



Where Innovation Begins

Reed Business Science & Medi

Reader Service Card | Free S

Product Search
 Company Search

SUBSCRIBE 
 Renew or register for your **FREE** subscription today.

2001 Scientist of The Year



The Challenge of Bioinformatics

An Essay by David Haussler

- Current Issue & Archives
- Calendar of Events
- Web Product Showcase
- Laboratory Design Newsletter
- Micro/Nano Newsletter
- Resources
- Awards
- Editorial Info/Contact Us
- R&D Industry Guide
- Career Center
- Homepage
- e-Newsletter Archive

The sequencing of the human genome has opened a new door on two very old questions: What makes us human, and what makes us different from each other? Although the genome sequence itself, and indeed the best efforts of the entire field of molecular biology, will not be nearly adequate to answer these questions, in the next decade it is here that we will find the most intriguing new data pertaining to them.

Biology approaches these questions in a narrow but precise sense. At the level of genes and biomolecular pathways, it asks what makes one person's physical make-up different from another's, what makes us all different from the other great apes, what makes the great apes different from the other primates, and what makes the primates different from the other mammals. Currently we know relatively little about the answers to these key questions. In 10 years, we will know a lot more. It may change the way we think about ourselves and our place in the natural world. Molecular investigation of humanness will never and should never replace psychological, philosophical, social, and religious soul-searching. But its facts should inform these broader dialogues.

How will this new knowledge come about? The first steps have already been taken. We have developed the technology to sequence entire genomes, and we have applied it to our own genome. In the next few years we will have public versions of not only our genome, but the genomes of the chimpanzee, the laboratory mouse, the zebrafish, and many other species (see "Science," v. 294, Oct. 2001, 82-85). We will have a genome zoo. The assembly and analysis of the data from this zoo will be an enormous task for the relatively new applied science field of bioinformatics.

Pulling it together

Bioinformatics interfaces with basic science in biology and chemistry--in particular biochemistry, molecular biology, cell biology, and developmental biology--with molecular and population genetics, and with several other applied fields including medicine, computer science, and biotechnology. It forms a kind of interdisciplinary glue as all of these fields become more information intensive, and find they cannot easily separate themselves from the vast web of molecular entanglements that make life possible and that respect no disciplinary boundaries.

- Sp
- P
- U
- UPD
- COM
- x
- x

Larl
 Inc.
 Qua
 Seq
 Ban
 Gen

 E-mail Newsletter

 Register



**LABORATORY
INFORMATICS
EXPO2003**
ONLINE CONFERENCE AND CELEBRATION
November 5, 2003

Despite the cautions mentioned above, these boundary-breaking molecular pathways are an important part of what makes us human, and knowing them will enrich us. Yet they are too vast and complex to fit in a textbook. They are far too intricate, too baroque, too nonintuitive, and too damn quirky to be deduced from simple mathematical first principles. They can't in any way be reduced to an organized theory such that they all fit entirely within the head of a single scientist. So specialists must be able to tie into a common infrastructure of knowledge, data, visualization, analysis by algorithm, model by machine, and simulation *in silico*. The task of bioinformatics is to build this infrastructure.

The genes have it

Who are the molecular players in this new web of knowledge? One important set of players is the genes. These have evolved together in complex communities--sheltered, nurtured, and propagated by the various species on this planet. Genes have fascinating family histories within the evolution of the species. But the history of each family of genes is distinct in important ways from the history of the species they live in, a fact that creates challenges for comparative genomics. It remains a significant and challenging bioinformatic task just to pick out and properly match up all the primary DNA components of the genes in a gene family, the codons that make amino acids, and the regulatory sites that control the expression of the gene, so that we can begin to study the molecular evolution of a gene family. This is especially true of rapidly evolving gene families, which may be the most important in distinguishing us from the great apes (see "Nature," v. 413, Oct. 2001, 514-519).

