The GENIQUEST (GEN*omics* I*nquiry through* QU*antitative Trait Loci* E*xploration with* S*AIL* T*echnology)* Project: Bringing STEM Data to High School Classrooms

Introduction

The public currently has unprecedented access to vast collections of high-quality scientific data, providing a tremendous opportunity for K-12 educators and students. As the national cyberinfrastructure expands, stores of these data and the potential for their use can be expected to only grow over time. New scientific findings will also depend increasingly upon these data stores. (Bell 2004, NSF 2003). The opportunity for creating applications allowing K-12 educators to access and utilize resources from the scientific community has been largely under-realized. While scientific data and discovery are rapidly growing, the K-12 education materials community continues to revise curriculum and instruction materials through traditional means that struggle to keep pace. Revisions to this curriculum rarely take into consideration the newest scientific findings, and the complexity of the interfaces used by professional scientists and translation of the data to the secondary school level only act to compound the difficulty of increasing educators' access to cutting-edge science (Bell, 2004). The potential for using the existing body of scientific data in education remains largely untapped.

The field of computational biology is one area of cutting-edge science that has responded particularly strikingly to the development of the national cyberinfrastructure. A large portion of the field has developed within the lifetimes of most of its scientists, as well as of today's K-12 students. Despite being a young field, computational biology shows no signs of diminishing in significance. Instead, the developing national cyberinfrastructure is accelerating the growth and importance of this field of study and of the public data stores that fuel it. (NSF, 2003) In light of this shift, many biologists and educators have called for education to place a greater emphasis at all levels on the principles underlying computational biology. Well developed skills of scientific investigation and data manipulation will be vital to prepare today's students properly for later work in this and other scientific fields.

The GENIQUEST Project (GEN*omics* I*nquiry through* QU*antitative Trait Loci* E*xploration with* S*AIL* T*echnology*): Bringing STEM Data to High School Classrooms captures cutting-edge STEM data with research-based educational software frameworks. The development, exploration and trial of the GENIQUEST interface is designed test and learn about the potential for an application for accessing and utilizing current scientific data sets so it can be applied to other data sets. The GENIQUEST project aims to bring current digital science data and effective science teaching to secondary science students. The focus of this cyberinfrastructure project is the development of an application enabling students and teachers to investigate biological data sets using a research-based instructional model. It will integrate the publicly shared data set from the Jackson Laboratories, powerful analysis tools from and innovative approaches in science instruction to build a biology computing environment spurring student investigation and inquiry. In doing so, the GENIQUEST project will help the students of today develop skills necessary to succeed in the laboratories of tomorrow.

The Maine Mathematics and Science Alliance (MMSA), the Concord Consortium and Jackson Laboratories are partnered to conduct the *GENIQUEST* project. This is an Exploratory project responding to the *Discovery Research K-12* program solicitation *NSF 06 593*, Category B. *Development of Resources and Tools*, and subcategory 2; *Instruction of K-12 Student and Teachers.* The project directly addresses the issues in presented in NSF Grand Challenge 3: *Cutting- Edge STEM Content in K-12 Classrooms*.

Goals and Objectives

The GENIQUEST project directly addresses Grand Challenge 3 by bringing cutting-edge scientific data and techniques to the high school classroom in an approachable manner. The project's long-range goal is to improve students' understanding of science, scientific research, and the use of evidence in reaching scientific conclusions.

The Exploratory Project goals for GENIQUEST are to:
1. Develop and test a tool enabling students to import, manipulate and analyze genomics data.

2. Increase secondary students' knowledge and skills at data analysis and scientific research.
3. Improve secondary school teachers and students' access to cutting-edge STEM data sets.
4. Determine the viability of online data as a tool for engaging secondary school teachers and students in scientific inquiry.

As a result of this project, students will be able to:
- Understand basic principles of quantitative inheritance and QTL analysis
- Perform basic steps for analysis of QTL genomics data
- Choose appropriate methods for visualizing and analyzing QTL datasets
- Generate questions raised by data analysis
- Further investigate these questions by exploring data or analyzing new datasets
- Collaborate with other students locally or remotely on data analysis
- Report results of their analyses to a teacher, class or larger audience

As a result of this project, teachers will be able to use this tool in their classrooms to:
- Expose students to cutting-edge genomics concepts and data
- Introduce and improve student data investigation and analysis skills
- Guide students through the process of scientific inquiry
- Develop custom data investigations tailored to specific students or courses

The GENIQUEST project will use the expertise of three institutions, the Principal Investigators, and the organizational staff to increase teacher and student learning about genomics and research while testing an online application that may have more generalizable applications to other online scientific data sets. Although the three partners have had experience working with each other separately, they have not all worked together before, making this collaboration between scientists, educators and students a unique opportunity to combine their expertise in a way that will benefit the institutions themselves, teachers and students.

Maine provides an excellent location for a test bed for an exploratory technology project like GENIQUEST. Maine is geographically large and rural, but nonetheless possesses high quality classroom access to technology and research expertise. Every 7th and 8th grade student in the state and many high school students have an individual laptop provided by the Maine Department of Education. This geographic distribution of classrooms with a high quality access to the internet through schools and libraries sets a useful stage for the piloting of electronic educational materials that can later be expanded to educators throughout the nation.

Rationale

With the rapid development of a national science and engineering cyberinfrastructure, the increasing ease of access to online data can facilitate its exploration by students, teachers and researchers in a variety of learning and research environments (Zia, 2001; Manduca, et al., 2001). Although scientific data are increasingly accessible, teachers and students alike generally lack the capacity to ask testable questions of it (Manduca & Mock, 2002). Secondary educators require assistance in investigating age-appropriate questions related to data and need tools that provide easy access to cutting-edge data sets from science, technology, engineering and mathematics (STEM) research. (Hanson & Carlson, 2006) Because the interfaces used by scientists to access and manipulate these data are generally too complex for students, learner-centered design issues need to be taken into account when creating interfaces to assist students investigating authentic data (Bell, 2004).

