

## **BIOCOMPUTING EDUCATION II: FURTHER CHALLENGES AND OPPORTUNITIES “The Workshop Setting”**

Session Chairs: SUSAN J. JOHNS and STEVEN M. THOMPSON  
*Center for Visualization, Analysis, and Design in the Molecular Sciences (VADMS)*  
*Washington State University*  
*Pullman, WA 99164-4660*

The PSB '97 Biocomputing Education Session will focus on “the workshop format: How to fund, organize, and implement for success.” It will be structured to encourage the open exchange of information regarding teaching philosophies, teaching materials, and teaching experiences — “how” to teach biocomputing — especially using the workshop format, between all participants. The session will consist of an hour long discussion forum with a panel consisting of individuals who have presented successful biocomputing workshops and other interested participants. We will discuss the ins and outs of teaching biocomputing using a workshop format, describe some of the methodologies employed, and explore curricular issues. Furthermore, Bruno Gaeta, Tim Littlejohn, Christopher Dubay, and Susan Johns will recount the ideas espoused in their respective poster presentations. The panel will be kicked off with an introduction by Mitchell L. Sogin in which he will relate some of the details and rationale behind the highly successful Marine Biological Laboratory (MBL) Workshop on Molecular Evolution that he and Dan Davison have directed for nearly the past decade.

The MBL Molecular Evolution Workshop provides interdisciplinary training in molecular phylogeny/evolution and promotes collaborative efforts between molecular biologists, cell and developmental biologists, organismal biologists, and theoreticians. Leaders in the fields of Molecular Evolution, Population Biology, Developmental Biology, and Computational Biology present a lecture series that is integrated with a computer laboratory. During the initial phase of the workshop, sixty students attend an intensive two week lecture/laboratory course. In the second phase of the workshop, as many as fifteen students participate in a one week special projects session. Topics covered during the two week lecture series include, but are not limited to, the theory of origins of life, phylogenetic reconstructions from molecular data, macromolecular sequence alignment techniques, current trends in molecular systematics, molecular approaches to studies of population biology and biodiversity, comparative molecular biology, evolution of development, and comparative genomics. In the computer laboratory, high speed machines running a variety of operating systems are used to test the theoretical concepts described in the lectures series. Guided by the authors of phylogenetic analyses software and the designers of computer packages for analysis of gene sequences, the students learn different retrieval and evolutionary analysis techniques for molecular data. The goals of the workshops include:

1. To promote the application of molecular techniques to questions about origins of life, evolution of organisms, population biology, ecology and developmental biology.
2. To train molecular biologists in the use of phylogenetic techniques as a laboratory tool.
3. To establish interdisciplinary collaborations between organismal biologists, molecular biologists, developmental biologists, and mathematicians.
4. To provide a strong background in evolutionary theory for molecular biologists, mathematicians, and undergraduate educators.
5. To train students in the use of software for data base searching, sequence analysis, phylogenetic inference, and analysis of genome data implemented on a variety of computer platforms.

To maximize return on investments in comparative molecular studies, it is important to train graduate students, postdoctoral students, and principal investigators in the techniques of molecular evolution and phylogenetics. In 1988, the Marine Biological Laboratory at Woods Hole recognized the importance of molecular evolution and organized a workshop that spanned the interests of organismal biologists, molecular biologists, computational biologists, and theoreticians. Co-directed by Mitchell L. Sogin and Dan Davison from the University of Houston, the workshop received funding for the period 1988 through 1990 from the Alfred P. Sloan Foundation. The National Science Foundation assumed financial support for the workshop (1991-1993) through interdisciplinary funding initiatives. In 1994 NSF funding was renewed for 1994-1996 with a significant increase in the budget request. The partially funded increase was used to offer a special topics session which proved to be very popular. In 1996, NASA provided funds to supplement the workshop so that the special topics session could be offered in 1996, 1997 and 1998. The course was offered for the ninth consecutive year in August of 1996. Over the past nine years, more than 500 students have attended the workshop.

