

A PLACE FOR EVOLUTIONARY SYNTHESIS IN NORTH CAROLINA'S RESEARCH TRIANGLE

1. RATIONALE

For the past two decades almost every facet of evolutionary biology has been in a state of scientific revolution. This revolution has been fueled by theoretical and conceptual advances in evolutionary disciplines together with advances in molecular biology, genomics and information technology. The scale of molecular data generation and analysis has made ambitious projects like Assembling the Tree of Life suddenly tractable. At the same time, studying the genomic basis of phenotypic evolution and speciation is now possible beyond model organisms.

The advances in information technology have not only brought the potential for keeping pace with the rapid flow of empirical research, but have increased the potential for large-scale synthesis. Without comprehensive databases it would have been impossible for paleontologists to identify the cyclicity of mass extinction, much less the causes of individual extinctions. The Tree of Life will remain a skeleton without the morphologists, behaviorists, biomechanicists, and ecologists who bring flesh to the characters whose evolution is being traced. This reciprocal illumination can be ensured by including organismal biologists of all kinds in the research networks as they form.

We propose a Center for Synthesis in Biological Evolution to be established in the Triangle Region of Raleigh-Durham-Chapel Hill in North Carolina (see section 4). The CSBE will be a collaboration among three leading universities in the Triangle: Director Cliff Cunningham from Duke University; Associate Director Greg Gibson from North Carolina State University (NC State), and Associate Directors Joel Kingsolver and Daniel Reed at the University of North Carolina, Chapel Hill (UNC).

We will build upon successful elements of the National Center for Ecological Analysis and Synthesis (NCEAS) model, expanding both the range and types of activities that are needed for evolutionary synthesis. The Center for Synthesis in Biological Evolution will emphasize:

- 1- Assembly, integration and analysis of evolutionary databases
- 2- Cross-disciplinary catalysis of grand-scale evolutionary research
- 3- Cross-disciplinary synthesis of existing data
- 4- Application, education and outreach of evolution understanding

Although cross-disciplinary synthesis is our ultimate goal, it is placed third on this list to emphasize the immediate steps we must take to prepare our field for this opportunity for grand synthesis. Most of the databases necessary for synthesis have been conceived but are simply not ready. With important exceptions such as angiosperm phylogenetics and paleontology, our scientific culture has generally not included the large cooperative ventures necessary to collect primary data in the coordinated fashion necessary for synthesis — hence the need for active catalysis by the center.

2. SYNTHESIS ACTIVITIES

A- THE TRIANGLE CSBE VISION FOR MOTIVATING SYNTHETIC RESEARCH

The heart of the NCEAS model is the "working group" which seeks grand synthesis. Although this is the central goal of CSBE as well, in many cases the data has simply not yet been collected in a manner conducive to synthesis. We also feel a need for a series of cross-disciplinary "research catalysis meetings in the first two years of CSBE's existence. These meetings (described below) will bring together diverse groups of scientists on a wide range of subjects, not only to inspire cross-disciplinary collaboration, but to inspire the larger scale of scientific vision that can only come from cooperation and coordination. Our experience has led us to realize that large cooperative ventures need not require individuals to subsume their individual interests for the greater good.

For example, Director Cunningham has a long-term interest in achieving a synthesis between ecology and phylogenetic approaches. To achieve this in a comprehensive fashion is beyond the scope of any individual laboratory. To this end he founded CORONA (Coordinating Research on the North Atlantic), an NSF-funded Research Coordination Network bringing together over 100 European, Canadian, and American field ecologists, phylogeographers and paleontologists, studying everything from algae to fish. This network encompasses the entire trans-Atlantic marine biota: those taxa that are found on both coasts of the North Atlantic.

By dramatically increasing the scale of their scientific vision to encompass the historical ecology of an entire biota, members of CORONA quickly realized that this vision offered them a chance to make rapid progress on their own scientific agendas as well as advancing their scientific careers. In analogy to levels of selection, a group operates best when the advantages to the group and the individual are not in conflict. CORONA has deliberately avoided imposing a grand synthesis on the group, but there are many opportunities for individuals to express their own vision for collaboration. So far, CORONA has inspired European scientists to form two major networks based on the CORONA framework, and have already submitted proposals for large-scale funding both from Britain's NSERC and the EU.

Finally, before the working groups can operate efficiently, we must rapidly improve the state and integration of evolutionary databases, and begin entering the data itself in a systematic way.

B- CSBE IN ITS YOUTH: ASSEMBLY, INTEGRATION AND ANALYSIS OF EVOLUTIONARY DATABASES

The synthesis of existing data depends on the right data being easily accessible. The potential of databases for spurring synthesis has not been lost on the evolutionary community. Along with the shining example of the amazing patterns revealed by paleontological databases, many of us have found that we often structure our questions and research strategy depending on the information available in GenBank.

Many outstanding scientists have committed a great deal of time and energy to nascent database projects. These include (among many others) Maddison's Tree of Life, Donoghue and Sanderson's PhyloBank, Neigel's Population Genetics Database, Martins' and Clark's Etho-Source, Kingsolver's Selection Synthesis Database, Ackerly's Plant Functional Ecology Database, and Ronquist's MorphBank. Although many of us feel that genomic databases are far ahead in many respects, there is much work to be done to promote access to and utilization of BAC maps, EST (and eventually genome) sequences, microarray data, secondary metabolite profiles, cross-species comparison of QTL data, gene expression arrays, access to online museum catalogs, and to the kind of taxonomic databases being created by the NSF PEET program.

The CSBE will work to centralize the assembly, integration and analysis of evolutionary databases. For each database, the CSBE will promote development of ontologies for annotation, as well as uniform guidelines for promoting free exchange of data. Other areas of concentration might include assistance in the integration of online museum catalogs and cross-species comparison of QTL data. We will also explore the potential of object-based databases in addition to relational database designs.

CSBE will support three paths to databasing. 1) It will begin by taking a direct role in bringing committed scientists together to form a multi-disciplinary, federated database to be housed at CSBE, and thereafter in perpetuity at Duke. The priorities for this federated database will be set by the evolutionary community, both during our planning groups described below, and in the form of "white paper" proposals evaluated by our scientific advisory board (see below). 2) As our infra-structure develops for this centralized database, we will emphasize integrating our database with those seeking external funding to develop additional databases. 3) Throughout, we will work to incorporate "orphan" databases that will remain in perpetuity on Triangle servers.

