it was determined that 12 pixels represented 10 palindromes. This information was used as a scale to convert the distance between peaks in Table 1 from pixels to palindromes (see Table 2).
Figure 2. Palindromes in the SMAD4 mRNA and its predicted RNA structure (Free Energy of Structure = -1497.4 kkal/mol).

Figure 3. Palindromes in the SMAD4 mRNA 668 - 1659bp deletion mutant and its predicted RNA structure (Free Energy of Structure = -1309.0 kkal/mol.). This mutant has been shown to be involved in Juvenile Polyposis Syndrome.

Figure 4. Palindromes in the SMAD4 mRNA 539 - 2197bp deletion mutant and its predicted RNA structure (Free Energy of Structure = -1181.3 kkal/mol.). This mutant has been shown to be involved in pancreatic cancer.
Table 1. SMAD4 deletions with peaks measured in pixels.

<table>
<thead>
<tr>
<th>Location</th>
<th>100 seq</th>
<th>Peak 1</th>
<th>Peak 2</th>
<th>Peak 3</th>
<th>Peak 4</th>
<th>Peak 5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>1 - 1659 bp</td>
<td>12</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>B.</td>
<td>1659 - 1669 bp</td>
<td>12</td>
<td>86</td>
<td>86</td>
<td>86</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>C.</td>
<td>1690 - 1702 bp</td>
<td>12</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
</tr>
</tbody>
</table>

Table 2. SMAD4 deletions with peaks converted from pixels to palindromes.

<table>
<thead>
<tr>
<th>Location</th>
<th>100 seq</th>
<th>Peak 1</th>
<th>Peak 2</th>
<th>Peak 3</th>
<th>Peak 4</th>
<th>Peak 5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>1 - 1659 bp</td>
<td>12</td>
<td>1659</td>
<td>730</td>
<td>730</td>
<td>730</td>
<td>730</td>
</tr>
<tr>
<td>B.</td>
<td>1659 - 1669 bp</td>
<td>12</td>
<td>370</td>
<td>370</td>
<td>370</td>
<td>370</td>
<td>370</td>
</tr>
<tr>
<td>C.</td>
<td>1690 - 1699 bp</td>
<td>12</td>
<td>730</td>
<td>730</td>
<td>730</td>
<td>730</td>
<td>730</td>
</tr>
</tbody>
</table>

Discussion

The linear regression model found that, in mRNA, the SMAD4 gene exhibited approximately 4200 palindromes per 1000 bp. Given this, the Primary Assembly has about 1.14 times as many mRNA palindromes (and assuming palindromes are fairly evenly distributed); about 475 palindromes were calculated to exist in every 1000 bp. This model was found to be accurate within a range of about 13 palindromes (~5%).

The shift in pixels between graphs was the desired measurement rather than simply counting palindromes of each mutated sequence and subtracting that number from the wild type number of palindromes (3,738) because one trial might not provide sufficient information and there might be new palindromes formed as a result of a deletion.

According to the SMAD4 database, the three large mutations ended at bp 1659. This might be caused by unique tendencies of sequences in that exon/CDS to form cruciform or hairpin loops and become cut from the strand. The majority of deletion mutations (10 out of 15) were randomly chosen exons/coding sequences. These sequences might possess a certain quality that renders them unviable deletions.

Technically, the methods are accurate though pixel recording has to be as accurate as possible. If a pixel is only slightly off for the pixel-palindrome scale, entire results could end up being several palindromes inaccurate. Errors and variation in measuring pixel shifts are usually insignificant. However, large errors and inconsistencies demonstrate changes in palindrome distribution.

The distribution of palindromic sequences is best described as a linear regression model which assumes that there is no correlation between palindromic occurrences and there is a constant variance (Poisson distribution). The linear equation: p=round(-0.242996+0.425309*l), which follows a linear regression, was determined.

Table 2 shows the same exact data as Table 1 but the information is converted from pixels to palindromes. Length and Average deleted palindromes were plotted and the line of best fit would determine a mathematical model of the number of palindromes in a sequence of base pairs.

Each length of each deletion in bp by the resultant average number of palindromes deleted were plotted (see Figure 5) and the line of best fit, p=round(-0.242996+0.425309*l), which follows a linear regression, was determined.

In addition, to examine the changes in the RNA secondary structures, the wild type and mutated fasta files were analyzed in GeneBee Molecular Biology Server’s RNA secondary structure prediction program and the resulting structures and stem energy were recorded adjacent to each respective plot.

Figure 5. Number of palindromes by sequence length.
Ultimately, this study provides the first mathematical model of palindromic sequence distribution in the SMAD4 gene and found an accurate proportion of 475 palindromes (Primary Assembly) or 435 palindromes (mRNA) for every 1000 base pairs. By finding an accurate depiction of how palindromes are distributed throughout the SMAD4 gene, further research into palindromic sequences is possible. By knowing how many palindromes exist in however many base pairs, one can find patterns in uniformity of palindromes (or specific palindromes).

The Java program used could be revised to include approximate palindromes. In the future, distribution patterns between approximate and perfect palindromes may be compared to determine whether both approximate and perfect palindromes follow a linear regression or if they both follow a similar equation. A universal mathematical model for palindromic distribution has not yet been found. In addition, the mathematical model might be used in the future to compare and contrast numbers of palindromes of different genes, genomes, and randomly generated sequences. For example, by comparing the variance of various genes and genomes (which require certain palindromes to function) with randomly generated sequences (which should be purely random), evolutionary patterns in genes can be traced.