Technological approaches to biology have grown and flourished. Today, more than ever before, biomedical scientists are challenged to adopt advanced quantitative and computational methods. Computers are enabling researchers to improve data quality and laboratory efficiency, extend their ability to probe and model complex biological phenomena and enact or adjust to fundamental changes in the conduct of science. (Morris, Bean, Farber, Gallahan, Jakobsson, Liu, Lyster, Peng, Roberts, Twery, Whitmarsh, Skinner, 2005) Early efforts toward technology-based learning in schools treated computers

as yet another medium for displaying information. However, effective science learning needs to move to judging student understanding of concepts and processes and their ability to scientifically model or mathematize problem situations (Romberg, Carpenter & Kwako, 2005). New technology learning environments are emerging today that improve greatly upon their predecessors. Data sets, accompanying scientific publications, can serve as a valuable resource for education. Not only do students have an opportunity to work with different kinds of authentic data, the activity of deriving their own analyses and comparing the results to a published, peer-reviewed study can be a powerful tool for learning (Morris, et. al., 2005). The GENIQUEST application will acknowledge the existence of the varying levels of learning sophistication, expertise and knowledge involved in data investigation. This project will provide students access to data while building their capacity to understand and conduct scientific inquiry. In order to build this capacity for students, the application will support students with the use of embedded instructional scaffolds.

Using educative scaffolds addresses one of the main challenges by Branford, Brown and Cockling in, *How People Learn*: "The challenge for education is to design technologies for learning that draw both from knowledge about human cognition and from practical applications of how technology can facilitate complex tasks in the workplace. These designs use technologies to scaffold thinking and activity, much as training wheels allow young bike riders to practice cycling when they would fall without support" (2000, p. 214). Bransford *et al.* specifically identify scaffolding as an effective method for use in educational technology: "Scaffolding is suggested as one of five ways that technology can help establish effective learning environments." (2000. p. 243) In particular, scaffolding disciplinary explanation of complex data has been recommended to assist students in actively developing both conceptual understanding and understanding of data analysis processes (Bell, 2004).

This online application will allow teachers and students access to the same tools and techniques used by working scientists. As it introduces them to these tools and techniques, the application will guide them through the stages of scientific inquiry, helping them consider how to display their data and investigate relationships, then generate and explore questions raised by the data. This scaffolding will provide students with appropriate challenges in their investigation while supporting them in practicing new skills. Continued experience with the tool will lead students continually toward more open inquiry by building and reinforcing their scientific research skills.

The use of scaffolded student inquiry in software environments is well precedented and historically successful. For example; in the BGuILE project, students develop disciplinary-focused explanations of a culled data set about the micro-evolution of Galapagos finches (Reiser, 2001). The research demonstrated the necessity to scaffold student's data analysis processes (*e.g.*, students needed assistance in using interpretations of instances of data to draw claims and formulate hypotheses for subsequent interrogation). Other work has demonstrated that embedded scaffolds can be successful in helping students conduct more careful observations and analyses (Golan et al., 2001) and that pairing these scaffolds with proper teacher-student interactions is effective in supporting student learning (Tabak, Gurion & Baumgartner, 2004).

Using this technology to adapt tools from the scientific community for use in secondary classrooms brings potential benefits not just for the GENIQUEST Project, but also for additional projects attempting to address challenges such as NSF Grand Challenge 3. The products developed and lessons learned as a result of this exploratory project will provide valuable insight and reusable elements for future projects. Be demonstrating the development of network-enabled, rich and complex curricula that combine the power of real applications with the collaborative aspects of the web while tracking learner data and artifacts for future use, the GENIQUEST work will provide an open-source prototype for similar work in other content areas.

Content and Standards
The GENIQUEST project covers the science content - genomics, the nature of science - and mathematics content - Data Analysis and Probability, Problem Solving, Communication, Connections and

Problem Solving. These topics are represented in the National Science Education Standards (NSES) (NRC, 1996) and the NCTM Principals and Standards (NCTM, 2000) (Table 1).

Table 1: Examples Standards from NSES for Heredity, Nature of Science and NCTM Mathematics.

NSES, Grades 9-12	NSES, Grades 9-12
Heredity ➢ In all organisms, the instructions for specifying characteristics of organism are carried in DNA. ➢ Two copies of each chromosome, and therefore two copies of each gene, explain many feature of human heredity, such as variations that are hidden in one generation can be expressed in the next. ➢ The differentiation of cells is regulated through the expression of genes.	*Nature of Science* ➢ Identify questions and concepts that guide scientific investigations. ➢ Design and conduct scientific investigations. ➢ Recognize and analyze alternative explanations and models. ➢ Mathematics is essential in scientific inquiry. ➢ Scientific explanations must adhere to criteria such as; a proposed explanation must be logically consistent, it must abide by the rules of evidence; it must be open to questions and possible modification, and it must be based on historical and current scientific knowledge.
NCTM Standards, Grades 9-12	
Data Analysis and Probability ➢ Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them ➢ Select and use appropriate statistical methods to analyze data ➢ Develop and evaluate inferences and predictions that are based on data ➢ Understand and apply basic concepts of probability *Connections* ➢ recognize and use connections among mathematical ideas; ➢ recognize and apply mathematics in contexts outside of mathematics.	*Communication* ➢ organize and consolidate their mathematical thinking through communication; ➢ communicate their mathematical thinking coherently and clearly to peers, teachers, and others; ➢ analyze and evaluate the mathematical thinking and strategies of others; *Problem Solving* ➢ build new mathematical knowledge through problem solving; ➢ solve problems that arise in mathematics and in other contexts; ➢ monitor and reflect on the process of mathematical problem solving.