Targeted Database Planning Groups:

Our database efforts will begin with three parallel "targeted" planning groups to meet in the months before the Center opens. These will include 10-15 individuals each. These targeted working groups will devise specific strategies, and will devise specific staffing needs and infra-structural requirements to incorporate these plans. These plans will include the scope of the central "federated database" to be operated under the direct curation of the center. At the same time, they will devise a pathway to cooperate with and facilitate the development and integration of existing and future database efforts that will remain under the curation of their developers. In this way, the Center can ensure that the basic needs of the evolutionary community are met while

encouraging the many parallel efforts that will explore encompass the vast possibilities for databases that will remain beyond the capacity for a single center.

To ensure a rapid start, we will identify three areas for targeted working groups, as well as the individuals who will organize them. Hereafter, the targeted working groups will be proposed from the community and evaluated by the Scientific Advisory Board. These three groups will meet at the same time and the same location to allow inter-group coordination, but will mostly meet in separate rooms given the very different requirements.

Genomic and genetic databasing: Organizers: Associate Director Greg Gibson and Sudhir Kumar (Arizona State). Despite the prominence of such institutions as EMBL, GenBank, FlyBase and WormBase, there are large areas that require assembly to facilitate evolutionary comparisons, including availability of BAC clones and expression array data. This group will also plan the incorporation of population genetic data, as well as the preliminary design of the “evolutionary toolkit” as part of the front-end to the federated database (see below).

Phenotypic databasing: Organizer: Associate Director Joel Kingsolver. This group will be composed of the leaders of the many nascent phenotypic databases mentioned above, emphasizing morphology, development, and behavior. Members of the planning group will include paleontologists bringing their extensive database experience along with ideas about how the CSBE database might best enhance the questions they can ask in the future.

Phylogenetic and taxonomic databasing: Organizer: Director Cliff Cunningham. This group will be composed of the leaders of the many phylogenetic and taxonomic database initiatives, with the goal of achieving a common design, and incorporating the presentation and analysis of the genetic and phenotypic data to be organized by the other two targeted groups.

Targeted Sabbaticals:

During the Fall of 2004, as the targeted working groups are meeting, the *Scientific Advisory Board* (see section 5) will evaluate proposals for targeted sabbaticals from individuals especially interested in extending their previous databasing efforts. The idea is that many individuals have proven their commitment to databasing through their often heroic efforts in this area that is generally not rewarded by either grant funding or by University Promotion committees.

Unlike the “free-lance” sabbaticals to be described below, the Center will pay the full salary for targeted sabbaticals. This not only allows individuals to come without regard to their personal sabbatical schedules, but justifies the expectation that these individuals will devote effort to deliverable products associated with the federated databasing effort. These sabbatarians may or may not propose associated targeted working groups to enhance the coordination and collection of existing data. While an associated working group is not a requirement for a targeted sabbatical, all additional targeted working groups must be proposed together with in-house targeted sabbatarians to carry out the ideas of the group, and to coordinate data entry and database design.

CSBE will maximize the productivity of targeted sabbatarians with a database computing team exclusively devoted to their projects including 1 multi-media specialist working on web design; 2 expert database developers, and 2 programming professionals. In addition, each sabbatarian will be assigned a full-time individual to focus on data-entry, both from the literature and from cooperating individuals sending in their data. This individual will also bring in undergraduate work-study students as needed to help with necessary tasks.

The current budget envisions a standing population of approximately 5 targeted sabbatarians through the end of the second fiscal year of the grant. We will consider sabbatical requests ranging from a few months to a full year depending on the scale of the commitment. We will also consider proposals in lieu of a formal sabbatical from individuals wanting to make a series of visits to the center, working with the database team and their data-entry specialist. In short, if somebody wants to spend some focused time getting their databases entered and integrated, we will accommodate them.

To summarize, these targeted sabbaticals will not only give committed individuals the time they have long wanted to devote to their databases, but will bring them physically together at the center so that they will be able to coordinate their efforts. Targeted sabbaticals will be emphasized in the first two years of the Center. In the current budget, the third year brings in only the "free-lance" sabbaticals that are central to the NCEAS intellectual environment. This is flexible, though, and the advisory boards will help decide whether some level of targeted sabbatical support is desirable in the third year and beyond.

Beyond the Targeted Database Effort

CSBE will use its influence to encourage scientific journals to mandate database entry as a condition of publication. We have described how we propose to use targeted working groups and targeted sabbaticals during the first two years of CSBE's existence to speed the assembly of existing data. Throughout the lifetime of CSBE, the *Scientific Advisory Board* will evaluate proposals from the community to identify high-priority taxa and types of data to be incorporated. Although our team of programmers and specialists will shrink, it will continue to work with visitors and provide support to those pursuing external funding for new database projects that will be integrated with CSBE's federated database. This is similar to how current databasing proposals at NIH coordinate their efforts with FlyBase (whose board includes Associate Director Gibson).

But assembling existing data is not enough. For many questions, it is too early to attempt grand synthesis, and in many cases the data has not been collected in the coordinated manner needed for a grand synthesis. A mechanism is needed for networks of scientists to form among scientists in very different disciplines to coordinate their primary research efforts.

Cross-Disciplinary Research Catalysis Meetings

As many NCEAS working groups have discovered, the kind of data necessary for synthesis may not have been collected in a manner conducive to grand synthesis. Our center will benefit from integrated, cross-disciplinary research programs, and by dramatically increasing the ambition, scale and vision of primary research. To this end, CSBE will support proposals from the evolutionary community to hold short meetings to encourage formation of small networks of carry out synthetic primary research. These meetings will be larger than working groups (30-40), and will cast a wide net to bring together scientists with common interests, but very different world views.

The meetings will be based on the model PI Cunningham developed for CORONA (described above) where the emphasis is more on inspiring a vision of grand-scale collaborative research than achieving a group synthesis. These meetings will have the added benefit of spawning future working group proposals, as well as proposals for NSF research coordination networks and other collaborative and international grants. Director Cunningham will work with the Group Leaders to plan and moderate these meetings. Because these meetings are complementary to the working groups — and will help to pave their way — they will be de-emphasized as working groups form.