References


Acknowledgements
I would like to thank and acknowledge the University of Utah's ARUP Scientific Resource for Research and Education for providing the open-source SMAD4 database. I would like to also thank and acknowledge GeneBee Molecular Biology Server's RNA secondary structure prediction program for providing me with the stem energy and image of the RNA secondary structure.
Supplementary Data

Index 1
import java.util.*;
import java.io.*;
public class MutateGene { //arrays can be initialized in the constructor
  private String[] segment = new String[]
  mutationType = new String[]
  nucMutation = new String[]
  proteinChange = new String[];
  private int[] changeLoc1 = new int[]
  changeLoc2 = new int[];
  /**
   * Constructor
   */
  public MutateGene(String fileName) throws IOException{
    BufferedReader b = new BufferedReader(new FileReader(fileName));
    Scanner input = new Scanner(b);
    String info = ""
    while (input.hasNextLine())
      info += input.nextLine().trim() + "",
    StringTokenizer tokenizer = new StringTokenizer(info, "",");
    String cell[] = new String/*number of desired columns */6 *[];
    for (int i = 0; tokenizer.hasMoreTokens(); i++)
      cell[i] = tokenizer.nextToken().trim();
    for (int i = 0; i < cell.length; i++)
      if ((i%6)==0)
        segment[i/6] = cell[i];
      else if ((i%6)==1)
        mutationType[i/6] = cell[i];
      else if ((i%6)==2)
        changeLoc1[i/6] = Integer.parseInt(cell[i]);
      else if ((i%6)==3)
        changeLoc2[i/6] = Integer.parseInt(cell[i]);
      else if ((i%6)==4)
        nucMutation[i/6] = cell[i];
      else
        proteinChange[i/6] = cell[i];
  }
  void mutation(String DNA) throws IOException{
    Mutate m = new Mutate();
    for (int i = 0; i < /*number of columns*/; i++)
      if (mutationType[i].equals("Deletion"))
        m.deletion(DNA, segment[i], nucMutation[i], changeLoc1[i], changeLoc2[i], proteinChange[i]);
    // Other unused commands
  }
}

Index 2
import java.io.*;
import java.util.*;
public class FASTAFile {
  public String readFASTA(String fname)throws IOException{
    BufferedReader r = new BufferedReader(new FileReader(fname));
    Scanner input = new Scanner(r);
    String dna = ""
    String sDescription = input.nextLine();
    while (input.hasNext())
      dna += input.nextLine().trim();
    return dna;
  }
  public void writeFASTA(String fName, String sDescription, String sSequence, String seqLoc)throws IOException{
    String fLoc = //desired location
    PrintWriter out = new PrintWriter(new BufferedWriter(new FileWriter(fLoc + fName)));
    int nStart=0;
    int nNum = 70;
    out.println(sDescription);
    while (nStart < sSequence.length())
      if (nStart < sSequence.length() - nStart < nNum)
        out.println(sSequence.substring(nStart));
      else
        out.println(sSequence.substring(nStart, nStart + 70));
    nStart = nStart + nNum;
    out.close();
  }
}

Index 3
import java.io.*;
public class DnaPalindromes {
  // Finds palindromic DNA sequences and counts them.
  // The variable i is the first boundary of the array and starts at
void findPalindromes(String dna) {
    for (int i = 0; i < dna.length() - 1; i++) {
        for (int j = i + 1; j < dna.length(); j++) {
            String palSeq = dna.substring(i, j + 1);
            int count = 0;
            if (isPalindromes(i, j, dna) && palSeq.length() >= 3) {
                // Necessary command
            }
        }
    }
}

boolean isPalindromes(int i, int j, String palArr) {
    DnaCompare dnaComp = new DnaCompare();
    if (i > j - 1)
        return true;
    if (!dnaComp.correctPair(palArr.charAt(i), palArr.charAt(j)))
        return false;
    return isPalindromes(i + 1, j - 1, palArr);
}

public class DnaCompare {
    // Finds the correct base pairings only for exact palindromes
    boolean correctPair(char base1, char base2) {
        if ((base1 == 'A' && base2 == 'T') ||
            (base1 == 'T' && base2 == 'A') ||
            (base1 == 'C' && base2 == 'G') ||
            (base1 == 'G' && base2 == 'C'))
            return true;
        else
            return false;
    }
}

Index 4

SetDirectory["C:\Users\andrew\Documents\Bioinformatics\SMAD4\SMAD4_grapher"]
Show[ListLinePlot[Take[Import["name.csv"], 1], PlotRange -> {0, 20}, PlotStyle -> Color],
    ListLinePlot[Take[Import["name.csv"], 1], PlotRange -> {0, 20}, PlotStyle -> Color]]
Thymoquinone as a Novel Antibiotic and Chemotherapeutic Agent: a Natural Therapeutic Approach on Staphylococcus aureus, Bacillus anthracis, and Four NCI-60 Cancer Cell Lines

Syed Hossain1, Erin Sikes-Thurston2, Stephen H. Leplla3, and Alexander N. Wein3

Student1, Teacher2: Herndon High School, 700 Bennett Street, Herndon VA 20170
Mentor3: Microbial Pathogenesis Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 33 North Drive, Bethesda, MD 20892

*Corresponding author: hossain.23.s@gmail.com

Abstract
Recently, alternative medicine has received more attention in mainstream science as the pharmacology of many ancient medicinal herbs is understood. Various naturally occurring compounds have shown to work effectively as therapeutics including the antimicrobial drug penicillin and paclitaxel, a widely-used chemotherapeutic. Thymoquinone (TQ), a phytochemical compound found in the seeds of the plant Nigella sativa, has shown potential as a novel therapeutic agent. The seeds have been used as a folk remedy for various ailments for thousands of years. The activity of TQ was tested on two pathogenic bacteria: Staphylococcus aureus, and Bacillus anthracis, and four cancer cell lines: Colo205, SK-Mel-28, PC3, and A549. TQ was applied to the bacteria and cancer cells at seven different concentrations and concentration-response curves were generated. The minimum bactericidal concentration of TQ to Staphylococcus aureus was 31 µg/mL (188.8 µM), and the minimum inhibitory concentration of TQ for Bacillus anthracis was 3 µg/mL (18.3 µM). The minimum bactericidal concentration of TQ on B. anthracis was not determined. TQ promoted the growth of cancer cell lines at low concentrations, but was found to be cytotoxic at concentrations ranging from 50-100 µg/mL (304.5-609.0 µM). While the growth-promoting effects of TQ on cancer cell lines at low concentrations is worrisome for its future development as a chemotherapeutic, TQ deserves further exploration as an effective alternative to customary medicines for chemotherapy and diseases caused by pathogenic bacteria.