GENIQUEST is not designed to provide the foundation for knowledge about heredity, but instead to move teachers and students into exploration of real data concerning related advanced topics. Before students begin to explore current genetic research, they need to have a firm grasp of both Mendelian and non-Mendelian heredity. There already exists a range of options available to classroom teachers for introducing these ideas including classic textbooks at grades 9-12 and supplemental instructional materials such as, *Genes, Environment and Human Behavior.* Additional materials include; *Bioinformatics and the Human Genome Project* from BSCS, the *Genetic Education* modules for teachers from NIH, and Internet-based materials such as those from the Cold Spring Harbor Laboratory, www.cshl.org.

This project will address and extend learning beyond state and national standards in order to enable teachers and students to see how current science investigation is conducted. The extension helps address recommendations of the Advanced Placement (AP) Biology course syllabi. The MMSA conducted an analysis of these AP materials for the Maine Department of Education and found 25% of the current Advanced Placement Biology course has an emphasis on Heredity, Molecular Genetics and Evolutionary Biology (Maine Mathematics and Science Alliance, 2006).

One important aspect of the GENIQUEST project is its focus on the nature of science. The topics in this domain, the ability for teachers and students to think logically, construct and test questions, analyze and interpret data and use evidence to explain their findings, are not a part of most teachers' training. As a result, most secondary students receive little exposure to these processes. Scientific research involves these processes and skills so this content will be integrated into activities for the teachers and students. For example, in an activity focusing on examining data, the application will first lead students to consider

the source of the data, what the data might tell them and what questions/hypotheses they might have about the data.

The project will make use of techniques and data related to the determination of quantitative trait loci (QTL), a field of biology involved with determining the genetic basis for inherited traits that are determined not by one gene, but by the interactions of two or more of an individual's genes. The study of QTL is of significant interest to the biomedical field, as a great many diseases are inherited through this means. Use of this topic allows the project to integrate authentic scientific data with an emphasis on student inquiry, a focus that holds great potential for increasing relevance and student engagement and the possibility for original discovery. In the process, GENIQUEST will also help teachers learn about developing and testing questions and hypotheses. The link between ongoing scientific research and secondary school content through a scaffolded set of activities could be applied to a wide variety of data sets for secondary school educators.

Pedagogical Focus and Format

Data investigation is not often attempted in secondary science classrooms, and authentic investigation involving scientific processes is similarly infrequent. Gaps in these areas often arise because teachers lack the instructional strategies to approach investigating data by students. It can also be overwhelming, especially investigation of the large or complex data sets produced in cutting-edge STEM research. The GENIQUEST application will build students' and teachers' ability to access the data and their capacity for asking questions of real scientific data by embedding support for data investigation techniques. The application will provide varying levels of complexity and scientific inquiry. With these varying levels of complexity and support, the application can range from providing structured tutorial modules to serving as an open platform for importing and investigating data.

The lessons provided in the application will use the BSCS 5E instructional model (Bybee, et. al, 2006) as their foundational pedagogy, emphasizing the engagement, exploration, and explanation by the learner and providing opportunities for elaboration of their acquired knowledge and skills and evaluation of student learning. Instructional modules within the application guide students through data investigation according to the 5E model, and provide graduated levels of challenge to match students' skills with specific analysis tools. Table 2 outlines the various aspects of the framework and illustrates the connections among the parts of the GENIQUEST learning process.

Table 2: GENIQUEST Guiding Framework

Pedagogical Level	Purpose	Content
Introduction	To introduce students to specific science background knowledge and analysis tools	Related genomics instructional content, simplified datasets
Level I Inquiry with Selected Datasets	To allow students to practice techniques and reinforce learning from the Introduction level.	Expanded datasets similar to those from Introduction level
Level II Guided Inquiry	To provide opportunities to learn and practice scientific investigation skills with authentic datasets	Datasets that have yielded prior scientific publications
Level III: Student-led research work	To hone student investigation skills and aid student collaboration within and between classrooms	Recommended or student-selected public genomics data

There are four levels to the scaffolding of student inquiry within GENIQUEST. An introductory level will provide instructional materials for background content knowledge and will guide students through initial use of the available genomics analysis tools. Levels I, II and III of the application will guide students through the analysis of various genomics datasets and the steps of QTL analysis. Each of the three levels will move students one step closer toward using the application for open inquiry, in which they will select datasets and perform analysis largely independent of prompted assistance.

As shown in Table 2, these various levels of the application will have different, but related aims. Just as the lessons within them reflect the 5E instructional model, the experiences of students proceeding through the levels as a whole will evoke the same stages. The Introductory Level will focus on fundamental content, engage students with an initial scenario and guide students through exploration of

its various tools as they investigate the scenario. In Level I, students will work with an entire dataset, explaining and justifying their actions throughout the process. At the conclusion of this level, the application and their teacher will provide connections to scientific concepts as they move toward an extension of their learning in Level II to a dataset that has yielded previously published results. Finally, in Level III, students will push the elaboration of their learning to its natural conclusion as they investigate datasets using the skills they have honed in the previous levels. Student performance at scientific investigation in this level and the end of the previous level will demonstrate to them and their teacher that they have gained the proficiency needed for this open inquiry.

Anticipated Products

The GENIQUEST project will consist of a website and software package combination and selected supplementary online instructional materials. The supplementary online instructional materials will be adapted from existing, previously piloted materials used by Jackson Laboratories in a course developed jointly with the Maine School of Science and Mathematics. These materials and the archived course materials from the Jackson Laboratory will serve as an optional precursor to the classroom use of the software application. The materials will supply lessons bridging the gap between basic secondary curriculum about genetics and heredity and the more advanced background in inheritance of quantitative traits necessary for understanding the QTL analysis process.