C- CSBE IN ITS PRIME: *CROSS-DISCIPLINARY SYNTHESIS OF EXISTING DATA*

As a mature institution, NCEAS depends on its working groups and In-House community to achieve the kind of synthesis that is changing the science and culture of ecology writ large. By year three these activities will form the main focus at CSBE.

WORKING GROUPS

The most conspicuous success of the National Center for Ecological Analysis and Synthesis (NCEAS) model has been the working groups carrying out synthesis of existing data. Because these groups are proposed by the international community of scientists, they provide a constant flow of new ideas into the center. We see little reason to change the successful philosophy of these groups, which meet 2-3 times for 3-4 days each time.

FREELANCE SYNTHESIS: THE IN-HOUSE COMMUNITY

Perhaps less well-known are the in-house “freelance” scientists at NCEAS, including \approx 15 postdoctoral fellows and 6 sabbatarians. This inspired social experiment brings faculty — with the rare commodity of time on their hands — together with brilliant young post-docs; the track record of NCEAS post-docs in obtaining the best academic jobs is testament to the success of this model. Our center will emulate the NCEAS institutions of visiting working groups and in house “freelance” post-doctoral fellows and sabbatarians.

The Post-Doctoral positions will each last for two years, and will explicitly not include the collection of primary data. Like at NCEAS, though, Post-Docs will be able under special circumstances to take a brief leave to collect primary data if they have external funding. Each

Post-Doc application will identify a local mentor in the Triangle. Projects can include everything from simulations, synthesis from the literature, and improvement of evolutionary methods of analysis. The In-House community will hold regular seminars, beginning with the Triangle Evolutionary Community, and including visitors to the three Institutions.

Finally, I consider the post-docs at NCEAS as effectively part of the CSBE “in-house” community. Regular interactions between the post-docs at NCEAS and CSBE could include annual or semi-annual retreats. These retreats could begin with “short-courses” organized by the post-docs from one center to teach the other the basics about a major issue they see as particularly fruitful for interaction between the fields of ecology and evolution.

D- CSBE IN THE GREATER WORLD

THE EVOLUTIONARY TOOLKIT

NSF and the evolutionary community have already begun to prepare for the computational, algorithmic, and software improvements needed for the genomic era, notably in the \$11.6 million NSF Information Technology Research Grant “Cyber Infrastructure for Phylogenetic Research (CIPRES). CSBE will work closely with the CIPRES group to help disseminate information about CIPRES, as well as to provide a place to beta-test CIPRES’ latest developments. CIPRES involves 13 institutions, including NC State (Spencer Muse). David Swofford (see supporting letter) and his CIPRES group at FSU (including Peter Beerli and Fredrik Ronquist) have indicated their enthusiasm for working with CSBE (wherever it is eventually located) and using CSBE as a means to make its products available to the evolutionary community.

While we hope to work with these developers of the latest evolutionary methods in computation, we also see a need to provide a user-friendly platform providing many common approaches to analyzing data without endless cutting and pasting between formats. An example might be a single platform allowing you to do all of the analyses available in MEGA, MacClade, PAUP, Arlequin, COMPARE and Lamarc, just for starters. We look forward especially to collaborating with SAS, located in the Triangle. Associate Director Gibson has an especially close relationship with the leaders of SAS, and their supporting letter (from Russ Wolfinger) indicates their enthusiasm to work with us as we develop this toolkit.

CSBE would work with the developers of these methods — including considering targeted sabbaticals to develop evolutionary software — to help achieve a cross-platform “evolutionary toolkit”. This toolkit would ideally be incorporated directly into the front-end for CSBE’s databases, taking data of all kinds directly from the databases themselves. This could be similar to the ACeDB platform that unifies plant genomic databases (see Stein, 2003: Integrating biological databases. *Nat Rev Genet* 4: 337-345; www.ukcrop.net). Sudhir Kumar (see supporting letter), the developer of MEGA and several important databases, is especially enthusiastic about achieving this goal. In addition, the “white papers” describe above will allow members of the evolutionary community to propose new methods for inclusion in the toolkit. Our planning groups (see above) will include Kumar and others who have begun working towards this goal of a unified Evolutionary Toolkit, so that we can provide this service rapidly.

COLLABORATION WITH INTERNATIONAL PARTNERS

Although it is premature to specify specific details of partnerships, International cooperation is necessary for effective synthesis. Like NCEAS, CSBE will encourage including post-docs, sabbatarians, and group members from around the world. We have already spoken to David Lambert about the potential for post-doc and sabbatical exchanges with the Allan Wilson Centre for Molecular Ecology and Evolution in Massey, New Zealand. CSBE will need to work closely with this and other centers, particularly Oxford's legendary "Evolutionary Biology Group" who are consistently at the cutting edge of evolutionary methods and software. As an example of international outreach, Cunningham conceived of the "mini-PEET" program at the Society of Systematic Biologists when he was Executive Vice President. The many applicants for these small (\$5,000) awards were mostly from developing countries, and all but one went to South American Scientists. This level of interest suggests that small amounts of money from institutions such as scientific societies can go a long way. CSBE can work with scientific societies to coordinate efforts of this kind.

BROADENING THE SCOPE AND PRACTICE OF APPLIED EVOLUTION

Our discipline not only contributes to applied fields such as medicine, and agriculture, but benefits by ensuring that evolutionary methods are carried out appropriately. A major part of the mission of NC State is extension to the applied sciences, involving such activities as advice on how to deal with and identify pathogens or what to do to prevent build up of insecticide and herbicide resistance. Increasingly such activities rely on research in evolutionary biology encompassing QTL mapping, mechanistic understanding of symbiosis and mutualism, and genomic tools for species identification. The NC State genetics group will continue these active collaborations.

Similarly, Duke's Phylogenetic Consulting Center (a one-year experiment co-founded by PI Cunningham) helped increased the profile of evolutionary biology in the Duke Medical School genomics initiative. Huntington Willard leads the Institute for Genomic and Policy Sciences at Duke. He is very enthusiastic about working with CSBE, and has promised financial support for a post-doctoral "phylogenetic-consultant", to reside at the center (See Willard's supporting letter). This individual will begin working with medical centers and industry groups in the Triangle, but will eventually be available for consultation through the Web as well.