Introduction
Nigella sativa, a plant native to Asia, the Middle East, and Africa, has been used for centuries as a natural approach to promote health and fight various diseases1. The major product from this plant is the seeds, which are used as a spice and food preservative. The major bioactive constituent of this seed is thymoquinone (TQ) (figure 1), a phytochemical compound that has been reported to exhibit antimicrobial effects on Gram-positive and Gram-negative bacteria, though it has shown more activity against Gram-positive bacteria2.

One Gram-positive bacterium of concern is Staphylococcus aureus, which causes a variety of pus-forming infections in humans. It causes superficial skin lesions, more serious infections such as pneumonia, meningitis, urinary tract infections, and deep-seated infections3. S. aureus causes food poisoning by releasing enterotoxins into food and toxic shock syndrome by release of superantigens into the blood stream. Another Gram-positive bacterium of note is Bacillus anthracis, a large, spore-forming bacterium that causes anthrax4. Anthrax is a disease of domesticated and wild animals, particularly herbivorous animals, which can be fatal. The bacterium forms endospores that are very long lived in the environment5. If spores are ingested, inhaled, or come into contact with a skin lesion on a host, they reactivate and multiply rapidly. Because of the need for new therapeutics in treating these infections, we tested TQ for its growth inhibitory and bactericidal properties against these species of bacteria.

In addition to its effects on bacteria, TQ has shown activity against several cancers6-12. Cancer has a high mortality rate in the United States, and various types of treatment include chemotherapy, radiation therapy, and surgery; however, these treatments can have adverse side effects. Chemotherapy causes the immune system to weaken, and makes a patient more susceptible to other infections and diseases6. Many forms of chemotherapy also damage healthy cells7. Surgery carries serious risks and can have a long recovery time, and radiation therapy can lead to toxic side-effects. TQ has shown anti-proliferative effects on cell lines derived from cancers of the colon, ovary, breast, larynx, lung, myeloblastic leukemia, and osteosarcoma8-13. TQ has shown a degree of selectivity towards cancer cells, since normal cells such as human pancreatic ductal epithelial cells and mouse keratinocytes are resistant to the apoptotic effects of TQ14,15. Taken together, these studies suggest that TQ could be useful in intervening in the inflammatory cascade, which may cause the inhibition of cancer progression and therefore improve a patient's morbidity and mortality rates.

Based on the long tradition of using the seeds of N. sativa in alternative medicine, various groups have been working on TQ to study its biological effects. In this work, the efficacy of TQ was tested on pathogenic bacteria and cancer cell lines to see if it would limit their growth in vitro. Based on previous work with TQ, the compound should be able to inhibit the growth of the tested bacteria and cancer cell lines at micromolar concentrations. This study expands on previous work by examining the effectiveness of TQ against additional cell-based models of cancer and bacterial species.
Results

Four cancer cell lines, PC3, SK-Mel-28, A549, and Colo205, were treated with thymoquinone (TQ) to determine the compound’s ability to limit their growth. Cancer cell lines treated with low doses of TQ showed greater proliferation than the no-treatment condition; however, at higher concentrations, TQ was able to effectively kill the cells (figure 2). The EC50 for TQ on all four lines was between 50 and 100 µg/mL (304.5-609.0 µM). Colo205 showed no enhanced growth at low doses of TQ while the cell densities of the other three lines were 1.5 to 3 times higher than the no-treatment control.

TQ was tested for its ability to either inhibit the growth of, or kill, two bacterial species, S. aureus, and B. anthracis. TQ was able to inhibit the growth of both bacteria, but did not show bactericidal activity against B. anthracis at the concentrations tested (figures 3 & 4). S. aureus had both a minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) of 31 µg/mL (188.8 µM) of TQ. B. anthracis had a MIC of 3 µg/mL (18.3 µM) TQ.

Materials and Methods

Thymoquinone was purchased from Sigma (St. Louis, MO) and dissolved at 100 mg/mL in DMSO. PC3, Colo205, A549 and SK-Mel-28 cancer cell lines were purchased from the American Type Culture Collection (ATCC, Manassas, VA) and grown in Dulbecco’s modified Eagle’s Medium (DMEM) supplemented with fetal bovine serum to 10 % volume, 50-µg/mL gentamicin (all from Life Technologies, Grand Island, NY), and 10-mM Hepes, pH 7.3 (Quality Biological, Gaithersburg, MD) at 37 °C and 5 % CO2 in a humidified atmosphere. Staphylococcus aureus Cowan I was purchased from the ATCC. PC3 (prostate adenocarcinoma), Colo205 (colorectal adenocarcinoma), A549 (lung carcinoma), and SK-Mel-28 (malignant melanoma) cancer cell lines were harvested from T-75 flasks, counted, and plated in sextuplet in 96-well plates at 3 x 104 cells/well in 150 µL DMEM and allowed to grow overnight. The next day, serially diluted TQ in 50 µL DMEM was added at seven concentrations ranging from 0 µg/mL to 250 µg/mL (0 to 1522 µM) (final concentration) and the cells were incubated for 48 hours. For the last hour of the experiment, 3-(4, 5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT), was added at 0.5 mg/mL and the media was aspirated following this incubation. The MTT was dissolved in 91 % isopropanol containing 0.038 M hydrochloric acid and 0.5 % SDS. The absorbance of the wells were measured at 470 nm and 590 nm using a SpectraMax Pro 190 spectrophotometer and the difference of absorbances was used to determine viable cell density, which was compared to a no-treatment control to determine percent cell density. Frozen glycerol stocks of Staphylococcus aureus Cowan I and Bacillus anthracis Ames 33 (cured of pXO1 and pXO2) were streaked on Luria-Bertani (LB) agar plates and grown overnight at 37 °C. The following day, single colonies of each strain were inoculated into 3mL of LB broth and grown overnight in a shaker-incubator at 37 °C and 225 rpm. The next day, LB broth containing serially diluted TQ at concentrations of 1-1000 µg/mL (6.090-6090 µM) was prepared and the optical density of the culture at 600 nm (OD600) of the overnight culture was measured. To determine the minimum inhibitory concentration, the TQ LB broth was inoculated with S. aureus or B. anthracis at an OD600 of 0.05 and allowed to grow overnight. The following day, the OD600 of the TQ-treated cultures was measured and recorded. To determine the minimal bactericidal concentration, 200 µL of the highest concentration to show growth with TQ and the two lowest concentrations to show no growth were plated on LB agar and grown overnight.
Discussion