Genes play distinctly different functional roles in the species they support as well, and we cannot unravel these functions without understanding how they act in pathways and how these pathways form a system. Gene chips can measure the expression levels of all the genes at once in a wide variety of different individuals, different tissues, and different conditions--and can detect differences in the gene sequences themselves between individuals. They have enormous potential to help us understand the functions of individual genes, and how differences in particular genes can lead to different responses to treatment and different susceptibility to disease. Bioinformatics has a vital role to play in the assimilation, analysis, and integration of this crucial data into our collective web of knowledge about vital life processes.

The new medicine

According to the central dogma of molecular biology, genes beget RNA molecules, and (usually) these beget proteins, which do most of the work in the cell. So the study of genes is inexorably tied to the study of these other vital molecular players. A much-touted new field of proteomics has emerged to directly study these players en masse, including their structures, interactions, and life cycles within the cell. A large part of bioinformatics is devoted to proteomics, and this area has enormous potential. But we must be cautious in defining the evolution of the field of bioinformatics in analogy with the central dogma of molecular biology, as some have. Proteomics will not replace genomics. The fundamental object of molecular evolution, the logic of control, and the very definition of the species all lie in the genome.

In 10 years we will still not know what makes us human. We still won't know what really makes us different from one another. But with the help of bioinformatics, we will know a lot more about ourselves than we did before. A decade hence, when you go into a doctor's office, your doctor may first access massive amounts of information about your particular genetic make-up. He or she may then make massive numbers of molecular diagnostic tests in parallel to assess what is going on in the pathways within your body. And then, as part of final diagnosis and recommendation of treatment, he or she may undertake a computation on this

data that would definitely be considered massive in comparison to the common medical computations today. So bioinformatics may help dramatically “enlarge” medicine.

Society is just now waking up to the changes in store as we assimilate this new knowledge. People are asking the hard questions: Will my employers or my insurance company be allowed to obtain my genetic information without my consent, and will they be allowed to discriminate against me on that basis? Will the new medicine be shared widely or only be available to a privileged few? How far should individuals be allowed to go in selecting the genotypes of their offspring? We cannot reap the benefits of the coming revolution in health care without also addressing these societal questions. It is time for scientists, educators, religious leaders, policy makers, and indeed every citizen to be involved in an open discussion of these and other important questions. Perhaps the most important challenge of all for bioinformatics is to enlighten this discussion.

On the Road to Excellence

David Haussler, this year's R&D Magazine Scientist of the Year, credits some of his research accomplishments to the strong support provided during his college studies, first by his older brother during his undergraduate studies and then by his doctoral advisor.

Andrzej Ehrenfeucht was Haussler's doctoral advisor in the computer science department at the Univ. of Colorado in Boulder. Haussler specifically picked Colorado for his studies because of Ehrenfeucht's reputation and knowledge in the theory of computation and interaction between computation and logic. Currently a professor in computer science at Colorado, Ehrenfeucht's publications in the theory of computation go back nearly 30 years.

The breadth of Ehrenfeucht's studies are exemplified in one of his current research projects, a collaboration with Patricia Baggett at New Mexico State Univ., Las Cruces, on mathematics and technology in primary schools. A Web site demonstrating the results of this work can be found at



Andrzej
Ehrenfeucht



Mark
Haussler

A Web site demonstrating the results of this work can be found at <http://math.nmsu.edu/breakingaway/Lessons/lessons.html>. “My study with Ehrenfeucht was my first exposure to true genius, he's a brilliant interdisciplinary thinker,” says Haussler. “I had to work hard just to try to keep up with him.”

Haussler's older brother, Mark Haussler, is Regents Professor in the department of biochemistry at the Univ. of Arizona, Tucson. David Haussler credits his brother's early influence in the biological sciences and basic laboratory experimentation on helping him better understand the overall research process. “My brother invited me to come out to Tucson during my undergraduate days to help in his research on measuring the hormonal form of Vitamin D in human blood,” he says. No one had previously been able to design an assay that was sensitive enough to measure the Vitamin D levels in human blood. David Haussler also was given the opportunity to do the math analysis of his brother's research.

[Free Subscription](#) | [Advertising/Editorial Info](#) | [Contact Us](#)
