

# Embedding Computer Science Concepts In K-12 Science Curricula

Chi-Cheng Lin, Mingrui Zhang  
Department of Computer Science  
Winona State University  
Winona, MN 55987  
{clin, mzhang}@winona.edu

Barbara Beck  
Department of Biology  
Rochester Community and Technical College  
Rochester, MN 55904  
barbara.beck@roch.edu

Gayle Olsen  
Department of Nursing  
Winona State University  
Rochester, MN 55904  
golsen@winona.edu

## ABSTRACT

To engage a broader audience in computer science, we have developed a set of curriculum units embedded with computer science concepts for K-12 science education. We chose bioinformatics as a vehicle to deliver these units. Our curriculum development cycle began with the identification of a set of computer science concepts which are potentially relevant to life sciences. Problems in life sciences as well as bioinformatics tools to be used for solving these problems were carefully examined for the delivery of identified computer concepts. They were later presented to groups of regional K-12 science teachers in our summer workshop on bioinformatics. With their help, we adapted and polished these curriculum units to meet Minnesota state standards for K-12 science education. This paper describes our approach in developing the curriculum units.

## Categories and Subject Descriptors

K.3.2 [Computers and Education]: Computers and Information Science Education

## General Terms

Design

## Keywords

Bioinformatics, Interdisciplinary, K-12 science curriculum, Outreach

## 1. INTRODUCTION

Computer science has been extended to support other disciplines, in addition to being a field in its own right in the traditional way [4]. In spite of emerging opportunities for computer science and the continuous growth in the need for IT workforce, the computing society has witnessed a decline in enrollment in computer science programs for several years [10]. Though many have blamed the decreased demand for IT professionals for this decline, others have argued that the lack of early introduction to

computer science to students in secondary education [3, 8] also contributed to it.

Efforts have been made to battle this shrinking enrollment, such as the re-design of computer science curricula [12] and the outreach to K-12 education [2, 5, 6]. Both efforts are also pertinent to the goals of computer science education of engaging a broader audience in computer science concepts. Currently, the outreach activities are primarily in the format of workshops targeted to either high school computer science teachers or pre-college students. Those workshops are very valuable in recruiting students into the program. However, only a limited number of students are reached, given the fact workshop attendees are the ones who are already interested in computer science. Additional recruitment strategies for reaching a broader audience of K-12 students in North America still need to be identified.

In most states in the United States, computer science is offered as an optional program in K-12 science curricula. As a non-mandatory program, it is even more difficult to add sufficient content into an already packed schedule. In a survey of 13,000 high school teachers conducted by the Computer Science Teachers Association in 2007, there were eight challenges frequently faced by K-12 teachers ([csta.acm.org/Research/sub/CSTARResearch.html](http://csta.acm.org/Research/sub/CSTARResearch.html)). Among them is the “lack of curriculum resources.” In searching for a new strategy, we carried out a three-year project. Our project is to use bioinformatics as a vehicle and embed computer science fundamentals into K-12 life sciences.

Studies have suggested that making computer science relevant to real-world problems may stimulate more students, especially from under-represented groups such as females, to study computer science [6, 9]. A more recent survey conducted among regional high school students showed that the “use of computers in another field such as business or medicine” is a top motivator of prospective students, number one among females and number three among males, when they consider computer science as their major [3]. With help from regional K-12 science teachers, we have successfully identified some computer science concepts and embedded them in the existing subjects of K-12 science.

We chose bioinformatics as a vehicle to do it, as bioinformatics is an interdisciplinary field and uses computer technologies and computational tools to model, analyze, present, visualize, and manage biological data. Its primary goal is to retrieve information from the data and use the information to facilitate problem solving in life science and medicine. If done right, we may potentially engage a broader audience in computer science

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*SIGCSE'09*, March 3–7, 2009, Chattanooga, Tennessee, USA.

