UF Computational Biology:
Recommendations for Establishing National Prominence

November 13, 2008

Proposal

Establish Computational Biology as a center with a mission to develop and support world-class biological research requiring significant computational support. Search for and appoint an internal director tasked with development of UF Computational Biology. Develop training, collaboration and infrastructure for data-intensive biological research. Recruit faculty to computational biology. Seek a Legislative Budget Request (LBR) for on-going support.

Background

Disciplines of basic and applied biology, including medicine, plant and animal breeding, genomics, systems biology and ecology, are becoming ever more data-intensive. This trend will accelerate tremendously in the next several years as massive throughput DNA sequencing technologies, imaging, remote sensing, precision agriculture, and genomics information are applied to improving human health and nutrition, reducing human impacts on the environment, and enhancing productivity of agricultural systems. This new era of data-intensive science has fostered an urgent need for development of new computational tools and approaches that enable integration, interpretation and visualization of complex, multidimensional datasets (see Appendix A). In order to be effective, computational approaches need to be tailored to the unique problems and challenges presented by each field of application. To meet this challenge, this University-wide initiative will foster creation of new positions in computational biology research.

UF has tremendous strength in the biological sciences – in plant, animal and human -- as well as significant strength in computation applied to other disciplines, such as physics and chemistry. Computational biology leverages these two strengths by providing required research and support for informatics needed to answer important biological questions (see Appendix B) using massive amounts of genetic, molecular and other data, much of which is produced locally at UF.

The current model of funding research is challenged by the needs of computational biology. Biological scientists need significant partnerships and support to address problems requiring the manipulation of unprecedented amounts of data, as well as enhancements to fundamental computing infrastructures. Some universities are beginning to address these challenges (See Appendix C). UF is poised to become a leader with appropriate direction and resources.

Recommendations and Timeline

1. Conduct an internal search for a Director of Computational Biology (see Appendix D). The successful candidate will be appointed half-time as director. The director will administer the UF start-up funds for computational biology, expected to be $600K per year for 5 years.

2. Engage existing faculty, staff and students. The director should organize workshops for UF faculty, followed by a series of national workshops supported externally. A highly visible, open, renewable training program with national and international participation should be developed to continuously introduce new researchers to fundamental principles and methods of the discipline.
3. Build external funding through a competitive process of supporting seed projects. The center will support competitive funding for projects leading to external funding, travel support for investigation of other programs in computational biology, and support for a computational biology speaker series.
4. Recruit additional faculty to computational biology. Targeted hiring of two new positions per year over the next five years will build required strength in areas of need. The hires and start-ups should be funded by the Provost as a strategic initiative. The new hires can be at any level in any college supporting computational biology activities.
5. Identify and support finding and developing collaboration opportunities. UF should develop and support a faculty commons similar to the "VIVO" project at Cornell, which provides standardized searchable information on faculty interests and activities. UF can then leverage such a system to route information of interest to faculty members (see Appendix E).
6. Transition the Computational Planning Team to a Computational Biology Advisory Committee following the appointment of a director. The Advisory Committee provides an annual program evaluation and advises the director and the Research Advisory Council regarding program direction and effectiveness (see Appendix F).
7. A Legislative Budget Request (LBR) will be developed immediately upon the establishment of the Center to provide core IT and informatics support for computational biology investigators. The LBR should be for approximately $6M a year in recurring funding. The emphasis of the LBR is on molecular-level research support leading to advances in water quality, food production, and improvement in human health. The LBR addresses the fundamental funding problem in computational biology – individual awards cannot provide the research and support needed to build and maintain the collaborative infrastructure required for the solution of massive data problems in biology. The LBR should be named as the #1 UF priority and remain #1 until funded.

Appendices

A) The promise and challenges of representation and modeling of complex biological systems in computational biology
B) Example Problems. The Tree of Life, Translational Research
C) Experience of Other Institutions – Harvard/MIT, Idaho, Arizona
D) Job Description for the Director of Computational Biology
E) VIVO at Cornell
F) The Computational Biology Advisory Committee

Appendix A

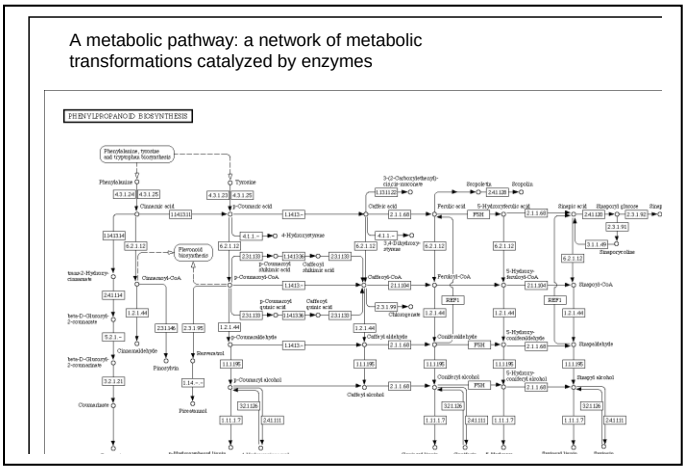
The promise and challenges of representation and modeling of complex biological systems in computational biology

