A Bioinformatics Approach to Systemic Lupus Erythematosus and Autoimmunity

Author
John Siefert, Conrad Weiser High School, Robesonia, PA

Mentor
Dr. Angelika Antoni, Kutztown University, Kutztown, PA

Funded by The American Association of Immunologists
High School Teachers Summer Research Program

Table of Contents

INTRODUCTION

Immunology is a topic rarely covered in high school biology courses. In states with required, course-specific graduation exams, it can be challenging to teach any topics outside the core standards. The Biology Keystone Exam is now a graduation requirement for all Pennsylvania high school students. The associated Pennsylvania Keystone Course Standards do not directly cover immunological topics and cover only content, not skills. Despite this absence, immunology is an important topic that deserves inclusion in high school biology curricula. Fortunately, immunology draws on many of the basic biological concepts covered on standardized tests like the Biology Keystone Exam and can be integrated into science curricula.

My research experience at Kutztown University with Dr. Angelika Antoni introduced me to new immunological content and laboratory skills. MRL mice, systemic lupus erythematosus (SLE) models and BALB/c mice, commonly used in immunology research were the focus of the experience. MRL and BALB/c macrophages were grown in culture to compare expression levels of RhoA and RhoH proteins in diseased versus normal mice. Results suggest a possible role of Rho proteins in the onset of autoimmune diseases like Type 1 diabetes and Systemic Lupus Erythematosus. My teaching experience and laboratory experience inspired me to design an immunology unit that will allow students to learn about SLE and autoimmunity through bioinformatics.

Bioinformatics is an emerging science that utilizes the power of computers and mathematics to make sense of biological data. Researchers from many fields of biological sciences utilize computer software and programs to find patterns and make sense of huge quantities of data. In a general sense, bioinformatics programs allow researchers to compare genetic information, from single nucleotides to entire genomes; from amino acids to proteomes. From these comparisons, insights in evolutionary relationships, disease susceptibility, gene expression, protein modifications, and much more can be made[1]. By learning some basic terminology and experiencing bioinformatics software, students can gain access to data previously inaccessible to them. Knowledge of bioinformatics can be useful in immunology applications. Teaching students the basic concepts of immunology and exposing them simultaneously to bioinformatics can give students an important blend of content and skill.

1. Hogeweg, P.. (2011). The roots of bioinformatics in theoretical biology. *PLoS Computational Biology 7*(3). doi: 10.1371/journal.pcbi.1002021

TEACHER GUIDE

SCIENCE BACKGROUND

Autoimmunity
The immune system is incredibly complex and effectively protects against a wide range of pathogens. Unfortunately, the complexity of the immune system sometimes results in undesirable responses. Response to environmental antigens can result in harmless allergic reactions and serious hypersensitivity. Another, often more serious, inappropriate response occurs when antigens expressed on the body's own cells are targeted by the immune system. An immune response to self-antigen is called autoimmunity and can result in serious diseases[1].

Though autoimmune disorders have wide-ranging symptoms, the general mechanism of autoimmunity is understood. As lymphocytes develop, some have affinity for self-antigen rather than foreign antigen. These cells are usually destroyed by varying mechanisms, creating "self-tolerance". A body's ability to determine self from non-self is critical. Autoimmune disorders occur when self-tolerance breaks down and self-reactive lymphocytes are not destroyed. Determining the precise genetic and environmental factors that cause a particular autoimmune disorder is much more challenging[1].

Autoimmune disorders are common and student awareness of these diseases is increasing. Rheumatoid arthritis, Crohn's disease, Multiple sclerosis and Type 1 diabetes are now household names. In each disease, self-reactive lymphocytes attack and destroy specific cell types in the body. A lesser known, though equally prevalent autoimmune disorder known as Systemic Lupus Erythematosus has been gaining awareness in recent years.

Systemic Lupus Erythematosus (SLE)
SLE is an autoimmune disorder that is still not well understood. Like many autoimmune disorders, the immune system produces antibodies against cells in the body. In SLE, the antagonized cells are widespread in the body. Self-reactive lymphocytes attack DNA and associated proteins involved in protein synthesis[1]. Inflammation and tissue damage, occurring throughout the body, can result in a wide variety of symptoms including rash, photosensitivity, ulcers, arthritis, seizures, blood abnormalities, and other manifestations in joints, skin, lungs, and kidneys[2]. The disease is especially vexing because periods of illness are divided by periods of remission. Because of the widespread symptoms, SLE is chronically misdiagnosed. In fact, a patient with SLE spends an average of four years and sees three physicians before being correctly diagnosed[3]. Due to the complexity of the disease, SLE is thought to be caused by a combination of genetics, environment, and hormones.

Genetics play a role in determining onset of the disease, however no single gene has been definitively linked with the disease. Several genome-wide association studies (GWAS) and smaller, cell-specific studies have pointed to genes relating to B cells, complement activation, macrophages, and interferon regulation factors (IRFs) as possible genetic causes of the disease[3,4,5]. As evidence of the indeterminate role of genetics in SLE, the likelihood of two identical twins having SLE is only 25%[2].

Women are much more likely to be diagnosed with SLE. Female patients outnumber male patients nearly 12:1. African Americans, Hispanics, Asian Americans, and Native Americans also have increased risk of SLE compared to Caucasians [2,3]. SLE also seems to occur concurrently with other autoimmune diseases, such as hemolytic anemia and thyroiditis. Because it is so difficult to diagnose, estimates of SLE prevalence vary widely. Studies estimate that between one and six Americans out of every 100,000 have this form of lupus[6].

Bioinformatics

Bioinformatics is the use of computational methods for comparing and analyzing genetic data[7]. The study originated in the 1970s but gained popularity in the 1980s and 1990s with the advent of affordable and powerful computers. The history of bioinformatics follows in the footsteps of molecular biology. One of the fundamental questions of molecular biology is how living organisms store, use, and express genetic information[8]. As molecular biology entered the "sequencing age" and massive quantities of data began pouring out of laboratories, tools to make sense of that data were necessary. Molecular biology and bioinformatics came together as genetics became a data-driven science[7].

Today, bioinformatics is being used in nearly all fields of biological science. From evolutionary biology to medical research, finding patterns and identifying anomalies in genetic data is key. Much of the genomic data being collected is published in databases like GenBank. Computer programs, often available to the public for free, allow users to quickly search and compare sequences of DNA, RNA and proteins. Software like BLAST and ClustalW have become standard tools used to align and compare sequences.

While the list of applications for bioinformatics is long, a few techniques are used most frequently. Sequence alignment is used to find similarities and differences between related species. This technique can be used to elucidate the genetic component of a disease or determine relatedness of two species in constructing phylogenetic trees. Sequence alignment can also be used to identify model organisms for basic research on genes or finding candidates for drug testing. Another application of bioinformatics is gene annotation, in which researchers identify the coding and regulatory regions of DNA. More complex genome analysis, such as maps of methylation sites, transcription factors and other genetic modifications can be done. Genome-wide Association Studies have become fundamental in the study of diseases, including cancer. Comparing genomes of thousands of patients has allowed researchers to pinpoint the mutations associated with diseases like Alzheimer's disease, breast cancer, and more[9].

1. Murphy, K. (2012). *Janeway's Immunobiology* (8th ed.) New York: Garland Science.
2. Systemic Lupus Erythematosus. (2014). Centers for Disease Control and Prevention. http://www.cdc.gpv/arthritis/basics/lupus.htm
3. Manzi, S. (2009). Lupus update: perspective and clinical pearls. *Cleveland Clinic Journal of Medicine 76*(2).
4. Kozyrev, S., Abelson, A., Wojcik, J., et. al. (2008). Functional variants in the B cell gene BANK1 are associated with systemic lupus erythematosus. *Nature Genetics 40*.
5. Antoni, A., Graham, L., Rauch, J., and Levine, J. (2009). Altered cell-cell and cell-matrix interactions in the development of systemic autoimmunity. *Autoimmunity 42*(4), 278-281.
6. Uramoto, K., Michet, C., Thumboo, J., et. al. (1999). Trends in the incidence and mortality of systemic lupus erythematosus, 1950-1992*). Arthritis and Rheumatology 42*(1).
7. Hogeweg, P. (2011). The roots of bioinformatics in theoretical biology. *PLoS Computational Biology 7*(3). doi: 10.1371/journal.pcbi.1002021
8. Nurse, P. (2008). Life, logic and information. *Nature 454*, 424-426.
9. Londin, E., Yadav, P., Surrey, S., et. al. (2013). Use of linkage analysis, genome-wide association studies, and next-generation sequencing in the identification of disease-causing mutations. *Pharmacogenomics 1015*, 127-146.

STUDENT OUTCOMES

Students will learn the basic terminology and techniques needed to use bioinformatics software including OMIM, BLAST, and ClustalW. As students learn about bioinformatics, they will simultaneously learn about the autoimmune disorder Systemic Lupus Erythematosus (SLE). Students will navigate sequence databases, searching for genes recently linked with SLE in research articles. Additionally, students will read a peer-reviewed article on SLE individually and then analyze the study as a group in a "Journal Club". This activity will expose students to primary research to demonstrate that it is possible to learn and apply content taught in the classroom with current research. Finally, in order to make connections between laboratory work and computer-based analysis, students will learn basic laboratory techniques and tools through the use of independently-driven "Instrument Training" assignments. Instrument Trainings will require that students learn background and safety information, experience equipment and procedures in the laboratory, and complete calculations for mixing and weighing, safely using glassware, diluting solutions, and using micropipettors.

LEARNING OBJECTIVES

After completing this unit, students will be able to:

- Explore OMIM (Online Mendelian Inheritance in Man) to learn more about autoimmune diseases and the associated genetic links.
- Use BLAST (Basic Local Alignment Search Tool) to compare gene sequences using GI and FASTA formats.
- Navigate BioServers and ClustalW to conduct multiple sequence alignment.
- Read and critique peer-reviewed research articles.
- Measure and mix chemicals to prepare solutions of various concentration.
- Identify and use appropriate glassware for specific laboratory purposes
- Dilute stock solutions to working concentration through standard and serial dilution.
- Appropriately use P20 and P1000 micropipettors to transfer small volumes of liquid in the laboratory.

STANDARDS

This unit aligns to the following Next Generation Science Standards:

- PS1A: Structure and Properties of Matter
- PS1B: Chemical Reactions
- PS4C: Information Technologies and Instrumentation
- LS1A: Structure and Function
- LS1B: Growth and Development of Organisms
- LS1D: Information Processing
- LS3B: Variation of Traits
- LS4D: Biodiversity and Humans

TIME REQUIREMENTS

Unit Implementation

This unit can be taught as a sequential, stand-alone unit. Because of the time restrictions imposed by national and state standards and the need to prepare students for graduation exams, it may not be practical to devote a week or more to bioinformatics and instrument training lessons. Instead, these lessons can be inserted individually into a pre-designed curriculum. More specifically, these assignments can be assigned once per unit or once a semester, as time permits. Additionally, if class time is spent introducing the necessary background information, the assignments can be assigned for homework so valuable classroom time can be devoted to lecture, group work, and laboratory activities.

One caveat to assigning the bioinformatics and instrument training assignments individually, rather than as a unit, is that students will need some prior knowledge. Prior knowledge for these lessons may include organization of the immune system, response to an infection, and autoimmunity for the bioinformatics assignments and laboratory safety for the instrument trainings. For additional information about prior knowledge, see page 10.

Scope and Sequence

Task	**Topic**	**Recommended Pacing: In-Class (82-minute block)**				**Recommended Pacing: Outside of Class (weeks)**			
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Bioinformatics #1	Exploring SLE in OMIM								
Bioinformatics #2	BLAST Terminology								
Bioinformatics #3	Finding Genes with GI Identifiers								
Bioinformatics #4	Comparing Genes in BioServers								
Journal Club	Analyzing Peer-reviewed Research								
Instrument Training #1	Glassware								
Instrument Training #2	Mixing and Weighing								
Instrument Training #3	Diluting Stock Solutions								
Instrument Training #4	Micropipetting								

ADVANCE PREPARATION

Bioinformatics
Advance preparation for bioinformatics assignments includes printing copies of bioinformatics assignments for students (Appendix II) and providing students with a general background of bioinformatics. The assignments are meant to be exploratory and should not be taught step-by-step. Teachers may want to give students a "tour" of each website when the activity is assigned. This may require teachers to familiarize themselves with each bioinformatics software prior to assigning.