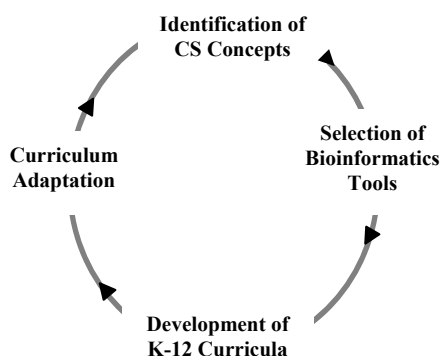
Copyright 2009 ACM 978-1-60558-183-5/09/03...\$5.00.

concepts than those taking high school computer science courses or participating in workshops. Using our existing bioinformatics track [12] as a recruitment platform, we have coordinated three summer bioinformatics workshops for K-12 science teachers, and have helped them develop curriculum modules with bioinformatics units.

This paper describes the strategies we used in our project as well as its status. We explain the approach we took in developing these curriculum units in Section 2. We then demonstrate how computer science concepts are embedded in K-12 life sciences in Section 3. We showcase an example of a developed high school curriculum unit in Section 4. We report the current status of our project in Section 5, and conclude with our future plan in Section 6.

## 2. APPROACH

Our curriculum development begins with identifying computer science concepts embeddable in K-12 life sciences. It consists of four major phases, identification of CS concepts to be delivered, selection of bioinformatics tools, development of K-12 life science curricula, and their adaptation. Though initiated by a group of professors and instructors from computer science, biology and nursing departments, our approach involves regional K-12 science teachers. The curriculum evolves based on the feedback from the teacher partners.



**Figure 1. Development of bioinformatics units for K-12 science curricula.**

### 2.1 Initiation Phase

To start, we asked a question to ourselves “What computer science concepts do we want to convey to the high school students?” Of course, we easily created a long list of concepts we thought are essential to computer science and worth exploration by students in middle or high schools. To narrow down this list, we asked a question “How is it connected to a subject in life sciences?” or “How is it used in solving a real problem?” for each item in this list. Programming is definitely not a candidate for two reasons. First, we would like to engage students in general, not just those who are already interested in CS, and second, we want to correct the misconception of CS as programming only. In the end, we want students to develop an appreciation of computer technology in solving life science problems. Using these strategies, we were able to finalize a list of concepts essential to CS and still usefully and clearly connected to life science. The concepts fall into three categories.

1. Data structures such as arrays and trees. How are they used in modeling and abstracting biological information;
2. Basic searching algorithms. Why is an algorithmic approach needed? Their applications in solving biological problems efficiently and effectively;
3. Databases and the query process. The available genomic database and query processes.

Once a list of CS concepts has been created, we started to research the bioinformatics tools using these concepts. A set of problems potentially relevant to K-12 science curriculum were also collected.

### 2.2 Engaging Regional K-12 Teachers

Throughout the whole process of curriculum development, we faced two critical issues. First, we needed to assure that the developed bioinformatics units aligned with Minnesota K-12 Academic Standards. As none of us had ever taught in K-12 schools, we had no first-hand knowledge about the standards. Second, we had to make sure that the units to be developed were attractive to K-12 students and that the computer science concepts are properly placed in the context of life sciences. During the initiation phase, we examined the curriculum units to be developed to see if they were adaptable by K-12 teachers, interesting to students, and relevant to both computer science and life sciences. We further developed the following criteria as a guideline:

1. Could the units be directly adapted by K-12 teachers for their science curricula? Are all needed bioinformatics education resources available in the public domain?
2. Are the units of interest to K-12 students?
3. Are subjects relevant to any real-world problems? If so, can we identify them?

To address both issues, we used our bioinformatics workshop as the mechanism to bring together the experiences of K-12 teachers and our knowledge about bioinformatics. Specifically, we designated the last phase of the workshop for workshop attendees to actually develop bioinformatics units and adapt these units in their teaching assignments. This phase constitutes the development of K-12 curricula and the curriculum adaptation steps in our approach (Figure 1).