THE TIMELINE

SUMMER 2004

ASSEMBLY OF ADMINISTRATIVE ADVISORY BOARD AND SCIENTIFIC ADVISORY BOARD
CALL FOR PROPOSALS FOR 5 POST-DOCS 5 TARGETED SABBATICALS 5 SYNTHETIC RESEARCH
CATALYSIS MEETINGS

FALL 2004

MEETINGS OF ADVISORY BOARDS
RENOVATING ERWIN MILL BUILDING SPACE
DATABASE PLANNING GROUP MEETINGS
SEARCH FOR ASSISTANT DIRECTOR AND
SCIENTIFIC DATABASE ADMINISTRATOR
HIRE FIRST OFFICE STAFF

JANUARY 2005

OPEN DOORS WITH 5 POST-DOCS, 5
TARGETED SABBATARIANS, DIRECTOR,
ASSISTANT DIRECTOR, DATABASE DIRECTOR
AND DATABASE TEAM

2005-2006

CONTINUE FOCUS ON DATABASING AND
CATALYSIS MEETINGS
AWARD FIRST 5 "TRADITIONAL" WORKING GROUPS,
REACH POPULATION ≈14 POST-DOCTORAL FELLOWS
(INCLUDING PHYLOGENETIC CONSULTANT)

2006-DECEMBER 2009

TRANSITION TO TRADITIONAL "FREELANCE SABBATICALS"
AWARD ≈ 9 WORKING GROUPS PER YEAR
BUDGETING ALLOWS FOR 5 MONTHS
BEYOND OFFICIAL END DATE OF SEPTEMBER 1,
MAKING 5 FULL YEARS OF OPERATION



The Erwin Mill Building on 9th Street (section 6)

3. EDUCATION AND OUTREACH

A- Involving Target Audiences in Synthesis Activities

As with most cross-disciplinary interactions – such as the research catalysis meetings described earlier – it is often possible for members in one field to slightly modify their collection and analysis of data in ways that make their work far more useful to those in another discipline. For example, the many areas of applied evolution are especially rich territory for evolutionary synthesis. Vast under analyzed datasets have been collected by professionals in agriculture and medicine that cast light on the evolution of pesticide and herbicide resistance, or the evolution of infectious disease. While these datasets can be used to test hypotheses of evolutionary principles, the evolutionary perspective may in turn suggest practical advice, for example on strategies for improved animal breeding, crop rotation during the introduction of transgenic plants, and restriction of antibiotic use to slow the evolution of drug resistance.

The most effective way to make sure that evolutionary synthesis is relevant to educators and professionals outside of major research universities is to directly involve them in the process of synthesis. CSBE will help the planners of synthetic activities – be they working group leaders or post-docs – to plan their activities to enhance communication with one or more specific audiences. Once this audience is identified, the CSBE education and outreach staff will develop mechanisms to involve members of this audience in the ongoing process of synthesis, and to bring useful information back to the target audience.

B- ENHANCING THE FLOW OF EVOLUTIONARY KNOWLEDGE THROUGH

LOCAL AND NATIONAL NETWORKS

There is no single way to connect the synthesis activities at CSBE with the many potential target audiences. Our outreach plan will combine the strengths of state-wide network of a land grant university (NC State) with a national organization with established channels of communication with communities of educators, academics, and policy makers (AIBS: American Institute of Biological Sciences). Associate Director Greg Gibson will lead the CSBE education and outreach team that includes a Extension Coordinator and an Education Coordinator. Both coordinators will work full-time at CSBE, with the Education Coordinator hired through a sub-contract to AIBS.

The Extension Coordinator: involving individuals from the local area

The Extension Coordinator will develop a working knowledge of the synthesis activities being planned at CSBE, and of the many mechanisms by which NC State educates the general public in practical aspects of biology. The Extension Coordinator will use this knowledge to help identify the individuals in agriculture, industry and medicine who are best suited to become directly involved in synthesis activities at CSBE (as described above). This individual will also help identify potentially interested faculty from local institutions such as Historically Minority Colleges and Universities, the North Carolina Museum of Natural History, the North Carolina Biotechnology Center, and the Medical Schools at Duke and UNC-Chapel Hill. The Extension Coordinator will arrange for representatives of interest groups to meet with visiting scientists, arrange public lectures, and identify opportunities for outreach that may not be obvious to those engaged in primary research. The Extension Coordinator will oversee the dissemination of useful information to the broader community through one of four mechanisms.

1) The Agricultural Research Service: NC State will commit responsibilities of a full time equivalent from its current staff of over 150 employees of the Agricultural Research Service to work with the CSBE. These individuals will be associated with individual faculty at NC State either on campus or at outlying field stations, and will work closely with the Extension Coordinator to spread information such as advice on crop rotations, strategies accounting for evolution of pesticide resistance, bioremediation of hog waste, and other practical deliverables emerging from CSBE synthetic activities.

2) The CSBE Phylogenetic Consultant: This individual (described earlier) is supported by the Duke Institute for Genome Sciences and Policy. He or she will work to demonstrate the utility of phylogenetic methodology in a variety of disciplines, ranging from functional genomics to

phylogenetic relationships of human pathogens to drug design. The phylogenetic consultant will give advice and training about the use of appropriate phylogenetic methodology, and how to interpret phylogenetic trees in complex situations such as the evolution of gene families, and identifying orthologous genes between human genomes and the genomes of model organisms.

3) Summer Institute of Evolutionary Biology: The CSBE will pay 1/4 salary of the Coordinator of the highly successful Summer Institute of Statistical Genetics at NC State each June. The CSBE will disseminate cutting-edge evolutionary thinking in the form of two new modules emerging from CSBE activities. These will be appended to the 18 existing modules that focus on statistical and quantitative genetics. This Institute provides a ready-made audience for the evolutionary modules, attracting over 300 graduate students, post-doctoral fellows, faculty, commercial sector scientists, and other professionals. These will be taught by CSBE sabbatarians and local faculty, and possible first year topics include Evolution and Development, Phylogenetic Analysis with Large Datasets, and the Genetics of Speciation. While predominantly lecture-based, the opportunity exist for hands-on computational workshops.