Journal Club
Teachers should print copies of the article at least one week prior to the Journal Club discussion. Additionally, it may be worthwhile to provide hints to the students regarding note-taking, critical reading, and summarizing of articles.

Instrument Training
Teachers should print copies of each Instrument Training. Additionally, all materials needed for laboratory protocols should be available for students (Materials and Equipment). Materials and equipment should be available for the entire duration of the assignment and easily accessible to students. Additional instructions for each Instrument Training includes:

Instrument Training #1 – Mixing and Weighing
- Teach proper procedure for mixing aqueous solutions with a magnetic stir plate.
- Teach proper procedure for measuring mass with a digital scale.
- Teach conversion of units (mass and volume).

Instrument Training #2 – Glassware
- Teach proper technique for cleaning up broken glass in your classroom/laboratory.
- Teach proper technique for lighting a Bunsen burner.
- Teach proper technique for using a magnetic stir plate.

Instrument Training #3 – Diluting Stock Solutions
- Teach concept of dilutions using $C_1V_1=C_2V_2$
- Teach proper technique for conducting 1/10 serial dilution.
- Discuss difference between stock solution and working solution with students.

Instrument Training #4 – Micropipetting
- Teach proper technique for using a micropipette.
- Teach concept of unit conversions in metric system.
- Teach proper format for labeling a photograph with title and caption.

Advance Preparation Timeline
None of the advance preparation is overly time consuming. Most of the advance preparation is teaching students background concepts and skills or familiarizing oneself with software and procedures. All solutions used during instrument trainings are safe and can be purchased at the indicated concentrations. Some solutions are made by each student, reducing teacher preparation.

Materials for each instrument training should be set out prior to student use. Once assigned, students will work independently, at their own pace.

MATERIALS AND EQUIPMENT

Bioinformatics

Equipment	Website	Substitute
Computers		Tablets, smart phones
Bioinformatics Software OMIM Website	http://omim.org/	Alternative SLE websites
Bioinformatics Software NCBI BLAST	http://blast.ncbi.nlm.nih.gov/Blast.cgi	PatternHunter, KLAST, BLAT, EMBOSS Needle
Bioinformatics Software BioServers (ClustalW)	http://www.bioservers.org/bioserver/	MUSCLE, T-COFFEE, DNA Subway

Journal Club

Equipment	Website	Substitute
RNA-Seq for Enrichment and Analysis of *IRF5* Transcript Expression in SLE	http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0054487	Any peer-reviewed SLE article in Open Access Journals: Public Library of Science, Emerging Infectious Disease, etc.

Instrument Training

Equipment	Cost – Carolina Biological Supply	Substitute	Activity
Pyrex Griffin 250mL Beaker	$3.90 each	Any other glassware	IT2
Pyrex 250 mL Erlenmeyer Flask	$4.75 each	Any other glassware	IT2
Wheaton Sampling (Media) 500mL Bottle	$11.35 each	Any other glassware	IT2
Corning 100x10mm Petri Dish	$60.70 per pack of 12	Any other glassware	IT2
Pyrex 10mL Graduated Cylinder	$6.40 each	Any other glassware	IT1,2
Pyrex Florence Flask 250mL	$9.55 each	Any other glassware	IT2
Carolina 0.1g Compact Balance	$99.00 each	Triple beam balance	IT1,2
Magnetic Stir Plate (Hanna Mini)	$107.00 each	Stir by hand	IT1,2,3
Magnetic Stir Bar (Spin Bar, 7/8")	$5.75 each	Stir by hand	IT1,2,3
Magnetic Stir Bar Retriever	$13.15 each	Stir by hand	IT1,2,3
Pyrex Vista 10x75mm Test Tube	$16.45 per pack of 50	Paper cups	IT1,2,3
Food Coloring Set of 4	$5.25 per pack of 4	Any dye	IT3,4
Carolina Micropipettes (P20, P200, P1000)	$149.95 each	Capillary pipettes, disposable pipettes	IT4
Carolina Micropipet Tips	$13.50 per rack of 100		IT4
4" x 125' Parafilm	$25.50 each	Wax paper	IT4
Microcentrifuge Tubes	$42.40 per pack of 1,000	Any small container	IT4
Ring Stand with Rings	$27.45 each		IT2
Heat Diffuser (wire gauze)	$2.55 each		IT2
Bunsen Burner	$25.90 each	Microwave, hot plate	IT2
Jello 6 oz (Local)	$1.18 each	Agar Agar, nutrient agar	IT1
3 oz. Dixie Cups (Local)	$15.00 per pack of 600		IT1
Disposable 1mL Pipettes	$1.75 per pack of 16	Small grad. cylinder	IT3
3% Hydrogen Peroxide	$8.95 per Liter	Any 3% solution	IT3
NaCl (Salt)	$6.65 per 500 grams	Any water soluble solid	IT3
Carolina Microcentrifuge	$349.00 each	Shake by hand, vortex mixer	IT4
95% Ethanol, 3.8 Liter	$26.25	Any 95% solution	IT3

**Because students work independently, a set of eight items will provide enough for a classroom of 24 students. All items above can be purchased from Carolina Biological Supply Company. Supplies can also be found at other scientific supply companies and many items can be purchased locally at a teacher's preferred stores.*

STUDENT PRIOR KNOWLEDGE

Prior Content

While the activities in this unit can be assigned without any prior content knowledge on the part of the students, the activities will be far more meaningful with some prior knowledge. Prior to assigning the bioinformatics assignments and the journal club assignment, students will benefit from some basic concepts of immune system organization, autoimmunity in general and systemic lupus erythematosus (SLE) in particular, and bioinformatics. For background information on autoimmunity, SLE, and bioinformatics, see pages 4 and 5.

Most high school biology textbooks will have a chapter of basic concepts in immunology. It would be worthwhile to teach several short lessons or assign readings and homework from this section of the text. If student textbooks do not have a chapter on immunology, there are many excellent online resources that can be accessed for free. Other textbooks can be purchased used. Janeway's Immunobiology by Kenneth Murphy (Garland Science) is a very thorough text on the subject. The following topics will be most useful to the student to complete the assignments in this unit:

1. Basic Concepts in Immunology
    a. Definition
    b. History

2. Principles of the Innate Immune System
    a. Immediate response to an infection
    b. Cells/structures of the innate response
        i. Neutrophils, eosinophils, basophils, monocytes, dendritic cells, macrophages
        ii. Cytokines
        iii. Inflammation
        iv. Complement Activation

3. Principles of the Adaptive Immune System
    a. Learned/adaptive response to an infection
    b. Cells/structures of the adaptive response
        i. B cells, T cells, dendritic cells, macrophages
        ii. Antibody/Antigen recognition
        iii. Major Histocompatibility Complex (MHC)

4. Response to an Infection
    a. Sequence of events in a normal immune response to an infection

5. Autoimmunity
    a. Self-tolerance
    b. Failure in self-tolerance
    c. Types of Autoimmune Disorders

Prior Technical Skill

The purpose of this curriculum is to introduce students to basic bioinformatics and laboratory techniques. Therefore, students need no prior technical skills, except for basic computer use and knowledge of laboratory safety.

STUDENT SECTION

RATIONALE

Significance and Introduction

The immune system is one of the most important biological processes for human beings and many other animals. The immune system keeps pathogens under control, every day, to ensure that an organism is able to do the important tasks to survive. When the immune system breaks down, the unpleasant symptoms of illness are often felt. At times, a breakdown in immunity can be fatal. It can be easy to ignore the immune system when everything is operating smoothly, but it is important to understand what happens when the immune response breaks down.

Every day, the media covers outbreaks of disease in some part of the world. From viral infections like HIV and Ebola to bacterial infections like *Salmonella* and parasites like Malaria, pathogens and the diseases they cause are household words. More recently, the general public has become increasingly aware of another type of disease; autoimmune disorders. Diseases like type I diabetes, rheumatoid arthritis, Crohn's disease, multiple sclerosis, and lupus are diagnosed in millions of Americans. Lifestyle changes, pharmaceutical drugs, and surgeries are needed to treat these diseases.

Autoimmune diseases are not as straightforward as infectious diseases. In diseases caused by an infection, symptoms are often easily traced back to the pathogenic organism. In autoimmune diseases, symptoms are caused by an over-reaction by the immune system. That is, the cells of the immune system inappropriately attack cells in the body. Understanding the causes of these diseases necessitates a complex understanding of the immune system, genetics, and the environment. Immunologists who study autoimmune diseases split their time in the laboratory, growing cells and working with organisms, on the computer examining DNA and proteins in bioinformatics programs, and in clinical hospitals, working with patients to alleviate symptoms.

SCIENCE BACKGROUND

Autoimmunity

The immune system is incredibly complex and effectively protects against a wide range of pathogens. Unfortunately, the complexity of the immune system sometimes results in undesirable responses. Response to environmental antigens can result in harmless allergic reactions and serious hypersensitivity. Another, often more serious, inappropriate response occurs when antigens expressed on the body's own cells are targeted by the immune system. An immune response to self-antigen is called autoimmunity and can result in serious diseases[1].

Though autoimmune disorders have wide-ranging symptoms, the general mechanism of autoimmunity is understood. As lymphocytes develop, some have affinity for self-antigen rather than foreign antigen. These cells are usually destroyed by varying mechanisms, creating "self-tolerance". A body's ability to determine self from non-self is critical. Autoimmune disorders occur when self-tolerance breaks down and self-reactive lymphocytes are not destroyed. Determining the precise genetic and environmental factors that cause a particular autoimmune disorder is much more challenging[1].

Autoimmune disorders are common and student awareness of these diseases is increasing. Rheumatoid arthritis, Crohn's disease, Multiple sclerosis and Type 1 diabetes are now household names. In each disease, self-reactive lymphocytes attack and destroy specific cell types in the body. A lesser known, though equally prevalent autoimmune disorder known as Systemic Lupus Erythematosus has been gaining awareness in recent years.

Systemic Lupus Erythematosus (SLE)

SLE is an autoimmune disorder that is still not well understood. Like many autoimmune disorders, the immune system produces antibodies against cells in the body. In SLE, the antagonized cells are widespread in the body. Self-reactive lymphocytes attack DNA and associated proteins involved in protein synthesis[1]. Inflammation and tissue damage, occurring throughout the body, can result in a wide variety of symptoms including rash, photosensitivity, ulcers, arthritis, seizures, blood abnormalities, and other manifestations in joints, skin, lungs, and kidneys[2]. The disease is especially vexing because periods of illness are divided by periods of remission. Because of the widespread symptoms, SLE is chronically misdiagnosed. In fact, a patient with SLE spends an average of four years and sees three physicians before being correctly diagnosed[3]. Due to the complexity of the disease, SLE is thought to be caused by a combination of genetics, environment, and hormones.

Genetics play a role in determining onset of the disease, however no single gene has been definitively linked with the disease. Several genome-wide association studies (GWAS) and smaller, cell-specific studies have pointed to genes relating to B cells, complement activation, macrophages, and interferon regulation factors (IRFs) as possible genetic causes of the disease[3,4,5]. As evidence of the indeterminate role of genetics in SLE, the likelihood of two identical twins having SLE is only 25%[2].

Women are much more likely to be diagnosed with SLE. Female patients outnumber male patients nearly 12:1. African Americans, Hispanics, Asian Americans, and Native Americans also have increased risk of SLE compared to Caucasians [2,3]. SLE also seems to occur concurrently with other autoimmune diseases, such as hemolytic anemia and thyroiditis. Because it is so difficult to diagnose, estimates of SLE prevalence vary widely. Studies estimate that between one and six Americans out of every 100,000 have this form of lupus[6].

Bioinformatics

Bioinformatics is the use of computational methods for comparing and analyzing genetic data[7]. The study originated in the 1970s but gained popularity in the 1980s and 1990s with the advent of affordable and powerful computers. The history of bioinformatics follows in the footsteps of molecular biology. One of the fundamental questions of molecular biology is how living organisms store, use, and express genetic information[8]. As molecular biology entered the "sequencing age" and massive quantities of data began pouring out of laboratories, tools to make sense of that data were necessary. Molecular biology and bioinformatics came together as genetics became a data-driven science[7].

Today, bioinformatics is being used in nearly all fields of biological science. From evolutionary biology to medical research, finding patterns and identifying anomalies in genetic data is key. Much of the genomic data being collected is published in databases like GenBank. Computer programs, often available to the public for free, allow users to quickly search and compare sequences of DNA, RNA and proteins. Software like BLAST and ClustalW have become standard tools used to align and compare sequences.