Biological systems are extremely complex. Biology deals with entities ranging in scale from single molecules to entire ecosystems, with dynamic processes involving thousands of components interacting in non-linear ways, with time spans ranging from fractions of a second to thousands of generations. Biological processes are in many cases not directly observable, and have to be analyzed through indirect, noisy and error-prone experimental methods; even providing a precise definition of a disease or of a human population is often very difficult. As a consequence, the representation of biological information and concepts in a form that can be automatically manipulated poses enormous challenges. An extremely wide spectrum of data representation formalisms is being employed, from purely numerical datasets to qualitative descriptions, to statistical and computational models, to text-based sources of information, to graphical visualizations. One of the most important challenges that modern biology will face is how to integrate an exponentially growing mass of extremely heterogeneous data that represents information at very different levels of abstraction and detail in order to generate new knowledge.

Among the most important types of representations used by biologists are various map, tree, pathway and network diagrams that describe relationships and interactions between elements in complex biological systems. The elements are typically species, genes, proteins and other molecules.

1. The maps are frequently based on natural coordinates of the system (size and weight, anatomical structures, and developmental time for species; genetic distance, physical distance measured in base pairs or angstroms, and evolutionary time for genes, proteins and other molecules, etc).
2. Trees summarize evolutionary relationships between species or genes.
3. Networks may include any set of relationships or interactions that can be summarized by some sort of graph containing nodes and connecting lines or arrows (called ‘edges’ in graph theory).

Some examples are illustrated in Fig. 1 on the next page.



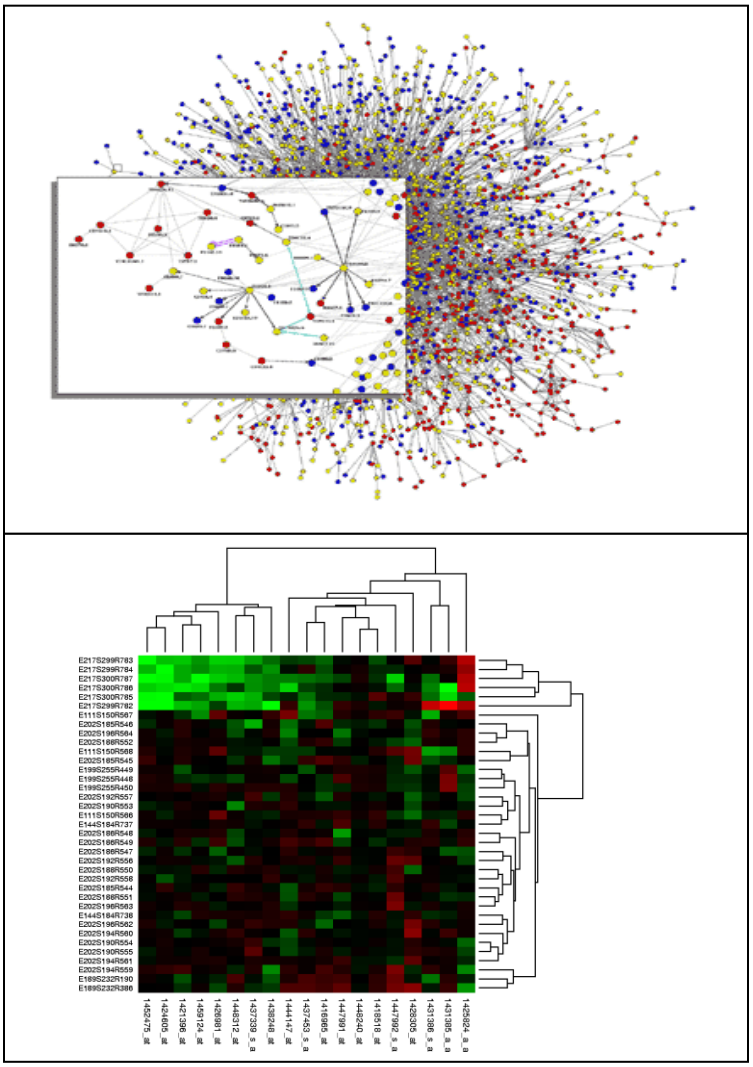


Fig 1. (Top) A metabolic network. (mid) A network of protein-protein interactions in a cell. (bottom) A cluster analysis showing correlations in gene expression from an expression profile.

Phylogenetic trees are usually calculated from alignments of multiple DNA or protein sequences. The tree is a model for how the given set of aligned sequences would most likely have derived from a single common ancestor. Typically an arrow in a pathway or network diagram summarizes a process at a deeper layer of complexity such as a biochemical reaction catalyzed by an enzyme, a physical interaction between proteins or a DNA-protein interaction that regulates transcription of a gene. Often the details of the underlying processes are not well understood, except for a few representative cases. In the case of a genetic network the type of molecular interaction involved may not be specified at all. For this reason, systems biologists have focused a lot of attention on seeing how far one can get by analyzing the topology, connectivity and scale properties of the networks. Networks can also be modeled quantitatively to explore predictions and hypotheses that can be tested experimentally. Mathematical models that exploit analogies to electrical circuits and Boolean logic networks, as well as, symbolic logic and formal language concepts derived from computer science have been applied to analysis of biological networks.