### 2.3 An Evolving Development

We used the workshop as a platform to network with regional K-12 teachers, and later used the feedback from teachers to initiate the next round of development and adaptation. Our workshop was first offered in summer 2006, and again in 2007 and 2008. Its topics covered three different areas, namely biology, computer science, and healthcare. The computer science unit included fundamental bioinformatics concepts, bioinformatics tools and databases, and a set of hands-on Web-based computer labs. More details of our outreach workshop can be found in [13]. The group of attendees was diverse in spite of its small size. Most were high school life science teachers with a wide range of teaching assignments in general biology, honors biology, advanced placement (AP) biology, physiology, anatomy, zoology, genetics, and biotechnology. There was one middle school earth science teacher and one biology instructor from a local community

college. It is noteworthy that our workshop also attracted a high school computer science teacher. A website about our workshop program is available at <http://cs.winona.edu/outreach>. It consists of the program, bioinformatics education resources, and example projects developed by our workshop attendees.

The development of curriculum units is not complete without their adaptation for K-12 teaching. Back in their schools, our workshop participants taught life science classes with the curriculum units they developed. They reported to us their experiences in adapting these units. We sought their comments and suggestions to improve the curriculum resources we provide in the workshop. Curriculum units were continuously added, deleted or redesigned based on their feedback. For instance, the attendees responded that the mathematics content for bioinformatics was beyond high school students. We made adjustments by revising those parts to explain the concepts in layman's terms.

Following the cycle of development, we are able to ensure all the criteria we set for the project were being met from beginning to end. For adaptability, we utilized proven bioinformatics curriculum resources, such as the National Center for Biotechnology Information (NCBI, <http://nlm.nih.gov>) and San Diego Supercomputer Center's Biology Workbench (<http://workbench.sdsc.edu/>). We designed fun activities including DNA isolation and an innovative molecular evolution simulation. We have also developed lectures and laboratories with subjects pertinent to human society, including the human genome, hereditary disorders, and cancer research and treatments with bioinformatics technologies. Example curriculum units showcasing how we embed the computer science concepts in the life science curriculum will be discussed in the next section.

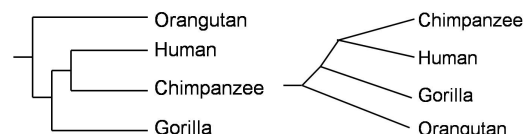
### 3. CONVEYING CS CONCEPTS

The initial design of curriculum units was done by a group of professors and instructors from computer science, biology and nursing. These units were later revised with the help of regional K-12 teachers and adapted in their teaching. The following two case studies demonstrate how computer science concepts are seamlessly blended into life science.

#### 3.1 Trees and Molecular Evolution

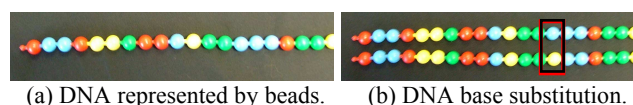
In computer science, the tree is an indispensable data structure and abstraction tool. It could be taught in the context of phylogenetic analysis. Phylogenetics is the study of relationships among biological sequences or among species, and it can be used to infer evolutionary relationships from shared features. The phylogenetic tree is a graphic that provides a visual representation to facilitate the analysis. The distance between two nodes on the tree is based on differences in DNA or protein sequences of the two species represented by the nodes. Tree terminologies such as root, leaf, parent, child, and subtree could be introduced to students. We have shown how the trees are used to represent data, and model and solve biological problems. In modeling biological problems, trees can be constructed in many different ways, yet they still represent the same set of relationships. It introduces the concept of tree isomorphism to students. Figure 2 illustrates two phylogenetic trees of several primates, where the human is inferred as more closely related to the chimpanzee than to the

gorilla. It also shows that there is more than one way to draw the tree.