The prerequisite knowledge for student use of the software package is within the scope of advanced secondary biology courses such as Human Physiology, Advanced Placement biology or student research courses. Some teachers may wish to use the supplementary materials for review or reinforcement of the underlying concepts, as a bridge to the use of the software, or within a unit of study involving use of the software package. While the project will include the adaptation of these materials to best assist its goals, the GENIQUEST project goals do not include design of curriculum sequence or units of study surrounding these supplementary materials. There are several high quality web sites and printed materials that provide a better core curriculum for genetics.

The principal product of the GENIQUEST project will be a software package and associated website to prepare students for QTL analysis and guide them in its steps. The software package will contain the tools necessary to import and analyze QTL data, the pedagogical framework necessary to guide students through the analysis process and the collaboration and networking ability to connect students with their teachers and each other.

The software package will ground its analysis power on tools designed at the Jackson Laboratories and in current use in research laboratories worldwide. In the various stages of a student's path through the software application, these tools will be provided selectively and bear a limited feature set. As students become more proficient with the use of the tools in QTL analysis, additional analysis capabilities will be made available. In the final stage of analysis, students will have access to the majority of the tools and features available to scientists performing similar analyses, with the tools presented in a familiar and approachable format.

The four-level pedagogical project framework described above will form the core of the software package and will be responsible for achieving the majority of the project's goals. This underlying framework will help students learn the fundamental scientific content behind QTL analysis within the context of the analysis itself. Guidance for students throughout all stages will be framed in the process of scientific inquiry.

Student Investigation with the Project Software

Students initially using the software will be presented with a scenario such as the condition of hypertension in humans and the difficulty in understanding its hereditary basis. Students will first complete a brief tutorial presenting the basic scientific principles behind an experiment designed to search for the condition's genetic determination. This tutorial will engage students in the concept and identify the task to be undertaken. Students will then load and explore a sample data set from an experiment designed to target the question presented in the tutorial.

Introduction Level

Students will initially experience the tools in the context of exploration, as they examine the data and begin to understand how each tool can help produce a solution to the initial research question. After importing the data, students will first examine its genotype plot (Figure 1). The tutorial will guide students through the tool, posing questions relating the plot and tool features to the underlying science presented in the introduction. Small tasks will introduce students to the features of the plot while connecting them more closely with the experimental process at hand. Similar tutorials will guide students through exploration their data using phenotype plots and the genetic map viewer.

In a second segment of the tutorial, students will explore the available analysis tools and learn their connection to the sample research question. Students will perform a one-QTL genome scan for the sample data set and explore the resulting plot and scan summary. As they do so, students will be prompted to match the scan results to their basic scientific underpinnings and describe how the plot results relate to the original research question. The guide will assist students in identifying relevant results and use these as a transition to a two-QTL genome scan. Students in this final stage of exploration will view the results of the scan and select relevant interactions from the scan summary. Additional guidance from the application will assist students in their exploration and selection.
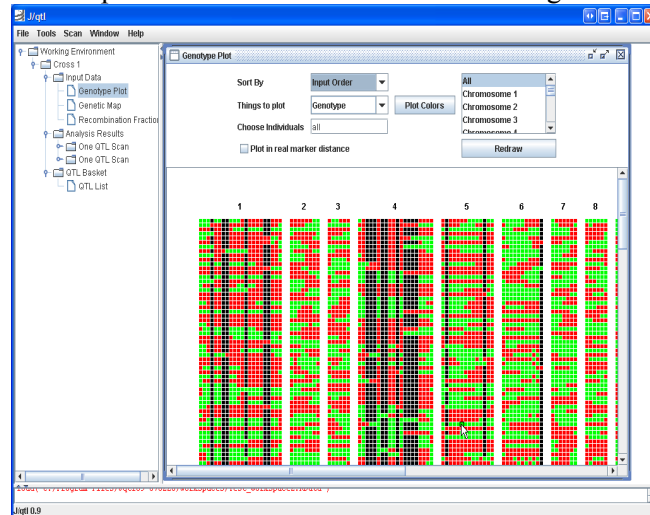

Figure 1: Genotype Plot of imported data

After the tool orientation phase, the application will highlight the scientific underpinnings of the various stages of their sample analysis and provide content-related connections and explanations of the steps students have undergone. This explanation phase of the orientation will contain personalized responses, so that the application links explanations and descriptions for a given student to the student's selections and responses in previous sections of the tutorial.

Level I: Inquiry with Pre-selected Datasets

Upon completion of the explanation and recap, students will enter the application's pedagogical Level I (Example screen shot Figure 2). At this level of investigation, students will perform an analysis of a pre-selected dataset similar to the tutorial set. This level will focus on helping students transfer their learning from the tutorial to a new, but closely related analysis scenario. In order to scaffold support of students' scientific investigation skills, the application will provide less initial
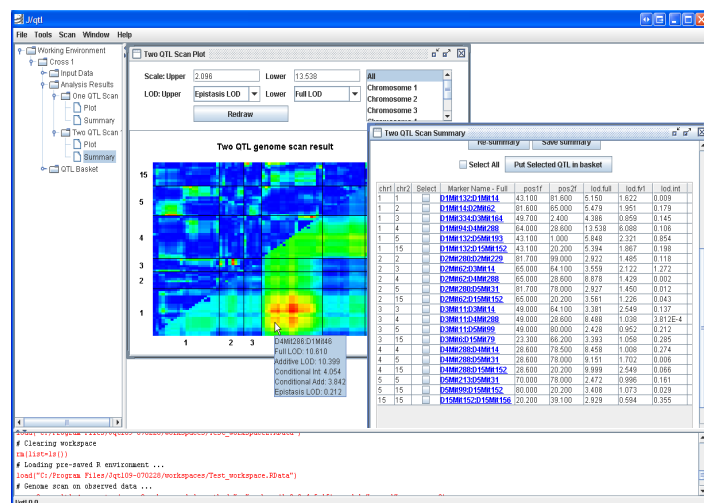

Figure 2: Level I investigation with collaboration forum

guidance during this phase than during the former phase, allowing students to exercise skills learned in the introductory phase.

