Collaborative Research: Digitization TCN: The Macroalgal Herbarium Consortium: Accessing 150 Years of Specimen Data to Understand Changes in the Marine/Aquatic Environment

Macroalgae are a diverse group of aquatic (marine, estuarine, and freshwater) organisms. The smallest species are simple filaments less than a few centimeters long, while giant kelp can reach more than 40 meters in length. Marine macroalgae are known collectively as seaweed, while freshwater species are sometime referred to as pondweeds, or simply as algae. A few species have common names, like sea lettuce, kelp, Irish moss, and rockweed. Some have been used for centuries as food in many regions of the world, especially in Asia and the Pacific Islands. In the U.S., there is increasing familiarity with the Japanese names for some macroalgae, such as nori, used to make sushi, and wakame, an ingredient in soups and other foods. There is a movement to use sea vegetables western and fusion cuisines. Most of the macroalgae eaten as sea vegetables are grown via aquaculture, primarily in Asia, but increasingly in other regions; a very successful kelp farm was recently established in Maine. Macroalgae are used in industry as a source of phycocolloids, a gelling agent in foods and cosmetics, and in agriculture as livestock feed and a soil amendment. Globally, aquaculture and wild harvest of macroalgae is a $7.4 billion industry (FAO 2010).

Fossil records show that some filamentous macroalgae were present on earth more than a billion years ago (Butterfield 2000). Scientific study and taxonomy of macroalgae predates Linnaeus (Linnaeus 1753, Dillenius 1741), but most of the conspicuous species occurring in accessible regions of the world were described by European phycologists during the 1800s. There are however, still many undescribed cryptic species that are difficult to distinguish morphologically. With the application of DNA sequencing there has been a recent dramatic increase in the number of recognized taxa.

Macroalgae occur in four divisions (phyla) spanning two kingdoms (*sensu* Cavalier-Smith, 2004). In the Plantae (*sensu lato*), the Rhodophyta, or red algae, comprise 10 orders with more than 6,300 recognized species (Guiry and Guiry 2012, Norton et al. 1996); they are mostly marine and estuarine. The Chlorophyta, or green algae, also in the kingdom Plantae, include over 4,600 described marine, estuarine and freshwater species in 15 orders (Guiry and Guiry 2012). The Charophyta are primarily freshwater and are divided into 8 orders with about 3,500 recognized species (Guiry and Guiry 2012). The brown algae make up the class Phaeophyceae in the division Heterokontophyta of the kingdom Chromista (Cavalier-Smith 2004). There are nearly 2,000 recognized species of brown algae in 18 orders (Guiry and Guiry 2012). They are almost exclusively marine and estuarine.

Macroalgae are the foundation of many marine, estuarine, and freshwater benthic ecosystems. They provide food, substrata and protection for a myriad of other aquatic organisms. As primary producers, macroalgae fix carbon, release oxygen, and play a critical role in nutrient cycling by taking up and sequestering massive amounts of nitrogen and phosphorus.

Macroalgal specimens have been collected as vouchers in floristic, ecological**,** and taxonomic studies. They are generally dried on herbarium sheets labeled with their species name, collector(s), date, and location of collection; other information such as habitat, substrata, and associated species may also be included. There are approximately 1.5 million macroalgal herbarium specimens in U.S. herbaria **(**Table 1). This estimate includes about 320,000 housed at the Smithsonian Institution (http://botany.si.edu/projects/algae/herbarium.htm), with most of the remaining specimens in herbaria at various universities, field stations, and natural history museums dispersed across the U.S.

This proposed project has two primary objectives:

1) To establish a Macroalgal Herbarium Consortium (MHC), a network of 49 U.S. universities and museum herbaria with small, medium and large macroalgal collections, that will develop and share tools, workflows, knowledge and experience that will streamline specimen digitization and data access.

2) To digitize more than 1.1 million macroalgal herbarium specimens in the MHC collections and to make the data electronically accessible in a way that will (a) facilitate research to document ecological changes in

marine, estuarine and freshwater environments; (b) engage the public and promote an appreciation of the importance of macroalgae and natural history collections.

PROJECT THEME AND INTELLECTUAL MERIT

The macroalgal specimens in U.S. herbaria have been collected over the past two centuries from broad geographical areas and a wide range of habitats. The collections are the legacy of generations of algal taxonomists, ecologists, and aquatic biologists, many of whom spent their professional lives discovering new species, resolving taxonomic issues, and documenting the distribution across the planet. The collective work and knowledge of these experts and many yet-to-be-made discoveries are sequestered in these collections, but the challenge of examining the physical specimens that are spread across the U.S. herbaria is an insurmountable obstacle for many potential studies. Digitization of specimens within herbaria will provide access to a wealth of information on biodiversity and the spatial and temporal distribution of macroalgal species that can help us understand how changes in aquatic environments have impacted marine, estuarine and freshwater ecosystems over the past 150 years. The insight gained from this information will be critically important for conservation management and developing environmental policy.

Macroalgae occur in lakes, ponds, streams, bays, estuaries and along seacoasts. Their communities are often assemblages of annual and perennial species that grow attached to benthic substrata. Many species are sensitive to environmental change, and as a consequence, community structure can be altered by the loss or gain of species in response to bioinvasions, climate change, and a wide range of human activities. Through its effect on other organisms, disturbances in macroalgal community structure can be amplified to impact entire aquatic, and potentially the surrounding terrestrial, ecosystems.

The proposed plan will digitize the macroalgal specimens housed in U.S. herbaria. The intellectual merit of the project is not so much in the digitization effort, but rather in the types of questions that could be answered through access to the information. Among the hypotheses that could be tested using easily accessible, digitized macroalgal specimen information are the following:

Bioinvasions - Hypothesis 1. *Access to temporal and spatial data on macroalgal distribution will help track the spread of invasive species, identify the dispersal vectors, assess the impact on native community structure and inform scientists and environmental agencies developing response plans*.

Marine, estuarine and freshwater ecosystem invasions by non-native species are reported in increasing numbers (Mathieson et al. 2008b. Invasive species frequently displace one or more, sometimes many, native species. The result can be a decrease in biodiversity, significant changes in community structure, and the collapse of ecosystems. A dataset of macroalgal species can provide a sensitive tool for assessing the progression of ecosystem changes that coincide with the appearance of algal and non-algal invasive species. Data from recent collections can provide insight into the dispersion of invasive species and the vectors of introduction. In turn, this information can aid regulators and environmental agencies developing containment, eradication and/or remediation plans.