While the list of applications for bioinformatics is long, a few techniques are used most frequently. Sequence alignment is used to find similarities and differences between related species. This technique can be used to elucidate the genetic component of a disease or determine relatedness of two species in constructing phylogenetic trees. Sequence alignment can also be used to identify model organisms for basic research on genes or finding candidates for drug testing. Another application of bioinformatics is gene annotation, in which researchers identify the coding and regulatory regions of DNA. More complex genome analysis, such as maps of methylation sites, transcription factors and other genetic modifications can be done. Genome-wide Association Studies have become fundamental in the study of diseases, including cancer. Comparing genomes of thousands of patients has allowed researchers to pinpoint the mutations associated with diseases like Alzheimer's disease, breast cancer, and more[9].

1. Murphy, K. (2012). *Janeway's Immunobiology* (8th ed.) New York: Garland Science.
2. Systemic Lupus Erythematosus. (2014). Centers for Disease Control and Prevention. http://www.cdc.gpv/arthritis/basics/lupus.htm
3. Manzi, S. (2009). Lupus update: perspective and clinical pearls. *Cleveland Clinic Journal of Medicine 76*(2).
4. Kozyrev, S., Abelson, A., Wojcik, J., et. al. (2008). Functional variants in the B cell gene BANK1 are associated with systemic lupus erythematosus. *Nature Genetics 40*.
5. Antoni, A., Graham, L., Rauch, J., and Levine, J. (2009). Altered cell-cell and cell-matrix interactions in the development of systemic autoimmunity. *Autoimmunity 42*(4), 278-281.
6. Uramoto, K., Michet, C., Thumboo, J., et. al. (1999). Trends in the incidence and mortality of systemic lupus erythematosus, 1950-1992*). Arthritis and Rheumatology 42*(1).
7. Hogeweg, P. (2011). The roots of bioinformatics in theoretical biology. *PLoS Computational Biology 7*(3). doi: 10.1371/journal.pcbi.1002021
8. Nurse, P. (2008). Life, logic and information. *Nature 454*, 424-426.
9. Londin, E., Yadav, P., Surrey, S., et. al. (2013). Use of linkage analysis, genome-wide association studies, and next-generation sequencing in the identification of disease-causing mutations. *Pharmacogenomics 1015*, 127-146.

APPENDIX I – DAILY PLANS

TEACHER DOCUMENTS

Bioinformatics Teacher

Assignment #1: Investigating the Genetic Component of Systemic Lupus Erythematosus through Online Mendelian Inheritance in Man (OMIM)

Background Information:

Online Mendelian Inheritance in Man (OMIM) is one of the most popular medical genetics websites. The website provides access to a database of all known diseases with a genetic component of some kind. The database is continuously updated with new information about the molecular relationship between genes (genotype) and disease (phenotype). The OMIM database links known information about a disease with genes in the human genome. All peer-reviewed research that has shown a link between genes and disease are cited for additional information. OMIM is operated by the National Center for Biotechnology Information (NCBI), the same organization that operates BLAST.

For a particular disease of interest, information about relevant genes, chromosomes, SNPs and other polymorphisms, and protein sequence can be easily accessed. Every disease with an entry in OMIM is given a unique identification called a MIM code. An MIM code is a six-digit number that describes the genetic disease. The first digit in the code reveals information about the genetic disease:

1: Autosomal Dominant	(before 1994)	4: Y-linked
2: Autosomal Recessive	(before 1994)	5: Mitochondrial
3: X-linked		6: Autosomal (after 1994)

Symbols before the MIM code can also reveal information about the disease:

***:** entry is a gene	**%:** loci is unknown, but disease is Mendelian
#: disease is controlled by multiple loci	**No Entry:** molecular basis is unknown
+: gene sequence and phenotype are known	**^:** entry no longer exists

One additional code is used to explain the location of a genetic component within the human genome. Each locus (specific location of a chromosome) is given a code consisting of letters and numbers. A chromosome has a short arm (abbreviated "p") and a long arm (abbreviated "q") separated by a centromere. When stained, a chromosome appears to have bands of light and dark. These bands can be counted and are labeled p1, p2, p3 or q1, q2, q3, etc. The counting begins from the centromere and moves outward. If the chromosome is magnified greatly, additional sub-bands and sub-sub-bands can be counted. For example, a locus with the code 7q31.2 is located on the long arm (q) of the 7^{th} chromosome in band 3, sub-band 1, and sub-sub-band 2.

Instructions:

1. Go to http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
2. In the search bar, type in the MIM code 612251.When the search is complete, click on the link to reveal more information about the disease Systemic Lupus Erythematosus.
3. Use the information of the page to answer questions 1-3 below.
4. On the right side of the screen, click the arrow to expand the "Clinical Resources" menu. Under "Genetics Home Reference", click the link "systemic lupus erythematosus."
5. Use the information on the Genetics Home Reference page to answer question 4 below.

Questions:

1. What is the name of the name of the gene/locus associated with SLE? *IRF5*
2. What is the chromosomal location of the locus? Include chromosome and band numbers.
   Chromosome 7, long arm, band 3, sub-band 2, sub-sub-band 1
3. Is this the only locus associated with susceptibility to SLE? Explain your answer.
   No, the MIM code is prefaced with a #, which suggests that multiple loci control the disease.
4. Write a one page essay on SLE. Use the attached rubric and be sure to cover the following topics:

a. Signs and symptoms
b. Body parts affected
c. Prevalence in countries and races/genders
d. Genetic components
e. Inheritance

Bioinformatics: Assignment 1 Rubric
Investigating the Genetic Component of Systemic Lupus Erythematosus through Online Mendelian Inheritance in Man (OMIM)

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
Heading – includes name, course, and date	**2**	**1**	**0**
Font – paper is typed with 1-inch margins, size 11-12 font in Times New Roman or Arial	**3**	**2**	**0**
Format – content is roughly 1 page in length, single spaced	**5**	**3**	**0**
Spelling/Grammar – paper shows editing and careful review. Spelling and grammatical errors are absent.	**5**	**3**	**0**
Content – Signs and symptoms of SLE discussed in detail.	**5**	**3**	**0**
Content – Body parts affected by SLE discussed in detail.	**5**	**3**	**0**
Content – Prevalence of SLE in various countries and races/genders explained.	**5**	**3**	**0**
Content – Genetic components of SLE discussed in detail.	**5**	**3**	**0**
Content – Inheritance patterns of SLE discussed in detail.	**5**	**3**	**0**
Column Total			
Total Score	**/40**		

Comments:

Bioinformatics Teacher
Assignment #2: BLAST Terminology

Background Information:
Bioinformatics is the use of computational methods for comparing and analyzing genetic data. The study originated in the 1970s but gained popularity in the 1980s and 1990s with the advent of affordable and powerful computers. The history of bioinformatics follows in the footsteps of molecular biology. One of the fundamental questions of molecular biology is how living organisms store, use, and express genetic information. As molecular biology entered the "sequencing age" and massive quantities of data began pouring out of laboratories, tools to make sense of that data were necessary. Molecular biology and bioinformatics came together as genetics became a data-driven science.

Today, bioinformatics is being used in nearly all fields of biological science. From evolutionary biology to medical research, finding patterns and identifying anomalies in genetic data is key. Much of the genomic data being collected is published in databases like GenBank. Computer programs, often available to the public for free, allow users to quickly search and compare sequences of DNA, RNA and proteins. Software like BLAST and ClustalW have become standard tools used to align and compare sequences.

While the list of applications for bioinformatics is long, a few techniques are used most frequently. Sequence alignment is used to find similarities and differences between related species. This technique can be used to elucidate the genetic component of a disease or determine relatedness of two species in constructing phylogenetic trees. Sequence alignment can also be used to identify model organisms for basic research on genes or finding candidates for drug testing. Another application of bioinformatics is gene annotation, in which researchers identify the coding and regulatory regions of DNA. More complex genome analysis, such as maps of methylation sites, transcription factors and other genetic modifications can be done. Genome-wide Association Studies have become fundamental in the study of diseases, including cancer. Comparing genomes of thousands of patients has allowed researchers to pinpoint the mutations associated with diseases like Alzheimer's disease, breast cancer, and more.

Instructions:
Use the websites below to explore the most popular bioinformatics database: NCBI BLAST

1. BLAST Help Glossary - http://www.ncbi.nlm.nih.gov/books/NBK62051/
2. NCBI Handbook Ch. 16 - http://www.ncbi.nlm.nih.gov/books/NBK21097/
3. NCBI Handbook Glossary - http://www.ncbi.nlm.nih.gov/books/NBK21106/
4. NCBI Statistics Tutorial - http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html

Questions:
Answer all questions in your own words.

1. What does BLAST stand for?
   Basic Local Alignment Search Tool
2. What is the process of alignment as it relates to bioinformatics?
   Making an alignment is lining up two DNA or amino acid sequences in order to measure the degree of similarity between the two sequences.
3. How is a "local alignment" different than a global alignment?
   A global alignment aligns the two sequences over the entire length. That is, from beginning to end. A local alignment aligns the two sequences of interest, but only over a short area of interest. The area of interest may be a gene, for example.
4. Define the following terms as they relate to BLAST:
   Similarity – *the degree of relatedness between two DNA or amino acid sequences.*
   Identity – *the degree to which two DNA or amino acid sequences have the same nucleotides or amino acids at a given position in an alignment.*
   Homology – *similarity due to descent from a common ancestor. Genes can be homologous.*

Paralog – *homologous genes within a species that exist due to gene duplication.*
Ortholog – *homologous genes from different species that exist due to descent from a common ancestor. The genes may share a function or not.*

5. What is the significance of an E value?
   Every search result in BLAST is given an E value. The E value gives an estimate of the statistical significance of a result. The lower the E value, the more significant the result. An E value of 0 means that there is a 0 in 100 chance that the result occurred due to chance alone. An E value of .05 means that there is a 5/100 chance that the result occurred due to chance.
6. Explain FASTA format
   FASTA is a description of a DNA sequence that begins with an identification number and concludes with the actual sequence.
7. Explain GI format
   GI is a "GenInfo Identifier". A GI number is an identification number for a DNA sequence. Any time a sequence is updated or changed, a new GI is created.
8. What is an Accession number?
   An accession number is a unique label given to a sequence when it is submitted to GenBank.

Bioinformatics Teacher
Assignment #3: Finding Genes with GI Identifiers

Background Information:
Searching bioinformatics databases to find information about specific genes is a fundamental skill for genetics researchers. Unfortunately, many of the popular databases and search programs are not intuitive or user-friendly. BLAST is a computer algorithm that allows researchers to compare sequences of DNA or proteins. Several types of searches can be made, depending on the original search or "query". If a nucleotide sequence is input, the search results can be filtered to show similar nucleotide sequences or related proteins in other organisms.

To input a query in BLAST, several formats can be used. Frequently used formats include GenInfo identifiers (GI), FASTA identifiers, or accession numbers. Each nucleotide sequence in the database will have all three types of identifiers. In this activity, nucleotide sequences will be input into BLAST using a GI identifier to search for a gene linked to Systemic Lupus Erythematosus (SLE).

Instructions:
Part A

1. Go to the official BLAST home site at: http://www.ncbi.nlm.nih.gov/BLAST/
2. Select "nucleotide blast"
3. Paste the GI number gi|629266059 into the search box and hit BLAST at the bottom of the page.
4. When the search results have loaded, scroll down to the section of the page showing "sequences producing significant alignments". Look at the search information for the first result. Answer questions 1 and 2 below, then click on the Accession number in the right column to answer the remaining questions.

Questions:
Part A

1. What is the E-value for the first result? What is the significance of this value?
   The E-value is 0.0. This suggests that the gene did not occur as a result due to chance. Rather, it appears because it is related to the query in some way.
2. What is the Identity % for the first result? What does this suggest about the sequence?
   The identity % is 100%. This means that 100% of the nucleotides are shared with the query.
3. Provide the following information for the sequence represented by the first result:
   a. Organism: *Homo sapiens*
   b. Gene Name: *Interferon Regulatory Factor 5*
   c. Length of gene (in base pairs): *2955 bp*
   d. Type of molecule (DNA, mRNA, amino acid, etc): *mRNA*
   e. Provide a brief summary of gene family to which this gene belongs: *Interferon regulatory factor proteins are transcription factors that control expression of proteins relating to cell growth, differentiation, apoptosis, and immune cells.*
   f. Based on your results, does it seem plausible that this gene could be linked to SLE? *Yes, it does seem plausible. IRF genes are known to be involved in the immune response*
   g. Write the first 20 nucleotides of the gene sequence: *gcagaaagcg gaactgagcc*

Instructions:
Part B

1. Go to the article "RNA-Seq for Enrichment and Analysis of *IRF5* Transcript Expression in SLE" at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0054487
2. Read the Abstract and Introduction before answering the questions. You may need to use additional sources to help answer the questions.