For mainstream biologists, a more fundamental role of these structures is that they establish associations between disparate elements of the biological system. For example, a group of genes might be associated by linkage in a genetic or genome map, by their functioning in a common pathway, by sequence similarity or evolutionary relationships represented in a tree, by their connectivity in a physical interaction network, or by a cluster analysis of co-regulated genes derived from a transcriptome profile. Each of these associations brings a different set of elements and information in to play. Hence, there is a lot of power in

1. constructing these representations (e.g. trees, genome assemblies, networks and clusters) from large datasets and
2. analyzing relationships and correlations between various representations. This is illustrated with a couple of examples below.

Genome maps

One of the simplest and most powerful examples is a genome map that shows the locations of genes and other sequence based information along chromosomes (large single DNA molecules). A genome map of an organism consists of one or more linear (or circular) DNA sequences, one for each chromosome. Typical chromosomes are several hundred million base pairs long (e.g. the human genome consists of 3 billion base pairs of DNA divided among 23 linear DNA molecules, whereas, the genome of the maize plant is 2.5 billion base pairs split in to 10 molecules). Given the complete DNA sequence of the genome, in principle, the exact location and sequence of every gene can be found. Plant and animal genomes typically contain 30,000 to 50,000 genes. In addition to genes, a multitude of other features can be mapped on the genome such as chromatin modifications, regulatory DNA sequences, microRNA's, mRNA splicing variants, genetic variants, etc. For extensively studied model organisms there may be hundreds of layers of data available that can be overlaid on the genome sequence. Hence, once a gene or sequence is implicated in a biological process – say by linkage to a phenotype of interest - the genome map automatically enables correlations with a host of other types of map-based data.

An ideal computational environment would enable a researcher to move facilely between various representations. A situation that might arise in quantitative genetics would be the sequence: genome map -> canonical pathway -> genome map. Suppose you have mapped six different quantitative trait locations (QTL's) that account for the variation in a phenotype of interest in some organism. Based on proximity in genome map, you identify gene A as a candidate for one of the QTL's. In another model organism, a homolog of gene A is implicated in a metabolic pathway that also includes genes B, C, D, E and F. If a similar pathway is responsible for your phenotype then a reasonable hypothesis is that other homologs in the pathway (B, C, D, E and/or F) will be linked in the genome to the other QTL's. Since you know the map locations of your QTL's this can be readily tested.

Another example, taken from the work of Prof. D. McCarty at UF, went something like this: pathway -> genome map -> multiple sequence alignment -> gene tree -> epigenome map -> decorated tree -> multiple sequence alignment -> protein structure -> protein structure model. Well, we are not quite there yet. In this case, we started with the regulatory network shown in Fig. 2 that includes two families of closely related genes that encode B3 type DNA binding proteins in plants, the AFL (ABI3/LEC2/FUS3) and VAL families, respectively. The AFL genes are positive regulators of the embryo development pathway in the plant seed; whereas, the VAL genes repress the embryogenesis pathway prior to seed germination. The set of B3 transcription factor genes were extracted from the genomes of rice and Arabidopsis and those sequences were aligned and used to construct a tree (Fig. 3 top). Among other things, the tree can be used to infer whether gene duplications within the family occurred before or after the evolutionary separation of rice and Arabidopsis 120 MYA or so. For example, the clade containing the OsCEB genes highlighted in blue is unique to rice. We have a hypothesis that the CEB clade is associated with a developmental process that occurs in rice and other grasses, but not in Arabidopsis.