As an example, macroalgal communities in the Great Bay Estuarine System (New Hampshire – Maine) have been studied extensively for the past 50 years (Mathieson and Hehre 1983, 1986, Mathieson et al. 1983, West et al. 2005). Very recently, there have been major biomass increases of the green macroalga *Ulva* (Hofmann et al. 2010**)** and the red alga *Gracilaria* (Nettleton 2012, Nettleton et al. in press). Blooms of these species have been attributed to increased nutrient input to the bay. Using molecular tools, it was discovered that the situation was due in large part to the appearance of previously unrecognized non-native cryptic species of both genera. DNA from herbarium specimens of both genera collected in Great Bay over the past 50 years was successfully used estimate the arrival time of the invasive species. A database of herbarium specimens would be invaluable in expanding such a study to track the invasion of these species along the entire northeastern coastline and to locate specimens for DNA confirmation.

In another example, Mathieson et al. (2008b) examined invasive species along the northeast coast of the U.S. using rapid field assessment techniques. They reported 7 non-native macroalgae, with these representing ~27% of the 26 introduced taxa known to occur in the Northwest Atlantic (Hofmann et al. 2010, Lyons et al. 2009, Mathieson et al. 2008b, 2008c, Schneider 2010, Villalard-Bohnsack and Harlin 1997), including 4 green, 4 brown, and 18 red algal species. Site selection in further rapid assessment studies could be guided by mapping the occurrence of suspected alien species or those species known to be associated with them. For example, some invasive species are found primarily as epiphytes on native species, such as the invasive red algal species *Neosiphonia harveyi*, which grows attached to common brown algal species *Ascophyllum nodosum* (Taylor 1962). Maps and specimen images of the native species would be helpful in a survey for the non-native epiphyte.

Climate Change - Hypothesis 2. *The dataset obtained by digitizing macroalgal specimens in U.S. herbaria will provide a sensitive tool for assessing the effects of climate change on marine, estuarine and freshwater ecosystems.*

Approximately 60% of the 1.5 million macroalgal specimens in U.S. herbaria were collected more than 50 years ago, many as early as the mid-1800s. Combined with the 40% that were collected more recently, the species distribution data will provide a valuable tool for elucidating temporal and geographic shifts in macroalgal communities associated with long-term changes in environmental conditions. Frequently, phycologists use collection information on herbarium specimen labels to locate sites for new studies. As a result, many locations have been collected repeatedly over a period of time (e.g. Dawson 1959, 1965, Doty 1948, Littler 1979, Littler and Murray 1975, Littler et al. 1991, Mathieson and Hehre 1986, Mathieson et al. 1998, 2008a, Widdowson 1971). A database of macroalgal collection information would make it easier and more efficient to identify specific sites with a higher sampling density in which to study changes in species presence/absence and community structure, and ultimately the effects of long term environmental changes. In the same way that Primack and co-workers (2003, 2004, 2009) have used flowering times of trees to assess climate change, the time of appearance or disappearance of annual macroalgal species or the time of reproductive maturity could be used to study the effect of climate change in aquatic environments.

Studies in several localities have demonstrated significant changes in community structure that coincide with even small changes in water temperature. For example, Parker and Dixon (1998) examined climate change implications in a study of fish communities in Onslow Bay (North Carolina), where a study is currently underway documenting similar changes in macroalgal biodiversity (D.W. Freshwater, pers. comm., 2012). Other long-term changes in macroalgal communities have been documented on Mount Desert Island and Casco Bay, Maine (Mathieson et al, 1998, 2008a) and in southern California (Littler et al., 1991). Long-term studies and comparisons of historical and recent herbarium records from Helgoland Island (North Sea) have provided compelling documentation of species losses through climate change, bioinvasions and of anthropogenic disturbance. (Bartsch and Kuhlenkamp, 2000). Such studies could be greatly facilitated by a freely available database of macroalgal specimens.

Human Impact - Hypothesis 3. *Temporal changes in the geographic distribution of macroalgae can be used to understand the impact of human activity on marine, estuarine and freshwater ecosystems*

In a study by Mathieson et al. (2008a), data from macroalgal specimens collected at a number of sites in Casco Bay, Maine by F.S. Collins in the late 1800s and early 1900s were compared to recent collections from the same sites. Changes in the presence/absence of species were most dramatic at sites that had experienced the greatest amount of human activity, including shipping (freight and oil), development, industrial and municipal discharge, and aquaculture operations. The more pristine sites exhibited relatively little change in species diversity. The research was complicated by the fact that Collins's specimens have been distributed across a number of herbaria. Collins's diaries were helpful, but some of his species concepts were different than those of today, making it impossible to confirm his species identifications without examining the actual specimens. Studies such as this would be much easier if images and data from U.S. macroalgal collections were readily available on-line.

Eutrophication of coastal waters, lakes and streams from point and non-point source inputs is a critical problem in many locations. The impacts of elevated nutrient levels include the frequent occurrence of nuisance, and in some cases toxic, micro- and macroalgal blooms (Valiela et al. 1992, 1997). Algal blooms can deplete oxygen in the water column during the night or through oxygen consumption by microorganism as the algae decomposes (Peckol et al. 1994). Anoxic conditions can lead to stress or death of fish, shellfish and other aquatic organism. In addition to ecosystem impacts, alga blooms create health threats as decomposing organism cover shorelines, or as toxic algal species are ingested by shellfish (FAO 2004). Algal blooms are frequently associated with changes in community structure marked by flourishing opportunistic algae including non-native invasive species (Maggs and Stegenga 1999, Nyberg 2007, Nyberg and Wallentinus 2005). Macroalgal collection data can be used to discover recent changes, such as the appearance or disappearance of indicator or keystone species, that are harbingers of problems. Such information can inform scientists and environmental managers of impending issues. In other cases, radical changes in community structure can be used as evidence to support regulated change such as the construction of more efficient wastewater treatment facilities or reduction of agricultural runoff.

Other Uses of Collection Data in Macroalgal Research

There are a number of other types of studies that can be supported by the availability of macroalgal specimen data. For example, herbarium specimens are indispensible in taxonomic studies. Recently, herbarium specimens have been used for molecular-based taxonomic studies (Brodie et al. 2007, Lane et al. 2006, Sutherland et al. 2011). In the same way that herbarium specimens provide morphological references for species identification, DNA sequences from type and authentic specimens can provide an authoritative reference for molecular identifications. Molecular data from macroalgal herbarium specimens have been used for studies that have lead to the description of new species (Neefus et al. 2002, Brodie et al. 2007, Mols-Mortensen et al. 2012), taxonomic synonymization (Neefus and Brodie 2009), range extensions discovery (Neefus et al. 2008; Mols-Mortensen et al. 2012), and extensive taxonomic revisions of genera (Choi et al. 2001), families (Lane et al. 2006) and entire orders (Sutherland et al. 2011). A specimen-level database for all U.S. collections would make it possible to locate and visually examine specimens that would be useful for taxonomic studies without having to visit or request loans from many herbaria. Another example is that anomalies on maps of georeferenced specimens can flag misidentified specimens, which may lead to further taxonomic discoveries.