This work explored the ability of thymoquinone (TQ), a phytochemical compound from the plant Nigella sativa, to inhibit the growth of bacteria and cancer cell lines. TQ showed growth inhibitory effects against S. aureus and B. anthracis at concentrations of 31 µg/mL (188.8 µM) and 3 µg/mL (18.3 µM), respectively. TQ was also bactericidal against S. aureus at 31 µg/mL (188.8 µM). Halawani found 3 µg/mL (18.3 µM) to be the MIC for S. aureus in his study. However, they used a different method of treating S. aureus with TQ. Halawani did not show TQ’s inhibitory effects against B. anthracis. Perhaps, if specimens of B. anthracis treated with higher concentrations were plated, an MBC could have been found. TQ successfully inhibited growth of both the pathogenic bacteria. There are novel applications using TQ against both the S. aureus and B. anthracis bacteria. The use of TQ against neoplasms also merits further investigation. In our work, TQ was growth-stimulating at low concentrations against cancer cell lines and only started inhibiting growth at concentrations in the 50-100 µg/mL (304.5-609.0 µM) range. Previous work has found that normal cells are resistant to the apoptotic effects of TQ. It would be essential to conduct further experiments using our methods against normal cells to conclude that TQ has no cytotoxic effects against normal cells. Such an experiment would further explore TQ as a potential chemotherapeutic. TQ also showed the ability to inhibit the tumor growth and block angiogenesis with almost no toxic side effects.

Based on the data in our work, and from the works of others, it can be concluded that the black seeds of the plant Nigella sativa (which contains the phytochemical compound thymoquinone) have potential applications in the alternative medicine field and thymoquinone itself, merits further research as a bioactive natural product. Given its anti-proliferative effects against bacteria and cancer cell lines in vitro, it has the potential of being used as a chemotherapeutic or antibiotic. Previous work has found that TQ is much more effective against Gram-positive bacteria than Gram-negative bacteria, and in our work, the compound worked best against S. aureus, since it was both growth inhibitory and bactericidal. In the context of alternative medical treatment, TQ can be administered systemically by consumption of the seeds or dishes containing them, or local infection can be treated by compounding a cream or balm containing the seeds. It could also be presumed, based on historical evidence and the results of this work that TQ can act as a preservative in foods containing seeds from N. sativa. Development of purified TQ as a pharmaceutical is outside the scope of this work, as the biological characteristics such as serum half-life and bioavailability have not yet been studied, but its use in alternative and complimentary therapy is supported by the results.

**Figure 4. Determination of the MBC of TQ on S. aureus and B. anthracis.** The overnight culture with the highest concentration of TQ to show growth along with the two lowest concentrations to show no growth for each species were plated on LB agar plates and grown overnight to determine if the effect was inhibitory or bactericidal.
References


Acknowledgements
This work was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Allergy and Infectious Diseases.
Mutations in Diaphyseal Medullary Stenosis-Malignant Fibrous Histiocytoma Related Gene MTAP Affects Expression of Splice Variants SV2 and SV5

Alla Uts, Olga Camacho-Vanegas, Catalina Camacho-Vanegas, and John A. Martignetti

Student*: High School for Mathematics, Science and Engineering, 240 Convent Avenue, NY, NY 10031
Intern*, Mentor*: Mount Sinai Medical Center, New York, NY
*Correspondence: allauts@gmail.com

Abstract
Diaphyseal Medullary Stenosis- Malignant Fibrous Histiocytoma (DMS-MFH) is a rare autosomal dominant bone syndrome. Thirty-five percent of patients diagnosed with DMS-MFH are at risk of developing a bone sarcoma, malignant fibrous histiocytoma (MFH). Symptoms of this bone disease include bone infarctions, bone pain, leg weakness, and the development of early-onset cataracts. Recently, studies in our laboratory have shown that either of two inherited mutations, IVS8 (-2) A>G or Ex9 (+72) A>G in the MTAP gene locus, results in the development of DMS-MFH. These mutations specifically target the methylthioadenosine phosphorylase (MTAP) gene, located on chromosome 9p21.3. Our laboratory has now shown that the MTAP gene encodes six alternative splicing isoforms. The effect that the two disease-causing mutations have on MTAP splice variant 1 (SV1) and MTAP splice variant 4 (SV4) is complete skipping of exon 9. This project focuses specifically on two additional MTAP isoforms: SV2 and SV5. We chose to work with these two splice variants because they have similar ends and according to our results, one mutation coordinately regulates the expression of both variants. Intriguingly, loss of MTAP activity has been reported in a number of cancers, suggesting that these splice variants may play a role beyond those becoming known for this syndrome. DNA sequence and quantitative real-time PCR (qRT-PCR) demonstrated that only mutation IVS8 (-2) A>G was found to have an effect in the development of DMS-MFH. The methylthioadenosine phosphorylase (MTAP) gene encodes an enzyme necessary in the salvage pathway of adenine and methionine. The MTAP gene encodes an enzyme necessary in the salvage pathway of adenine and methionine. The MTAP gene encodes an enzyme necessary in the salvage pathway of adenine and methionine. The MTAP gene encodes an enzyme necessary in the salvage pathway of adenine and methionine. The MTAP gene encodes an enzyme necessary in the salvage pathway of adenine and methionine.

Introduction
Rare diseases are often ignored simply because of the mistaken belief that only a few affected individuals suffer. What many fail to understand is that collectively, these rare diseases affect a significant portion of the world’s population. Diaphyseal Medullary Stenosis- Malignant Fibrous Histiocytoma (DMS-MFH) is a rare disorder with currently 6 families known to be affected. Those affected may suffer from debilitations, leg and arm weakness, pathologic fractures, irregularity in bone growth, and early onset cataracts. Using linkage analysis and positional gene cloning, our laboratory has mapped the disease locus to chromosome 9p21.3 and identified the disease-causing gene. Given that mutations in this gene result in bone dysplasia and bone cancer, understanding the function of this gene should provide an understanding of its role in normal bone health and how its dysregulation can result in cancer. Ultimately, we hope that understanding the functions of this gene may provide insight into better bone health and the treatment of bone cancer.