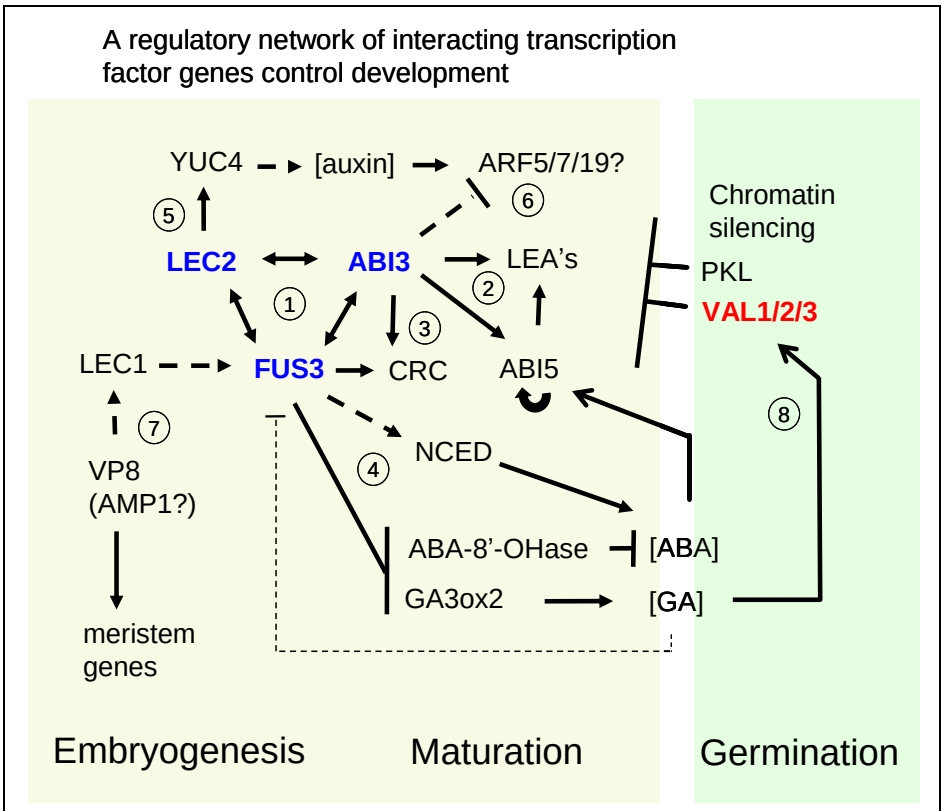
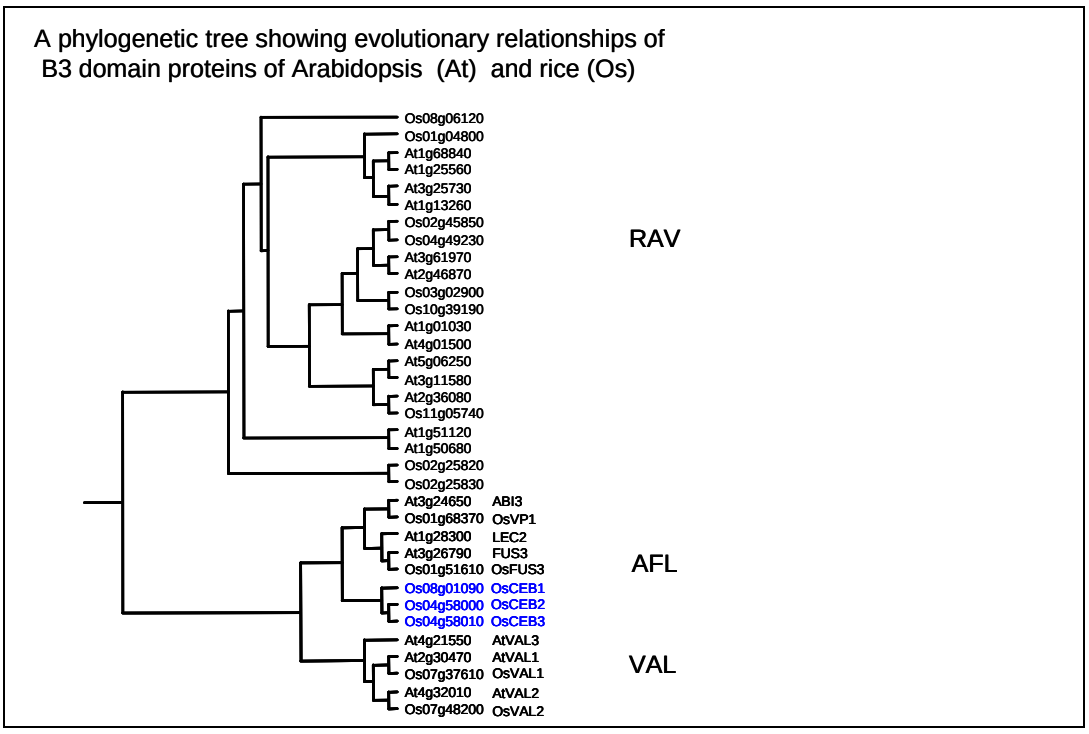


Fig 2. A transcription factor network regulating plant seed development. Arrow heads indicate positive regulation (transcriptional activation), flat headed lines indicate negative regulation (repression of transcription).