There will be synergy between this project and already-funded TCNs that are providing data to study the effects of environmental change on terrestrial communities, such as the Lichens, Bryophytes and Climate Change TCN (LBCC, 2011), and Macrofungal Collection Consortium (MCC, 2012). The strategic plan for the Network Integrated Collections Alliance (NIBA 2010) identifies the marine environment as region where large gaps exists in the availability of biodiversity information. The macroalgal specimen database, proposed here, would begin to fill the gap and would enable studies on the aquatic environments that would otherwise be difficult or impossible to accomplish. Integration of macroalgal data to other TCN data through iDigBio would allow large-scale comparisons over a greatly expanded range of habitats.

PROJECT PLAN & MANAGEMENT

Institutional Participation. Information about the institutions that comprise the Macroalgal Herbarium Consortium (MHC) is presented in Table 1. All U.S. herbaria with macroalgal collections were invited to join the MHC and the combined collections include nearly all macroalgal specimens in non-federal U.S. institutions. Institutions will be referred to by their *Index Herbariorum* codes (Thiers 2012), which are included in Table 1. Collections range from very large, such as NY, FH, F, UC, BISH and NHA, to mid-sized, such as NCU, MICH, PH and WTU, to quite small, such as WVA, YUO, and LSU. While some of the large herbaria have collections from a broad geographic range, some of the small- and medium-size collections represent intensive sampling in specific locations. In some collections, such as NY, BKL, and YU, 95% of the specimens were collected more than 50 years ago, while in other, such as NHA, MASS, and

WNC, the vast majority of the specimens are more recent. Since the aim of the study is to provide biodiversity information and spatial and temporal distribution data, all specimens are considered important.

There are three main categories of institutional roles in the digitization effort (Table 1): **Primary Digitization Centers** (PDCs) will conduct on-site digitization of their own collections plus those from other MHC institutions. PDCs include the lead institution (NHA) plus the 5 institutions submitting collaborating proposal. The second category, **Digitizing Institutions** (DIs), will conduct on-site digitization of only their own specimens. DIs will be funded through sub-awards on the lead or a collaborating proposal. **Specimen/ Data Contributing Institutions** (CIs) will send their collections to one of the PDCs for digitization and/or will share data from already-digitized collections through the MHC. CIs will not receive funding directly from the project. Institutional roles were determined by size of the collection, preference and experience of personnel, time commitment to other projects, geographic location and condition of the collection. It should be stressed that all members of the MHC will be full participants throughout the project.

Table 1. Institutions and their roles in the Macroalgal Herbarium Consortium digitization project by region and numbers of specimens. Boldfaced entries indicate institutions (PDCs) submitting collaborative proposals. *Index Herbariorum* (Thiers 2012) codes for each institution are given in parentheses. The institution where each collection will be digitized is listed under the heading Digitizing Institution. All institutions have agreed to their role. States EPSCoR states are labeled with a superscript [e].

				Macroalgal Specimens		
Organization (Herbarium Code)	State	Role	Digitizing Institution	Total Number	Already Databased	Already Imaged
Northeast/New England						
University of New Hampshire (NHA)	**NH[e]**	**PDC**	**NHA**	**85,000**	**3,000**	**3,000**
Harvard University (FH)	MA	DI	FH	150,000	0	0
University of Vermont (VT)	VT[e]	DI	VT	3,500	0	0
University of Massachusetts (MASS)	MA	CI	NHA	7,000	0	0
Yale University (YU)	CT	CI	NHA	6,600	0	0
Brown University (BRU)	RI[e]	CI	NHA	3,000	0	0
University of Rhode Island (KIRI)	RI[e]	CI	NHA	2,500	0	0
University of Connecticut (CONN)	CT	CI	NHA	2,350	2,350	1,000
Central East Coast						
New York Botanical Garden (NY)	**NY**	**PDC**	**NY**	**150,000**	**3,500**	**2,000**
Academy of Natural Sciences (PH)	PA	DI	PH	37,000	224	244
Brooklyn Botanic Garden (BKL)	NY	DI	BKL	9,200	0	0
Rutgers University (CHRB)	NJ	CI	NY	3,500	0	0
New York State Museum (NYS)	NY	CI	NY	500	0	0
West Virginia University (WVA)	WV[e]	CI	NY	150	0	0
Southeast/Gulf Coast						
University of North Carolina (NCU)	**NC**	**PDC**	**NCU**	**60,000**	**0**	**0**
University of South Florida (USF)	FL	DI	USF	20,000	0	0
Duke University (DUKE)	NC	DI	DUKE	19,000	16,000	0
Univ of North Carolina Wilmington (WNC)	NC	CI	NCU	6,000	6,036	200
Mote Marine Laboratory (MOT)	FL	CI	USF	5,000	0	0
Florida Museum of Nat. History (FLAS)	FL	CI	NCU	2,500	0	0
University of South Carolina (USCH)	SC[e]	CI	NCU	2,000	0	0
Univ. of Texas Marine Science Institute	TX	CI	UNH	2,000	0	0

University of Alabama (UNA)	AL[e]	CI	NCU	1,000	0	0
Texas A&M University (TAES)	TX	CI	NCU	1,000	0	0
University of Texas Pan Am. (PAUH)	TX	CI	UNH	100	0	0
Louisiana State University (LSU)	LA[e]	CI	NCU	40	0	0
Midwest						
University of Michigan (MICH)	**MI**	**PDC**	**MICH**	**60,000**	**250**	**250**
Field Museum of Natural History (F)	IL	DI	FH	109,505	0	0
Miami University (MU)	OH	CI	MICH	10,000	0	0
Butler University (BUT)	IN	CI	MICH	5,500	0	0
Michigan State University (MSC)	MI	CI	MICH	3,000	0	0
Youngstown State University (YUO)	OH	CI	MICH	150	0	0
Northwest and Pacific						
University of Washington (WTU)	**WA**	**PDC**	**WTU**	**25,000**	**0**	**0**
Bishop Museum (BISH)	HI[e]	DI	BISH	78,795	78,795	0
University of Guam Marine Lab (GUAM)	GU	DI	GUAM	13,600	13,600	0
University of Hawaii (HAW)	HI[e]	DI	HAW	12,000	0	0
Oregon State University (OSC)	OR	DI	OSC	9,000	0	0
University of Alaska SE (ALAJ)	AK[e]	DI	ALAJ	8,300	8,000	0
Univ. of Alaska Museum (ALA)	AK[e]	CI	WTU	7,182	0	0
Nat. Tropical Botanical Garden (PTGB)	HI[e]	CI	PTGB	300	0	0
South-Central Western US						
University of California (UC)	**CA**	**PDC**	**UC**	**202,000**	**102,000**	**102,000**
University of California (UCSB)	CA	CI	UC	7,500	1,936	0
San Diego Nat. Hist. Museum (SD)	CA	CI	UC	5,000	500	396
Hopkins Marine Station (GMS)	CA	CI	UC	4,000		
Humboldt State University (HSC)	CA	CI	UC	2,000	0	0
University of California (DAV)	CA	CI	UC	700	0	0
University of Utah (UT)	UT[e]	CI	UC	300	0	0
University of California (IRVC)	CA	CI	UC	200	0	0
Shasta College (RESC)	CA	CI	UC	100	0	0
Totals						
49	26			**1,143,072**	**236,191**	**109,090**

Project Activities and TCN Structure- The project will build on tools and workflows that have been developed by other digitization projects and will use a hybrid of those employed in the LBCC (Lichen and Bryophytes Climatic Change) TCN, the Macrofungi Collection Consortium (MCC), the New England Vascular Plant (NEVP) TCN and the Caribbean Plant and Fungus digitization project at NYBG. The project will also take full advantage of digitizing and online tools, workflow protocols, and training available through the iDigBio national resource. Some MHC members are involved in other digitization projects and will be able to share their experiences to increase efficiency and ensure data quality. The project will develop modified workflows and tools to streamline and more tightly integrate components produced by other TCN and U.S. and global digitization efforts.