4) Graduate and Continuing Education Initiatives: The Extension Coordinator will also be responsible for coordinating CSBE activities with those of two major NCSU educational resources. The Science House provides K-12 outreach high school programs across the State. Its activities include the new Centennial Middle School on Campus, and support for distance learning through web-based courses (such as Bruce Weir's internationally acclaimed Forensic Statistics modules), and the Kenan Fellows program for training of high school teachers. The McKimmon Center for Continuing Education's classes for mature students, provides a natural conduit for outreach.

Educational Outreach at the National Level Through AIBS

The Education Coordinator will reside full-time at the CSBE, and will be our primary connection to the considerable resources made possible by our partnership with the American Institute of Biological Sciences (AIBS). The Education Coordinator will be hired through a sub-contract to AIBS. This sub-contract will include a 1/4 salary to support CSBE activities at the AIBS office in Washington DC, whose staff will give CSBE a major role in synthesizing evolutionary knowledge useful to policy makers, government agencies, educators, media, and society. As with the Extension Outreach, each working group or post-doc at CSBE will choose a target audience most relevant to its activities.

Traveling to meetings and writing materials as necessary, the Education Coordinator will reach target audiences in collaboration with AIBS staff including Susan Musante, M.Ed., AIBS Education and Outreach Program Manager; Oksana Hlodan, M.Ed, ActionBioscience.org Editor; Robert Gropp, Ph.D., AIBS Senior Public Policy Representative; and Richard O'Grady, Ph.D. AIBS Executive Director.

1) Online Lesson Plans for K-12 and Undergraduate Teaching: The Education Coordinator will work with AIBS to develop lesson plans and K-16 curricular content based on CSBE synthesis activities. Results will be posted to the CSBE website as well as to AIBS's bilingual (English/Spanish) free access education website, **ActionBioscience.org**. This site receives 100,000 hits a month from 152 countries, and materials from the site are used by over 120

colleges and universities around the world. The lesson plans on this site are currently created in collaboration with the National Association of Biology Teachers and are keyed to National Science Education Standards. With CSBE involvement, the plans on this site will also be developed with the BioQUEST Curriculum Consortium and the Biological Sciences Curriculum Study, both members of AIBS, also with the National Research Council's *Bio2010* program.

2) Online and Print Dissemination of CSBE Activities: AIBS will produce a major article for lay readers and an *Eye on Education* column every year in the AIBS journal, *BioScience*, free online and with free reprints. These will cover aspects of the research and education taking place at the CSBE. Articles by, and interviews with, CSBE scientists in both English and Spanish will be published on ActionBioscience.org (described above). AIBS will also facilitate the dissemination of CSBE press releases, through the EurekaAlert media network of AAAS.

3) The Education Working Group: When appropriate, the Education Coordinator will identify appropriate members of the evolution-education community to attend the research catalysis meetings and to join CSBE working groups. Twice a year CSBE will convene working groups dedicated to educational issues, with topics proposed by the community as with other working groups. This idea has already generated formal expressions of interest from, among others: 1) Judy Scotchmoor and David Lindberg, of the Understanding Evolution website project at Univ. of California Museum of Paleontology; 2) Judy Diamond, of the Explore Evolution project at the University of Nebraska State Museum; 3) Yolanda George, AAAS Education Division; 4) Jay Labov, National Research Council, and 5) John Jungck, of the BioQUEST Curriculum Consortium, as well as several members of the AIBS Education Committee. BioQUEST is especially eager to share its education materials (simulations, databases, investigative case-based learning modules) in phylogenetics, population genetics, bioinformatics, and biocomplexity with CSBE scientists, and to involve undergraduate students in CSBE activities. All these groups have important minority-inclusion programs that CSBE will work with.

4) Visiting CSBE Scientists Program: Through this program the AIBS Public Policy Office will help CSBE scientists become effective communicators to policy-makers. This program will bring 3 to 5 CSBE scientists to events such as AIBS's annual Congressional Visits Day, AIBS National Press Club roundtables, or for special briefings with Administration or Congressional staff. In the course of these visits, AIBS staff will train CSBE scientists in effective communication with policy makers and the media; help identify public policy audiences that would benefit from learning about CSBE research; potentially leverage other national resources that could advance the CSBE effort; and help communicate CSBE's findings to the public policy community. AIBS public policy staff will also arrange for home district visits of NC politicians to CSBE.

5) Evaluation of Education Program: As with other CSBE programs, the evaluation of extension and outreach programs will take place under the authority of the Administrative Advisory Board (described below). This board will have resources to hire professional evaluators and consultants, who will attend CSBE leadership meetings and selected meetings and working groups. Evaluation criteria will include participant numbers and impact, publication of articles and reports, learning outcomes and the use of curriculum content and teacher resource tools. Criteria will include public understanding of evolution and of CSBE's role, and reports will be prepared for peer-reviewed journals in the areas of science education and public view of science.

4. COMMUNICATION, KNOWLEDGE TRANSFER AND INFORMATICS

Synthesis of evolutionary data will require that investigators worldwide have immediate and comprehensive access to databases as well as analytical tools. Our goals for databasing and analysis are described in detail in Section 2. These goals include

- Developing an in-house federated, integrated database that will exist in perpetuity on Triangle servers.
 - Integrating our database with those seeking external funding for additional databases
 - Providing a home for “orphan” databases
- Including basic analysis tools in the “evolutionary toolkit” that will form a front-end to the federated database

Our plan is to use the distributed computing Grid, initially at NC State University, to support parallel processing tasks, though we will work toward full integration of the CSBE computing environment with Internet II. The role of the **Associate Director for Informatics and Computation**, Dan Reed, will be to oversee and coordinate the activities of both in-house and off site personnel who will pursue our objectives.

Dan Reed’s staff will be led by a Master’s level Computing and Database Administrator who will have responsibility for connectivity of CSBE workstations to the internet and Grid, function of firewalls and antiviral software, and general maintenance of standard software. This will allow the Manager to focus his/her resources on communication of CSBE activities, liaison with the director and scientific advisory board, identification of needs for software implementation, and oversight of database development.