The application will monitor student actions and progress during this phase, offering assistance as student actions warrant. Assistance may take one of several forms. The application may provide additional prompts and questions to students and highlight salient features of the data or analysis tools in the process. The application may also search for other students in the class or remote classes who are at a similar point in the analysis process for these data and connect these students via its collaboration tools. In other cases, the application may notify the student's teacher that the student is in need of assistance.

As students complete this phase of analysis, the application may ask students for descriptions of their observations and prompt them to explain and justify their actions in the data analysis. The application will summarize the student's work and present the student's teacher with a categorized overview of the student's progress. The student responses and teacher summaries will provide evaluation of individual students' understanding and progress. Teacher responses to these student summaries will be provided to the student and serve as important feedback for the student's use of the application in the final two levels.

Level II: Guided Inquiry

The penultimate phase of the application will provide students with scientific data that has led to published scientific analyses. Students will analyze these same data themselves before being able to view the published results, with the goal of testing the investigation skills they have developed against those applied by working scientists. The supplied datasets will derive from Jackson Laboratories' QTL archive, and will be selected to provide an appropriate mix of approachability and challenge. These data may be minimally pre-processed to eliminate extraneous subtleties or constrain the solution set somewhat while duplicating as closely as possible the conditions of the original data analysis.

Datasets for students in this phase will be chosen from a small subset of candidate sets, so that a classroom of students will work on a variety of datasets with enough overlap to enable student collaborations to develop. Depending upon students' prior performance and demonstrated strengths, the application will provide the teacher suggestions of appropriate datasets for various students or for potential student groupings within a class. Students exploring and analyzing these datasets will have all previously explored tools at their disposal.

Prompting from the application will take on a secondary role during this phase, but will assist students in recalling the main phases of the data analysis. As desired, students may continue to call up hints and assistance at any time during the analysis. The application will monitor student actions during this phase to provide context-dependent assistance to students and student progress summaries to teachers.

As students complete their analyses, they will compare their work with those derived by scientists from the same datasets. For students who have not arrived at similar results, the application may prompt students to return to the analysis process and re-examine the data. Students completing this phase may have discovered connections not highlighted in the published results or may have additional questions they are interested in exploring. These questions and connections form the basis for the final, open analysis phase of the application.

Level III: Student-Led Research

In the final, open analysis level, students are free to explore their chosen dataset as desired, examining connections or correlations with full use of the analysis tools. In this final stage of the process, prompting scaffolds will be infrequent, although students may continue to access them as desired. For example, students might apply the tools they acquired to real world data sets looking for possible genetic links to disease by analyzing phenotype and genotype data from a variety of mouse strains.

The collaboration tools from the application take on a greater prominence for students in this phase. Students at this level of inquiry will examine new datasets, published or non-published, and compare their findings with those from other students. The distributed nature of the application framework will permit

collaboration between groups of any students using the application, so that students may find themselves analyzing data with students in other classrooms or even other schools.

The application's website bridge will facilitate the development of these collaborations by allowing students to "advertise" their work on a particular dataset, pairing students together and providing a means for students to publish findings to the audience of other students using the application. This central publication of findings will provide ownership and incentive for students conducting the analysis as well as enabling students to engage in peer review of other students' findings. Given the massive proliferation of genomics data, students with advanced skill and time may indeed uncover original findings of interest to working computational biologists. Biologists may also be able to identify promising students through examination of posted "results" and supply constrained datasets to these students for initial analysis.

Architecture and Development

The technical resources involved in this project satisfy these requirements well. The tool will owe much robustness and flexibility to its foundation on Concord Consortium's integration of its OTrunk Java component framework with the Scalable Architecture for Interactive Learning (SAIL). SAIL, an outgrowth of the NSF-funded Web-based Inquiry in Science Environment (Linn and Slotta, 2000; Slotta, 2001, 2004), has been developed by the NSF center for Technology Enhanced Learning in Science (TELS) as an optimal environment for the development of inquiry learning investigation (Slotta & Aleahmad, in press). SAIL represents the product of over a decade of prior research into technology-enhanced student learning. (Slotta and Linn, 2000; Horwitz and Christie, 2000). Technologists at the Concord Consortium and U.C. Berkeley have been leading SAIL development for the past several years.

Two key ideas form the basis for SAIL. One is an assembly architecture to knit together reusable pedagogically-aware Java components into curricular activities. These rich components already include:

- Computational models with rich visual representations. These include among others molecular, dynamics and biological models.
- Graphs for displaying both real-time and saved data.
- Sensor collection components for collecting and graphing real-time data from sensors as well as analyzing previously-collected data.
- Drawing tools which can range from a simple bitmapped painting, to object drawing, to concept mapping.
- Models written in general purpose modeling languages such as NetLogo.
- Assessments ranging from multiple-choice to open-response text input.
- Components which can render web content ranging from html, css, to flash and quicktime. While browsers are very capable at this task there are many times in which web content may need to be delivered in a more constrained environment which does not necessarily allow browsing to other sites. The integration of the many forms of web content and interaction with the more powerful modeling and analysis tools that are available in Java supports deeper learner exploration and inquiry and the creation of both richer explicit and implicit learner artifacts.

The second key idea is that each of these components is provided by SAIL a network-enabled pedagogically-aware persistence service which lets them load and save learner data. The underlying SAIL architecture takes care of storing a complete revision history of what has been saved and also makes sure that the data is associated with the correct student, workgroup, class, and teacher. This persistence is supported by the core SAIL framework that is included with the client application and the SAIL Data Service (SDS) web service. The SDS is designed to integrate with existing web portals to allow them to easily deliver SAIL-based activities to their learners, persisting the learner data and reporting back to the main portal. At this time the SDS is supporting the TELS WISE portal as well as Concord Consortium's TEEMSS2 Do It Yourself portal, two completely different portals with different underlying architectures that integrate with SAIL and the SDS to author and deploy SAIL-based curricula.