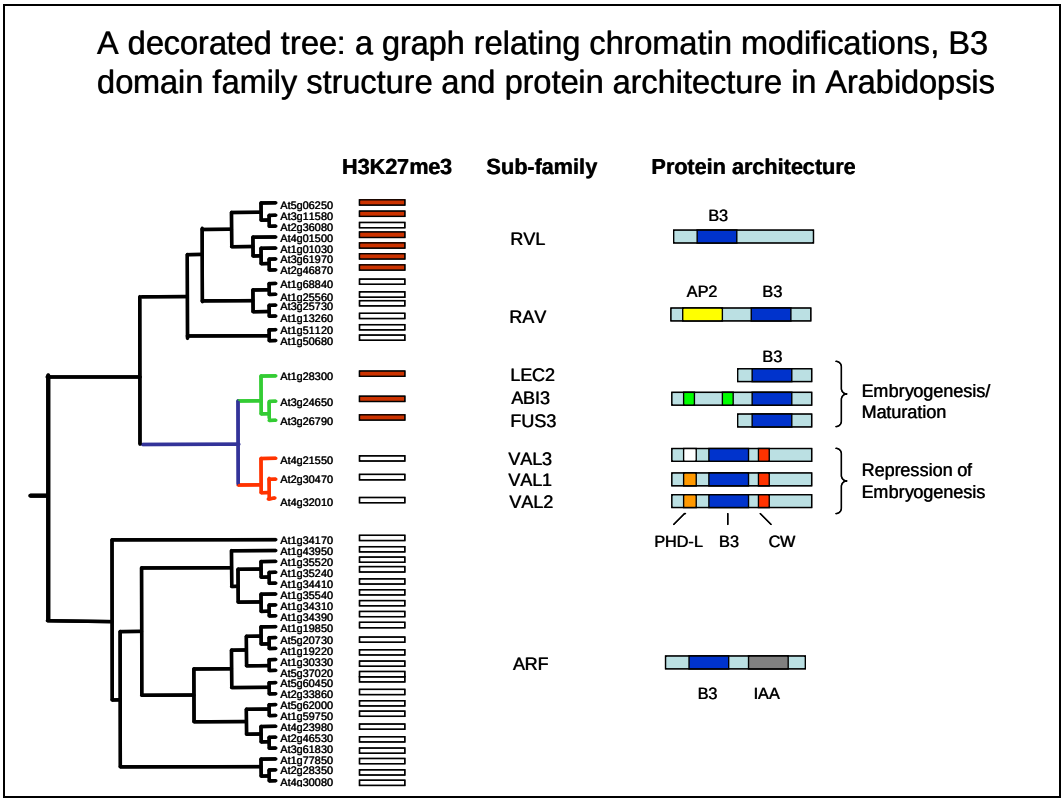


Fig 3. (Top) A tree of B3 domains from rice and Arabidopsis. (bottom) A tree of Arabidopsis B3 domain proteins overlaid with the pattern of H3K27me3 chromatin modification observed in seedlings.

Decorating the tree with other data

The tree of Arabidopsis B3 domain proteins can also be overlaid with other available data from the genome map or other sources (Fig. 3, bottom). In this case, key chromatin modifications have been mapped in the Arabidopsis genome. Among dozens of post-translational modifications of histone proteins found in eukaryotic chromatin, trimethylation of lysine (K) 27 of histone 3 is highly correlated with repressed (transcriptionally inactive) chromatin states. Decorating the B3 tree with data on the level of H3K27me3 modification in AFL and VAL genes in seedlings reveals a clear pattern, all three AFL genes carry the H3K27me3 modification, whereas, VAL genes do not. A similar pattern is evident in the H3K27me3 status of the RAV and RVL families suggesting that an analogous functional symmetry may exist between those clades, as shown in Fig. 3.

Finally, underlying the tree are the amino acid sequence differences contained in the multiple sequence alignment. The amino acid differences that distinguish the VAL and AFL clades are potentially correlated with functional differences in the proteins. One way to approach this is to map the locations of conserved and clade-specific amino acid variants in a known B3 domain structure (Fig. 4). This leads to hypotheses that can be tested experimentally. Ideally, given proper tools one can go a step further and use molecular modeling programs to create detailed structural models for each family of B3 domains.

Computational challenges

1. Automating the process of constructing multiple sequence alignments suitable for construction of high quality trees.
2. Computing trees for large numbers of sequences.
3. Identifying and assembling genes for canonical pathways in sequenced genomes.
4. Data integration across diverse datasets and sources.
5. Automation of protein modeling and visualization.
6. Interactive, virtual-reality immersive visualization of complex, multi-dimensional data structures.

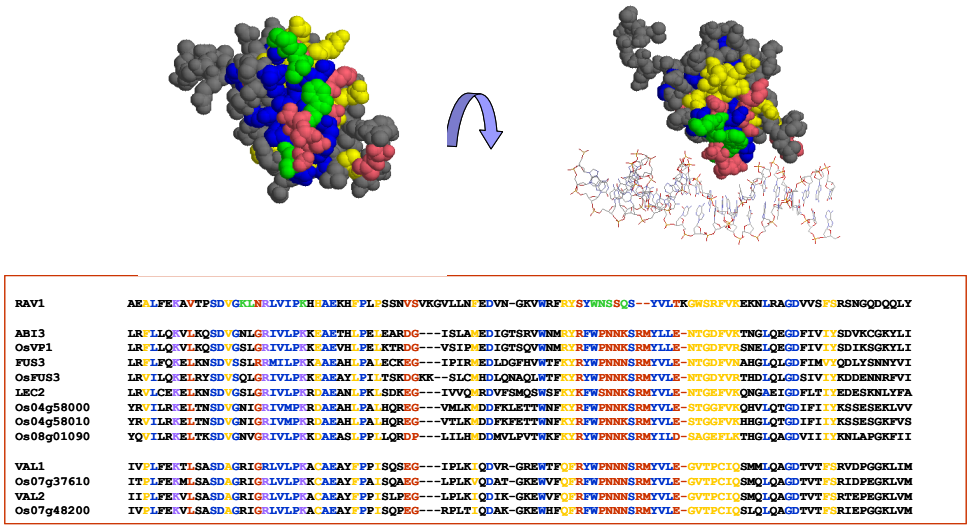


Fig 4. Relating alignments with protein structure. Conserved and clade-specific amino acids are highlighted on the putative DNA binding surface of the B3 structure. The RAV B3 domain is docked in the major groove of DNA.