The methylthioadenosine phosphorylase (MTAP) gene was linked to DMS-MFH after the region was narrowed down. Before the research taken on by Mount Sinai, the MTAP gene was believed to have eight coding exons. Dr. Camacho-Vanegas et al. (2012) found that the MTAP gene indeed contains three additional exons and can encode six different alternatively splice variants, in addition to the previously well-characterized wild type form (Figure 1). Novel terminal exons were found in all of the MTAP splice variants. As shown in Figure 1, some of the isoforms contain three terminal exons (9s, 10, 11) and others contain only one (9L). It was established that approximately 40 million years ago, two independent retroviruses integrated downstream of terminal exon 8 of MTAP and at some point, during primate evolution, became part of the gene. The MTAP gene encodes an enzyme necessary in the salvage pathway of adenine and methionine. Adenine is essential for the energy of the cell and DNA synthesis while methionine is critical in the process of protein synthesis. Therefore, mutations in the gene can potentially cause an irregularity in the salvage pathway and consequently, the production of adenine and methionine. The reduction of MTAP activity has been accounted for many other cancers, such as osteosarcoma, lung cancer, breast cancer, liver cancer, and other dire cancers.

Previous studies established that two heterozygous changes are associated with DMS-MFH. Five- hundred controls were then analyzed for polymorphisms. None of the controls had the heterozygous changes, strongly suggesting that the changes are pathogenic mutations and not present in unaffected individuals. Through computational analysis, using ESEfinder Release 3.0 program, the data suggested that these changes could potentially affect splicing of the exons. The intronic change at position IVS8 (-2) A>G was predicted to obliterate an acceptor splice site while the identical change at position Ex9 (+72) A>G potentially...
confirmed that the effects of Ex9 (+72) A>G and IVS8 (-2) A>G mutations cause complete skipping of exon 9 in MTAP SV1 and acceptor site. Previous experiments have demonstrated that both searches for a similar complimentary sequence, known as the weak base is mutated, the spliceosome does not recognize the signal and signal. Normally, the spliceosome splices the strong site. Once the adenine to guanine, the spliceosome searches for a weaker splicing site, resulting in the activation of a cryptic splice site. In exon 9. The spliceosome will not recognize the splicing acceptor be to generate isoforms that have a partial or a complete deletion SV2 and SV5, in which exon 9 is the last exon of the variant, will be to generate isoforms that have a partial or a complete deletion in exon 9. The spliceosome will not recognize the splicing acceptor site, resulting in the activation of a cryptic splice site.

When the mutations target the bases and change them from adenine to guanine, the spliceosome searches for a weaker splicing signal. Normally, the spliceosome splices the strong site. Once the base is mutated, the spliceosome does not recognize the signal and searches for a similar complimentary sequence, known as the weak acceptor site. Previous experiments have demonstrated that both mutations cause complete skipping of exon 9 in MTAP SV1 and SV4. After extensive studies on the two splice variants, results confirmed that the effects of Ex9 (+72) A>G and IVS8 (-2) A>G mutations on MTAP SV1 and SV4 are the same.