The vast majority of the resources committed to the project will be directed at digitization activities (see "Integrated Budget Summary" in the Supplementary Documents), although significant resources will be focused on portal development, training, education and outreach. The work plan distributes specimen digitization across the 18 PDCs and DIs. Each of these institutions will barcode, image, transcribe label information, georeference, and nomenclaturally update specimens. Other digitization projects have centralized major components of the process, but the more distributive model proposed here has several advantages: 1) Sharing the effort among more institutions will increase the throughput rate without requiring the development and construction of high speed “assembly line” equipment; 2) Distributing label transcription, georeferencing and annotation avoids process bottlenecks and backlogs; and 3) Perhaps most importantly, it provides infrastructure and capacity for the entire digitization process at a number of smaller institutions that do not already have this capability. At the end of the project, these institutions will be able to digitize their other collections independently or participate in other collaborative initiatives.

Several components of the project will be centralized, including overall project management and logistics, tool development, data integration and bioinformatics, portal management, and coordination with iDigBio and other resources. These tasks will be coordinated by the project PIs and an Information Management Team (IMT) that includes a full-time Informatics Specialist/Programmer (NHA), a full-time Portal Manager (NY) and a part-time Informatics Specialist (WTU). Additional support will be available from IT staff at NHA, NY and several of the other participating institutions, especially UC, and HAW.

Workflow

<u>Collections</u> – Collections at the specimen-contributing CIs will be shipped to one of the PDCs. Larger collections will be shipped in batches on a schedule determined by the capacity and current workload at the PDC. CIs will be provided with boxes and packing material and UPS shipping labels. Once the specimens have been imaged and QC performed at the PDC, they will be shipped back using the same shipping materials. On arrival at the PDC, incoming collections will be frozen for a period according standard herbarium procedure for incoming loans. Collections owned by the DIs and PDCs will be digitized on-site.

<u>Barcoding and Preliminary Data Capture</u> – Any specimen that does not already have a barcode will receive one prior to imaging. Barcodes used for each collection will be consistent with barcode specifications used at the institution that owns the collection. During the barcoding process, a skeletal database record will be created for each specimen by scanning its barcode and entering the “filed as” taxon name. Specimens in need of repair before further handling would be segregated at this point for repair. For barcodes that do not include a herbarium code, the code will be added as a prefix in the database record. In most herbaria, specimens are organized in folders by species. To speed the process, the preliminary data capture input screen will automatically carry the taxon name, and other optional fields, forward from record to record until changed. Based on experience of other digitization projects at The New York Botanical Garden (NY), the throughput rate for this task is 100 specimens per hour including time for repair of 1 specimen out of 50.

<u>Imaging</u> – Each PDCs and DIs will be equipped with one or more Imaging Stations similar to those developed for the LBCC and MCC TCNs. They consist of a Canon Eos 5D Mark II 21.1 megapixel digital camera with a Canon EF 50mm f/2.5 Macro lens, a Photo e-Box Plus 1419 (MK Direct), a Kaiser RS 1 copystand, a barcode reader, a laptop computer and several other smaller components (detailed specifications are provided in the Budget Justifications). A specimen is placed inside the lightbox, which sits on the copystand. The camera is attached to the copystand arm where it has a view of the specimen through a port on the top of the light box. The camera is connected via a USB cable to the laptop computer. Canon utility software allows the operator to view the specimen image in real-time and to make any adjustments necessary to exposure, color balance and focus before capturing the image. In practice, the camera autofocus works well and exposure and color balance adjustments need to be checked only periodically and at the start of an imaging session. The image is saved to disk using the barcode as the filename. Alternatively, the image can be saved using the software’s default file name (a prefix plus incremented number). The files are then automatically renamed to the barcode in batch mode using Bardecodefiler (Softek Software, Leighton Buzzard, UK). Based on experience at NY, the latter method increases the throughput significantly. Some algal herbarium sheets contain multiple specimens or packets of duplicate specimens that require handling. For specimens that do not require special handling, NY has found an imaging rate exceeding 100 per hour. Because a number of specimens will require special handling, a rate of 50 per hour is more realistic.

Label Transcription - There are a number of existing tools that can help in transcribing the information on specimen labels into the project database. Many recently collected specimens have typewritten or computer printed labels. Others, including older specimens, have handwritten labels. Specimens with computer printed labels are sometimes already databased and where available, these records will be added to the project database. Computer printed labels where digital records cannot be located, and typewritten labels, will be transcribed with the help of OCR. Augmenting OCR for specimen label transcription is an area of active development. It was the subject of a recent iDigBio workshop (October 1 & 2, 2012). A "Hack-a-thon" to develop solutions for issues related to the process has been organized by iDigBio for February 2013; several of the MHC members plan to participate. The Informatics Specialist/Programmer hired for the proposed project will work in this area. Pending developments that further improve efficiency, ABBYY FineReader (v. 11, Professional Edition, ABBYY, Moscow, Russia) will be used in batch mode to convert typewritten and printed labels to text. The text files containing OCR results from multiple labels will be edited manually in MS Word (Microsoft, Redman WA) to remove spurious characters and to add line breaks to separate fields and remove line breaks to join multiple lines of single fields. The manual editing has been found to greatly improve the accuracy and efficiency of SALIX (Arizona State University), which will be used for semi-automated parsing of label text into appropriate database fields. Using ABBYY FineReader in batch mode plus SALIX, throughput rates for transcribing all label information should exceed 35 per hour. Although some progress has been made in methods for transcribing handwritten herbarium labels (Gehrke et al. 2011, Steinke 2012, Steinke et al. 2010, 2011) at this point it is more efficient to do this task manually. If the data transcribed from handwritten labels is limited to collector, collector's number, determiner and locality (taxon name and barcode will have been captured during barcoding), 30 labels per minute is a conservative estimate. Records for labels requiring interpretation by an expert will be flagged.