Along with his staff, the administrator will have ultimate responsibility for connectivity of CSBE workstations to the internet and Grid, function of firewalls and antiviral software, and general maintenance of standard software. The Administrator will oversee the database team that will support the databasing efforts by sabbatarians described above. At the beginning this team will include one multi-media specialist for web interface and design, two expert database developers, and two programming professionals, although the number will change as CSBE reduces its effort on database development. Sabbatarians, post-docs, and the Biocomputing Manager at the CSBE will have weekly meetings to consider data warehousing and processing needs as they arise, in a communal manner.

A- FOSTERING DATABASE INTEGRATION

The database planning groups (Section 2) will begin work to promote adoption of ontologies for annotation that will facilitate cross-talk between databases and more efficient browsing by end-users, and to promote uniform guidelines for free exchange and rapid deposition of data. Similar to the MIAME standards for deposition of microarray data in the Gene Expression Omnibus, we envisage adoption of a set of compliance standards for various types of evolutionary data, whether phenotypic, genetic, or taxonomic. This workshop will also explore the virtues of object-oriented and relational design concepts specifically as they pertain to evolutionary biology, and address issues relating to open access to software.

In addition to the evolutionary toolkit described above (Section 2) there are emerging areas of evolutionary analysis, such as experimental design for microarray studies and annotation of EST sequences, that would benefit from the development of new software. The CSBE will actively attract sabbatarians and post-doctoral fellows who wish to engage in software development, and will employ a full-time analytical consultant to respond to enquiries relating to existing tools.

B- CONNECTING TO THE NCSU GRID AND BEYOND (THIS SECTION BEING SCALED UP)

AN EXPERIMENT IN ACCESS TO COMPUTING POWER

On an experimental basis we will provide access to the computing power available through CSBE's connection to the NC BIOGRID, and its eventual larger computing grids and clusters (see section 4 below). Users around the world will be able to submit jobs too big for their own computing facilities, with priority being assigned relative to the computing resources already available to those individuals, and their prior usage of the resource. While we realize this service could quickly be swamped with requests, we hope that careful monitoring of usage and careful limits on the maximum amount of computation time allowed will make this a successful experiment, and will lead to an expansion of our capability.

As described in Section 2, the CSBE will experiment in providing computational support for advanced data analyses that cannot be carried out with the computing resources available at most institutions. Examples include phylogenetic analysis of datasets consisting of tens of kilobases of DNA sequence from 50 taxa; or Bayesian analysis of several thousand microarrays. The appropriate framework for such tasks is grid computing, in which multiple high performance nodes are clustered virtually, and middleware is used to distribute tasks efficiently across the computing grid. The Duke, UNC and NCSU grids are in the process of ever tighter integration, and together with nodes at various institutes in the Research Triangle, constitute a computing environment as powerful as any in the world.

CONFIGURATION

The NC State configuration (see www.ncsu.edu/itd/hpc/Grid/Grid.php) currently consists of 160 processors and almost 190 GFlop peak performance, with 3.5 TBytes of attached memory. Importantly, the SAS Institute in Cary, NC, is situated in the heart of the Triangle and is also an active participant in emerging Grid computing here, and is an active collaborator on bioinformatics programs in the Triangle (see letter of support). The CSBE will buy into the grid through the purchase, with assistance from the SUN Center of Excellence (see supporting letter) – a key member of the NC Bioinformatics and Genomics Consortium, who have also supplied a letter of support - of a node to be housed at NCSU. The Center will have priority access to the node, in return for access to parallel nodes in the NC State grid (and eventually the NC grid as it evolves), as well as administrative and information technology support. The structure of the NC Grid is continually evolving, but has at its hub the server, middleware and IT support provided by a not-for-profit company in Research Triangle Park, MCNC, illustrated at <http://www.ncsu.edu/itd/hpc/Images/Network4.JPG>.

The NCSU high performance computing cluster is on the same VLAN (IP subnet) as the MCNC and fiber connected to it at up to 10 Gbps. An Avaki Grid Platform provides a meta-scheduler capable of distributing jobs across a heterogeneous set of computers at multiple, separately administered sites, and simplifies running applications on the grid without modifying code, while Sun Microsystems is providing supporting Grid engines and Directory Servers. Dr Henry Schaffer in the Department of Genetics at NCSU and a member of the NC Supercomputing initiative, will advise the center on its connection with the emerging grid.

SERVERS

In addition to workstations for use by in-house and visiting scholars, CSBE computation will be facilitated by purchase of a node likely consisting of a Sun Fire V880 configuration with eight 1.05 GHz processors backed up by a Sun StorEdge 3310 fiber channel 876 GB storage array. This will be more than adequate to handle most day-to-day in house activities, but from time to time access to processors will have to be prioritized. Early in the funding period, the CSBE Scientific Advisory Board will appoint a panel of local computer scientists and users of high end evolutionary computation to outline a structured protocol for assigning priority base on computational time required, urgency, and the interests of broad community access.

The protocol will be made public and community feedback sought to ensure that particular evolutionary biology communities, or even individuals, gain unfair access advantages. The full time Biocomputing Manager will administer this protocol with consultation from the Director as needed to resolve disputes, particularly involving applications from off-site, and the SAB will review usage and conflict resolution on an annual basis. Sabbatarians, post-docs, and the Biocomputing Manager at the CSBE will have weekly meetings to consider data warehousing and processing needs as they arise, in a communal manner. Our data storage will begin with almost a terabyte of memory – though we note that purchase of extra memory in later years will be within budgetary flexibility.

TELE-CONFERENCEING?

NCEAS has deliberately avoided introducing tele-conferencing, since the interpersonal interactions in working groups are so important, and because of the obvious “slippery slope” of attendance by video is so potentially attractive. Since we are sympathetic to NCEAS’ conviction that a virtual center is not desirable, we would prefer to have this important decision about tele-conferencing be made in consultation with the two Advisory Boards (Section 5). Nevertheless, there are a number of initiatives that could be made possible by a tele-conferencing facility. These include allowing people elsewhere to “attend” regular discussion groups at CSBE either interactively, or through a web-broadcast. Often, hearing the reactions of others to the latest papers, or reading books along with a group is as important as other collaboration for individuals at isolated institutions. Anyway, most people in discussion groups don’t speak up anyway, so a web-broadcast should work!