Concord Consortium has been the technical lead on the SAIL and SDS persistence integration as well as the author of many of the modeling components. Concord has also developed a scripting environment

and framework called Pedagogica that supports dynamic adaptation of component presentation and interaction to learners based on learner actions and data. Recently Concord Consortium has also integrated their OTrunk framework into SAIL, supporting the creation and modification of rich Java component-based interactive content with persistence of the state of both author and learner content. OTrunk uses SAIL's network-enabled pedagogic data persistence and in turn provides SAIL the flexibility and agility associated with its extensible XML-based declarative language. Concord Consortium has also taken the lead on adapting outside Java-based components and tools such as NetLogo or PhET's Circuit Construction Kit into SAIL, OTrunk, and Pedagogica.

The work Concord Consortium and the TELS team have been doing integrating SAIL, OTrunk, and Pedagogica architectures is designed to be easily scalable and accessible to teachers and developers alike. Its extensibility is intended to provide for dynamic evolution and its open source design provides a ready response to the sustainability issues plaguing many software learning environments. By basing its applications on these frameworks, the GENIQUEST project ensures that its products can respond to the rapid evolution of cutting-edge computational biology and be adapted easily to reflect advances in technology and understanding about student learning. The open source nature of the GENIQUEST project will also allow future developers to expand upon its achievements while retaining the value of the project's core work. The rapid expansion of the field of computational biology and increasing interest in bioinformatics education make this project's open expandability especially valuable.

The curricula for this project will be authored in the Concord Consortium's integration of SAIL and OTrunk. These frameworks will allow students using the application to carry their data across multiple work sessions and continue to work with their investigations when offline. The framework will also permit students to collaborate by sharing datasets and work within a classroom or with other classrooms at remote locations.

Concord Consortium will integrate the Jackson Laboratories' Java application "J/qtl (Java for Quantitative Trait Loci)" into the SAIL/OTrunk authoring, deployment and reporting system. This involves four elements of work:

1. Extending the SAIL leaner-data persistence architecture to support integration of components that require read-write access to a file system-like collection of data
2. Integrating the J/qtl (and its associated R-based scripts and system) into OTrunk and performing any necessary modification to support effective learner-data persistence
3. Supporting the creation of scripted report generation tools within J/qtl to produce retrievable data about learner work with the materials
4. Developing and maintaining an authoring, deployment, and reporting portal to support students and teachers using the application

The J/qtl application was developed by Jackson Laboratories and is in use by scientists worldwide for QTL analysis. Its use in the project will open the power of true professional-level scientific analysis software for student use.

All of the underlying software frameworks for the project are available under an open-source license that allows integration with both open-source and closed-source codebases. This allows the GENIQUEST project to integrate the Java-based J/qtl tools easily with the Concord Consortium's frameworks to many of the powerful features of this software available to students. Because the SAIL architecture demands only a simple, automated installation no user configuration and because the application's framework will be Java-based, the resulting application will be ideal for use across all platforms and any variety of classroom configurations.

The application framework will also create exceptional potential for learning and educational research. The combination of the SAIL/OTrunk frameworks with Pedagogica-style scripting will permit the collection of large amounts of fine-grained data about student actions. In addition, the system supports analysis of these data to produce and integrate more semantically meaningful reports on these data. As the

authors and researchers find more ways of extracting pedagogically-significant reports from these data, these reports can be included into the activities themselves for use by both teachers and students.

Work Plan

The project materials will be piloted and field-tested in secondary classrooms in Maine and Massachusetts. The target population for GENIQUEST is secondary teachers and students with a significant background in biology. It is likely that Advanced Placement programs or existing courses in Biology and research courses in target schools will provide a portion of the user base for the project application, especially as many texts traditionally used for such courses do not contain in-depth or up-to-date information about the genomics content or techniques addressed by the application. The identification of specific student and teacher user base will be part of the proposed work and documented in the project evaluation.

The project will test the developed materials in 8 classrooms (5 in ME and 3 in MA), beginning with 4 pilot classrooms during the first year of application development and extending to a field test in 8 classrooms the following year. Pilot and field test teachers will receive 1½ days of training on the use of the materials prior to implementing the application in the classroom. Teachers will then use the application with their students and provide feedback in an additional one-day feedback session between project years one and two. Two classroom visits to each teacher during each project year will provide information about student use of the application. These observations will be combined with teacher feedback to inform application revisions.

Because of the monitoring capabilities provided by the SAIL framework, additional information on student use of the project software will be provided continuously during the implementation in secondary classrooms. The project software will collect selected information about student actions during its use, from logs of student progress through the software tools to descriptions of time intervals separating mouse clicks and window openings. These data together with student responses to application prompts will form a comprehensive picture of student activity within the project application. This information will be used as a secondary source of pilot and field test information, and will directly inform revisions of the project software. After development of the software is complete, the collection capabilities can be used to conduct educational research on student use of the tool.

Table 3: Project Timeline of Activities

Year One
Sept.- Oct. 2007: PIs meet for project organization and establish meetings and scope of work. Partners review existing Jackson Laboratory content/software and create development schedule.
Oct.-Jan 2007: Lesson outlines created and parameters for adapted tool interfaces determined. Tools adapted to fit into OTrunk/SAIL architecture. Lesson infrastructure programming begun. Optional precursor classroom materials selected and adapted.
Feb.-Apr. 2007: Pilot teachers recruited and trained. Beta version of application readied.
May-2007: Application piloted in 4 classrooms. Pilot visits and interviews conducted. Revision based on pilot data.
June-Aug. 2007: Workshop held with 8 field test teachers. Feedback on software and classroom materials gathered. Software revisions prepared for field test trials.