Appendix B

Computational biology is replete with challenging fundamental scientific problems. Here we present sample problems requiring cross disciplinary teams of biological, computational, medical and other scientists.

The Tree of Life

The Tree of Life is a metaphor for connecting all living organisms in their historical context. However, it is more than a metaphor, and large-scale efforts are underway around the world to reconstruct the evolutionary history (phylogeny) of life on Earth, the Tree of Life. This Tree, when reconstructed, will serve as a major reference for all other areas of biology. When the Tree is linked with the traits of organisms (whether molecular, biochemical, ecological), new perspectives and inferences will emerge. For example, association of the Tree with locality and habitat data for the millions of museum specimens held worldwide (the Florida Museum of Natural History is home to > 25 million specimens and artifacts) will yield information relevant to climate change and extinction. Furthermore, the methodologies developed for reconstructing the evolutionary history of all species can be equally applied to the evolutionary history of genes (or any other trait), permitting new approaches to understanding the history of specific gene families and enabling functional genomics.

The Tree of Life can be built from DNA sequences. However, phylogeny reconstruction is a complex computational problem because the number of possible trees increases exponentially with the number of sequences/species included in the analysis. For even a modest number of DNA sequences, each representing a single species, for example, the number of possible trees is mind-boggling: for 10 species, there are roughly 2 million possible trees that must be compared to select the one that is best supported by the data, for ~200 species, there are more possible trees than there are atoms in the universe. Given that the estimated number of living species today is somewhere between 1 million and 10 million (or more), the scope of the computational problem associated with linking all of life in an interconnected framework is immense. Despite significant progress in recent years, the magnitude of the task that remains is daunting. Data-mining pipelines are underdeveloped, as are tools for the assembly, analysis, and visualization of larger and larger trees. Proper tools do not yet exist to integrate and display data from multiple sources (e.g., molecular sequences, genomic data, expression data, morphological and developmental data, fossil evidence), and the results of most phylogenetic studies have not been assembled and made readily accessible, and, therefore, remain effectively unavailable to the wide variety of potential user communities in research and in education.

An expanded intersection between phylogenetic biology and computer science, computational biology, and bioinformatics is clearly needed. Some important attempts have been made recently to bridge this gap, such as the NSF-funded CIPRES project, an $11-million effort for CyberInfrastructure for Phylogenetic RESearch (B. Moret, University of New Mexico, original PI). This effort has highlighted the enormous needs that exist, and has also emphasized the need for phylogenetic biologists and their colleagues in the relevant computer sciences to craft real-

world solutions both to deliver better phylogenetic inferences and to render this information truly useful and enabling for research and education.

Building an effective cyberinfrastructure for phylogenetic biology will require solutions to a host of practical and theoretical problems in computer science, bioinformatics, and phylogenetics itself. Some of these problems are well characterized but computationally daunting, such as developing good heuristic solutions for several NP-complete problems for which very large data inputs are now at hand (multiple sequence alignment; construction of phylogenetic trees; assembly of synthetic "supertrees"). Others are not so well characterized but are emerging as problems unanticipated before the availability of massive quantities of DNA sequences and other data (phylogenetic incongruence between different regions of the genome; the complexities of gene family diversification; the role of whole-genome duplication in species diversification). Still other problems arise because of the breadth and heterogeneous nature of data that can be brought to bear on building phylogenetic trees: morphology, development, gene expression and other post-genomic data, etc. The informatics problem of data integration is exemplified by the vast array of diverse data sets that retain footprints of evolutionary history.

Few of these problems currently have off-the-shelf solutions. There is, however, substantial prior experience in the phylogenetics community for building a next-generation phylogenetics cyberinfrastructure. First, the diversity of mathematical and computational methodologies for solving specific problems in phylogenetics has never been higher. This has been enabled in large part by substantial buy-in from the math and computer science communities. Through its Assembling the Tree of Life and other programs (3 AToL projects funded at UF), NSF has supported several smaller infrastructure-related projects, including TOLWeb, TOLKIN, PhyLoTA, pPod, and others (TOLKIN is housed at UF, pPod has a UF co-PI). Moreover, many bioinformatics resources have added phylogenetic components in recent years (e.g., the phylogenetic trees in PFAM and GenBank's BLAST server). However, an overarching infrastructure fostering true high-level integration is still largely missing. These efforts need to be focused and coordinated, and much more attention needs to be given to synthesizing knowledge and making it truly useful for research and education. With UF's current expertise in phylogenetic biology, genetics, statistics, and computer science, UF is poised to take a leading role in the development of the tools required to reconstruct the Tree of Life and in training a new generation of computational biologists who are equally comfortable with evolutionary history and computation.