Georeferencing – Data records containing locality information and that do not already have geographic coordinates will be georeferenced using GeoLocate in batch mode and, as the project database of collection locations increases, the collaborative community-based version. Refinement of collecting locations is an excellent opportunity for crowd-sourcing especially if it can be done by the collectors themselves or for older specimens by a researcher or other person familiar with the original collector's study sites. Improvements to georeferencing accuracy, precision and efficiency will be an area of active development by project personnel. Many herbarium sheets in the collection are of different species collected same site, or reflect multiple collecting trips to the same site. It would not be unusual to find dozens or even hundreds of specimens from the same location in one herbarium, or across several. Through-put estimates for georeferencing are about 30 locations per hour for an experienced person. Considering there will be multiple specimens from some locations, it should be possible to process 100 specimens per hour.

Nomenclatural Annotation - Like all groups of organisms, macroalgae have gone through an escalating amount of taxonomic and systematic revision. Based on molecular evidence, species, genera, families and entire orders have been revised and specimens in most herbaria are in desperate need of annotation to apply new names. Algaebase (Guiry and Guiry 2012) is used as a taxonomic authority database by phycologists and algal taxonomists around the world. It is continually updated to reflect recent taxonomic changes. Algaebase has recently become part of the World Registry of Marine Species centered at VLIZ Flanders Marine Institute in Belgium (http://www.marinespecies.org). WoRMS provides an online tool (http://www.marinespecies.org/aphia.php?p=match) to lookup the currently accepted name of any taxon using Algaebase as the authority. In batch mode, the WoRMS system will accept a list of taxon names and will return the list with additional fields for the status of the name, the currently accepted name, the authorities, the date published, classification, and citation. Through a tool developed for the proposed project, the WoRMS system will be used to check the status of the taxon names on all digitized specimens. Where the name on the label differs from the currently accepted name, the MHC developed tool will add this information to the project database and at the request of the institution that owns the specimen, it would generate a file of annotation labels for their specimens. For specimens that have been georeferenced during the project, coordinates will be added to the annotation. The WoRMS system operation in the background, so the actual personnel time required to update the records is expected to be no more than an hour per 1000 records. Printing and attaching annotation labels to specimens are not part of the estimate and would be the responsibility of the curatorial staff at each institution.

Informatics

Data management, tool development, and information dissemination will be the primary responsibility of the full-time Informatics Specialist/Programmer to be hired at NHA, and the full-time Portal Manager to be hired at NY, with the assistance of existing computer and IT staff at both institutions.

Data Pipeline - Additional details are provided in the Data Management Plan, but an outline of the starting configuration for data flow follows. This informatics pipeline will be continually modified as new tools and methods are developed to increase efficiency and streamline the process.

1. Preliminary Data Capture at each institution will create a skeletal record for each specimen using a simple data capture application created for the project. Data will reside in a local database at the DI, with backup on the project server at NHA and one of the other PDCs.

2. Specimen images (21.1 megapixels) captured in the Canon EOS utility will be saved in Canon's proprietary RAW format (~25.8MB). Images will be automatically renamed corresponding to their barcode in batch mode using Bardecodefiler. Renamed image files will be batch processed in Adobe Lightroom v4.1 (Adobe Systems Inc, San Jose CA) to produce DNG format files for archival, 8MB RGB JPEG files for web access, and cropped grayscale files for label transcription. All data will be stored onsite at the PDCs and DIs and in repositories at NHA, NY, UC, HAW, or WTU.

3. Grayscale images of printed and typed labels will be batch OCR processed in ABBYY Finereader v.11 with output to a text file, which will be manually edited in MS Word. The edited file will be saved and backed up locally. Edited text files will be parsed into Darwin Core (TDWG 2009) compliant fields semi-automatically in SALIX. CSV output files from SALIX will be saved and backed up locally and used to populate records in the Preliminary Data Capture database corresponding to the OCR'd labels.

4. The MHC Digitizing Institutions are already using NSF funded Specify 6 (Specify Software, University of Kansas), commercial KE Emu 4 (KE Software), or alternative software for their collections database. The database with the skeletal and SALIX populated records will be imported into Specify, KE Emu, or other database software. Information from grayscale images of handwritten labels will be manually transcribed directly into Specify or KE EMu to take advantage of lookup tables and entry verification features.

5. Specify, KE Emu, or alternative databases will be backed up regularly on- and offsite. Periodically, new project records will be exported for batch georeferencing via GeoLocate and nomenclatural updating against Algaebase using WoRMS. Completed records will be sent in Darwin Core (DwC) compliant format to NHA for backup and the Project Portal Manager at NY for dissemination through the Macroalgal Collection Portal, the iDigBio HUB and global portals like GBIF. A copy of the data and images for collections from each of the CIs will be sent to the contributing institution in a format compatible with their needs.

Information Dissemination – As it is generated, project data will be sent to iDigBio for dissemination through the national HUB where it will join the other TCN data. This will enable broad biogeographic hypotheses to be tested, including comparisons of terrestrial and aquatic responses to disturbance and long term climate change. The project will also create a Macroalgal Collection Portal using a Symbiota instance tailored to the needs of the MHC, the outreach and educational goals of the project, and the needs of phycological community as a whole. Information about the project will be communicated at scientific venues including the annual meetings of the Phycological Society of America (PSA) and the Society for the Preservation of Natural History Collections (SPNHC). Information about macroalgae, the importance of herbarium collections and the project will be disseminated through outreach efforts (see Broader Impacts).

Tool Development – Tool development in the project will focus on applications that will increase efficiency by streamlining the digitization workflow and on development of additional functionality in the Macroalgal Collection Portal. A number of the components in the workflow are open-source software, maintained and continually improved by a community of programmers. The Informatics Specialist/Programmer and the project Portal Manager will be actively involved in the evolution of these tools. They will concentrate on augmented OCR, georeferencing, areas that the DIs identify as digitization bottlenecks, and on adding portal features identified by the CIs, educational participants, and the project Advisory Committee.

The project will make a significant contribution to georeferencing technology, especially for use with aquatic organisms. Macroalgae occur in or along the edges of bodies of water. It is also true that marine macroalgae do not occur in freshwater and freshwater macroalgae cannot tolerate saltwater. For georeferencing, these factors greatly reduce the number of places from which a specimen could be collected. It also presents an opportunity for modifying existing georeferencing software to allow it to be “tuned” for macroalgae, other aquatic organisms, or other groups with restricted distributions. Many species have limited geographic distributions; there are relatively few Pacific seaweeds that occur in the Atlantic and vice versa. By building species awareness and intelligence into georeferencing tools, the software would learn where a given species is more likely to occur and where it shouldn’t be. These modification will increase the accuracy, precision and efficiency for this and other georeferencing efforts.

<u>Quality Assurance</u> – Quality control will be embedded in every aspect of the workflow and will be a combination of human and automatic processes. In general, automated processes in the workflow will be checked by humans and vice versa. Any quality issues detected will be immediately fixed and any type of error that appears to be systematic will be addressed through workflow modification. Overall, the proposed methodology has been guided by best practices collected by the iDigBio national resource and by the experience of members of the MHC who are participating other digitization projects.