5. MANAGEMENT PLAN

A. ORGANIZATION AND MANAGEMENT

The CSBE will be a partnership between NSF and a consortium of three leading universities in the Research Triangle: Duke University, North Carolina State University (NCSU) and the University of North Carolina at Chapel Hill (UNC). Duke will act as the lead institution for the grant. CSBE will be physically housed off-campus in Durham (see Section 6).

B. THE DIRECTORATE

Management for CSBE will consist of a four-person directorate that represents each of the three CSBE universities. Members of the directorate will be involved in science and administration of CSBE, and will act as liaisons to their home universities. The management structure of the directorate directs the three-pronged mission of the proposed center in terms of scientific synthesis, informatics/computation, and education/outreach. Because of the breadth and integration needed among the various activities at the Center, the directorate will consist of a full-time Center Director and three part-time Associate Directors. The Director and Associate Directors will serve renewable 5-year terms. In addition, an Assistant Director will direct the day-to-day administration and operation of the CSBE.

The **Center Director** will be responsible for the overall administration and successful operation of the CSBE, including management and staffing, management of funds and resources and communication with the research, educational, and public communities. The Director will provide leadership for all aspects of the Center's activities. The **Center Director** is Associate Professor of Biology at Duke. With the support of Duke University (see section 6), Cunningham will be relieved of his teaching responsibilities and will reduce his research commitments in order to give his full attention to the Center. Duke will also pay his summer salary, and provide technician salary and supply money for his laboratory in order to reduce the need to pursue external funding. Cunningham has made a five year commitment as center director,

The proposed director studies marine invertebrate phylogenetics, population genetics and biogeography. His most relevant administrative experience includes founding and directing the CORONA (CoOrdinating Research on the North Atlantic) research coordination network which has brought together over 100 American and European geologists, evolutionists, and experimental ecologists to study the trans-Atlantic marine biota. Cunningham is currently Associate editor at *Evolution*, *Systematic Biology*, and *The American Naturalist*, and just completed his term as Executive Vice President of the Society of Systematic Biologists. Cunningham was the co-founder of the Duke Phylogenetic Consulting Center, and leads and coordinates two multi-institutional teams of scientists carrying out NSF-funded projects: Hydrozoan taxonomy (PEET), and a phylogenomic survey of arthropod relationships (Biocomplexity:AToL).

The **Associate Director for Science and Synthesis** will be Joel Kingsolver, William Rand Kenan, Jr., Distinguished Professor for Teaching Excellence in Biology, UNC. Kingsolver will provide leadership for the scientific and synthetic activities of the Center, including the working group, post-doctoral, and sabbatical programs (Section 2). Kingsolver will serve as chair of the

Scientific Advisory Board, and will oversee the nominations to and the activities of this board and the Administrative Advisory Board (see below). Kingsolver is a leading evolutionary ecologist who combines experimental and theoretical approaches to study selection and evolution of quantitative traits. A fellow of the AAAS, he is former Editor-in-Chief of *The American Naturalist* and former Vice-President of the Society for the Study of Evolution. He is also actively involved in developing multi-media and software for teaching genetics and evolution, including *Genetics Today* and *EvoBeaker*.

The **Associate Director for Informatics and Computation** will be Daniel Reed, Kenan Eminent Professor at the University of North Carolina, Chapel Hill. Reed comes to Carolina from the University of Illinois at Urbana-Champaign, where he served as director of the National Center for Supercomputing Applications (NCSA), a 400-person research institute with a mission to develop computing infrastructure in support of scientific research. NCSA was the birthplace of the modern Web browser that sparked the Internet revolution.

At Illinois, Dan Reed directed both NCSA and the National Computational Science Alliance, a nationwide partnership of more than 50 institutions to advance scientific discovery via high-performance computing. Reed is a principal investigator for the National Science Foundation's TeraGrid project, an effort to build and deploy the world's largest, most comprehensive computing system for open scientific research.

The **Associate Director for Education and Outreach** will be Greg Gibson, Associate Professor of Genetics, NCSU. Gibson will direct the activities of the Center in education and outreach in collaboration with Susan Musante and Richard O'Grady at AIBS. Gibson is an evolutionary geneticist whose research integrates genomic, quantitative and developmental approaches to the study of genetics of complex traits, mainly in *Drosophila* but increasingly in dogs and humans. He is the lead author of *A Primer of Genome Science*, a leading textbook of genomics and bioinformatics. He has a consultant to the SAS Institute on development of the SAS Microarray Solution and SAS Genetics software, and has interests in analysis and management of very large gene expression datasets. Gibson is also a member of the FlyBase Advisory Board, a faculty of the NCSU Bioinformatics Research Center, currently serves as Chair of the NC Governor's Taskforce on Genomics and Public Health, and is Assistant Director for Life Sciences for the North Carolina Agricultural Research Service.

The **Assistant Director** will reside full time at the Center, and will oversee the day-to-day operations of the CSBE. This individual will have primary responsibility for accounting and logistical operations for the Center, and for supervising administrative and other staff at the Center (see below). The Assistant Director will be hired by the Director in consultation with the Associate Directors and will report to the Director.

C. ADVISORY BOARDS

It is essential that the decision-making process of the Center be transparent. This will be accomplished by very different two advisory boards which will each meet twice yearly. For the sake of continuity, the 5 members of the **Administrative Advisory Board (AAB)** will serve for the entire five year period covered by this proposal. This board will provide independent oversight and evaluation of the Center activities and its impact on the evolutionary community. Its members will each represent a major field of evolutionary biology, and will be chosen by the Directorate in consultation with relevant program officers at NSF, and might include recognized leaders such as those who have held the elected Presidency of a major evolutionary society.