Translational Science

Biological research is undergoing a fundamental transformation in its methodological and scientific approach, in response to the rapid and widespread adoption of *high-throughput experimental technologies*. The most visible consequence of these advances is an exponential increase of the amount of data produced by each experiment, at all levels (from DNA sequencing to genotyping, to gene expression analysis, to proteomics, to high-level observations on genotype/phenotype correlations), and at a steadily decreasing cost. This scenario opens up unprecedented new opportunities for studying biological systems on a large

scale, with a holistic perspective that promises to expand our understanding of biological processes and of their connections with clinically relevant findings. Fulfilling the goals of translational science, the discipline that tries to bridge the gap between investigation at the molecular level and high-level medical findings, will require the ability to turn experimental data into new knowledge, by automatically combining, transforming and interpreting them in novel ways.

We are therefore witnessing a shift from hypothesis-based to hypothesis-free research, in which the data, rather than being used to confirm or disprove preexisting hypotheses, are used both to generate hypotheses and to validate them, in an iterative process of successive refinement and analysis. The consequence of this paradigm-changing evolution is that researchers increasingly need to handle very large volumes of heterogeneous data, including both the data generated by their own experiments and the data retrieved from publicly available repositories of genomic knowledge. Integration, exploration, manipulation and interpretation of data and information therefore need to become as automated as possible, since its scale and breadth is, in general, beyond the limits of the competences of any individual researcher and of the basic data management tools in normal use. The “traditional” data inspection and analysis methods are quickly becoming inadequate in a scenario in which an investigator can sample hundreds of thousands of variables in parallel. Ad-hoc analysis methods need to be developed in order to address the obvious problems with statistical significance of results, and new data storage and retrieval systems are needed in order to handle the unprecedented volumes of data and information being generated in an efficient and productive way. But more importantly, all phases of the scientific discovery process (background knowledge gathering, experiment design, hypothesis generation and testing, interpretation of results, generation of new knowledge) will have to adapt to this new reality. In an era in which an entire new genome can be sequenced and annotated in a matter of days, it will become essential to be able to automatically link new observations and findings to preexisting knowledge, and to quickly establish relationships between heterogeneous datasets.

Computational biology will therefore play an increasingly central role in translational science, as our understanding of the deep disease-causing mechanisms at the molecular level grows. It will allow us to discover new links between genetic factors and pathologies, to make better predictions on the consequences of mutations and on the effectiveness of therapies, to design drugs that are tailored to specific classes of patients, and to elucidate the role of environmental factors in the development of diseases. UF researchers are at the forefront of these efforts, thanks to the close proximity and strong partnerships between its clinical enterprise and its first-class basic science departments.

Appendix C

Experiences of Other Institutions

Harvard-MIT Division of Health Sciences and Technology

The HST was established in 1970, when MIT and Harvard Medical School agreed to develop a joint program in medical science. Since then, HST has expanded to include doctoral, master's, and training programs. HST today brings together MIT, Harvard Medical School, Harvard University, Boston area teaching hospitals, and multiple research centers in a unique collaboration that integrates science, medicine, and engineering to solve problems in human health. Overall, 400 graduate students of science, medicine, engineering, and management receive their training under the supervision of 60 full-time and 200 affiliate faculty.

The HST's training activities are organized over seven distinct training programs, four at the doctoral level (Bioinformatics and Integrative Genomics, Neuroimaging, Bioastronautics, and Speech and Hearing Bioscience), two at the post-doctoral level (Biomedical Informatics Program and Clinical Investigator Training Program), and the HHMI-funded Graduate Education in Medical Sciences program. The three primary research focus areas are: Biomedical Imaging, Biomedical Informatics and Integrative Biology, and Regenerative and Functional Biomedical Technologies.

The distinguishing trait of the HST community is that its members are “fluent in multiple languages. [...] The languages of medicine, engineering, and the life sciences merge and evolve to unlock biomedical innovations.” The HST strives to provide “an environment where unorthodox avenues of inquiry can be embraced, enabling breakthrough solutions to some of the most intractable challenges to human health.” The strength of its research and training program derives from the partnership between two world-class institutions (Harvard and MIT) and from access to the resources of the HMS teaching hospitals, leading to partnerships.

Details on selected programs:

BIG - Bioinformatics and Integrative Genomics. Funded by NIH to support students in Bioinformatics (computational analysis and mathematical modeling of biological data) and Functional Genomics (high-throughput basic research to discover and characterize functional dependencies in biological systems). Emphasis is on the mathematical and biophysical modeling of complex biological systems, and experimental validation of computational predictions.

