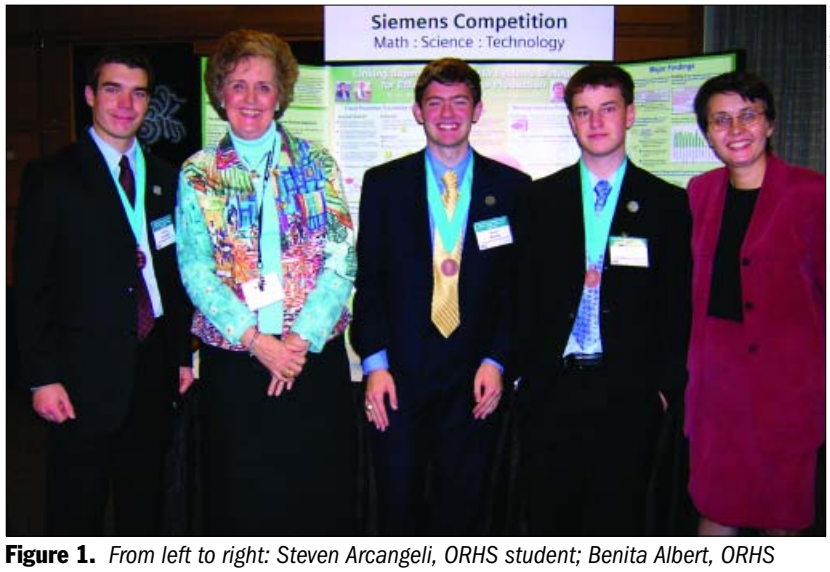
Mentoring SCIENTISTS of the FUTURE

Dr. Nagiza Samatova and Christopher Symons of ORNL write about their experience mentoring students from a local high school. Three students made important contributions to a project funded by DOE's Office of Advanced Scientific Computing Research, and their hard work was recognized when they received a top award at the 2006–07 Siemens Competition in Math, Science, and Technology.

The use of the term "mentoring" at a research institution might evoke the notion of a seasoned researcher imparting years of hard-earned knowledge to a post-doc or young scientist. Rarely would the word bring to mind high school students working alongside researchers at a national laboratory. Yet that is precisely what takes place at ORNL under the Math Thesis course offered by the Oak Ridge High School (ORHS) Mathematics Department.

Through this Math Thesis course, we have had the pleasure of mentoring several ORHS students who have conducted research projects with us at ORNL. Included in this group are seniors Steven Arcangeli, Scott Horton, and Scott Molony, who worked in conjunction with our research team on the project entitled "Data Intensive Computing for Complex Biological Systems (BioPilot)," funded under the DOE Office of Advanced Scientific Computing Research (ASCR). For their contributions to this important research, these ORHS students recently won the top team award at the 2006–07 Siemens Competition in Math, Science,



R. ARCANGELI

Figure 1. *From left to right: Steven Arcangeli, ORHS student; Benita Albert, ORHS Mathematics Department; Scott Molony, ORHS student; Scott Horton, ORHS student; and Dr. Nagiza Samatova, ORNL researcher and mentor.*

More Mentoring Experiences

Over the past two years, the team of researchers led by Dr. Nagiza Samatova and consisting of Dr. Tatiana Karpinets, Ramya Krishnamurthy, Dr. Hoony Park, and Christopher Symons has successfully mentored several ORHS students in various research projects.

Last year, the mentoring team worked with three ORHS students—Patricia Brent, Nicholas Grabenstein, and Tarik Umar. These students were national team finalists in the 2005–06 Siemens Competition. They worked on a project in the field of Natural Language Processing, and their choice to enter the Siemens Competition and their subsequent success helped inspire this year's students to enter the competition and to work as hard as they did.

The mentors also worked with Alice Gu, another ORHS senior who became a semifinalist in this year's Intel Science Talent Search competition. She was one of 300 semifinalists chosen from a pool of over 1,700 students.