Questions:

Part B

1. Define the following terms:
    a. Polymorphism: *variation in a DNA sequence, either by single nucleotides (single nucleotide polymorphism) or multiple nucleotides (restriction fragment length polymorphism).*
    b. Haplotype: *a set of polymorphisms in a DNA sequence that are usually inherited together.*
    c. Alternative Splicing: *a process during protein synthesis that allows a single gene to code for multiple proteins.*
2. Write a short summary of the article, explaining how polymorphisms in *IRF5* may increase a person's risk of SLE.

    The study shows that polymorphisms in the IRF5 gene can increase a person's risk for developing SLE. Several haplotypes have been identified in humans for the IRF5 gene. IRF5 experiences alternative splicing in humans, which makes it harder to find a link between DNA sequence and disease. Instead, it is important to determine which mRNA transcripts are present in healthy vs SLE patients. The study revealed that SLE patients do, in fact, express different IRF5 mRNA than healthy individuals. People with the H2 haplotype are at increased risk.

[20]

Bioinformatics Teacher
Assignment #4: Comparing DNA Sequences in BioServers with ClustalW

Background Information:

Humans share roughly 99.9% of their DNA with every other living human being. The 0.1% that is different accounts for the differences in appearances between each person. Sometimes, these differences can be responsible for causing disease. Interestingly, when the same type of analysis is used to compare human and our nearest primate ancestor, the chimpanzee, the difference increases to 1.2%. In other words, 98.8% of human DNA is identical to chimpanzee DNA. Clearly, small changes in DNA can result in big differences in an organism.

The tiny differences in DNA sequence that result in the variation seen among humans are called polymorphisms. Literally, the word translated to "many forms". More specifically, polymorphisms are nucleotides that vary from human to human. These nucleotides can occur in short stretches or as individual nucleotides. Short stretches of DNA that are variable are known as Restriction Fragment Length Polymorphisms (RFLPs) while individual variable nucleotides are known as Single Nucleotide Polymorphisms (SNPs). If two gene sequences are lined up and compared, it is possible to identify and count the differences between the two sequences.

This technique of aligning and comparing gene sequences can be used to link specific polymorphisms with disease. If the sequences of diseased and healthy patients are compared, polymorphisms unique to diseased individuals are said to be linked to disease. A bioinformatics program called BioServers hosted by Cold Spring Harbor Laboratory can be used to compare any gene sequences of interest.

Instructions:

1. Go to the BioServers home site at: http://www.bioservers.org/bioserver/
2. Click on "Sequence Server". You can work without registering.
3. Click on "Manage Groups" in the upper, right corner of the page.
4. Select "Public" from the drop down menu with sequence sources.
5. Check the box next to "AAI SLE Genes" created on 04/13/2015 and click "OK".
6. Drop-down menus will appear in the work page. Select "H. sapiens IRF5 variant 3" in the first drop-down and "H. sapiens IRF5 variant 5" in the second drop-down.
7. Make sure both boxes are checked on the left side of the screen and then click "Compare – Align: Clustal W".
8. Search results will appear in a new window. This will become your new work page.
9. Enter "4000" in the box to change the number of nucleotides displayed per page. Click "Redraw" to refresh the page.
10. IRF5 variant 3 should appear as the sequence on top and IRF5 variant 5 should be lined up immediately below. Any nucleotides highlighted in yellow are SNPs.
11. Count the number of differences between the two sequences following the directions below:
    a. Count all yellow-shaded polymorphisms (highlighted letters represent SNPs; highlighted dashes (-) represent insertions/deletions).
    b. If several nucleotides in a row are highlighted, count it as a single polymorphism.
    c. Be careful of polymorphisms that "wrap-around" the end of one line of sequence text to the next. These should still be counted as a single polymorphism.

Questions:

1. What is the total sequence length of the two IRF5 sequences? *3003 base pairs*
2. What is the total number of differences between variant 3 and variant 5? *32 differences*
3. Calculate the Percent Difference between the two sequences. Use the following equation to solve:
   % Difference = (# of nucleotide differences/total sequence length) = 1.06%
4. Is your calculated percent difference similar to the 0.1% difference estimated for humans overall?
   The calculated percent difference of 1.06% is significantly different than 0.01% for humans overall.
5. Explain your answer above, based on your knowledge of the IRF5 gene and its link to SLE.

The IRF5 gene has SNPs that have been linked to increased risk for developing SLE. It would make sense that genes with documented variability would have more variation than other regions of DNA. Genes that encode essential proteins, those necessary for survival in a particular functional shape, likely have far less variation than genes encoding proteins that work in a variety of forms.

Journal Club Teacher
Analyzing Peer-Reviewed Research

Background

When scientists publish their research, the research article must first go through a process called peer review. Peer review ensures that articles published in scientific journals are of scientific interest and involved valid science. Rather than trusting the integrity of the authors, scientists in the same field of research (peers) provide input on the quality of the work and significance to the field of science.

A research paper is first sent to an editor of a scholarly, or peer-reviewed, journal. The editor will decide if the article is appropriate for the journal. If so, the article will be forwarded to several other peer-reviewers who will judge the methods, analysis, and significance of the work. The comments are returned to the editor and the original author. The research may be accepted or rejected, or more frequently, accepted with revision. The researchers will revise the experiment or the paper as the comments require and the paper is submitted again for publication.

Peer review is an important process in the scientific community. It is a form of self-policing to minimize bad science or personal bias from being published. For scientists in a particular field, it is important to stay up-to-date with peer-reviewed articles. Scholarly journals are at the very front of scientific knowledge. While it may take years for new research to reach a textbook, it may only take weeks for the most current research to reach journals. While the articles are technical and often complicated, it is important for any science student to learn how to find peer-reviewed articles and practice reading technical scientific writing.

Instructions

1. Go to http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0054487 to access the article "RNA-Seq for Enrichment and Analysis of *IRF5* Transcript Expression in SLE.
2. Click "Download PDF" and read pages 1-11, including the Abstract, Introduction, Results, and Discussion.
3. While reading, take notes in a notebook. If the article is printed, highlight and underline terms, processes, methods, data, results, and conclusions that are important in the article.
4. If terms and techniques that are not familiar are used, be sure to look them up in a scientific dictionary or trustworthy internet source.
5. Bring the original PDF article and notes to class on the day of the Journal Club. Be prepared to share thoughts and ideas about the article in small groups and then as a class.

Questions

1. Generate a list of keywords from the article. What are 6-8 of the most important terms from the article? Define each term.
2. What is the author's main purpose? What is their overall goal for the research?
3. List 3 laboratory techniques that the researchers used in their experiment. Briefly describe each method.
4. Did the researchers successfully accomplish their purpose? Explain.
5. What questions do you still have about the research after reading the article?
6. Write a summary of the article that is no more than a single page. Use the formatting guide and sample as a guide.

Stone R., Du P., Feng D., Dhawan K., Rönnblom L., et al. (2013) RNA-Seq for enrichment and analysis of *IRF5* transcript expression in SLE. *PLoS ONE 8*(1): e54487. doi:10.1371/journal.pone.0054487

Journal Club Teacher
Additional Instructions

Reading the Article and Summarizing
After assigning the journal article to students, they should be given sufficient time to read the article several times before discussing as a class. It may be worthwhile to give students one week to two weeks to read, take notes, and write their summary. If all of the work will be completed in class, it may require two full blocks for students to accomplish the same task. The article will be very challenging for students to read.

Begin each class day asking students if they have questions about the reading. Answer any questions if possible or look up the answers together. It is important to show students that the only way to learn while reading is by stopping and looking up unknown words. Additionally, review the document "Format for Journal Summary" with students, which provides tips for reading, note-taking, and writing the summary. Students may also want to look at the "sample journal summary" provided.

Journal Club
Students sometimes have the impression that scientists work independently and reading scientific articles is an independent process. In laboratories, researchers often gather together to read and discuss current scholarly research articles. Hosting a journal club is a way to show students a social, collaborative side of science. When students go to a movie with friends, they immediately share their ideas and thoughts; likes and dislikes. It is important to show students that the same can be done after reading a book or articles.

If school policy allows, provide food and drinks for students on the day of the summary. Water, juice, coffee, tea, or soda can relax a setting and encourage sharing of ideas that may not happen in a formal school setting. Additionally, food options like pastries, donuts, muffins, crackers, vegetables, fruit or other light foods can do the same.

Start the journal club by breaking students into small groups of 3 or 4. Have students share their answers to questions 1-5 in the assignment. Ask students to generate a list of questions that they still have, even after reading the article, writing a summary, and discussing with peers.

After sharing in small groups, reconnect as an entire class. Ask the students to share their groups' answers to discussion questions. Poll the class for questions that remained unanswered during the small group discussion. Ask students for the thoughts about the research and the article. Was it challenging to read? Do they have a better understanding after discussing the article? What further research would be interesting? Were there flaws in the experimental design that they can see?

After the journal club, give students an extra day to revise their journal summary. They should take the feedback from peers to edit their summary before submitting for grading.

Format for Journal Summary

Use the following format and tips when writing your journal summary.

Name
Date
Course

Article Title

Keywords:
4-6 words crucial to the experiment

Summary:
3-4 paragraphs

First Paragraph

- Background problem/question being addressed
- Specific research problem being addressed/purpose of experiment
- Very brief summary of results

Second Paragraph

- Introduction of researchers who conducted the experiment
- Description of the experimental methods

Third Paragraph

- Explanation of the results of the experiment
- Broader significance of the experiment to science and society as a whole

References (with annotation)
APA format

In the annotated bibliography, share your thoughts about ideas presented in the article. Be sure to include questions generated by your reading, the relation between what you already knew and what you read. Also, critique the article. Is it well written? Is the data presented in a straightforward manner? What further research would you like to see?

Hints:

- Consistency is key; whatever you do, do it every time.
- Write a summation of the article. Be sure to avoid a copy/paste approach or anything that might hint at plagiarism. I recommend reading the article and taking notes. Then, write your summary from your notes.
- Use a formatting guide for your reference. I recommend Noodle Tools but others are available.
- Technical writing is typically written in active voice as opposed to passive voice. For example, a passive sentence would say "The experiment was conducted by the researchers." The same sentence, worded in active voice would say "The researchers conducted the experiment."
- Spell out numbers from zero to ten. Use the numerical # for numbers 11+.
- Avoid using the words referring to yourself, like "we, our, us, etc". Summarize the article as it is written; you did not have a part in the research.
- Use past tense when referring to the research being conducted.
- Use descriptive and simple words rather than "showy" words. Be as straightforward as possible.
- Use italics for scientific names and capitalize proper nouns.
- Avoid pronouns, especially to begin a sentence. Use the noun itself instead.

Sample Journal Summary

Student Name
Course Name
Date

Occurrence and antibiotic Resistance of Escherichia coli O157:H7 in a Watershed in North-Central Indiana

Keywords: immunomagnetic separation, Polymerase chain reaction, *E. coli* O157:H7

Summary:

Although *Escherichia coli* is most commonly found in animal sources, the bacteria can be a serious problem when waterborne. *E. coli* O157:H7, an enterohemorrhagic strain, produces the shiga toxin and can cause kidney damage. Scientists investigated the presence of *E. coli* O157:H7 in an Indiana watershed. Researchers screened samples for fecal coliform bacteria and O157:H7, and also investigated the antimicrobial resistance levels of bacteria in the watershed. Scientists found *E. coli* O157:H7 in various points along the river, and also found some of these bacterial colonies to be resistant to different antimicrobial drugs.