Year Two
Sept.- Oct. 2007: Content-based revisions to lessons completed. Dissemination begun.
Oct.-Jan 2007: Associated website completed. Application field tested in 8 classrooms in ME and MA. Field test classrooms visited and results collected.
Feb.-Apr. 2007: Final field testing complete. Application revised based on field test data.
May-2007: Feedback workshop held with field test teachers.
June-Aug. 2007: Application revised based on teacher field test feedback. Project disseminated at conferences.

The development scheme for the software itself will assist the revision process. Because the software will be built using agile development frameworks, information collected from the various feedback mechanisms can be turned into software improvements without the need for extended revision and release

schedules. This agile development philosophy will assist developers in making the most of a short development cycle.

Partners and Project Staff

The Maine Mathematics and Science Alliance (MMSA) is a private, non-profit organization that provides professional development, teacher materials, digital resources and district assistance for the improvement of science and mathematics education. The MMSA focuses its work on the implementation of high quality content, instruction, professional development, and instructional material development for improving student achievement in science and mathematics. Some recent examples of digital materials development include: 1) the Phenomena and Representations for Instruction of Science in Middle Schools (PRISMS) project, an NSF funded National Science Digital Library project that focuses on analyzing digital resources. (An overview of the PRISMS project can be found at the Project 2061 web site, *http://www.project2061.org/publications/2061Connections/2005/2005-01b.htm* and at the MMSA web site; *http://www.mmsa.org/science/science_NSF_projects.php#nsdl*), 2) three units for GLOBE at *www.globe.gov* that are available online, 4) *www.nasalearn.org*, (instructional materials and standards aligned from NASA materials) and 5. *www.maptasks.org* (Assessment materials for district wide use). The MMSA partners with Maine schools, research laboratories, and universities to accomplish its goals. Specifically for GENIQUEST, the MMSA will provide the research based instructional model, alignment of content to standards, links to research on student learning, pilot sites support and professional development of the GENIQUEST project. Three people will be involved in this project; Dr. Francis Eberle as Principal Investigator and Executive Director of MMSA, will provide oversight and management support to the project. Chad Dorsey, Project Manager and Science Associate at MMSA, will provide the day to day management and support for the project, and Brianne VanDeBossche will serve as a project assistant.

The Concord Consortium (*www.concord.org*) is a non-profit research and development organization dedicated to realizing the educational potential of information technologies. Concord has developed the core technology used in this project through a series of NSF-funded projects that support guided student learning through inquiry-based explorations using probes and sensors, sophisticated models, and software tools. Concord's combination of expertise and resources has resulted in the development and dissemination of dynamic and successful educational tools nationwide. The Concord Consortium's list of at least nine current national projects include the Calipers Project, in collaboration with SRI International, and the Molecular Literacy Project, developing new secondary materials to support interest in biotechnology careers.

In the fall of 2003, the Concord Consortium co-founded the Technology Enhanced Learning in Science (TELS) Center (www.telscenter.org), a Teaching and Learning Center funded by the National Science Foundation. By providing major funding for an Education Accelerator, TELS supports applied research on the educational impacts on science of information and computer technologies. SAIL, the principal architecture for the GENIQUEST project, was developed through the work of Concord and the TELS center.

Four people from the Concord Consortium will be involved in the project. Dr. Stephen Bannasch as Co-Principal Investigator will provide general oversight and project management. Paul Burney, Scott Cytacki and a junior programmer from Concord will serve as programmers and interface developers.

The Jackson Laboratory is a leading mammalian genetics research institution in Bar Harbor, Maine with a research staff of 36 Principal Investigators. Jackson Laboratories is a pioneer in developing the mouse as the leading mammalian model for genetics research, and houses more than 2,800 mouse strains —the most extensive and diverse resource of genetically defined mice in the world—as well as state-of-the-art scientific services such as statistical analysis, high-performance computer facilities, distance conferencing facilities, microinjection, DNA sequencing, confocal microscopy and image analysis, and advanced flow cytometry. The Laboratory is also a center for mouse bioinformatics databases, including the Mouse Genome Database and Gene Expression Database.

The Jackson Laboratory has an 83-year history of educating and training science students through inquiry-based research mentorships. Two Nobel Prize winners attribute their lifelong pursuit of research to a start as high school students in Jackson Laboratories' historic student programs. These mentorships have grown dramatically in size and diversity in recent years, with an average of 27 different laboratories each year over the past three years enthusiastically accepting summer students as research interns. Through student mentorships, an intership program for degree candidates in the UMaine Masters of Science in Teaching program, and a sabbatical program for Maine in-service teachers, Jackson Laboratory provides students and teachers with solid grounding in modern computational biology, genetics, and developmental biology research methods.

The Jackson Laboratory is currently conducting a course with the Maine School of Science and Mathematics a state supported magnet school, including over 20 videoconference classes and several days of on site instruction at the MSSM campus. The students are continuing their work during the spring semester, where they will be applying the tools they acquired to real world data sets. The materials and processes developed from this and prior work (Blaisdell, 2004) will form the backbone of the curriculum and supplementary materials for the GENIQUEST project, and the experiences of students in this joint course will inform the development of tutorials and workflow for the GENIQUEST project application.

Dr. Jon Geiger, Director of Educational Programs for the Jackson Laboratory will serve as the liaison to the scientists and the project. Dr. Gary Churchill will oversee the project's science content and Randy Von Smith will manage the daily work of the project and oversee the work of the teacher translating existing course materials to a level more accessible to high school students. This bank of materials will form the basis for the GENIQUEST Project's supplementary instructional resources.

Evaluation

The evaluation of the GENIQUEST program will be focused on determining whether the project succeeds. This includes:

1. Develop and test a tool enabling students to import, manipulate and analyze genomics data.
2. Increase secondary students' knowledge and skills at data analysis and scientific research.
3. Improve secondary school teachers and students' access to cutting-edge STEM data sets.
4. Determine the viability of online data as a tool for engaging secondary school teachers and students in scientific inquiry.