Another student mentee is Andrew Jallouk. An ORHS graduate and freshman at Vanderbilt university, Andrew worked alongside Steven Arcangeli, Scott Horton, and Scott Molony on complementary projects, and thus helped contribute to their success.

"I think we would all like to see more collaborations of this sort in the U.S. It feeds interest in math and science and provides valuable research experience that is normally not available until college."

Steven Arcangeli
ORHS student

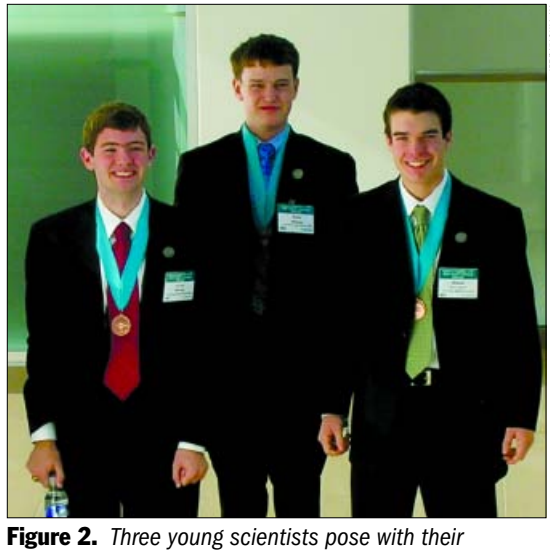
and Technology. This prestigious award carries with it $100,000 in scholarship money that the three students will share.

Mentoring Young Students in Research

With the constant focus on the diminishing competitive edge the United States currently holds in science and technology over much of the rest of the world, we are gradually seeing an increased emphasis on preparing young students for career fields, such as scientific research, that are deemed important to the U.S.'s future ability to compete technologically. At the highest funding levels, this emphasis is often made by requiring proposal writers in scientific fields to outline how a portion of the proposed research grant would be used to promote teaching or training that will help in the development of the next generation of scientists. This requirement is commonly addressed by including compensation for postdocs or graduate research assistants within the budget, but rarely does the educational impact extend beyond those who have already decided not only to do research, but also to focus on the field in which they will be working.

If we are to hope that more of our brightest young students will consider careers in research, we need to provide them with a glimpse into what such a career can offer in terms of intellectual stimulation and other intangible rewards. And this needs to be done before a career decision has been made, that is, before a student enters graduate school, and even earlier than that.

Many young students who enter science and engineering fields do so based on the fact that these fields offer good job prospects upon graduation. A very small percentage of high school and undergraduate students actually know what career path they would like to take with a strong



R. ARCANGELI

Figure 2. *Three young scientists pose with their Siemens Competition prize medals.*

degree of certainty. Thus, we almost certainly lose many potentially avid, intelligent researchers simply because such a career path is never even considered. A career in scientific research is not for everyone, but we should not be keeping some of our brightest young minds from pursuing research professions simply because we neglect to introduce them to such a possibility.

As things stand now, the U.S.'s advantage in science and technology relies heavily on the fact that many of the best and brightest from other countries come to the U.S. to learn and work. This may not always be the case, as recent declines in international applications to U.S. graduate schools forewarn. As of now, the decline is often blamed on student visa issues arising from security concerns, but other countries are taking advantage of this to attract these same students that would usually come to the

Figure 3. *Proud mentors and teachers, Dr. Nagiza Samatova (far left) and Benita Albert (far right) stand with successful students—ORHS seniors Steven Arcangeli, Scott Molony, and Scott Horton.*

U.S. As other countries prosper in the global economy, the lure to leave home for better opportunities in the U.S. decreases.

One potential method for addressing the closing competitive gap is to encourage more of our best domestic students to enter careers in science and technology. However, this may need to be done at an early age to have an impact. As a researcher, it is a difficult choice to spend time and/or money on students below the graduate level. Such efforts are more likely to hinder than assist in the completion of proposed deliverables, and so such efforts, when they are included in proposals, are listed as deliverables themselves. But there is nothing wrong with this model, and it is only proper that funding agencies understand that money put toward such goals is targeted at long-term prospects, rather than immediate results. Of course, there will also be rare cases when younger students do indeed make an immediate impact.