Materials and Methods

Previously, our lab generated three MTAP minigene constructs from exon 6 - exon 11 of the MTAP gene that included either the control, IVS8 (-2) A>G or Exon 9 (+72) A>G (Camacho-Vanegas et al.2012). These constructs were then transiently transfected into MCF7 cell lines in order to generate RNA transcripts for analysis. The isolated mRNAs were then reverse transcribed into cDNA, and each variant was amplified using forward primer Exon9F: 5’ GACAGATTATGACTGCTGGGA 3’ and reverse primer Exon9Rev 5’ GTTGTGGGAAGCAGTATCCAG 3’. After that, products were separated on the basis of size by gel electrophoresis, bands containing the correctly sized amplicons were cut from the gel and purified, and isolated nucleic acids were cloned and subsequently sequenced. qRT-PCR primers for each variant were then designed and lastly, the levels of each variant were quantified. MCF-7 breast cancer cells are functional in vitro models that were used for minigene transfection. Because these cells are MTAP-deficient, the extracted mRNA should come from only the minigenes. The following vectors were used: Minigene 1- pcDNA 3.1 (MTAP-Control), Minigene 2- pcDNA 3.1 (MTAP-IVS8 (-2) A>G), Minigene 3- pcDNA 3.1 (MTAP-Ex9 (+72) A>G). Dr. Camacho-Vanegas et al. (2012) previously published the vectors that were used for this work. About 10 x 105 MCF-7 cells were plated in each well of a 12-well plate. The minigenes were transfected in triplicates using lipofectamine 2000 when the cells were ~70% confluent. Specifically, 1 ug of plasmid and 4.2 ul of lipofectamine was diluted in serum-free medium and put in each well. After 24 hours, cells were collected for mRNA extraction. RNA was extracted by following the given protocol of the RNeasy Mini Kit (Qiagen). After the elution of RNA, the nucleic acid was quantified using the Thermo Scientific Nanodrop 1000. 500 ng concentration of RNA was used. To reverse transcribe the given RNA samples into cDNA, we used the iScript cDNA synthesis kit. The final mix of each reaction included Buffer 5x, RT Enzyme, 500 ng of RNA and H2O to complete a final volume of 20ul. PCR was preformed twice using two different procedures: one with PCR Master Mix (Promega) (containing dNTPs, 10X Buffer, MgCl2 and Taq) and the other using the AmpliTaq Gold kit (Invitrogen) and adding each reagent separately. A total volume of 50ul was used for the PCR. The PCR reactions were prepared as follows: After PCR, 6x loading dye is added to each cDNA sample and then loaded into a 2% agarose gel (2 grams of agarose in 100 mL of TBE buffer) that is already mixed with 5 ul of ethidium bromide. The 1 Kb Plus DNA Ladder (Invitrogen) was loaded in the first and the last wells of the gel. The gel ran for 2 hours at 150 volts, which was enough to separate the bands. After gel electrophoresis, the appropriately sized bands were cut out and specifically categorized; upper bands were labeled as ‘a’ and lower bands were labeled as ‘b’ (illustrated in Figure 3). The Qiaquick Gel Extraction Kit (Qiagen) was used and its standard protocol was followed. Once the DNA was eluted, the nucleic acid was quantified using the Thermo Scientific Nanodrop 1000. 50 ng of eluted DNA was then ligated. We used the PCR TOPO Cloning Kit (Invitrogen) and followed its given protocol. 1 ul of ligation and 15 ul of Top10 competent cells were pipetted into six new 1.5 ml centrifuge tubes. The tubes were then held in ice for 30 minutes and after, were put in a 42°C incubator for 30 seconds. S.O.C medium was poured into each centrifuge tube and then put in the shaker at 37°C for
one hour. After one hour, cells were plated into ampicillin petri dishes because vector pCR 4- TOPO carries the ampicillin resistance gene. The petri dish cultures were then incubated at 37°C overnight to allow colonies to grow. The following day, colonies were picked from all six samples and then individually put into 15 ml tubes. Each tube consisted of 2 ml of LB, 2 ul of ampicillin (50ug/ul) and a single colony. The tubes were then put in the shaker machine at 200 rpm and 37°C overnight. A minimum of 20 colonies were picked for each band. The LB and bacteria solution was centrifuged after it was dispensed into a 1.5 ml centrifuge tube. Subsequent to centrifuging the tubes, the standard protocol for the Qiaprep Spin Miniprep Kit (Qiagen) was used for the purification of plasmid DNA. 0.2 ul of enzyme ECORI was added to the samples in order to have the insert released. 2 ul of Buffer 10x was poured into each of the tubes as well as 15.80 ul of sterile water. 2 ul of DNA, from the miniprep, was added to each tube as it was then placed in the incubator at 37°C overnight. The digestions that had two bands in the gel electrophoresis were successful in showing both the linearized vector and the insert. Once we verified that the minipreps contained inserts, the samples were then prepared for sequencing. At least 10 miniprep samples were sequenced for each band. The samples were sent for sequencing and then analyzed and aligned with the help of Sequencher v3.0 software application (Gene Codes Corporation). Each sample was compared with the control exons: 6, 7, and 9 of the MTAP gene. The chromatogram of the samples was traced and examined for any unusual patterns. We designed qRT- PCR primers for each splice variant. Each set of primers was tested for linearity, sensitivity, and specificity. Sequences for the specific primers are as follows: SV2DelForward Primer CATAACCTGAAGTTCCAGATGATT, SV5 DelForward Primer GGAGGAAGCATTCAGATGATC, and SV2 / SV5 Reverse Primer GGCAGGATTTCACCTCTGCA. Del represents the deletion of the first nine bases of exon 9. The dilution series of cDNA (Figure 5 (x-axis of line graph)-1: 1/10, 2: 1/100, 3 :1/1000, 4: 1/10,000, 5: 1/100,000) tests the linearity of the primers. The aim is to get a line closest to the regression coefficient approaching 1, avoiding contamination. The specificity of a primer is to determine what the primer is binding to. Good specificity means that the primer binds to only the selected sequences and nothing else. The sensitivity of a primer is how well the primers can detect the needed sequences. Once the primers passed all of the requirements, we quantified the level of each splice variant using them.

Results
According to figures 4a and 4b, the space that is filled with gap marks between exon 7 and exon 9 is a portion of the exon 9 sequence that the spliceosome skips in order to find its complementary splice site. The deletion of the first nine bases of exon 9 is found in samples that are derived from the IVS8 (-2) A>G mutation. However, such a deletion is not seen in sequences derived from the control and the Ex9 (+72) A>G mutation. The illustrated samples are only from the first experiment. Two more experiments were thereafter conducted with another set of samples to confirm the results. According to our results, the weaker splicing site is after the first nine nucleotides of exon 9. Only mutation IVS8 (-2) A>G affected the genetic sequence of the splice variants. Mutation Ex9 (+72) A>G, however, produced no visible changes in the sequence. Table 1 highlights the number of colonies analyzed by DNA sequencing and the number of splice variants that have the deletion in the first nine bases of exon 9. In RT-PCR, the forward primer attaches to the beginning of exon 6 while the reverse primer attaches to the exon 9 end.

The four experiments shown in Figure 5 all have very good linearity with a regression coefficient approaching 1. The results were normalized by averaging the values of beta-actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) housekeeping genes. The first two experiments are testing the levels of MTAP SV2-deleted while the following two are testing the levels of MTAP SV5-deleted. According to Figure 5, the splice variants SV2-deleted and SV5-deleted were overexpressed when compared to normal only in cells transfected with the construct containing the IVS8 (-2) A>G mutation. SV2-deleted and SV5-deleted variants, however, were not present in cells transfected with the normal and +72 constructs.
Discussion

Based on the results, only mutation IVS8 (-2) A>G affected the splicing of MTAP SV2 and SV5. This mutation was the source of the first nine nucleotide deletion of exon 9 in MTAP SV2 and MTAP SV5. In agreement with our hypothesis, the spliceosome looked for a weaker signal to splice and recognized the following AG sequence in exon 9. Figure 4a illustrates the control (ctttAg) and the mutated (ctttGg) splice acceptor site. With one mutation-affected nucleotide change, the spliceosome did not recognize the acceptor splice site and continued scanning to find a different site with the corresponding AG sequence. According to Figure 4b, the spliceosome found the complementary AG pair and spliced off the first nine bases of exon 9.