The specific tasks of the AAB will be to ensure that appropriate and independent evaluations of CSBE and the Directorate are performed, and for advising on any changes in the Directorate. The AAB will assess the success of the CSBE in achieving its four central goals: 1) Cross-disciplinary synthesis of existing data; 2) Catalysis of cross-disciplinary primary research; 3) Assembly, integration and analysis of evolutionary databases; 4) Application, education and outreach of evolutionary understanding. The board will be provided the resources they deem necessary, such as the services of researchers and paid consultants, to evaluate the impact of the Center on the field of evolutionary biology.

In contrast, the turnover imposed by staggered three-year terms of the 18 members of the **Scientific Advisory Board (SAB)** will ensure a constant flow of new ideas and opinions about the Center's direction. Its membership will be chosen under the key guiding principle of being inclusive of scientific disciplines, types of institutions, under-represented groups and career stages. The Associate Director for Science and Synthesis (Kingsolver) will chair the board, and its initial membership will be chosen by the Directorate in consultation with the AAB. The specific tasks of the SAB will be to guide the scientific and synthetic activities of the Center. Specifically, they will solicit and evaluate applications for support in each of the following areas described in detail in Section 2: cross-disciplinary research catalysis meetings, working groups, targeted and free-lance sabbaticals postdoctoral fellowships. The Scientific Advisory Board will also evaluate "white papers" from the evolutionary community to focus the resources of CSBE databasing efforts with respect both to kind of data being collected groups of organisms to receive priority in entering data, and analysis software to be included in the "evolutionists' toolkit". These white papers will be evaluated both with regard to the likely productivity of the group that will be using the data, and commitments of time by the proposers themselves to complement those of the CSBE team.

D. STAFF

The Staff will support the Directorate and the programs of CSBE. The staff will be supervised by the Assistant Director. The administrative staff will include an administrative manager (staff specialist level 9), responsible for accounting and overall supervision; an administrative assistant (staff specialist level 8) responsible for logistical arrangements (travel, housing, reimbursement); and an office assistant (staff specialist level 7) to provide office support. The IT staff is described above (Section 4).

6. CAPABILITIES OF THE INSTITUTIONS TO HOST AND MANAGE THE CENTER

We propose a Center for Biological Synthesis in Evolution to be established in the Triangle of Raleigh-Durham-Chapel Hill in North Carolina. This Research Triangle of North Carolina is one of the most dynamic, affordable, and attractive growth areas in the United States, nicely bridging the traditions of the northern and southern regions of the country. The Raleigh-Durham airport has excellent connections nationwide, and is rarely closed due to weather.

The Center will be located in the Erwin Mill Building in downtown Durham (see photo above), within easy walking distance of restaurants, shops and entertainment. Culturally, these three towns within 30 minutes of one another provide a unique diversity of opportunities, which combined with the proximity to the Carolina Coast and the Appalachian mountains, with Washington DC and Atlanta just a few hours further away, makes it a very attractive (and affordable) place to recruit faculty and post-docs for sabbaticals and fellowships. For scientific, educational, and institutional reasons, the Triangle will be an ideal location for the CSBE.

A. SCIENTIFIC

The three consortium universities of the Center (Duke, NCSU and UNC) have over 80 faculty with exceptional breadth in evolutionary biology, spanning the full range of relevant disciplines from genomics and molecular genetics through functional biology and evolutionary physiology to systematics, paleontology, evolution development, and a very strong group studying the philosophy of evolutionary biology. These faculty will participate in database administration, working groups, and will be available as official mentors to incoming post-doctoral fellows.

At all institutions there is a strong applied evolutionary perspective to the major genomic and bioinformatics initiatives at our agricultural, veterinary, and medical schools. This intellectual environment, will be especially important for recruiting sabbatical faculty and postdoctoral fellows to the Center.

The Triangle includes the North Carolina Museum of Natural History and Research Triangle Park, a premier biotechnology and computer technology incubator. Our institutions have a close relationship with the North Carolina Center for Biotechnology and the North Carolina Super-computing Center; our institutions also jointly run the NSF-supported Statistical and Applied Mathematical Sciences Institute in Research Triangle Park. Duke in particular has very close ties with the botanists and zoologists at the Smithsonian National Museum of Natural History, with many co-funded grants and shared students. Duke is the home to a major herbarium, with the largest collection of lichenized fungi in the world.

B. EDUCATIONAL

The Triangle is home to the three major research Universities that will host the CSBE. These represent the Liberal Arts, Private, and Land Grant missions (UNC, Duke, and NC State), and an equal number of prominent Historically Minority Universities (NCCU, Shaw and

NCA&T). Several other Colleges are located within an hour's drive. Finally, the Triangle is home to innovative initiatives in biology education and outreach. These include Associate Director Bollenbacher's initiatives, described above, which combine post-doctoral research opportunities with teaching at North Carolina's Historically Minority Colleges.

C. INSTITUTIONAL COMMITMENTS

The combined resources of the three host Universities offer the potential of substantial institutional commitments to recruit the CSBE to our area. The two Institutions receiving funding from this grant have both agreed to Institutional commitments that roughly equal the amount they will be taking in overhead.

Duke University Commitments:

As the lead institution, Duke made the major concession to allow an off-campus overhead rate of 28% (down from 54%) that is normally reserved for projects located far away. In addition, the University made three classes of commitments (see letter from Vice Provost Siedow).

Supporting the Director's Laboratory

The first is to help Director Cunningham spend less time grant-writing and more time devoted to the Center. He intends to continue a small research operation in his laboratory, with two graduate students (down from his usual five). Duke has agreed to pay for the entire salary of Cunningham's lab director, as well as to provide supply money for each of the five years.

Supporting the Director's Salary

Duke has agreed to pay the entire director's nine-month salary as well as three months of summer salary throughout the period covered by the grant. This is especially notable since the University has agreed to suspend Cunningham's teaching commitments entirely for two full years, with a limited return to Undergraduate teaching (6 weeks per year) thereafter.

Supporting for Office Renovation

We anticipate negotiating approximately \$100,000 for renovation costs from the owner of the Erwin Mill Building. Duke has agreed to match those costs up to a maximum of \$100,000. This should cover the expenses of renovation, since the existing space is already generic office space.

Supporting Building Management Costs

A considerable portion of the 220,000 rent we anticipate being charged per year is composed of management costs (utilities, janitorial, and etc.). Duke has agreed to pay these costs, reducing our rent by about 1/3.