Researchers collected bacterial samples from three points along the Wildcat Creek and two located in Kokomo, a city that uses the creek as a water source. In order to collect fecal coliform samples, researchers used membrane filtration to separate colonies for counting. Scientists used immunomagnetic separation to detect *E. coli* O157:H7. This separation method is often used to isolate bacterial cells from fluid samples. To confirm the presence of shiga toxins (stx1 and stx2), researchers used Polymerase chain reaction (PCR) to amplify bacterial DNA in order to identify stx1 and stx2 genes. Also, scientists used the Kirby-Bauer method to determine the antimicrobial resistance levels in *E. coli* O157:H7. This experimentation method uses antibiotic discs placed on agar plates to determine resistance using bacterial zones of inhibition.

Scientists found that fecal coliform counts increased as water traveled downstream. Also, 58.7% of water samples contained *E. coli* O157:H7. Researchers found O157:H7 at every test sight, including those in the city of Kokomo. Out of all isolates tested for shiga toxin presence, 100% of isolates contained the shiga toxin genes stx1 and stx2. Scientists used 21 isolates to test for resistance against eight antimicrobial agents. Researchers found thirteen of twenty-one isolates to be resistant to at least one antibiotic. The highest amount of resistance occurred with tetracycline, kanamycin, and streptomycin. Overall, researchers found that the water present in Wildcat Creek is contaminated with *E. coli* O157:H7 and coliform bacteria, and more must be done to determine pollution sources to prevent further contamination.

References:

Fincher, L., Parker, C., & Chauret, C. (2009). Occurrence and antibiotic Resistance of Escherichia coli O157:H7 in a Watershed in North-Central Indiana. *Journal of Environmental Quailty, 38*, 997-1004. doi:10.2134/jeq2008.0077

This study shows that more must be done to prevent further contamination of Wildcat Creek. The water, testing positive for E. coli *O157:H7 and coliform bacteria, is used by Kokomo's citizens and could potentially cause illness to those who use it. I found this study interesting because most research focuses on* E. coli *in animals, but these bacteria are also a serious problem in water sources around the world.*

Instrument Training Teacher
IT 1: Mixing and Weighing
All answers should be recorded in a notebook devoted solely to lab work.

Prior Knowledge

1.1 Define the following terms as they relate to mixing and weighing in the laboratory:

Weigh Boat – *an open container, usually disposable, used for weighing objects on a scale.*
Spatula – *a laboratory utensil used for manipulating dry chemicals*.
Magnetic Stir Plate – *a laboratory device with a rotating magnet used to mix solutions.*
Magnetic Stir Bar - *a magnet coated in Teflon that can be placed inside a solution to be mixed on a magnetic stir plate.*
Homogenous – *mixtures that have the same composition throughout the entire volume.*
Solution – *an aqueous mixture that is homogenous.*
Mixture – *a substance made of two or more substances, but not chemically combined.*
Tare – *the weight of an empty container.*
Solvent – *a substance that dissolves a solute.*
Solute – *a substance that is dissolved in a solvent.*

Safety Issues

1.2 What safety concerns must be considered as the investigation is conducted? Provide an extensive list.

Never use broken glassware, wear appropriate safety clothing and eyewear in the lab, never eat/drink out of lab glassware, be cautious using electronic devices around water, etc.

Prior Skills

1.3 Write a paragraph explaining the proper procedure to mix a solid (powder) into solution with water.

A beaker or flask containing half of the final volume of water should be placed on a magnetic stir plate. A magnetic stir bar should be placed into the beaker gently. Slowly turn up the speed of the magnetic stir plate so the stir spins slowly, creating a gentle vortex. While the stir plate is stirring, slowly pour the powder into the solution. When the powder has dissolved, pour the remaining volume of water into the beaker or flask until completely mixed.

1.4 Write a paragraph explaining the proper procedure to measure the mass of a solid using a digital scale.

Turn on the scale ensure that the correct units are displayed. Place an empty weigh boat on the scale and press "tare" or "zero" to recalibrate the scale, accounting for the weight of the weigh boat. Slowly add the desired solid onto the scale, little by little. When the desired volume has been reached, wait several seconds to ensure that the scale is not adjusting.

1.5 To make a batch of Nutrient Agar for culturing bacteria, 32 grams of dehydrated agar must be mixed to make a liter of media.

a. How much dehydrated agar must be measured to make 500 mL of media?
16 grams
b. How much dehydrated agar must be measured to make 300 mL of media?
9.6 grams
c. How much dehydrated agar must be measured to make 50 mL of media?
1.6 grams

Laboratory Protocol

1.6 To make Jello, 3 oz. of dehydrated Jello powder is mixed with 1 cup of water.

a. Calculate the amount of dehydrated Jello powder needed to make 50 milliliters of Jello.

85 grams/236.6mL = x grams/50mL = 17.96 grams of Jello

b. After calculating 1.6a, measure the dehydrated Jello powder using a digital scale and hot water using a graduated cylinder. Mix the ingredients using a magnetic stir plate and pour into a paper cup to set. Include a photograph of the completed Jello in your lab notebook.

Provide photograph

c. Write a paragraph explaining why hot water is needed to mix the Jello.

Jello contains gelatin, which is a mixture of proteins extracted from collagen in the skin, bones, and connective tissues of animals. Gelatin is somewhat hydrophobic, meaning it does not readily bond with water. Gelatin is more water soluble in hot water. Gelatin is soluble in polar solvents, like oil.

Name_____

IT1: Mixing and Weighing

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
1.1 Terms are accurately defined and relate to laboratory use	**4**	**2**	**0**
1.2 At least 4 safety issues are described in detail.	**2**	**1**	**0**
1.3 Paragraph accurately describes proper procedure for mixing a solution.	**3**	**2**	**0**
1.4 Paragraph accurately describes the proper procedure for measuring a solid.	**3**	**2**	**0**
1.5 Agar calculations are accurately provided.	**3**	**2**	**0**
1.6a Jello calculation is accurately provided	**1**		**0**
1.6b Jello is measured and mixed appropriately. A photograph is included.	**2**	**1**	**0**
1.6c Paragraph thoroughly explains the need for hot water in mixing Jello.	**2**	**1**	**0**
Column Total			
Total Score	**/20**		

Comments:

Instrument Training Teacher
IT 2: Glassware
All answers should be recorded in a notebook devoted solely to lab work.

Prior Knowledge

2.1 Define the following terms as they relate to glassware in the laboratory:

Adhesion – *the attraction between molecules or particles of two different types.*
Cohesion – *the attraction between molecules or particles of the same type.*
Density - *the mass of an object per unit volume.*
Meniscus – *the concave upper surface of a liquid in a glass tube.*
TC (To Contain) – *a designation given to glassware that is accurate to contain a given volume, but will not pour exactly the volume. (Usually flasks or graduated cylinders)*
TD (To Deliver) – *A designation given to glassware that will accurately dispense a given volume. (Usually pipettes or burets)*
Hydrogen Bond – *a weak, intermolecular attraction between hydrogen atoms and electronegative atoms like nitrogen, fluorine, or oxygen.*
Polarity - *an alignment of atoms in a molecule such that opposite sides of the molecule have opposite charges.*

2.2 Write a paragraph about the proper procedure for cleaning up broken glass in the laboratory.

First, announce to the class that glass has broken and alert the teacher. Next, clean visible glass shards with a broom and dustpan. Place these shards into a sharp object container in the lab. Finally, using a wet paper towel, gently wipe down the floor or table where the glass broke to remove any small slivers of glass.

Safety Issues

2.3 What safety concerns must be considered as the investigation is conducted? Provide an extensive list.

Never use cracked or chipped glassware, never eat/drink out of laboratory equipment, always wear protective clothing and eyewear, tie back long hair in the laboratory, do not grab objects that may be hot, never leave an open flame unattended, frequently check to ensure that gas is turned off when not in use, wear close-toed shoes in the laboratory, avoid wearing clothing with baggy sleeves, etc.

Prior Skills

2.4 For each of the following pieces of glassware, describe the appropriate use, the location in the laboratory (cabinet # or lab bench) and draw the shape:

Beaker – *glassware commonly used to sit, mix, and heat liquids.*
Erlenmeyer Flask – *glassware used to mix liquids without risk of spilling or boiling liquids to reduce evaporation.*
Graduated Cylinder – *glassware used to measure the volume of a liquid.*
Test Tube – *glassware used to hold, mix, or heat small quantities of solid or liquid chemicals.*
Florence Flask – *glassware with a round bottom and long neck designed for uniform heating or boiling of liquid chemicals.*
Petri Dish – *a shallow, circular dish with a covered lid used to culture microorganisms.*
Media Bottles – *glassware used to store solid or liquid chemicals*

Laboratory Protocol

2.5 a. Accurately determine the volume of water able to fit in a test tube. DO NOT measure the volume of the test tube itself.

Answers will vary.

b. Accurately measure the mass of water that fits into the test tube used above. DO NOT measure the mass of the test tube itself.

Answers will vary.

c. Now that you know the mass and volume of water in your test tube, determine the density of the water, expressed in grams/milliliters.

Divide mass of water/volume of water.

d. What is the "theoretical" density of water at room temperature? Why might your calculated density be slightly different?

Density is dependent of temperature. At room temperature (21 Celsius), the density of pure water is 0.998 grams/milliliter. If students' calculated densities vary, there could be several reasons. If water is colder or warmer than 21 Celsius, volume will change. Additionally, human error is a factor when measuring mass and volume. Limitations in the accuracy of scales and graduated cylinders may also result in slightly different numbers. Finally, if tap water is used rather than distilled water, solute in the water may result in higher densities.

2.6 Complete the set-up shown below using the proper equipment. Once set up, boil exactly 100 mL of water and record the time to reach a vigorous boil. Include a picture of the set up in your lab book.

** place a heat diffuser (wire mesh) over the ring stand to avoid shattered glass (not shown)*

Name_____

IT2: Glassware

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
2.1 Terms are accurately defined and relate to laboratory use.	**4**	**2**	**0**
2.2 Paragraph accurately describes clean-up procedure used in our lab. Paragraph is written clearly and with proper grammar and spelling.	**3**	**2**	**0**
2.3 At least 4 safety issues are described in detail.	**2**	**1**	**0**
2.4 All equipment is listed with its proper use and location. The glassware is also drawn to the best of the student's ability.	**4**	**2**	**0**
2.5a/b. Volume and mass of water are accurately calculated.	**2**	**1**	**0**
2.5c/d. Density is calculated accurately. Student provides "theoretical" density and thoroughly explains any differences.	**3**	**2**	**0**
2.6 Set-up is proper. Photograph is included. Time of boiling is included. Set-up is disassembled and returned to proper place.	**2**	**1**	**0**
Column Total			
Total Score	**/20**		

Comments:

Instrument Training Teacher
IT 3: Diluting Stock Solutions
All answers should be recorded in a notebook devoted solely to lab work.

Prior Knowledge

3.1 Define the following terms as they relate to microscopes in the laboratory:

Stock Solution – *a concentrated solution used to store a chemical, but must be diluted before use*
Working Solution – *a diluted solution that is used in a laboratory procedure*
Dilute – *to make less concentrated by adding water or another solvent.*
Weight-in-Volume – *a method of explaining concentration in weight of solute/volume of solvent.*
Volume Percent – *a method of explaining concentration in volume of solute/volume of solvent, usually when the solute is a liquid.*
Serial Dilution – *a series of dilutions used to dilute a solution by a known factor.*

3.2 The equation $C_1V_1=C_2V_2$ is used to dilute solutions from higher to lower concentration. Define each of the variables in the equation:

C_1 = *starting concentration* V_1 = *starting volume*
C_2 = *ending concentration* V_2 = *ending volume*

Safety Issues

3.3 What safety concerns must be considered as protocol is conducted?

Never use cracked or chipped glassware, never eat/drink out of laboratory equipment, always wear protective clothing and eyewear, tie back long hair in the laboratory, do not grab objects that may be hot, never leave an open flame unattended, frequently check to ensure that gas is turned off when not in use, wear close-toed shoes in the laboratory, avoid wearing clothing with baggy sleeves, etc.