The overarching question in the evaluation is; *What is the viability and best approaches for using scientific data with students for improving learning of science?* To answer this question the project will collect data from both teachers and students. The piloting teachers will be surveyed as to their understanding of the tool, ease of use, content, applicability to the high school Biology program, and whether this project has increased their knowledge of genomics. The online application will have a tracking mechanism that will help the PIs measure the use of the application by students. The project will document the time in particular areas of the web site, student responses, functionality of the web site, movement to higher levels of the application and persistence of students. This will be carried out by the PIs and Project Manager during the project and examined as formative data to inform the on-going improvement and work of the project as well as measurements of success for the accomplishment of the project.

The GENIQUEST Project is an exploratory project designed to investigate the feasibility of bringing genomics data and scientific investigation into secondary classrooms. The limited resources and exploratory nature of the project naturally reduce the scope of the evaluation that may be conducted. The primary evaluation role for this project will be performed by the project's external advisory board, which will convene at the beginning, middle and end of the project duration.

An external Advisory Board will be established and comprised of an education technology specialist, a higher education faculty member, a curriculum specialist, and a secondary biology teacher for the purpose of intellectual support and project monitoring. External Advisory Board members will include:

Dr. Paul Horwitz is a Senior Scientist who directs the Concord Consortium's Modeling Center and also runs Concord's Modeling Across the Curriculum, Models and Data, and Fostering Transfer projects. Paul is a theoretical physicist with an A.B. from Harvard College and a Ph.D. in physics from New York University. His educational research interests center around helping students use mental models to learn and apply scientific principles.

Pamela Van Scotter is the director of the Center for Curriculum Development at BSCS, where she has worked on curriculum development and professional development projects for ten years and oversees the work of all curriculum projects. Pam previously directed a Design Study on Multidisciplinary Science at the High School Level and the revision of Middle School Science & Technology, BSCS's three-year, integrated science program for grades 6–8, as well as acting as a curriculum developer and writer for a number of projects during her tenure at BSCS. Pam has a master's degree in anthropology from Washington State University with an emphasis in physical anthropology and linguistics.

Dr. Molly Schauffler has a doctoral degree in Plant Science from the University of Maine, and currently works with active and pre-service teachers to develop science and math curricula that require students to use real scientific data using guided inquiry-based pedagogy. She has taught numerous workshops (funded by EPA and NSF) giving teachers the necessary skills to work with online environmental data, and has developed a Web-based directory of sites containing data from environmental monitoring programs in Maine. Molly teaches a graduate-level course "Monitoring Environmental Change" (developed through an NSF Postdoctoral Fellowship), which engages teachers in real environmental monitoring projects.

Biology Teacher(s). Several candidates have been identified for this slot on the Advisory Board. They include Luci Leveque, Cony Regional Vocation Center, Augusta ME: , Jeff Cook, Cony High School, Augusta, ME: and Beth Chagrasulis, Lake Region High School Naples ME. These teachers may volunteer to be GENIQUEST piloters so the final selection has not been made at the time of this proposal.

Dissemination

The GENIQUEST project is being developed to be available worldwide on the MMSA website. Once the application proves to be viable based on its goals, then sharing of the site (application) will begin. If successful, GENIQUEST will begin to present at state (Maine Science Teachers Association, and Massachusetts Association of Science Teachers), regional (National Science Teachers Association regional conference) and then national (National Science Teachers Association, National Association of Biology Teachers) science conferences. The MMSA has a science education list serve of about 1,000 science educators that will be utilized for sharing the information about the application. Jackson Laboratories and Concord Consortium also have their own networks of teachers who will be notified about the application. Since this is an exploratory development project, the dissemination of the project is not a big part of the work as the quality and effectiveness of the product is the first priority. However, as the materials prove to be successful there are avenues to spread the work about the materials.

One intent of this project is to develop an open-source prototype platform that can be used with other online data sets so the GENIQUEST developed tools and application can be transferred to other data sets with minor programming adjustments. This will be explored to the extent is possible near the conclusion of this project.

Prior NSF Support

Dr. Francis Eberle: PI on ESI-0435217 (2005-2008) Phenomena and Representations for the Instruction of Science in Middle Schools (PRISMS) and NSDL science project. The web site is being developed and presentations have been given at two NSDL conferences and one NSTA conference. And has served as Co-PI on ESI-0101928 (2001-05): *The Northern New England Co-Mentoring Network*

(NNECN.) Evaluations of this mentoring project for reported a 90 % retention rate for the new teachers in ME, NH, and VT. The NNECN model has been presented for three years at the New Teacher Center's national conference in San Jose, CA, 3 NSTA conferences, one NCTM conference. Francis is PI on Co-PI on NSF Grant #ESI-0353315 Curriculum Topic Study. Francis has coauthored 2 books: *Uncovering student Ideas in Science Vol.1* and *Vol. 2* (2005 and 2007) published by NSTA Press, as a result of the CTS project. Dr. Francis Eberle was PI and Co-PI/PD on ESI-9819468 (1999-04) Broadening Educational Access to Mathematics in Maine (BEAMM), a K-8 Local Systemic Change project for seven low performing school districts. Students reported performance gains on the state test at a rate much higher than students in comparison schools and the state average. Has Co-authored an article that has been accepted in the Journal of Mathematics Education Leadership called *Local systemic change project: Professional development for improving student performance in seven low performing districts.* Presentations have been given at NCTM and the state Mathematics Teachers conferences.

Dr. Stephen Bannasch: Technology enhanced Elementary and Middle School Science (TEEMSS) ESI-9986419 and ESI-0352522. This pair of projects addresses the low unitization of probes in grade 3-8 by developing excellent student materials, resources and online courses. Initial research documented learning gains and was published. (Metcalf & Tinker. 2004 Probeware and handhelds in elementary and middle school science. *Journal of Science Education and Teaching.*)