Based on the data in Table 1, IVS8 (-2) A>G mutation caused the deletion of the first nine bases of Exon 9 in SV2 and SV5 in 84% of the sequenced minipreps. We determined that the Ex9 (+72) A>G mutation does not affect the splice variants because the deletion was not present in the miniprep sequences with Ex9 (+72) A>G. Due to a small pool of patients diagnosed with DMS-MFH, it is difficult to give a forward-looking diagnostic. However, a genotype-phenotype correlation was confirmed by previous research in the DMS-MFH project and was used as an agent in determining the mutations that are important in the phenotype of the syndrome. Based on this correlation, the deletion of the nine bases does not affect the phenotype because in order to correlate with the syndrome, both mutations should produce the same genetic effect, leading to similar symptoms. The genotype-phenotype correlation is only one way of analyzing our results. Another way is taking into account protein. The deletion of the first nine nucleotides of exon 9 may very well be significant. Nine bases are three codons, indicating that the reading frame is conserved. The three codons make up three amino acids; therefore, the IVS8 (-2) A>G mutation causes a deletion of the first three amino acids of exon 9. We can do in vitro experiments to test if the protein can be translated and is functional. To see if such a deletion of amino acids affects the phenotype of the syndrome, we must first discover a specific phenotypic quality that is exclusive to only the IVS8 (-2) A>G mutation. It is only then that we can correlate the deletion of the three amino acids with a specific symptom that is just found in patients with the IVS8 (-2) A>G mutation.

In this project, we used MCF-7 breast cancer cell lines for transfection. MCF-7 breast cancer cells may not express some of the tissue-specific splicing factors that are in bone osteosarcoma cells. In the future, we can avoid negative results by transfecting with an osteosarcoma cell line. We can also analyze why and how deregulation of the levels of MTAP splice variants is involved in carcinogenesis. In order to do so, we will take different tumors and follow a similar methodology by measuring the expression of the splice variants via qRT-PCR. We will test the presence of the deleted variants in the various tumors and determine how overexpression or underexpression of variants correlates with protein expression and ultimately the phenotype of DMS-MFH.
Figure 5. Quantitative RT-PCR Analysis of MTAP SV2-deleted and SV5-deleted Expression in Cells with Each of the Minigene Constructs.

Using specific primers, I quantified the levels of MTAP SV2, SV5, SV2-deleted, and SV5-deleted by qRT-PCR. SV2-deleted and SV5-deleted are splice variants with the nine nucleotide deletion in exon 9.

References


Genetic Analysis of the Brown, Brook, and Tiger Trout Populations in the Lake Champlain Basin

Markie Palermo1,2*, Sandra Latourelle3, and Nancy L. Elwess3

Student1: Essex High School, Essex Junction, Vermont
Intern2, Mentor3: State University of New York at Plattsburgh
*Corresponding author: mpalermo@ccsuvt.org

Abstract
The purpose of this pilot project was to assess both the brook trout (Salvelinus fontinalis) and the brown trout (Salmo trutta) populations in the Lake Champlain Basin using six microsatellite DNA loci. Additionally, the DNA was tested to determine whether hybridization between brook trout and brown trout was occurring to produce a proposed tiger trout (Salmo trutta X Salvelinus fontinalis). Moreover, the hypothesis was that the further apart the test sites, the greater the genetic diversity within the trout populations. DNA samples were taken from adipose or caudal fin clippings through live capture and release at various locations in the Lake Champlain Basin. Furthermore, in the field, the rate of the current was determined, GPS longitude and latitude coordinates were found, the different trout species caught were measured in length, and water and air temperature were recorded. Preliminary data suggested that there were genetic differences in the trout populations at the various locations.

Introduction
The Lake Champlain Basin covers 8,234 square miles in Vermont and New York (Figure 1). The New York State Department of Environmental Conservation (NYDEC) placed genetic monitoring of brook trout, Salvelinus fontinalis as a priority, due to their declining numbers. In addition, brook trout also serve as indicators for the quality of coldwater habitats. Several agencies (NYDEC, U.S. Department of Interior-Fish and Wildlife Division, Lake Champlain Research Institute, and Trout Unlimited) support the genetic monitoring of these fish in order to produce quantitative indicators concerning the health of this ecosystem. Data produced from this pilot project can contribute to planning decisions and protection of habitat that will contribute to a healthy economy and ecosystem. At the same time, non-native brown trout, Salmo trutta are being stocked in streams and tributaries of the Lake Champlain Basin. The concern by numerous regional and state agencies is that the native brook trout populations within the Lake Champlain Basin are declining, due to the mating of the female brook trout with the male brown trout (hybridization), thus producing a sterile tiger trout, Salvelinus fontinalis X Salmo trutta. It is known that tiger trout exist in other parts of the country because of this interbreeding, but it is unknown if this hybridization is occurring in the Lake Champlain Basin. This is an important community issue for two reasons: the need to maintain a healthy ecosystem and because fishing has a significant economic impact within this geographical region. It was hypothesized that the further apart the test locations, the more genetically diverse the trout populations were. It was also hypothesized that the brook trout and the brown trout populations are interbreeding to produce tiger trout. The purpose of this pilot project was to genetically test the trout from tail clippings, using microsatellite markers to determine if and where this hybridization had occurred. Microsatellite markers are increasingly being used to analyze population genetics. Microsatellites are a group of tandem repeated DNA sequences consisting of di-, tri-, and tetranucleotides. The more tandem repeats within a tested DNA sample, the longer the DNA segment will be. This project also determined genetic diversity using six microsatellite genetic markers for the brown trout and brook trout samples located within the Lake Champlain Basin. The lengths of these segments were measured in terms of base pairs. This genetic survey will serve as a baseline for further genetic monitoring of brook trout, brown trout and possibly tiger trout within the Lake Champlain Basin.

Materials and Methods
Both brook and brown trout were caught and released in True Brook in Saranac, New York and Great Brook in Plainfield, Vermont by electroshocking the water. The trout (N=26) were scooped up in nets and small tail clippings were taken from the caudal fin or adipose fin and the samples were put in sterile collection tubes. Each fish was measured for length; and weather conditions, GPS satellite positioning and water temperatures were...
Discussion

The improvement of microsatellite genetic markers in recent years has supplied a source of polymorphism needed for genetic identification in fish. The two main purposes of this pilot project was to first use six microsatellite DNA markers to assess the genetic diversity of brook trout and brown trout populations in the Lake Champlain Basin. Second, these six microsatellites were also used to determine whether hybridization between brook trout and brown trout was occurring and producing a proposed tiger trout.