BMI – Biomedical Informatics. Goal: to form “leaders in the development and application of information technology in health and biomedical science.” The field includes traditional biomedical disciplines, computer science, biostatistics, epidemiology, decision sciences, health care policy and management. The emphasis is on the application of informatics to manage data and information in health care, bioinformatics, public health, or biomedical computing.

Computational Biology at the University of Idaho

The University of Idaho established a graduate program in Bioinformatics and Computational Biology earlier this decade. The emphasis is on interactions between biology and math/statistics, but the program takes a broad perspective, with core faculty drawn from nine departments (Biochemistry, Biological Sciences, Computer Science, Fish and Wildlife, Forest Resources, Mathematics, Plant Sciences, Rangeland Ecology, and Statistics) in four colleges and one institute (Science, Natural Resources, Agricultural and Life Sciences, Engineering, and the WWAMI medical education program), plus additional faculty from Physics, Philosophy, and History. The training program grew out of the Initiative for Bioinformatics and Evolutionary Studies (IBEST (http://www.ibest.uidaho.edu/home/), which combines expertise from biology, biochemistry, mathematics, statistics, and computer science to examine the patterns and processes of evolution and their relevance to biomedicine and to develop the analytical tools needed to do so. The program is funded through COBRE grants, the NIH equivalent of NSF's EPSCoR, at $10 million for 5 years; they have just started their second period of funding. See http://www.bcb.uidaho.edu/default.aspx?pid=85454 for more information on the faculty affiliation and http://www.bcb.uidaho.edu/default.aspx?pid=97845 for details on the requirements of the graduate program.

Computational Biology at the University of Arizona

Research and training in computational biology at the University of Arizona are distributed among several departments and programs, including the medical school, Computer Science, and Ecology and Evolutionary Biology, and several interdepartmental programs. For example, the BIO5 Institute brings together faculty and other researchers from five disciplines—agriculture, medicine, pharmacy, basic science, and engineering—to tackle complex biology-based problems and train students. Computational Biology is a key element of BIO5. Computational Biology is also a research focus area for faculty in the Life Sciences, another Computational Biology group is part of the Biomedical Engineering Program, and a small Computational Biology research cluster is part of the Computer Science Department. An NSF IGERT grant (Integrated Graduate Education and Research Training) supports a program in Comparative Genomics, encompassing functional, evolutionary, and computational genomics (http://www.genomics.arizona.edu). Additional training in computational biology occurs through specific tracks in existing graduate programs, such as the Biomedical Engineering Interdisciplinary Program, rather than through a Computational Biology program per se. The personnel involved in the NSF-funded iPlant Collaborative ($50 million over 5 years) come from Plant Sciences, the BIO5 Institute, Computer Science, Management Information Systems, Ecology and Evolutionary Biology, and Mathematics; additional personnel from other institutions (Cold Spring Harbor Labs, UNC-Wilmington, University of Louisville, Purdue, and Arizona State) are also participants in iPlant.

Appendix D

Director of Computational Biology

The Director of Computational Biology at the University of Florida will build a new program engaging existing faculty members across the university as well as developing new external funding for work associated with the analysis, management, and visualization of cellular information at the molecular level.

The director will be responsible for

- Engaging existing faculty, staff and students from across the university in multi-disciplinary teams to address problems in plant, animal and human biology resulting in a culture of cooperative action to address biological science at the molecular level.
- Develop and coordinate interdisciplinary training in computational biology and informatics.
- Building external funding by administering seed money, recruiting new faculty, working with deans and directors to develop start-up packages.

The director will report to the Vice president for Research who will receive an annual evaluation report from the Computational Biology Advisory Committee regarding progress in program development.

Appendix E

VIVO at Cornell

Faculty at UF often have difficulty locating other faculty with related and/or complimentary research interests. Computational Biology problems often require a multi-disciplinary approach in which potential collaborators must be identified from across UF. Cornell has addressed this problem with a system of their design named VIVO. Their VIVO system is maintained by the Cornell Library.

From the Cornell web site http://vivo.cornell.edu/about:

> VIVO (not an acronym) brings together in one site publicly available information on the people, departments, graduate fields, facilities, and other resources that collectively make up the research and scholarship environment in all disciplines at Cornell.
>
> Search VIVO for information about faculty, departments and research units, undergraduate majors, graduate fields, courses, research services and facilities --- anything related to academic and research pursuits at Cornell.

A search for "Computational Biology" using VIVO at Cornell returned 109 faculty. A sample screen shot is below:

The UF Library has started a project in collaboration with Cornell to implement the VIVO system at UF for UF research and scholarship. With appropriate development and support, the UF system could provide researchers with critical information about related activity for collaboration.

The Computational Biology Advisory Committee

Purpose

The Computational Biology Advisory Committee advises the Director of Computational Biology regarding the direction of the program and its effectiveness in meeting its goals.

Reporting

The Computational Biology Advisory Committee reports to the Research Advisory Committee.

Membership

The group currently has ten members: Mike Conlon, Alberto Riva, Don McCarty, Jed Keesling, Joseph Glover, Pam Soltis, Erik Deumens, Jose Fortes, Rob Ferl and Glenn Morris.