**Figure 2. Phylogenetic tree example.**

Active learning has been recognized as an effective pedagogical technique. Russell Hulse, a Nobel Laureate in Physics and a strong advocate for K-12 interdisciplinary science education, has emphasized the importance of learning-by-doing in preparing K-12 students for interdisciplinary work [7]. Working with a biologist, we created a hands-on activity for students to develop an appreciation of the use of computer technology in problem solving. This activity was developed using Pop-It beads to simulate molecular evolution in a classroom setting. Molecules are represented by strings of different colors of beads. The objective is for students to gain an understanding of how phylogenetic trees display evolutionary relationships.



**Figure 3. Using Pop-It beads to simulate molecular evolution.**

Beads come in four colors representing the four DNA bases A, T, C, and G as shown in Figure 3(a). Base substitutions (i.e., mutations) result from substituting one color of bead for another as shown in Figure 3(b). The ancestor (original sequence) molecule diverges into two lineages, each of which undergoes an independent mutation. Each of these lineages also diverges, with the descendants undergoing independent mutations, until a population of eight lineages is created. The resultant bead data is converted into DNA sequence data, and then processed by Web-based bioinformatics tools at Biology Workbench to create the phylogenetic tree.

A Java program with a graphical user interface was developed to convert the bead data to DNA data automatically. We also showed that without the computer program the conversion would take several steps requiring a lot of involvements from participants. We used it to demonstrate the benefit of automation in problem solving. Furthermore, it is made clear to students that the problem of phylogenetic trees could not be solved without computer even for short sequences, not to mention dealing with longer gene sequences in the real-world cases.

At the end of this activity, a teacher may request students use the beads to represent real DNA sequences. Apparently, for longer sequences, both representation and manual alternation of DNA bases become difficult, if not impossible. It is obvious that computer is needed to perform the simulation. A computer program may be adapted by students who are interested in programming to simulate the evolution process conveniently. It could be adapted as a high school CS project.

## 3.2 Searching Algorithms and Databases

Sequence alignment, the comparison of DNA or protein sequences, is considered as the cornerstone of bioinformatics. It can help determine the function of a newly discovered gene sequence. Basic Local Alignment Search Tool (BLAST) [1], the Google™ of bioinformatics, is an application of sequence alignment. It is a “search engine” that allows users to enter DNA or protein queries to find similar sequences in biological databases. We illustrated how BLAST is used to search biological databases and how the result obtained from the search is interpreted. BLAST can be abstracted as a string matching problem, where both query and database are character strings. The BLAST algorithm, however, is not trivial, as the size of the databases to search is enormous and the exact match is not required to obtain results with biological significance. Consequently, the algorithm and data structure need to be carefully crafted to perform the search efficiently. The algorithm can be summarized as the following steps:

1. Prepare a list of short, fixed-length segments from the query.
2. Find highly similar or exact matches for each segment. A data structure, such as a hash table, is utilized to achieve an efficient search.
3. Extend each match to potentially obtain a longer match. As the optimal solution cannot be discovered until the whole database is explored, a heuristic is needed to decide when to stop the extension.

Even though the details of the BLAST algorithm are beyond the scope of high school science, we described the steps listed above to demonstrate the concepts we want to convey. Specifically, we pointed out the use of a heuristic to obtain a suboptimal solution as a compromise when dealing with a hard problem which is commonly found in bioinformatics. We used it to demonstrate the essence of algorithmic approaches in solving real-world problems. It allows us to explain that a computer program without a careful design would take an impractically long time to complete. We further discussed that even an “efficient” algorithm might become infeasible for searching a biological database due to the magnitude of the genomic data.

Sequence alignment was applied to studying human mitochondrial DNA (mtDNA) using bioinformatics in our workshop. Human mtDNA is an intriguing subject relevant to numerous real-world problems in many aspects. Researchers have used it for analyzing human evolution and for determining the haplogroup of individuals. To make this module interesting to the participants and their students, we developed wet labs to give workshop participants hands-on experience in isolating their own cheek cell DNA. Participants amplified the mtDNA sequences using the polymerase chain reaction, and then ran gel electrophoresis to see the results. Follow-up Web-based computer labs were developed for processing the mtDNA sequence data. BLAST was used to determine a person’s ancestral lineage, for a given mtDNA sequence. Trees were brought up again, where the mtDNA sequences were also used to build phylogenetic trees.