Sixty students are admitted to the workshop according to their backgrounds and research interests. Fifty percent of the students in the workshop are graduate students with the remaining positions being filled by postdoctoral students and established investigators. As part of the admissions process, a balance is sought between organismal and molecular biologists. Applications from biometrics/mathematics students are more rare and are given special consideration. The evolution of development and the comparative genome sciences are also considered to be important communities that must be served by the molecular evolution workshop. Finally, participant positions are reserved for several faculty who represent the interest of undergraduate education.

The general format of the Workshop in Molecular Evolution has adopted a Gordon Conference style with morning and evening lectures. Most afternoons are reserved for demonstrations and for the student use of computer facilities. Sogin is responsible for the general organization of the workshop including recruitment of faculty. Davison is well known in the genomics community and is largely responsible for assembling the complex array of high performance machines used in the computer laboratory. In a departure

from earlier versions of the workshop, a third week of special topics is now offered for fifteen students. Students are accepted to this extended format on a competitive basis. The first day of the special topics session is devoted to students making short presentations describing their own research projects. Each of the subsequent mornings, the students meet as a group for two or three hours and discuss their progress. According to the student evaluations, the success of the special topics session is tied to the nearly constant interchanges between the faculty and other students as well as the access to high performance machines.

Each student is provided with a computer environment relevant to hardware configurations that might be available at their home institution. Access to cutting edge technology in the form of Alpha machines from Digital is also offered. The integration of the computer laboratory with the lecture series is a distinguishing feature of this workshop. In addition to sixteen Macintosh computers (Power PC's) and ten pentium based PC's, the 1996 workshop laboratory was equipped with five Digital Alpha machines running OSF/1 (kindly provided by the Digital Equipment Corporation), five SUN Sparc Stations, a Silicon Graphics workstation, and a VMS cluster consisting of a VAX 4200, a VAX Station 4000/90, a VMS Alpha workstation and a Digital Alpha 2100 Sable server equipped with multiple EV-5/300 mhz processors. All of the machines are configured to allow file sharing/transfers and telnet connections. The high performance machines can be accessed via the PC's or numerous NEC-X terminals.

Participants in the course are divided into small, platform-dependent, teams led by a student with experience in the given computer environment. The students are provided with size limited data sets for exercises in the alignment of macromolecular sequences and their phylogenetic analyses. They are challenged with problems that can only be solved using the computer tools made available. The workshop faculty describe and demonstrate alignment techniques and multiple sequence alignment editors. The students learn how to import data sets into the interactive alignment editors, generate output formats that are compatible with phylogenetic analyses packages, run conversion programs for changing formats, and use the phylogenetic inference packages provided. The most current versions, often beta releases, of PHYLIP and PAUP are loaded on all of the machines and demonstrated to all students. Students who are not familiar with internet communications are instructed in the general use of telnet and ftp software for making remote connections and transferring files.

Based upon prior experience, the workshop organizers have learned how to optimize use of the computer resources so that the students can gain an appreciation of the computational limits of various phylogenetic analysis packages. However, one of the driving factors in upgrading the computer laboratory has been the increased demand by students to use higher performance processors. The students are required to demonstrate proficiency and understanding of computational limitations of algorithms before they are allowed to work with data sets brought to the course. After gaining an appreciation of sequence alignment tools and phylogenetic inference packages, the workshop considers strategies for searching molecular data bases and extracting sequences for further analyses.

Since its inception, the workshop has undergone major changes in response to the suggestions and ideas of students and faculty. Sessions on the inference of molecular phylogenies have been expanded and now represent a broader range of methodologies and philosophical points of view. Over the last three years, the speed of computers in the laboratory have dramatically improved through collaborative arrangements with vendors and use of machines from the Program in Molecular Evolution at the MBL. The course has evolved from a series of theoretical lectures in molecular systematics and organismal evolution (1988-1990) to a broad based survey of molecular evolution that is now augmented by a computer based laboratory.