Prior Skills

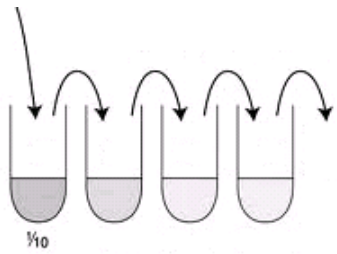
3.4 For a laboratory exercise, you are asked to make 200 milliliters of 8% (w/v) NaCl solution. How much NaCl needs to be added to the water? *16 grams NaCl*

3.5 You have 95% ethanol (stock solution) in the laboratory. You need 100mL of 70% ethanol (working solution).

a. How much 95% ethanol should you start with?
$C_1V_1=C_2V_2$
(95%)(x) = (70%)(100mL)
x = 73.68mL 95% ethanol

b. How much water should you add to the 95% ethanol to make 100mL of 70% ethanol?
100mL total volume – 73.68mL stock = 26.32mL water

3.6 Write a paragraph explaining the purpose of a serial dilution.



Serial dilutions are common in many laboratory settings. Serial dilutions repeatedly dilute a solution by a known amount. Serial dilutions are useful when very small quantities of a chemical are needed but scales cannot accurately measure the quantity. Additionally, serial dilutions can also be used to reduce the number of microbial cells in a liquid culture to obtain a convenient concentration with which to work.

Laboratory Protocol

3.7 The goal of this activity will be to accurately measure the mass of NaCl, mix it with a known amount of water, and dilute the solution.

a. Start by measuring 1g of NaCl. Add water until you reach a total volume of 9mL. Add 4 drops of food coloring and mix well. You now have a 1/10 or 1% (w/v) concentration of NaCl. Perform 3 serial dilutions (as shown above). Photograph your serial dilution tubes in a test tube rack and include the photograph in your lab notebook.

Photo should show 4 tubes, each one successively lighter in color.

b. How many grams of NaCl are present per milliliter in each tube?

Tube 1 = 0.1g/mL
Tube 2 = 0.01g/mL
Tube 3 = 0.001g/mL
Tube 4 = 0.0001g/mL

3.8 Obtain a bottle of 3% hydrogen peroxide (H_2O_2). Make 10 mL of 1% H_2O_2. Mix the solution with a magnetic stir plate. Have your teacher initial your rubric as confirmation.

$C_1V_1=C_2V_2$
(3%)(x) = (1%)(10mL)
x = 3.33mL of 3% hydrogen peroxide

Add 3.33mL hydrogen peroxide to 6.67mL distilled water.

Name_____

IT3: Mixing and Separating Solutions

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
3.1 Terms are accurately defined and relate to laboratory use.	**3**	**2**	**0**
3.2 All variables of the equation are accurately defined.	**2**	**1**	**0**
3.3 At least 2 safety issues are described in detail.	**2**	**1**	**0**
3.4 Mass of NaCl is calculated.	**1**		**0**
3.5 Starting volumes of ethanol and water are accurately calculated.	**2**	**1**	**0**
3.6 Paragraph accurately explains the purpose of serial dilution in detail.	**2**	**1**	**0**
3.7a Serial dilution is properly executed and a photograph is included.	**4**	**2**	**0**
3.7b Concentrations of NaCl/mL are accurately calculated.	**2**	**1**	**0**
3.8 1% solution is properly made and rubric is initialed	**2**	**1**	**0**
Column Total			
Total Score	**/20**		

Comments:

Instrument Training Teacher
IT 4: Micropipetting
All answers should be recorded in a notebook devoted solely to lab work.

Prior Knowledge

4.1 Define the following terms as they relate to pipetting volume in the laboratory:

liter – *a unit of volume in the metric system.*
microliter – *a unit of volume in the metric system equal to one millionth of a liter.*
micropipette – *a small pipette that uses aspiration to dispense small volumes of liquid.*
volume – *the quantity of three dimensional space taken up by an object.*
aspiration – *the process of drawing in or out using suction.*
precision – *the degree to which repeated measurements change each repetition.*
accuracy – *the proximity of a measurement to the actual or true value.*

4.2 Explain the relationship between the units of measurement for volume in the metric system. Be sure to include the terms liter, milliliter, and microliter.

Volume units in the metric system are connected to length. One liter is defined as the volume of a cube 10 cm x 10 cm x 10 cm. Sub-units of a liter are divided into units of 10. So, a deciliter is defined as one tenth (0.1) of a liter. A centiliter is one hundredth (0.01) of a liter, a milliliter is one thousandth (0.001) of a liter, and so on. A microliter is one millionth (0.000001) of a liter.

Safety Issues

4.3 What laboratory safety concerns must be considered as the investigation is conducted? Provide an extensive list.

Never use cracked or chipped glassware, never eat/drink out of laboratory equipment, always wear protective clothing and eyewear, tie back long hair in the laboratory, wear close-toed shoes in the laboratory, avoid wearing clothing with baggy sleeves, etc.

Prior Skills

4.4 Convert the following volumes between metric units.

1 L = _____*1,000*_____ mL
1 L = _____*1,000,000*____µL
50 mL = ___*0.05*_____L
5,000 µL = _*0.005*_____L
675 µL = __*0.675*_____mL

4.5 Selecting a micropipette is important for accurately measuring small volumes of liquid. The list below explains the volume range for three common micropipettes:

P20	1 µl to 20 µl
P200	20 µl to 200 µl
P1000	200 µl to 1000 µl

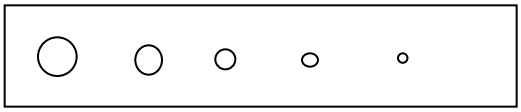
You are asked to pipette several solutions into a microcentrifuge tube in the following order: 750 µL of distilled water, 25 µL of DNA, and 1 µL of restriction enzyme. Write a paragraph describing the sequence of events necessary to transfer these liquids without cross-contamination.

First, a P1000 pipette should be obtained a fresh tip applied. The water should be dispensed into a microcentrifuge tube and the tip discarded. Next, a P20 pipette should be obtained with a fresh tip and the DNA dispensed into the tube. The droplet should be placed on the side of the tube to confirm transfer. The tip should be discarded and a new tip applied. Finally, the P20 pipette should be used to transfer the restriction enzyme. Again, the droplet should be applied to the side of the tube to confirm transfer. The tube should be centrifuged briefly to mix ingredients.

Laboratory Protocol

4.6 Obtain a small piece of Parafilm, a P20 micropipette with tips, and a small microcentrifuge tube with water. Using proper technique, pipette drops of water along the Parafilm in the following volumes. Then, compare with a friend to ensure that the proper volume has been pipetted and photograph. Include the photograph, labeled with a title and caption, in your notebook.

20 µl, 15 µl, 10 µl, 5µl, 2µl



4.7 Label a clear microcentrifuge tube with your initials and #1. Transfer 1 mL of distilled water, 2.0 µl blue food coloring, 25 µl yellow food coloring and 3.0 µl red food coloring into the tube. Cap the tube, microcentrifuge to mix, and photograph your tube. Include a photograph, labeled with a title and caption, in your notebook.

Colors will vary based on specific food coloring dye used. Make a "standard" tube with the measurements described above. Student tubes should match the standard exactly.

Name_____

IT4: Micropipetting

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
4.1 Terms are accurately defined and relate to laboratory use.	**3**	**2**	**0**
4.2 Paragraph accurately describes relationship between units of measurement. Paragraph is written clearly and with proper grammar and spelling.	**3**	**2**	**0**
4.3 At least 2 safety issues are described in detail.	**2**	**1**	**0**
4.4 Unit conversions are accurately calculated.	**4**	**2**	**0**
4.5 Paragraph accurately describes how to pipette samples, including selection of pipettes and changing of tips. Paragraph is written clearly and with proper grammar and spelling.	**4**	**2**	**0**
4.6 Water droplets are pipetted on Parafilm and reflect correct volumes. Photograph is included with title and captions	**2**	**1**	**0**
4.7 Photograph is included. Color matches prepared sample. Picture is included with detailed title and caption explaining the contents of the photograph.	**2**	**1**	**0**
Column Total			
Total Score	**/20**		

Comments:

APPENDIX II – DAILY PLANS

STUDENT DOCUMENTS

Bioinformatics

Assignment #1: Investigating the Genetic Component of Systemic Lupus Erythematosus through Online Mendelian Inheritance in Man (OMIM)

Background Information:

Online Mendelian Inheritance in Man (OMIM) is one of the most popular medical genetics websites. The website provides access to a database of all known diseases with a genetic component of some kind. The database is continuously updated with new information about the molecular relationship between genes (genotype) and disease (phenotype). The OMIM database links known information about a disease with genes in the human genome. All peer-reviewed research that has shown a link between genes and disease are cited for additional information. OMIM is operated by the National Center for Biotechnology Information (NCBI), the same organization that operates BLAST.

For a particular disease of interest, information about relevant genes, chromosomes, SNPs and other polymorphisms, and protein sequence can be easily accessed. Every disease with an entry in OMIM is given a unique identification called a MIM code. An MIM code is a six-digit number that describes the genetic disease. The first digit in the code reveals information about the genetic disease:

1: Autosomal Dominant	(before 1994)	4: Y-linked
2: Autosomal Recessive	(before 1994)	5: Mitochondrial
3: X-linked		6: Autosomal (after 1994)

Symbols before the MIM code can also reveal information about the disease:

***:** entry is a gene	**%:** loci is unknown, but disease is Mendelian
#: disease is controlled by multiple loci	**No Entry:** molecular basis is unknown
+: gene sequence and phenotype are known	**^:** entry no longer exists

One additional code is used to explain the location of a genetic component within the human genome. Each locus (specific location of a chromosome) is given a code consisting of letters and numbers. A chromosome has a short arm (abbreviated "p") and a long arm (abbreviated "q") separated by a centromere. When stained, a chromosome appears to have bands of light and dark. These bands can be counted and are labeled p1, p2, p3 or q1, q2, q3, etc. The counting begins from the centromere and moves outward. If the chromosome is magnified greatly, additional sub-bands and sub-sub-bands can be counted. For example, a locus with the code 7q31.2 is located on the long arm (q) of the 7th chromosome in band 3, sub-band 1, and sub-sub-band 2.

Instructions:

1. Go to http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
2. In the search bar, type in the MIM code 612251.When the search is complete, click on the link to reveal more information about the disease Systemic Lupus Erythematosus.
3. Use the information of the page to answer questions 1-3 below.
4. On the right side of the screen, click the arrow to expand the "Clinical Resources" menu. Under "Genetics Home Reference", click the link "systemic lupus erythematosus."
5. Use the information on the Genetics Home Reference page to answer question 4 below.

Questions:

1. What is the name of the name of the gene/locus associated with SLE?
2. What is the chromosomal location of the locus? Include chromosome and band numbers.
3. Is this the only locus associated with susceptibility to SLE? Explain your answer.
4. Write a one page essay on SLE. Be sure to cover the following topics:
    a. Signs and symptoms
    b. Body parts affected
    c. Prevalence in countries and races/genders
    d. Genetic components
    e. Inheritance

Bioinformatics: Assignment 1 Rubric
Investigating the Genetic Component of Systemic Lupus Erythematosus through Online Mendelian Inheritance in Man (OMIM)

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
Heading – includes name, course, and date	**2**	**1**	**0**
Font – paper is typed with 1-inch margins, size 11-12 font in Times New Roman or Arial	**3**	**2**	**0**
Format – content is roughly 1 page in length, single spaced	**5**	**3**	**0**
Spelling/Grammar – paper shows editing and careful review. Spelling and grammatical errors are absent.	**5**	**3**	**0**
Content – Signs and symptoms of SLE discussed in detail.	**5**	**3**	**0**
Content – Body parts affected by SLE discussed in detail.	**5**	**3**	**0**
Content – Prevalence of SLE in various countries and races/genders explained.	**5**	**3**	**0**
Content – Genetic components of SLE discussed in detail.	**5**	**3**	**0**
Content – Inheritance patterns of SLE discussed in detail.	**5**	**3**	**0**
Column Total			
Total Score	**/40**		

Comments:

Bioinformatics
Assignment #2: BLAST Terminology

Background Information:
Bioinformatics is the use of computational methods for comparing and analyzing genetic data. The study originated in the 1970s but gained popularity in the 1980s and 1990s with the advent of affordable and powerful computers. The history of bioinformatics follows in the footsteps of molecular biology. One of the fundamental questions of molecular biology is how living organisms store, use, and express genetic information. As molecular biology entered the "sequencing age" and massive quantities of data began pouring out of laboratories, tools to make sense of that data were necessary. Molecular biology and bioinformatics came together as genetics became a data-driven science.

Today, bioinformatics is being used in nearly all fields of biological science. From evolutionary biology to medical research, finding patterns and identifying anomalies in genetic data is key. Much of the genomic data being collected is published in databases like GenBank. Computer programs, often available to the public for free, allow users to quickly search and compare sequences of DNA, RNA and proteins. Software like BLAST and ClustalW have become standard tools used to align and compare sequences.