Initial screening of all DNA samples using each of the six microsatellite genetic markers produced a variety of results, some more easy to interpret than others. The thermal cycler was programmed for 35 cycles of 94°C for 30 sec, 56°C for 30 sec, followed by 72°C for 45 sec. The samples were assayed for allelic diversity using six microsatellite markers which were known to be polymorphic; they were Sfo-C129, Sfo-C79, Sfo-C113, Sfo-292, Sfo-262 and MST-85 (Table 1). After PCR, 5 µl of 5x loading dye was mixed with 20µl of each DNA sample for each marker and loaded into a 1.25% agarose gel containing ethidium bromide. Following DNA gel electrophoresis, each gel was visualized and photographed using a gel documentation system (Figures 2-4). Once the DNA was analyzed through DNA gel electrophoresis, further analysis was done using an Agilent 2100 bioanalyzer to determine the exact size of each PCR product (data not shown).

Results

The trout samples examined demonstrated a wide range of values for both the number of alleles and size in base pairs of the PCR products (Table 2). All the microsatellite genetic markers except the MST-85 fell within the range of previously published results for these markers. The genetic marker Sfo-292 had the greatest number of alleles for the brown trout (15), brook trout (11) and tiger trout (5), as well as the largest size range in base pairs (197-327 bp) for all samples (Table 2, Figure 2). In contrast, the Sfo-C79 marker had the least number of alleles (2) for each trout tested and also had the smallest size range (100-108 bp) (Table 2, Figure 3). The Sfo-C113 marker exhibited a size range for the tested samples (129-155 bp) with brown trout having three alleles, brook trout six alleles and the unknowns three alleles (Table 2, Figure 4). Finally, the MST-85 marker was not a useful marker due to all the DNA fragments it produced (data not shown).
studies making this particular marker unreliable for this study.

Considering the results from this project, microsatellite markers Sfo-C113, Sfo-262 and Sfo-292 showed promise for identifying genetic diversity as well as identifying tiger trout. The Sfo-C113 marker showed the least number of alleles (Table 2) but did produce two DNA bands for one of the unknown trout samples thought to be a proposed tiger trout (Figure 4, lane 5). The bottom band in lane 5 (Figure 4) for the proposed tiger trout fell within the size range for brown trout, while the top band fell within the range for brook trout (Table 2). On the other hand, the Sfo-292 produced the greatest number of alleles (Table 2) but it also produced two distinct DNA bands for one of the proposed tiger trout samples (Figure 2, lane 8). Here again, one of the two DNA bands (from Figure 2, lane 8) matches both the brook and brown trout DNA bands. Finally, the Sfo-262 genetic marker showed that the brook trout alleles fell within the range of 320-356 bp and 392-412 bp; while the brown trout alleles were measured to be 356-375 bp. The proposed tiger trout had alleles which matched two alleles from brook trout (338 and 350 bp) and one allele from brown trout (375 bp). Further testing on an increase sample size and from more collection sites will need to be done to determine consistent reliability of these markers. As mentioned, the trout were only collected from two sites.

There were several areas of concern with this study. First, it was hoped that more trout samples could be collected, but due to the record snowfall during the winter and record rainfall during the spring, this produced swollen and fast moving streams which had an impact on the number of trout within the stream. Second, most of the brown trout were collected from the Plainfield, Vermont location, while most of the brook trout samples were collected at the True Brook stream near Saranac, NY. This made it difficult to compare any genetic differences between the brook trout between the Plainfield, VT and Saranac, NY locations as well as the brown trout populations.

To expand this project, more trout samples need to be collected and tested to truly evaluate the brook and brown trout populations. Also, more collection sites need to be included from the Lake Champlain Basin. Furthermore, the DNA samples collected for the suspected tiger trout should be sequenced to confirm the findings. These results will be shared with the Lake Champlain Research Institute and the U.S. Department of Fish and Wildlife. It could influence the management and stocking of the streams within the Lake Champlain Basin. Finally, the oldest and most established markers (Sfo8, Sfo12, Sfo18, and Sfo23) should be included, since they have been widely used and have generated familiarity within the 'trout community' to promote consistency.

Table 2. Summary of variations for microsatellite markers

<table>
<thead>
<tr>
<th></th>
<th>Brown Trout (N=12)</th>
<th>Brook Trout (N=11)</th>
<th>Unknowns (N=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sfo-C113</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allele range</td>
<td>131-137 bp</td>
<td>137-155 bp</td>
<td>129-149 bp</td>
</tr>
<tr>
<td>Number of alleles</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td><strong>Sfo-262</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allele range</td>
<td>356-375 bp</td>
<td>320-412 bp</td>
<td>323-375 bp</td>
</tr>
<tr>
<td>Number of alleles</td>
<td>5</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td><strong>Sfo-292</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allele range</td>
<td>205-327 bp</td>
<td>197-292 bp</td>
<td>197-293</td>
</tr>
<tr>
<td>Number of alleles</td>
<td>15</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td><strong>Sfo-C79</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allele range</td>
<td>100-104 bp</td>
<td>104-108 bp</td>
<td>100-108 bp</td>
</tr>
<tr>
<td>Number of alleles</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Sfo-C129</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allele range</td>
<td>239-247 bp</td>
<td>241-256 bp</td>
<td>249-259 bp</td>
</tr>
<tr>
<td>Number of alleles</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 5. The percent allele frequency for the Sfo-C113 genetic marker is shown for the brook (N=11), brown (N=12), and unknown trout (N=3) samples. The brook trout alleles (red bars) fell within the range of 137-155 bp; the brown trout (blue bars) alleles were measured to be 131, 135, and 137 bp. The unknown trout (green bars) had alleles which matched one allele from brook trout (149 bp) and two alleles from brown trout (131 and 137 bp).
References


Acknowledgements
Special thanks to wildlife biologist Madeleine Lyttle from the U.S. Department of Fish and Wildlife for assisting with the collection of trout in Vermont; and to Dr. Timothy Mihuc, Director of the Lake Champlain Research Institute for assisting with the collection of trout in New York. All permits and permission for the capture and release of the trout were granted and followed.
Your theory is crazy...
but not crazy enough to be true.