## 4. AN EXAMPLE CURRICULUM UNIT

Our approach worked out effectively. The workshop participants showed enthusiasm in adapting our curriculum materials. As a result, we started to see bioinformatics initiatives in place in

regional high schools. After the first workshop, we supported one of our 2006 workshop participants in developing and implementing a bioinformatics project for her biology classes at Hastings High School in Minnesota. The project, “*Using the Web to Analyze DNA*”, is a hands-on computer laboratory adapted from our summer workshop. The laboratory handout begins with a humorous DNA-related cartoon, followed by a summary of how online biological databases and bioinformatics tools are used to pose and answer biological questions. It consists of a list of online bioinformatics resources and three activities.

Its first activity is to identify a source for a DNA or protein sequence. A BLAST tutorial is provided in the handout followed by a set of exercises using fictional dinosaur DNA sequences and a realistic amino acid sequence. Its next activity is to compare two or more DNA or protein sequences. Biology Workbench is introduced next; tools are used to align the sequences and to build the phylogenetic trees. Questions are asked based on the trees the students construct. The last activity is to find a DNA sequence of a species that the student wants to study, using NCBI’s Taxonomy Browser. Bioinformatics related concepts are introduced in these activities.

While developing this curriculum for her school, she frequently communicated with us. Through the exchange of information, we have learned from her that taxonomy is an important subject in the high school life science curriculum. Even though we did not include it in the workshop she attended, she was able to explore NCBI’s website and included Taxonomy Browser in her project. Consequently, we adapted her project into our second workshop. She also presented her project at the Minnesota Science Teachers Association Spring Conference in 2007.

What is significant is that the teacher revised her curriculum a year later because her students “didn’t seem to understand how difficult it is to compare DNA sequences and how great it is to have a computer do this for us.” She modified the original activity to include an additional hour in which students try to line up the DNA sequences and then build the phylogenetic tree themselves manually first. The hope is that students’ neutral reaction to the original activity will change when they see how difficult comparing DNA sequences can be without the help of computers, and consequently that they will better appreciate the contributions of computer science to solving real-world problems.

## 5. CURRENT STATUS

It has been three years since we started this project in late 2005. We have followed up with our workshop participants and collected feedback from them for assessing the curriculum units we have developed. Among the fifteen participants, thirteen of them have adapted what they have learned and developed in our workshops into their teaching. The following is a brief summary.

1. Eleven teach biology primarily, one teaches earth science, and the other teaches math/computer science.
2. Eleven teach either 9 to 12 grades at high school, two teach 7 or 8 grade at middle school, and one also works with elementary schools.
3. Bioinformatics has been used in teaching DNA/RNA, evolution, genetic engineering, genetics, taxonomy, and human migration in various levels of biology, earth science, AP CS project, and independent study.

4. On-line databases, BLAST, phylogenetic trees, sequence alignment algorithms are adapted.

When surveyed on the greatest obstacles, not surprisingly, it was reported the lack of time in preparing and planning. It was also hard for them to find time to teach new subjects in the already crowded curriculum. Other barriers included money, equipment, and expertise. Only one teacher said that there was no obstacle preventing her from doing it. Several responses indicated that their students felt the subject of bioinformatics very interesting. Seven out of 13 teachers thought it was a “very good idea” to include bioinformatics in the science curriculum, two a “good idea”, and one “neutral”. The survey demonstrates that our approach of introducing computer science concepts in the context of life sciences using bioinformatics as a vehicle is not only feasible but also well-received by teachers and their students. We quote some of their comments here.

“... they love it. The kids are better at figuring this all out than me!”

“I hope to incorporate more this year. ... plan to incorporate more in the next few years.”

“I strongly believe that Bioinformatics should be included in the science curriculum. It has proven applications in various aspects of science ...”