While the list of applications for bioinformatics is long, a few techniques are used most frequently. Sequence alignment is used to find similarities and differences between related species. This technique can be used to elucidate the genetic component of a disease or determine relatedness of two species in constructing phylogenetic trees. Sequence alignment can also be used to identify model organisms for basic research on genes or finding candidates for drug testing. Another application of bioinformatics is gene annotation, in which researchers identify the coding and regulatory regions of DNA. More complex genome analysis, such as maps of methylation sites, transcription factors and other genetic modifications can be done. Genome-wide Association Studies have become fundamental in the study of diseases, including cancer. Comparing genomes of thousands of patients has allowed researchers to pinpoint the mutations associated with diseases like Alzheimer's disease, breast cancer, and more.

Instructions:
Use the websites below to explore the most popular bioinformatics database: NCBI BLAST

1. BLAST Help Glossary - http://www.ncbi.nlm.nih.gov/books/NBK62051/
2. NCBI Handbook Ch. 16 - http://www.ncbi.nlm.nih.gov/books/NBK21097/
3. NCBI Handbook Glossary - http://www.ncbi.nlm.nih.gov/books/NBK21106/
4. NCBI Statistics Tutorial - http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html

Questions:
Answer all questions in your own words.

1. What does BLAST stand for?
2. What is the process of alignment as it relates to bioinformatics?
3. How is a "local alignment" different than a global alignment?
4. What does sequence similarity searching mean?
5. Define the following terms as they relate to BLAST:
   Similarity
   Identity
   Homology
   Paralogy
6. What is the significance of an E value?
7. Explain FASTA format
8. Explain GI format
9. What is an Accession number?

Bioinformatics
Assignment #3: Finding Genes with GI Identifiers

Background Information:
Searching bioinformatics databases to find information about specific genes is a fundamental skill for genetics researchers. Unfortunately, many of the popular databases and search programs are not intuitive or user-friendly. BLAST is a computer algorithm that allows researchers to compare sequences of DNA or proteins. Several types of searches can be made, depending on the original search or "query". If a nucleotide sequence is input, the search results can be filtered to show similar nucleotide sequences or related proteins in other organisms.

To input a query in BLAST, several formats can be used. Frequently used formats include GenInfo identifiers (GI), FASTA identifiers, or accession numbers. Each nucleotide sequence in the database will have all three types of identifiers. In this activity, nucleotide sequences will be input into BLAST using a GI identifier to search for a gene linked to Systemic Lupus Erythematosus (SLE).

Part A
Instructions:
1. Go to the official BLAST home site at: http://www.ncbi.nlm.nih.gov/BLAST/
2. Select "nucleotide blast"
3. Paste the GI number gi|629266059 into the search box and hit BLAST at the bottom of the page.
4. When the search results have loaded, scroll down to the section of the page showing "sequences producing significant alignments". Look at the search information for the first result. Answer questions 1 and 2 below, then click on the Accession number in the right column to answer the remaining questions.

Questions:
1. What is the E-value for the first result? What is the significance of this value?
2. What is the Identity % for the first result? What does this suggest about the sequence?
3. Provide the following information for the sequence represented by the first result:
    a. Organism
    b. Gene Name
    c. Length of gene (in base pairs)
    d. Type of molecule (DNA, mRNA, amino acid, etc)
    e. Provide a brief summary of gene family to which this gene belongs
    f. Based on your results, does it seem plausible that this gene could be linked to SLE?
    g. Write the first 20 nucleotides of the gene sequence.

Part B
Instructions:
1. Go to the article "RNA-Seq for Enrichment and Analysis of *IRF5* Transcript Expression in SLE" at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0054487
2. Read the Abstract and Introduction before answering the questions. You may need to use additional sources to help answer the questions.

Questions:
1. Define the following terms:
    a. Polymorphism
    b. Haplotype
    c. Alternative Splicing
2. Write a short summary of the article, explaining how polymorphisms in *IRF5* may increase a person's risk of SLE.

Bioinformatics
Assignment #4: Comparing DNA Sequences in BioServers with ClustalW

Background Information:
Humans share roughly 99.9% of their DNA with every other living human being. The 0.1% that is different accounts for the differences in appearances between each person. Sometimes, these differences can be responsible for causing disease. Interestingly, when the same type of analysis is used to compare human and our nearest primate ancestor, the chimpanzee, the difference increases to 1.2%. In other words, 98.8% of human DNA is identical to chimpanzee DNA. Clearly, small changes in DNA can result in big differences in an organism.

The tiny differences in DNA sequence that result in the variation seen among humans are called polymorphisms. Literally, the word translated to "many forms". More specifically, polymorphisms are nucleotides that vary from human to human. These nucleotides can occur in short stretches or as individual nucleotides. Short stretches of DNA that are variable are known as Restriction Fragment Length Polymorphisms (RFLPs) while individual variable nucleotides are known as Single Nucleotide Polymorphisms (SNPs). If two gene sequences are lined up and compared, it is possible to identify and count the differences between the two sequences.

This technique of aligning and comparing gene sequences can be used to link specific polymorphisms with disease. If the sequences of diseased and healthy patients are compared, polymorphisms unique to diseased individuals are said to be linked to disease. A bioinformatics program called BioServers hosted by Cold Spring Harbor Laboratory can be used to compare any gene sequences of interest.

Instructions:
1. Go to the BioServers home site at: http://www.bioservers.org/bioserver/
2. Click on "Sequence Server". You can work without registering.
3. Click on "Manage Groups" in the upper, right corner of the page.
4. Select "Public" from the drop down menu with sequence sources.
5. Check the box next to "AAI SLE Genes" created on 04/13/2015 and click "OK".
6. Drop-down menus will appear in the work page. Select "H. sapiens IRF5 variant 3" in the first drop-down and "H. sapiens IRF5 variant 5" in the second drop-down.
7. Make sure both boxes are checked on the left side of the screen and then click "Compare – Align: Clustal W".
8. Search results will appear in a new window. This will become your new work page.
9. Enter "4000" in the box to change the number of nucleotides displayed per page. Click "Redraw" to refresh the page.
10. IRF5 variant 3 should appear as the sequence on top and IRF5 variant 5 should be lined up immediately below. Any nucleotides highlighted in yellow are SNPs.
11. Count the number of differences between the two sequences following the directions below:
    a. Count all yellow-shaded polymorphisms (highlighted letters represent SNPs; highlighted dashes (-) represent insertions/deletions).
    b. If several nucleotides in a row are highlighted, count it as a single polymorphism.
    c. Be careful of polymorphisms that "wrap-around" the end of one line of sequence text to the next. These should still be counted as a single polymorphism.

Questions:
1. What is the total sequence length of the two IRF5 sequences?
2. What is the total number of differences between variant 3 and variant 5?
3. Calculate the Percent Difference between the two sequences. Use the following equation to solve:
   % Difference = (# of nucleotide differences/total sequence length)
4. Is your calculated percent difference similar to the 0.1% difference estimated for humans?
5. Explain your answer above, based on your knowledge of the IRF5 gene and its link to SLE.

Journal Club Teacher
Analyzing Peer-Reviewed Research

Background

When scientists publish their research, the research article must first go through a process called peer review. Peer review ensures that articles published in scientific journals are of scientific interest and involved valid science. Rather than trusting the integrity of the authors, scientists in the same field of research (peers) provide input on the quality of the work and significance to the field of science.

A research paper is first sent to an editor of a scholarly, or peer-reviewed, journal. The editor will decide if the article is appropriate for the journal. If so, the article will be forwarded to several other peer-reviewers who will judge the methods, analysis, and significance of the work. The comments are returned to the editor and the original author. The research may be accepted or rejected, or more frequently, accepted with revision. The researchers will revise the experiment or the paper as the comments require and the paper is submitted again for publication.

Peer review is an important process in the scientific community. It is a form of self-policing to minimize bad science or personal bias from being published. For scientists in a particular field, it is important to stay up-to-date with peer-reviewed articles. Scholarly journals are at the very front of scientific knowledge. While it may take years for new research to reach a textbook, it may only take weeks for the most current research to reach journals. While the articles are technical and often complicated, it is important for any science student to learn how to find peer-reviewed articles and practice reading technical scientific writing.

Instructions

6. Go to http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0054487 to access the article "RNA-Seq for Enrichment and Analysis of *IRF5* Transcript Expression in SLE.
7. Click "Download PDF" and read pages 1-11, including the Abstract, Introduction, Results, and Discussion.
8. While reading, take notes in a notebook. If the article is printed, highlight and underline terms, processes, methods, data, results, and conclusions that are important in the article.
9. If terms and techniques that are not familiar are used, be sure to look them up in a scientific dictionary or trustworthy internet source.
10. Bring the original PDF article and notes to class on the day of the Journal Club. Be prepared to share thoughts and ideas about the article in small groups and then as a class.

Questions

7. Generate a list of keywords from the article. What are 6-8 of the most important terms from the article? Define each term.
8. What is the author's main purpose? What is their overall goal for the research?
9. List 3 laboratory techniques that the researchers used in their experiment. Briefly describe each method.
10. Did the researchers successfully accomplish their purpose? Explain.
11. Write a summary of the article that is no more than a single page. Use the formatting guide and sample as a guide.

Stone R., Du P., Feng D., Dhawan K., Rönnblom L., et al. (2013) RNA-Seq for enrichment and analysis of *IRF5* transcript expression in SLE. *PLoS ONE 8*(1): e54487. doi:10.1371/journal.pone.0054487

Format for Journal Summary

Use the following format and tips when writing your journal summary.

Name
Date
Course

Article Title

Keywords:
4-6 words crucial to the experiment

Summary:
3-4 paragraphs

First Paragraph

- Background problem/question being addressed
- Specific research problem being addressed/purpose of experiment
- Very brief summary of results

Second Paragraph

- Introduction of researchers who conducted the experiment
- Description of the experimental methods

Third Paragraph

- Explanation of the results of the experiment
- Broader significance of the experiment to science and society as a whole

References (with annotation)
APA format

In the annotated bibliography, share your thoughts about ideas presented in the article. Be sure to include questions generated by your reading, the relation between what you already knew and what you read. Also, critique the article. Is it well written? Is the data presented in a straightforward manner? What further research would you like to see?

Hints:

- Consistency is key; whatever you do, do it every time.
- Write a summation of the article. Be sure to avoid a copy/paste approach or anything that might hint at plagiarism. I recommend reading the article and taking notes. Then, write your summary from your notes.
- Use a formatting guide for your reference. I recommend Noodle Tools but others are available.
- Technical writing is typically written in active voice as opposed to passive voice. For example, a passive sentence would say "The experiment was conducted by the researchers." The same sentence, worded in active voice would say "The researchers conducted the experiment."
- Spell out numbers from zero to ten. Use the numerical # for numbers 11+.
- Avoid using the words referring to yourself, like "we, our, us, etc". Summarize the article as it is written; you did not have a part in the research.
- Use past tense when referring to the research being conducted.
- Use descriptive and simple words rather than "showy" words. Be as straightforward as possible.
- Use italics for scientific names and capitalize proper nouns.
- Avoid pronouns, especially to begin a sentence. Use the noun itself instead.

Sample Journal Summary

Student Name
Course Name
Date

Occurrence and antibiotic Resistance of Escherichia coli O157:H7 in a Watershed in North-Central Indiana

Keywords: immunomagnetic separation, Polymerase chain reaction, *E. coli* O157:H7

Summary:

Although *Escherichia coli* is most commonly found in animal sources, the bacteria can be a serious problem when waterborne. *E. coli* O157:H7, an enterohemorrhagic strain, produces the shiga toxin and can cause kidney damage. Scientists investigated the presence of *E. coli* O157:H7 in an Indiana watershed. Researchers screened samples for fecal coliform bacteria and O157:H7, and also investigated the antimicrobial resistance levels of bacteria in the watershed. Scientists found *E. coli* O157:H7 in various points along the river, and also found some of these bacterial colonies to be resistant to different antimicrobial drugs.

Researchers collected bacterial samples from three points along the Wildcat Creek and two located in Kokomo, a city that uses the creek as a water source. In order to collect fecal coliform samples, researchers used membrane filtration to separate colonies for counting. Scientists used immunomagnetic separation to detect *E. coli* O157:H7. This separation method is often used to isolate bacterial cells from fluid samples. To confirm the presence of shiga toxins (stx1 and stx2), researchers used Polymerase chain reaction (PCR) to amplify bacterial DNA in order to identify stx1 and stx2 genes. Also, scientists used the Kirby-Bauer method to determine the antimicrobial resistance levels in *E. coli* O157:H7. This experimentation method uses antibiotic discs placed on agar plates to determine resistance using bacterial zones of inhibition.