The success of high school students Steven Arcangeli, Scott Horton, and Scott Molony is indeed an exceptional occurrence, but with the attention it has drawn, perhaps it will allow others to glimpse the more common rewards that such mentoring can bring.

ORNL and ORHS Cooperative Research Program

ORHS in Oak Ridge, Tennessee offers their students the course entitled "Math, Science, Computer Science Thesis," which is a research cooperative with ORNL. The Math Thesis course advisor is Benita Albert of the ORHS Mathematics Department. Under this program, ORNL research scientists host senior students and serve as their mentors throughout the course. The course duration usually covers the Fall and Spring semesters, and the students typically spend the previous summer working closely with their research mentor(s) at the lab.

The Learning Environment

As part of their Math Thesis course, the three ORHS seniors joined our team within the ORNL Computer Science and Mathematics Division this past year. The students arrived early in the summer, and quickly showed their dedication and willingness to work long hours, despite the fact that they were not required to do anything over the summer. They quickly became accustomed to the work environment and developed a comfortable rapport working together. In addition to our mentorship, Dr. Tatiana Karpinets and Dr. Hoony Park served as co-mentors for the students.

The students were instructed in the fundamentals of graph theory, statistical theory, systems biology, bioinformatics, artificial intelligence, and programming in C/C++. Having a large, diverse research team, we were able to use the teaching opportunities to more formally expose the members of our team to the basics of the others' research areas. In an interdisciplinary group such as ours, this turned out to be quite valuable for the team as a whole. The students naturally did a lot of learning on their own as well, typically while applying what they learned to complete small subtasks for their project. They were also assigned research papers on topics related to their project that they presented to the team. Their

"To immerse high school students in a rich research project and to observe their fresh, eager, and risk-taking young minds produce valued results has been a highlight of my 39-year teaching career. The opportunity for team research work, state-of-the-art scientific investigations, writing, and presenting scholarly results before expert juries will make both the students and the high school curriculum stronger. The students in Thesis at ORHS are now local heroes, and Thesis has become a 'hot' course."

Benita Albert
ORHS Math Department

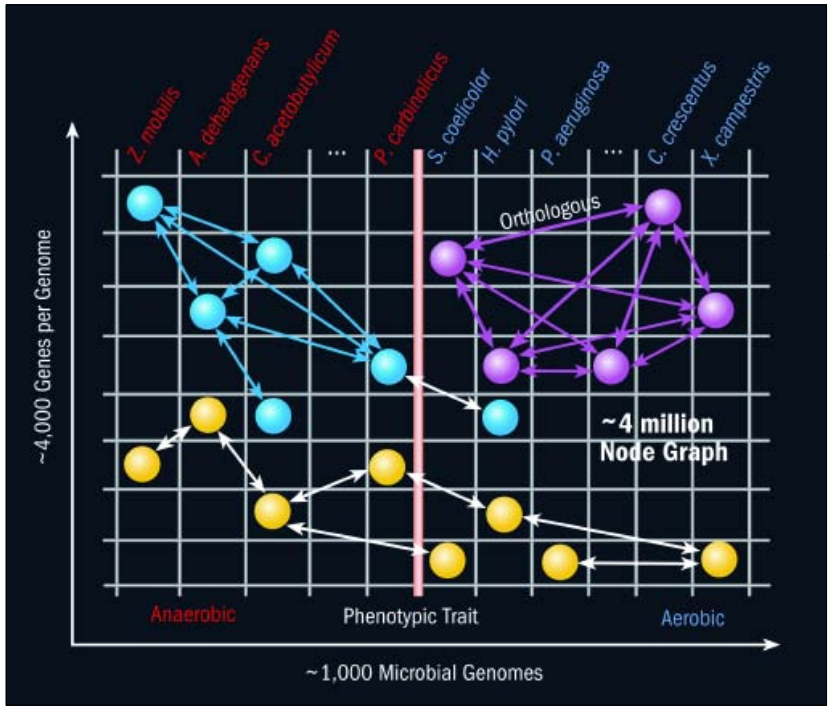


ILLUSTRATION: A. TOVEY SOURCE: N. SAMATOVA, ORNL

Figure 4. *One of the students' innovations was the linking of desirable microbial traits—such as resistance to high temperature, multi-sugar uptake, and high ethanol yield—with specific genes and biochemical pathways that are likely to be important for these traits. To find these genes, they divided the organisms according to the presence or absence of a trait. Their underlying intuition was that if a gene was critical to a trait, then Mother Nature would preserve it, and it therefore would be conserved by evolution. Hence, genes that are crucial to the trait will cluster on one side of the divide. They formulated the problem as a graph-theory problem of finding the cliques of genes (pair-wise connected) that are conserved throughout evolution among the organisms that possess a beneficial trait.*

J. HORTON

Figure 5. *Award-winning team members Steven Arcangeli and Scott Horton take a break from the competition and work on some molecular models.*

ability to comprehend new material published in high profile research journals was impressive, especially given their limited backgrounds in these fields.

The students were given plenty of support, but were essentially left to work on their own much of the time. This was apparently ideal, as they worked very closely with one another and demonstrated extreme dedication to each other and to their project.

Major Accomplishments

As part of their thesis project, the students decided to evaluate how some of the tools developed by the ORNL team under the BioPilot project could be applied to better understand the mechanisms underlying the biomass-to-bioethanol conversion by microbes.

One of the main hopes of the students was to be able to link desirable microbial traits, such as resistance to high temperature, multi-sugar uptake, and high ethanol yield, with specific genes and biochemical pathways that are likely to be important for these traits. To find these genes, they divided the organisms according to the presence or absence of a trait. The underlying intuition was that if a gene were critical to a trait, then it would be conserved by evolution. Hence, genes that are crucial to the trait will cluster on one side of the divide. The problem was formulated as a graph-theory problem of finding the cliques of genes that are conserved throughout evolution among the organisms that possess a beneficial trait.

In addition to the algorithmic benefits of applying graph theory, using these methods allowed the students to think more intuitively about very complex problems by casting those problems into a simpler space. Quite impressively, the students seemed to quickly grasp the necessary mathematics behind the graph theoretical techniques that they would employ. After that, as they progressed towards evaluating the basic idea, they were able to come up with several intuitive alternatives to overcome some remaining computational complexities involved in the problem.

While investigating the utility of the method, the students discovered that even for 28 microbial genomes (see figure 4) divided into aerobes (growing with oxygen) and anaerobes (growing without oxygen), the existing clique finding codes running on the ORNL SGI RAM supercomputer took an impractical amount of time. Even though the size of the graph was only 66,000 vertices (genes), the problem required scaling the codes to potentially millions of genes (1,000s of microbes with about 4,000 genes each). The challenge was in the data intensive nature of the problem, which

"Performing research under our mentors has impacted decisions I'll make about what I would like to do with my life in the future."

Scott Horton
ORHS student

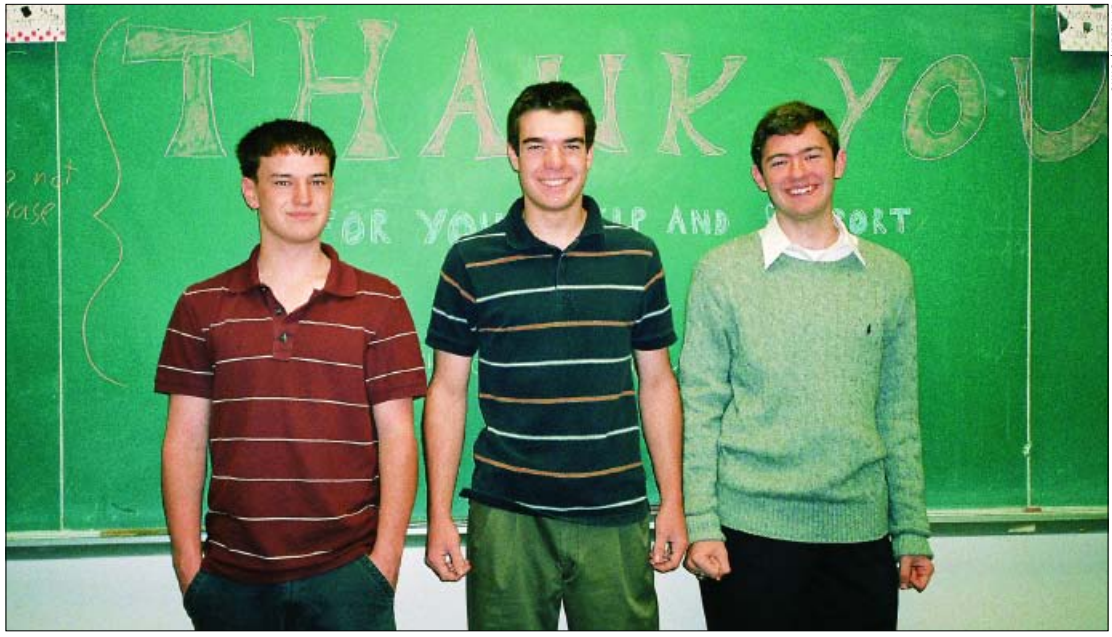


Figure 6. *Scott, Steven, and Scott, grateful for the research experience made possible by the ORNL/ORHS mentoring program.*

required access to terabytes of memory. To address this technical challenge the students extended the existing codes by prototyping techniques from the scientific literature.

Promising Results

While the results obtained by the students are preliminary, they are quite promising. By looking at a reasonably well-studied trait—aerobic versus anaerobic growth—and using only a small number of microbes, the students identified a target list of important genes. Many of these genes were known, thus confirming the validity of their idea, while many others offer a basis for future hypothesis formulations and experimental validation.

Initial work comparing biochemical pathways enriched by these genes using sub-graph matching is also interesting for understanding how the topology of pathways might affect their contribution toward phenotypic function. When applying this methodology to the industrially accepted bioethanol producer, *Z. mobilis* bacterium, their results were consistent with a recently published paper in *Nature Biotechnology* that used experimental methods alone. Future work that will look into a database larger than the original 28 microbes and include a richer set of traits directly related to ethanol production could take the students a step closer to engineering biofuel.

The Impact on the Students

All of the students involved now express a strong interest in scientific research as a possible career goal. They have learned a great deal and gained a tremendous amount of confidence in their own abilities. When students are self-motivated, truly involved, and capable, the experience can be rewarding for both sides on many levels. If even one of these students who previously may have chosen a different career path moves on to become an avid researcher (and this is more than likely, in our opinion), then we think the investment in time and energy will clearly have been worth it.

Even if it is only once or twice a decade, we hope that every researcher will have the same opportunity to work with and inspire, and potentially be inspired by, these young researchers that can benefit from exposure to real research. And hopefully, if given the chance to contribute and experience the intellectual stimulation possible in a research environment, such students will commit themselves to something, perhaps as they've never done before, and thus reap the rewards that such dedication can bring.

We call this mentoring experience a success because of the interest in future research possibilities that the project has fostered in the students. The recognition that the students have received through Siemens is a great reward for them, and hopefully the exposure that they have received will inspire others to create more opportunities for young people to become exposed to research. But the project would have been a success regardless of its outcome or the recognition involved, simply because of the interest it has engendered in these bright young minds. ●

Contributors: Dr. Nagiza F. Samatova and Christopher T. Symons, ORNL, Computer Science and Mathematics Division

"My exposure to scientific research has dispelled many illusions. It is accessible, messy, wrought with frustration, but is also the most fulfilling experience imaginable. Nothing can compare to the satisfaction of making a contribution, however small, to the greater body of knowledge."

Scott Molony
ORHS student