Training

Training will be provided at several levels using a combination of workshops, one to one instruction and web based media. Digitization at the DIs and PDCs will be done by a combination of undergraduate and graduate students, museum interns, and herbarium staff. Since the member institutions are widely distributed, regional training opportunities will be scheduled. A distributed model for training will be used. There will be train-the-trainers sessions at NHA and/or NY early in the project, which would train personnel from the PDCs. In turn, the PDCs will hold sessions to train personnel at the DIs in their region. PDC and DI personnel will train their students and interns. MHC members will also take advantage of training sessions, workshops, and online resources offered by iDigBio.

Project Oversight

<u>Day-to-Day Management</u> – Overall project management and planning will be coordinated by the lead Principal Investigator (PI). Day-to-day project operations and logistics will be managed by the Project Coordinator at NHA. Two management-level positions will be hired for the project, the Portal Manager, to be based at NY and an Informatics Specialist/Programmer at NHA. PIs on the collaborating proposals from the PDC and sub-award PIs at the DIs will oversee operations at their respective institutions. Collectively, the PIs, the Project Coordinator, the Informatics Specialist and the Portal Manager will form the Project Management Team. The focus of the Project Management Team in Year 1 will be on start up functions, such as training and support, and initiating a simple but effective communication strategy using Google Groups for sharing training materials, facilitating discussions among participants and providing status updates on the work on individual data sets. Training and support will remain key responsibilities in Year 2, when many additional institutions will join the project. By Year 3, the focus will turn toward making sure that the promised deliverables are on track, supplying additional support for aspects of the project that may have fallen behind schedule, and promoting the project with the goal that it will be broadened after the completion of the project to include a wider range of organizations, ideally with a global scope.

<u>The Consortium</u> - All participating institutions are members of the Macroalgal Herbarium Consortium for the entire duration of the project and beyond. As such, they are eligible to represent the project at scientific, technical, citizen scientist meetings, and other venues. MHC meetings will be scheduled in association with

annual meetings of the Phycological Society of America (PSA) and the Society for the Preservation of Natural History Collections (SPNHC). These meetings will give participants the opportunity to report on progress, share tips and techniques, and suggest additional ways to enhance the project. Between in-person gatherings, Consortium members can communicate via email, or in discussions on the Google Groups site.

Advisory Committees – A Science Advisory Committee will help make certain the project fulfills its goals relative to the Theme. It will contain phycologists engaged in systematic or ecological research (Drs. Arthur Mathieson, UNH; Clinton Dawes, USF; Craig Schneider, Trinity College; Juliet Brodie, Natural History Museum London; and Michael Wynne, UMich) and PIs from other TCNs and biodiversity projects (Dr. Barbara Thiers, NYBG; Patric Sweeney, Yale). The Technical Advisory Committee will include people with expertise on digitization workflow, tools and infrastructure (Deb Paul, iDigBio; Anne Barber, SALIX; Ed Gilbert, Symbiota). Education and Outreach Committee members will include staff from the education departments of the Field Museum, the Bishop Museum and others. Members of the Advisory Committees are welcome to all annual project meetings, to participate in the on-line Google Groups discussions. Meetings of the Advisory Committee will take place using Skype conferencing as needed.

Table 2. Timetable of project activities. Institutions omitted from the Digitization Equipment Installation row already have equipment in place.

	Year 1	**Year 2**	**Year 3**	**Year 4**
Informatics Infrastructure	Set-up informatics infrastructure and workflows	Continued development and refinement	Continued refinement	Continued refinement
Digitization Equipment Installation	At NHA, BISH, GUAM, HAW, NY, MICH, WTU, OSC and ALAJ	At FH, NCU, USF and DUKE	At BKL	
Training	Training-the-trainers workshop. DI training at PDCs. Online training.	Training for DIs scheduled to start digitizing in Yr 2 or 3	Training for DIs scheduled to start digitizing in Yr 3 or 4	
Specimen Digitization	Begins at NHA, BISH, DUKE, GUAM, HAW, NY, MICH, WTU, OSC, ALAJ. Finishes at HAW, OSC, ALAJ	Begins at FH, F, NCU, USF, Finishes at BISH, DUKE, GUAM, WTU	Begins at VT, BKL, UC. Finishes at FH, VT, NCU, USF, MICH	Begins at PH. Finishes at NHA, F, NY, PH, BKL, UC
Portal Development	Set-up Symbiota instance; customize for the MHC project	Module development based on MHC and end-user feedback	Module development based on MHC & user feedback	Module development based on MHC & user feedback
Outreach	Internships at NY . Public engagement development at BISH. Presentations to end-users at PSA and SPNHC	Internships at NY and F. Public engagement development at F. Presentations to end-users at PSA and SPNHC	Internships at NY and F. Public engagement development at F. Presentations to end-users.	Internships at NY and F. Public engagement development at F. Presentations to end-users at PSA and SPNHC
MHC Meetings	At PSA and/or SPNHC	At PSA and/or SPNHC	At PSA and/or SPNHC	At PSA and/or SPNHC

BROADER IMPACTS AND OUTREACH

Advance Discovery and Understanding While Promoting Teaching, Training and Learning
Digitization will be done largely by undergraduate and graduate students who will be trained to image specimens, transcribe labels, georeference locations, and update taxonomic names. In addition to technical training, students and interns will have opportunities to participate in macroalgal collecting trips and learn to make herbarium specimens. Students wishing to develop relevant independent research or education projects that involve macroalgal collections will be mentored by the senior personnel, staff and/or graduate students. For example, a student planning to teach science, might develop curricular modules on the uses of macroalgae as sea vegetables (e.g. sushi, kelp pickles, rockweed soup). Students would be able to apply to the project for funding to attend regional scientific meetings, like the Annual Northeast Algal Symposium, to present their results; they would be mentored in developing an oral or poster presentation.

Broaden Participation of Underrepresented Groups
The MHC included institutions from 26 states and of these, 10 are EPSCoR states. The project's distributed model extends digitization training and infrastructure to smaller institutions that have not been part of TCNs or similar programs. Four of the DIs are in EPSCoR states. Internships at the Field Museum will recruited from the Northeastern Illinois University (NEIU), Student Center for Science Engagement (SCSE), which is committed to engaging students from diverse backgrounds in STEM fields with a focus on underrepresented and low-income students. Similar opportunities will be provide for underrepresented groups through the New York Botanical Garden internships where they will help with digitization and participate in museum programs and field trips. The MHC institutions in Hawaii and Guam will engage Pacific-Islanders directly in the project and through the public education modules developed for the Bishop Museum Science Adventure Center.

Enhance Infrastructure for Research and Education

One of the stated goals of the NSF ADBC program is to foster the development of tools, workflows and other infrastructure that will increase digitization efficiency. One focus of the project will be on workflow optimization through better integration of the components. For example, OCR assisted label transcription will be improved by developing tools to semi-automate what is now a manual editing step between OCR and field parsing. We will also focus on increasing georeferencing throughput by tuning the process for aquatic organisms, which occur only in bodies of water and along coastlines. The Macroalgal Collection Portal will include a crowd-sourcing module to allow scientist, citizen scientists, and others with local knowledge to refine collection locations. Improved tools and workflows will be made available for use by other specimen digitization projects.

Broad Dissemination to Enhance Scientific and Technological Understanding
Eight of the 49 members of the MHC are museums or botanical gardens. NYBG and The Field Museum will establish internship programs around the project. The Field Museum's Education Department reaches 650,000 educators, students, adults and families annually. Moreover 250,000 educators and students visit the Museum on field trips each year. The Field Museum Education Department will participate in the project and include macroalgae in their mission, which is to bring scientific discoveries to the general public, integrating them into all aspects of their public programming—from lectures, to distance learning, to educator professional development. Their efforts will include K-12 Curriculum Development in Chicago High Schools, teacher training workshops on how to use the project's Symbiota portal in the classroom to teach about macroalgae and concepts of taxonomy, development of laminated guides to common macroalgae that explains their importance as bioindicators of environmental stress, they will develop modules on macroalgal biofuel production as part of their Biofueled Vehicle program, and initiate a macroalgal LifeDesk project in which all MHC members can participate.

The Bishop Museum is visited by 450,000 students, adults and families each year. The Bishop Museum's Education Department makes 12 presentations each day that reach 225,000 visitors annually, not including their travelling exhibits. As part of the proposed project, they will develop programming for their Science Adventure Center focusing on ethnobotanical uses of macroalgae, algal biodiversity, and the use of macroalgal herbaria. The modules will be seen by every visitor and will act as a template for other museums, aquariums and science centers. Other special programs at the Bishop Museum where macroalgae can be featured include the *Mad About Science Festival*, the *Communicating Climate Change* program, and the Science and Culture of Art program, which guides students in hands-on activities that might include creating art from some of the most beautiful common red, green and brown seaweeds they can find washed up in the shore.

Citizen science groups will be engaged to assist with project outreach. Examples include the Friends of the Farlow Herbarium at Harvard, similar volunteer groups at other herbaria, and the Marine Docents at UNH, who provide weekly cruises aboard the RV Gulf Challenger for the public to learn about marine biology. Project personnel will provide field training, demonstrations, materials and information to support their efforts. The Macroalgal Portal will include a variety of modules for public education and involvement, such as a virtual herbarium tour, how to make a herbarium specimen, sea vegetable recipes from famous chefs, and a "Name that Seaweed" module where K-12 students can suggest scientific names for newly discovered species.

The data available through the Macroalgal Portal tools will have an enormous impact on the scientific community including phycologists, conservation biologists, and ecologists. It will provide valuable resource for environmental managers and regulators that potentially can influence future policies and help protect our aquatic ecosystems. A project status tool on the Portal will map the MHC institutions and provide digitization schedule and status for each collection. Portal tools will allow accurate mapping of species distributions, and allow for selection of biodiversity survey sites (e.g. high priority, low collection density).

Project awareness will be promoted through Herbarium and Algal-L listserves, presentations at regional, national, and international conferences including annual PSA and SPNHC conferences, scientific publications, and demonstrations/workshops for educators and environmental management groups. The PIs and project personnel will be enthusiastic participants in iDigBio meetings, discussions and workshops.

Benefits to Society

Macroalgae, their role in the ecosystem and their importance to society are poorly understood by most people outside the phycological community. As previously stated, macroalgae are the foundations of many aquatic ecosystems. The ability to test the hypotheses that form the theme of this project will provide an understanding of how bioinvasions, climate change and human activity have altered the distribution of macroalgal species and ecological community structure over the past 150 years and predict how they may do so in the future. The analysis, interpretation and synthesis of this information can have a profound effect on how we manage our aquatic resources to prevent ecological catastrophes, public health problems, and economic challenges.

PROJECT DELIVERABLES

The proposed project has major 4 components: 1) Establishment of the Macroalgal Herbarium Consortium; 2) Digitization of the MHC macroalgal collections; 3) Providing access to the digitized collections; and 4) Educational enrichment and public engagement. Project deliverables for each category are as follows.

1) Establishment of the Macroalgal Herbarium Consortium

The proposal development for the MHC project has already brought together people from 49 universities, museums, and field stations in 26 states and U.S. possessions. The consortium members share a set of common goals centered on their collections and their scientific and educational missions. The project will build a strong and enduring relationship among the participants that will foster the sharing of ideas, knowledge and experience. Evidence of success will be the active involvement of members in MHC meetings, discussions, and other activities

2) Digitization of the MHC collections

The most quantifiable and "tangible" product of the project will be the database of digitized macroalgal specimens. As detailed in Table 1, MHC collections contain in excess of 1.1 million macroalgal specimens. Determined by the size of their collection, digitization resources, commitments to other projects and institutional preference, digitization activities at various DIs will begin and be completed in different years.

Based on institutional estimates of their specimen number, a total of 197,597 specimens will be digitized during Year 1 at NHA, BISH, DUKE, GUAM, HAW, NY, MICH, WTU, OSC and ALAJ. In Year 2 a total of 314,092 specimens will be digitized at NHA, FH, F, BSIH, GUAM, NY, NCU, USF, DUKE, MICH and WTU. During Year 3 a total of 378,694 specimens will be digitized at NHA, FH, VT, F, NY, BKL, NCU, USF, MICH and UC. In the final year of the project 252,689 specimens will be digitized at NHA, F, NY, PH, BKL and UC.

3) Providing access to the digitized collections
The most visible product from the project will be the online Macroalgal Collection Portal, which will provide access to the digital collections for scientists, environmental managers, and the public. The portal, an instance of Symbiota, will be created at the beginning of the project and data will be added continuously. Specimen records will be accessible as soon as they are completed by the DIs and checked by the institution that owns the collection. With input from Consortium members and feedback from portal end-users, the Project Management Team will design and implement new features that improve portal functionality. Features might include such things as automatic generation of annotation labels for collections to reflect taxonomic revisions, enhancements to modules that allow macroalgal taxonomists and other experts initiate changes, tools to generate herbarium loan requests, or educational modules for K-12 and public engagement. A number of other macroalgal databases are available via the internet, including the collections at the Smithsonian Institution. The portal MHC portal will integrate access to these additional resources. Utilization of the portal will be quantified through Google Analytics, and by the utilization of project data in presentations and publications

4) Education, public engagement and other outreach efforts will be quantified by the number of students, teacher, adults and families that participate in education programs developed by the Education Departments at the Field Museum and the Bishop Museum; by the number of citizen sciences that help with outreach and the number of people who participate in their activities; and by the utilization statistics for the educational and public engagement modules on the Macroalgal Portal.

RESULTS FROM PRIOR SUPPORT

Barringer: NSF-BRC-0346619. *AILANTHUS Grows in Brooklyn: Curation, Data Capture, and Presentation of Historical Specimens from New York, New Jersey and Connecticut*. (2004 – 2007). Label data from 83,000 specimens were databased and are now searchable on the Brooklyn Botanic Garden website. Specimens were reidentified, repaired, and reorganized in archival folders. The data has been used by regional flora projects and conservation. The PI made a poster presentation on the project at the 2007 SPNHC meeting. **Barrington**: NSF-EF-1120897. *Collaborative Research: Digitization TCN: Mobilizing New England Vascular Plant Specimen Data to Track Environmental Changes (TCN-NEVP)*. ($106,0000; 2012 – 2016). This project was only recently funded. We attended the kick-off meeting of all participating institutions at Yale on September 6. Assembly of materials for data and image retrieval is underway; we have purchased barcode labels and a laptop computer for use with a label maker to barcode specimen folders and do preliminary data capture in years one and two. **Dick:** NSF-DEB-0640379 – *Comparative phylogeography of 12 widespread tropical rainforest tree species* ($494,000; 2007-2012). Supported 27 peer-reviewed publications, laboratory training for four undergraduate students, and four REU supplements for fieldwork in Panama. The project provided resources for mentoring of seven PhD students and two post-docs. **Giblin:** NSF-TCN-1115161 – *Digitization TCN Collaborative Research: North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change* ($107,925; 2011-2015). Digitized 30,000 lichen and bryophyte specimens at the University of Washington Herbarium; coordinating the digitization of 50,000 lichen and bryophyte specimens at other herbaria in the Pacific Northwest; 15 undergraduates have been involved in the project to date. **James**: NSF-DEB-1057453 – *Collaborative Research: Consortium of Pacific Herbaria (CPH) - A Research and Collections Network for Oceania* ($691,466; 2011-2014). The *Herbarium Pacificum* is actively digitizing collections from the Polynesia-Micronesia-Fiji region, and to date has databased label information for 286,000 specimens, of which more than 30,000 specimen sheets have been imaged. **Karol:** NSF-DEB-1036466 – *ATOL: Collaborative Research: Assembling the Green Algal Tree of Life (GrAToL)* ($561,000 to P.I. Karol; 2010-2015). Primary activities have been collection and analysis of gene sequence data and organellar genomic data of more than two hundred representative green algae to generate a taxonomically broad phylogeny and classification for green algae. Thus far three graduate students and several undergraduate students have gained training and experience through this grant. **Liston:** NSF-DBI-0955887 – *Pacific Northwest Herbaria Portal: Developing a taxonomically diverse online resource from large and small collections* ($339,390; 2010-2013). More than 197,000 Oregon State University Herbarium specimens have been digitized and made available via the Consortium of Pacific Northwest Herbaria data portal. A total of 29 undergraduates, two high school students, and five graduate students gained herbarium experience through the DBI grant. **Livshultz:** NSF-EF-1115131 – *Digitization TCN Collaborative Research: North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change* ($105,788 to co-P.I.; 2011-2015). PH will digitize all of its North American lichens and bryophytes in the fourth year of this collaborative project. We are now curating our collections in preparation. **Lücking (P.I.) and Von Konrat (co-P.I.):** NSF-EF-1115002 – *Digitization TCN Collaborative Research: North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change* ($143,252; 2011-2015). This project aims at digitizing approximately 130,000 lichen and bryophyte collections housed at the Field Museum and several collections from other institutions. The project started in early 2012 and thus far approximately 30,000 lichen collections from the Field Museum and Morton Arboretum plus roughly 10,000 bryophyte collections have been digitized and core databased. **McCourt**: NSF-DEB-1112040 – *Collaborative Research: Phylogeny and Systematics of the Characeae (Charales)* (co-P.I. with a K. G. Karol at The New York Botanical Garden for $468,000; 2010-2013). This project is in its third year. Primary activities have been collection and analysis of DNA sequence and genomic data from several hundred taxa of Characeae, collections in the US, Europe, and Australia, and isolation, culture, and specimen preparation for hundreds of Characeae. Outreach to science festivals, schools, and public exhibits are underway. **Mishler, Miller**: NSF-DEB-1057385 – *Digitizing Pacific Coast Seaweeds: Documenting the Past to Interpret the Future* ($551,696; 2011-2013). Imaging of all the specimens funded to be digitized has been completed (73,034 specimens), and we are in the second phase of capturing selected fields for the database (12,527 of the specimens have been databased so far). **Neefus, Sullivan:** NSF-EF-1208829 – *Collaborative Research: Digitization TCN: Mobilizing New England Vascular Plant Specimen Data to Track Environmental Changes (TCN-NEVP)* ($118,100; 2012-2016). This project was only recently funded and two students have been hired on an hourly basis to barcode folders and do preliminary data capture. **Pfister:** NSF-EF-1114957. *Digitization TCN Collaborative Research: North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change.* ($317,967; 2011-2015). Our portion of this project has only recently begun. Two students have been hired to being the digitization and a project coordinator will soon begin. **Pryer**: NSF-DEB-0347840 – *CAREER: Reconciling Patterns of Phylogenetic Rate Heterogeneity in Ferns with Morphology, Ecology, and Life History* ($688,779; 2004-2010). This project contributed to the training of five postdocs (all now have permanent university or museum positions), two graduate students, and three undergraduates. The funding resulted in 25-refereed publications and allowed for the development of an online Fern DNA Database. **Schils**: NSF-DBI-0646290 – *Renovations to the GUAM Herbarium* ($62,158; 2007-2009). In conjunction with restoring herbarium specimens, label information of the terrestrial and marine plant collections of the University of Guam Herbarium were entered in databases. **Sherwood:** DEB-0542608. *Biodiversity of Hawaiian Rhodophyta: morphological vouchers, DNA archival and sequence diversity assessment.* (PI $599,999; 2006 - 2009, plus a one-year NCE).. We characterized a total of 1,946 red algal specimens from the main and Northwestern Hawaiian Islands, representing 24 orders, 152 genera and 252 species of red algae. Thus far 10 manuscripts have been published from the survey, with one in revision, and one in press. **Stekoll:** NSF-0960285 – *MRI-R2: Acquisition of Instrumentation in Support of Terrestrial, Nearshore, and Marine Ecosystems Collaborative Research at the University of Alaska Southeast* ($323,023; 2009-2012). This project was acquisition of new instrumentation to enhance collaborative research at UAS. Several instruments were purchased and are being utilized in a number of different research projects involving marine biology and environmental science ecological studies. **Thomas:** NSF-BRC-1057303. *Collaborative Research: Consortium of Pacific Herbaria - A Research and Collections Network for Oceania*. ($306,599; 2011– 2014). Six students have been hired on an hourly basis to barcode folders and do preliminary data capture and imaging. Internet portal has been launched to serve regional herbaria specimen data. Regional workshops have been completed and several more planned.