## 6. CONCLUSIONS

The seminal article *Computational Thinking* [11] emphasizes that “We should expose pre-college students to computational methods and models.” To engage a broader population of pre-college students, we carried out a three-year project for embedding computer science concepts into K-12 life science curricula. We used bioinformatics as a vehicle. We hope that our outreach effort will counter the negative stereotypes about computer science among K-12 students and demonstrate the contribution of computer science to life science. We chose to work with regional K-12 teachers, instead of directly with students, with the intent that they can help us deliver the message to their students more effectively.

Our approach begins with identifying computer science concepts which show clear relevance to life sciences. Problems in life sciences and bioinformatics tools used in solving these problems are carefully examined and identified. They are adapted and presented to groups of regional K-12 science teachers in our summer bioinformatics workshop. With help from those teachers, we developed a set of bioinformatics curriculum units for K-12 sciences. The approach has produced successful outcomes.

Our outreach efforts demonstrated that computer science concepts can be embedded into K-12 life science curricula. We plan to maintain collaborations with our workshop participants by providing our expertise in bioinformatics. Via communication with the regional K-12 teachers, we will learn about their teaching experiences with bioinformatics. We will continuously follow up with them and adapt those curriculum components.

## 7. ACKNOWLEDGEMENT

This project is funded by a grant through HealthForce Minnesota, a “Center of Excellence” designated by Minnesota State Colleges and Universities and funded by the State of Minnesota.

## 8. REFERENCES

- [1] Altschul S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, “Basic Local Alignment Search Tool,” *J Mol Biol* 215(3): 403–410, 1990.
- [2] Blum, L. and T. J. Cortina, “CS4HS: An Outreach Program for High School CS Teachers,” Proceedings of the 38th SIGCSE technical symposium on Computer science education, pages 19-23, March 2007.
- [3] Carter, L., “Why Students with an Apparent Aptitude for Computer Science Don’t Choose to Major in Computer Science,” Proceedings of the 37th SIGCSE technical symposium on Computer science education, pages 27-31, March 2006.
- [4] Denning, P., “Crossing the Chasm: The Keys to a Coherent Profession are Bridges between Computing Technologists and the Multitude,” *Communication of the ACM*, Vol 44, No. 4, 21-25, April 2001.
- [5] Dodds, Z. and L. Karp, “The Evolution of a Computational Outreach Program to Secondary School Students,” Proceedings of the 37th SIGCSE technical symposium on Computer science education, pages 488-452, March 2006.
- [6] Doerschuk, P., J. Liu, J. Mann, “Pilot Summer Camps in Computing for Middle School Girls: From Organization Through Assessment,” Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education, pages 4-8, June 2007.
- [7] Hulse, R. A., “Preparing K–12 Students for the New Interdisciplinary World of Science,” *Experimental Biology and Medicine* 231:1192-1196 (2006).
- [8] Klein, A., “K-12 Education Shrinking Future College Graduate Population In Computer Studies,” *Journal of Computing Sciences in Colleges*, Volume 21 Issue 4, pages 32-34, April 2006.
- [9] Peckham, J., L. L. Harlow, D. A. Stuart, B. Silver, H. Mederer, and P. D. Stephenson, “Broadening Participation in Computing: Issues and Challenges,” Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education, pages 9-13, June 2007.
- [10] Vegso, J., “Continued Drop in CS Bachelor’s Degree Production and Enrollments as the Number of New Majors Stabilizes,” in *Computing Research News*, March 2007, Vol. 19, No. 2.
- [11] Wing, J. M., “Computational Thinking,” *Communication of the ACM*, Vol 49, No. 3, 33-35, March 2006.
- [12] Zhang, M., E. Lundak, C.-C. Lin, T. Gegg-Harrison, and J. Francioni, “Interdisciplinary Application Tracks in an Undergraduate Computer Science Curriculum,” Proceedings of the 38th SIGCSE technical symposium on Computer science education, pages 425-429, March 2007.
- [13] Zhang, M., C.-C. Lin, G. Olsen, and B. Beck, “A Bioinformatics Track with Outreach Components,” Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education, pages 186-190, June 2007.