Scientists found that fecal coliform counts increased as water traveled downstream. Also, 58.7% of water samples contained *E. coli* O157:H7. Researchers found O157:H7 at every test sight, including those in the city of Kokomo. Out of all isolates, tested for shiga toxin presence, 100% of isolates contained the shiga toxin genes stx1 and stx2. Scientists used 21 isolates to test for resistance against eight antimicrobial agents. Researchers found thirteen of twenty-one isolates to be resistant to at least one antibiotic. The highest amount of resistance occurred with tetracycline, kanamycin, and streptomycin. Overall, researchers found that the water present in Wildcat Creek is contaminated with *E. coli* O157:H7 and coliform bacteria, and more must be done to determine pollution sources to prevent further contamination.

References:

Fincher, L., Parker, C., & Chauret, C. (2009). Occurrence and antibiotic Resistance of Escherichia coli O157:H7 in a Watershed in North-Central Indiana. *Journal of Environmental Quailty, 38*, 997-1004. doi:10.2134/jeq2008.0077

This study shows that more must be done to prevent further contamination of Wildcat Creek. The water, testing positive for E. coli *O157:H7 and coliform bacteria, is used by Kokomo's citizens and could potentially cause illness to those who use it. I found this study interesting because most research focuses on* E. coli *in animals, but these bacteria are also a serious problem in water sources around the world.*

Instrument Training
IT 1: Mixing and Weighing
All answers should be recorded in a notebook devoted solely to lab work.

Prior Knowledge

1.1 Define the following terms as they relate to mixing and weighing in the laboratory:

Weigh Boat	Solution
Spatula	Mixture
Magnetic Stir Plate	Tare
Magnetic Stir Bar	Solvent
Homogenous	Solute

Safety Issues

1.2 What safety concerns must be considered as the investigation is conducted? Provide an extensive list.

Prior Skills

1.3 Write a paragraph explaining the proper procedure to mix a solid (powder) into solution with water.

1.4 Write a paragraph explaining the proper procedure to measure the mass of a solid using a digital scale.

1.5 To make a batch of Nutrient Agar for culturing bacteria, 32 grams of dehydrated agar must be mixed to make a liter of media.

a. How much dehydrated agar must be measured to make 500 mL of media?

b. How much dehydrated agar must be measured to make 300 mL of media?

c. How much dehydrated agar must be measured to make 50 mL of media?

Laboratory Protocol

1.6 To make Jello, 3 oz. of dehydrated Jello powder is mixed with 1 cup of water.

a. Calculate the amount of dehydrated Jello powder needed to make 50 milliliters of Jello.

b. After calculating 1.6a, measure the dehydrated Jello powder using a digital scale and hot water using a graduated cylinder. Mix the ingredients using a magnetic stir plate and pour into a paper cup to set. Include a photograph of the completed Jello in your lab notebook.

c. Write a paragraph explaining why hot water is needed to mix the Jello.

Name_____

IT1: Mixing and Weighing

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
1.1 Terms are accurately defined and relate to laboratory use	**4**	**2**	**0**
1.2 At least 4 safety issues are described in detail.	**2**	**1**	**0**
1.3 Paragraph accurately describes proper procedure for mixing a solution.	**3**	**2**	**0**
1.4 Paragraph accurately describes the proper procedure for measuring a solid.	**3**	**2**	**0**
1.5 Agar calculations are accurately provided.	**3**	**2**	**0**
1.6a Jello calculation is accurately provided	**1**		**0**
1.6b Jello is measured and mixed appropriately. A photograph is included.	**2**	**1**	**0**
1.6c Paragraph thoroughly explains the need for hot water in mixing Jello.	**2**	**1**	**0**
Column Total			
Total Score	**/20**		

Comments:

Instrument Training
IT 2: Glassware
All answers should be recorded in a notebook devoted solely to lab work.

Prior Knowledge

2.1 Define the following terms as they relate to glassware in the laboratory:

Adhesion	TC (To Contain)
Cohesion	TD (To Deliver)
Density	Hydrogen Bond
Meniscus	Polarity

2.2 Write a paragraph about the proper procedure for cleaning up broken glass in the laboratory.

Safety Issues

2.3 What safety concerns must be considered as the investigation is conducted? Provide an extensive list.

Prior Skills

2.4 For each of the following pieces of glassware, describe the appropriate use, the location in the laboratory (cabinet # or lab bench) and draw the shape:

Beaker	Florence Flask
Erlenmeyer Flask	Petri Dish
Graduated Cylinder	Media Bottles
Test Tube	

Laboratory Protocol

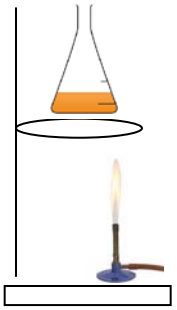
2.5 a. Accurately determine the volume of water able to fit in a test tube. DO NOT measure the volume of the test tube itself.

b. Accurately measure the mass of water that fits into the test tube used above. DO NOT measure the mass of the test tube itself.

c. Now that you know the mass and volume of water in your test tube, determine the density of the water, expressed in grams/milliliters.

d. What is the "theoretical" density of water at room temperature? Why might your calculated density be slightly different?

2.6 Complete the set-up shown below using the proper equipment. Once set up, boil exactly 100 mL of water and record the time to reach a vigorous boil. Include a picture of the set up in your lab book.



** place a heat diffuser (wire mesh) over the ring stand to avoid shattered glass (not shown)*

Name_____

IT2: Glassware

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
2.1 Terms are accurately defined and relate to laboratory use.	**4**	**2**	**0**
2.2 Paragraph accurately describes clean-up procedure used in our lab. Paragraph is written clearly and with proper grammar and spelling.	**3**	**2**	**0**
2.3 At least 4 safety issues are described in detail.	**2**	**1**	**0**
2.4 All equipment is listed with its proper use and location. The glassware is also drawn to the best of the student's ability.	**4**	**2**	**0**
2.5a/b. Volume and mass of water are accurately calculated.	**2**	**1**	**0**
2.5c/d. Density is calculated accurately. Student provides "theoretical" density and thoroughly explains any differences.	**3**	**2**	**0**
2.6 Set-up is proper. Photograph is included. Time of boiling is included. Set-up is disassembled and returned to proper place.	**2**	**1**	**0**
Column Total			
Total Score	**/20**		

Comments:

Instrument Training
IT 3: Diluting Stock Solutions
All answers should be recorded in a notebook devoted solely to lab work.

Prior Knowledge

3.1 Define the following terms as they relate to microscopes in the laboratory:

Stock Solution	Weight-in-Volume
Working Solution	Volume Percent
Dilute	Serial Dilution

3.2 The equation $C_1V_1=C_2V_2$ is used to dilute solutions from higher to lower concentration. Define each of the variables in the equation:

C_1 =	V_1 =
C_2 =	V_2 =

Safety Issues

3.3 What safety concerns must be considered as protocol is conducted?

Prior Skills

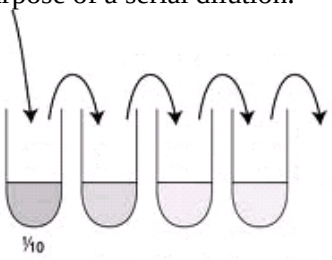
3.4 For a laboratory exercise, you are asked to make 200 milliliters of 8% (w/v) NaCl solution. How much NaCl needs to be added to the water?

3.5 You have 95% ethanol (stock solution) in the laboratory. You need 100mL of 70% ethanol (working solution).

a. How much 95% ethanol should you start with?

b. How much water should you add to the 95% ethanol to make 100mL of 70% ethanol?

3.6 Write a paragraph explaining the purpose of a serial dilution.



Laboratory Protocol

3.7 The goal of this activity will be to accurately measure the mass of NaCl, mix it with a known amount of water, and dilute the solution.

a. Start by measuring 1g of NaCl. Add water until you reach a total volume of 9mL. Add 4 drops of food coloring and mix well. You now have a 1/10 or 1% (w/v) concentration of NaCl. Perform 3 serial dilutions (as shown above). Photograph your serial dilution tubes in a test tube rack and include the photograph in your lab notebook.

b. How many grams of NaCl are present per milliliter in each tube?

3.8 Obtain a bottle of 3% hydrogen peroxide (H_2O_2). Make 10 mL of 1% H_2O_2. Mix the solution with a magnetic stir plate. Have your teacher initial your rubric as confirmation.

Name_____

IT3: Mixing and Separating Solutions

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
3.1 Terms are accurately defined and relate to laboratory use.	**3**	**2**	**0**
3.2 All variables of the equation are accurately defined.	**2**	**1**	**0**
3.3 At least 2 safety issues are described in detail.	**2**	**1**	**0**
3.4 Mass of NaCl is calculated.	**1**		**0**
3.5 Starting volumes of ethanol and water are accurately calculated.	**2**	**1**	**0**
3.6 Paragraph accurately explains the purpose of serial dilution in detail.	**2**	**1**	**0**
3.7a Serial dilution is properly executed and a photograph is included.	**4**	**2**	**0**
3.7b Concentrations of NaCl/mL are accurately calculated.	**2**	**1**	**0**
3.8 1% solution is properly made and rubric is initialed	**2**	**1**	**0**
Column Total			
Total Score	**/20**		

Comments:

Instrument Training
IT 4: Micropipetting
All answers should be recorded in a notebook devoted solely to lab work.

Prior Knowledge

4.1 Define the following terms as they relate to pipetting volume in the laboratory:

liter	aspiration
microliter	precision
micropipette	accuracy
volume	

4.2 Explain the relationship between the units of measurement for volume in the metric system. Be sure to include the terms liter, milliliter, and microliter.

Safety Issues

4.3 What laboratory safety concerns must be considered as the investigation is conducted? Provide an extensive list.

Prior Skills

4.4 Convert the following volumes between metric units.

1 L = _____ mL
1 L = _____µL
50 mL = _____L
5,000 µL = _____L
675 µL = _____mL

4.5 Selecting a micropipette is important for accurately measuring small volumes of liquid. The list below explains the volume range for three common micropipettes:

P20	1 µl to 20 µl
P200	20 µl to 200 µl
P1000	200 µl to 1000 µl

You are asked to pipette several solutions into a microcentrifuge tube in the following order: 750 µL of distilled water, 25 µL of DNA, and 1 µL of restriction enzyme. Write a paragraph describing the sequence of events necessary to transfer these liquids without cross-contamination.

Laboratory Protocol

4.6 Obtain a small piece of Parafilm, a P20 micropipette with tips, and a small microcentrifuge tube with water. Using proper technique, pipette drops of water along the Parafilm in the following volumes. Then, compare with a friend to ensure that the proper volume has been pipetted and photograph. Include the photograph, labeled with a title and caption, in your notebook.

20 µl, 15 µl, 10 µl, 5µl, 2µl

4.7 Label a clear microcentrifuge tube with your initials and #1. Transfer 1 mL of distilled water, 2.0 µl blue food coloring, 25 µl yellow food coloring and 3.0 µl red food coloring into the tube. Cap the tube, microcentrifuge to mix, and photograph your tube. Include a photograph, labeled with a title and caption, in your notebook.

Name_____

IT4: Micropipetting

	Excellent	**Satisfactory**	**Unsatisfactory/ Absent**
4.1 Terms are accurately defined and relate to laboratory use.	**3**	**2**	**0**
4.2 Paragraph accurately describes relationship between units of measurement. Paragraph is written clearly and with proper grammar and spelling.	**3**	**2**	**0**
4.3 At least 2 safety issues are described in detail.	**2**	**1**	**0**
4.4 Unit conversions are accurately calculated.	**4**	**2**	**0**
4.5 Paragraph accurately describes how to pipette samples, including selection of pipettes and changing of tips. Paragraph is written clearly and with proper grammar and spelling.	**4**	**2**	**0**
4.6 Water droplets are pipetted on Parafilm and reflect correct volumes. Photograph is included with title and captions	**2**	**1**	**0**
4.7 Photograph is included. Color matches prepared sample. Picture is included with detailed title and caption explaining the contents of the photograph.	**2**	**1**	**0**
Column Total			
Total Score	**/20**		

Comments: