



At ISMB, a Focus on Bottlenecks for Systems Approach to Biology, Bioinformaticians Becoming 'Embedded Systems'

July 07, 2009

Newsletter: [BioInform](#)
[BioInform - July 10, 2009](#)

By [Vivien Marx](#)

Over 1,500 participants gathered in Stockholm, Sweden, last week for this year's Intelligent Systems for Molecular Biology conference and European Conference on Computational Biology.

Although attendance was a few heads lower than last year's meeting in Toronto, it was "a little higher than what we expected," particularly in light of the current economic downturn, said Burkhard Rost, a Columbia University researcher and president of the meeting's organizer, the International Society for Computational Biology.

Conference talks, keynotes, birds-of-a-feather gatherings, and spontaneous meetings of scientists with similar interests swept the breadth of the field, from algorithm development, to statistical methods and computational pipelines, to ontology development and text mining, to GPUs and cloud computing.

One consistent theme in talks was the need for collaboration and integration of both tools and people — particularly the need for bioinformaticians to become "embedded" within experimental groups rather than segregated in a separate department.

Gunnar von Heijne, director of Stockholm University's Bioinformatics Center and honorary chair of the conference, said that in Sweden, like elsewhere, bioinformatics started out as the activity of a "few scattered groups" in biology and computer science departments. "There was a time when it seemed to make sense to collect the bioinformatics people in separate institutes to give them an identity and to create community," he said.

Times are changing, however, particularly due to large-scale biology and the realization by wet-lab scientists that computational colleagues are important for their projects and must be "integrated very closely" with the wet lab. "We're dispersing back out into the experimental departments," von Heijne said.

Oliver Hofmann, associate director of the Harvard School of Public Health's Bioinformatics Core, said that bioinformaticians have traditionally been "sheltered" in separate centers, but that has given way to a "natural embedding" of bioinformaticians into the work of scientific colleagues, where they become not only part of a given project and its analysis phases but also of planning and study design, as well as project evaluation.

Embedded Systems

"I'm already an embedded system," said Gene Myers, ISMB conference co-chair. Several years ago, he changed his career trajectory by joining Howard Hughes Medical Institute's Janelia Farm Research Campus, where he has been "deeply immersed with biologists."

He said that as research takes more and more of a systems approach, "computation is becoming the key bottleneck." Data interpretation is computationally intensive "simply because of the scale of experiments," he said.

Myers recommended that all project participants remember to have the bioinformatics team "in place at the beginning," an approach that not all scientists have fully embraced, he said.

At the same time, Myers said, the "level of interaction and sophistication on both sides of the table" has increased a "great deal."

Keynote speaker and developmental biologist Eugenia Maria del Pino Veintimilla from the Pontifical Catholic University of Ecuador said that "as microscopes did in their time," computational advances will provide "new ways to allow the biologist to approach questions that were not possible to approach in the past."

She noted that many biologists may not be aware of the possibilities computational tools and approaches offer and acknowledged that she had first turned down the offer to deliver a keynote at ISMB, since she "didn't feel she had anything to say" to computational scientists. She said she changed her mind, however, and determined that discussing the challenges of her field might help open the door for new computational approaches.

"Computational biology is essentially biology," Rost said. As computational scientists develop methods that are increasingly integrated into biology, the field will evolve, he added.

Allyson Lister, a PhD student at Newcastle University's School of Computing Science, said that even though her center was devised to bring together bioinformatics and systems biology researchers, on a five-year grant that is just drawing to an end, it took "a year or so" for experimentalists to "understand the merits of the other side," and begin collaborating with bioinformaticians and statisticians early in a project.

Rost explained that younger scientists will have it "embedded in themselves" to connect the two fields, which could help to resolve the existing computational bottleneck he sees "in many projects."

Clouds Ahead

Speaking for himself and not as the ISCB president, however, Rost said that

collaborations between researchers in computing and biology still face some challenges. "I see trouble ahead, I see clouds," he said.

For example, he noted that it is difficult to publish a bioinformatics method in a major journal, which is key for obtaining recognition and possibly a prestigious position in academia. "So we already have a built-in conflict," he said. "I don't know how to resolve it."

Myers agreed that it is "very hard" to get computational methods published in the biological literature. Even if these approaches are biologically relevant, there is still a tendency to favor the experiment, he said. For example, in genome-wide association studies, "methods there are everything," since the findings are about flow of information.

Marie-France Sagot, director of research at the Laboratory of Biometry and Evolutionary Biology at the University of Lyon and the French National Centre for Scientific Research, agreed that there is often tension between the experimental and computational disciplines because these scientists must apply for grants in the same places, but "do not have the same CVs" and differ in their career advancement prospects.

Biologists and bioinformaticians "still don't think together," said Sagot. "There is still a problem of dialog."

Another challenge is that many small and medium-sized labs can't afford to have bioinformaticians as part of the group from the onset of a study. What these researchers might end up doing is hire "low-level IT people, and they can't do the job," Rost said.

Reinhard Schneider, group leader at the European Molecular Biology Laboratory in Heidelberg, Germany, said he and his colleagues are often called "too late" — for example, when scientists have already had a sequencing instrument in place for six months.

"They plunk in one USB drive after the other, they lose data, and then they say, 'Oh, maybe we need to change something,'" as opposed to grappling with IT questions such as storage and backups ahead of time. Although he believes attitudes are changing, he said, "right now we are called more as the fire brigade after [the fact]."

At Newcastle University, some researchers with smaller informatics needs write into their grants that they would like to draw on the university's bioinformatics support unit, which is funded with grants from multiple projects across the university, Lister said. Scientists can build, for example, one tenth of a bioinformatician's services into their grants, she said.

Harvard's Hofmann said that in his experience, many scientists are not adding "the IT side" to their grants, which leads to insufficient funding for data post-processing, for example in second-generation sequencing experiments.

Remember Design

Besides having the right equipment, software, and addressing pipeline and storage questions, researchers often underestimate the role of experimental design in facilitating the computational analysis of large-scale experiments and yielding the information being sought, said Myers. "Sometimes, as an informatician, I am thinking, 'Gee if I just had this additional piece of information — if it had been thought about, it would have made a

difference."

Second-generation sequencing is empowering scientists, "but we do have economy of scale issues with respect to the informatics," he said. Prior to the advent of next-generation sequencers, a large genome center might have had 10 or 15 informaticians and 100 staffers running the sequencers. While the number of people needed to do sequencing has "gone from 100 to 1" with next-gen systems, the number of informaticians is "still at 15," Myers said.

The challenge in academic projects is that bioinformatics has not "figured out how to shrink-wrap code," he said. While not all of bioinformatics in genomics should be shrink-wrapped, the "day-to-day" tasks shouldn't require a staff of informaticians, he said.

This could offer an opportunity for the commercial sector, he said, "for carrying and supporting the less research-oriented aspects of the problem."

Stockholm University's Von Heijne pointed out that not all of biology or bioinformatics is made up of high-throughput pipelines. For example, at his center for membrane biology, a computational group takes on molecular dynamics and modeling, while a bioinformatics group works on topology and homology searches, working "on a small scale" with techniques that do not involve daily gigabyte storms.

Having these teams is "extremely important" for both biologists and informaticians, he said, because "they start talking about the same types of problems, small-scale problems," and combine perspectives to tackle challenges.

Genomeweb system

These settings are generally managed by the web site so you rarely need to consider them.

Issue Order: 2

