**1. OVERVIEW |** "A Biological Approach to Documenting Traditional Ecological Knowledge in Synchronic and Diachronic Perspectives" is an international, interdisciplinary, and multiethnic effort that applies an innovative technology in evolutionary biology to significantly enhance linguistic and anthropological research in traditional ecological knowledge. The synergistic collaboration across multiple disciplines (linguistics, anthropology, biology) will allow the project to meet collective research objectives that would otherwise be out of reach for each field. The project has the following major goals:

1. Transform ethnobotanical research by creating a plant DNA barcode reference library for the targeted region, the Sierra Nororiental de Puebla. A DNA barcode reference library comprises a series of short gene sequences extracted from universal regions in the genome for all plant species in a determined geographic region (Kress et al., 2005, 2009). These referenced DNA sequences are directly linked to fertile (flowering or fruiting) voucher specimens, which have been positively identified to species based on morphological characteristics by expert taxonomists. Once the reference library (voucher specimens + DNA sequences) is established, subsequent identifications to species can be based simply on the DNA sequencing of small amounts of targeted ethnobotanic vegetative plant material (in this project, primarily leaf tissue), which can be compared to those samples in the library (see section 4). This methodology establishes a mechanism for determining to species a previously unattainable level of botanical specimens that in this project will be directly linked to endangered linguistic and cultural information obtained through ethnographic research.
2. Develop a substantive set of language documentation material on the actual state of traditional ecological knowledge (TEK, including the nomenclature, classification and economic and symbolic use of regional flora) among Nahuat and Totonac communities in the Sierra Nororiental de Puebla. This knowledge will be linked to plants that through the DNA barcode methodology will be efficiently and accurately linked to species in Western scientific nomenclature.
3. Approach ethnobotanical research from the perspective of language documentation and discourse by recording, transcribing, and analyzing how native speakers *talk about* local flora. The target is 100 hours of transcribed Nahuat (20 hours from each of 5 communities) and 20 hours of transcribed Totonac (8 from one community and 12 from another) discourse. All material will be recorded at 48KHz, 16-bit with headworn microphones (in two-person conversations each speaker will be separately miked)
4. Situate the results of this ethnobotanical research in a comparative and diachronic perspective, addressing theoretical issues in historical linguistics and cultural history.
5. Create a model for collaboration that stretches from DNA sequence laboratories at the Smithsonian Institution and Institute de Biología (National Autonomous University of Mexico, henceforth UNAM) to a grass-roots effort by an Indigenous cooperative, Tosepan Titataniske, to develop and sustain an education, exhibition, and research center in Cuetzalan, state of Puebla, Mexico.
6. Provide for the rapid dissemination of project results locally (through written, illustrated field guides; educational tools; and public exhibits in the Tosepan center) and internationally (through archiving on open access websites such as AILLA and DobeS of all project data and results as they emerge).

As recent studies have shown, documentation and analysis of nomenclature, classification, and use of flora and fauna can shed significant light on important facets of the cultural history of linguistic groups, such as the ecology of ancestral homelands and historical patterns of contact (reflected in loans, calques, and shared attitudes and knowledge). Comparative ethnobiological studies among dispersed villages belonging to a single group of languages can also address questions of diachronic lexicosemantics: the factors that may affect relative retention, loss, and semantic shift in biotaxa nomenclature.

The aforementioned cultural historical and linguistic issues, to be addressed in this study, clearly go beyond those typically treated in single-community ethnobotanical studies, studies (be they ethnographic or linguistic) that are not only limited spatially (to a single village) but that often rely on a small set of "well informed informants" to provide insights into ecological knowledge that are at least implicitly held to typify a system of shared knowledge. The major problem in addressing the macro issues mentioned in the previous paragraph has always been the accurate documentation of TEK across a dispersed set of related and non-related linguistic communities. This project will efficiently accomplish this level of documentation both through the innovative application of an emerging technology (DNA barcoding) to a novel situation (ethnobotanical research) and through the synergistic efforts of a highly diversified team of individual and institutional collaborators (see sect. 8).

At one end of the research spectrum is the participation of the Western scientific botanical community. Seventy-four taxonomists, all renowned for their expertise in neotropical flora, have committed to identify specimens in the plant families they study, ensuring that the plants used to create the DNA barcode reference library (both those newly collected and those already deposited in herbaria) will be accurately determined to species by those most qualified to do so (see supplementary documents, henceforth SD, Part D for commitment letters). The U.S. and Mexican national herbaria are the home institutions for processing voucher specimens and the state herbarium of Puebla, actively involved in this state's floristic studies, is an active participant (SD, Part C: 1, 2, 3, 14). Four additional herbaria (Missouri Botanical Garden, New York Botanical Garden, and the University of Texas and California Academy of Sciences herbaria) have extended letters (SD, Part C: 15–20) expressing support and agreeing to lend voucher specimens when needed. Two leading state-of-the-art molecular labs, the Smithsonian Institution and the Instituto de Biología (UNAM), will process the plant material and develop a four-locus (*rbcL*, *matK*, *ITS*, *trnH-psbA*) DNA barcode that is both economically viable and scientifically precise enough to warrant development as a tool to identify vegetative plant material to species. The use of a DNA barcode reference library increases the efficiency of ethnobotanical fieldwork by making identification to species in many cases a technical process not solely dependent upon the taxonomic expertise of systematists.

At the other end of the research spectrum is the remarkable support for ethnographic and linguistic research from three Indigenous cooperatives in the Sierra Nororiental: Tosepan Titataniske, Tosepan Kali, and Tosepan Pajti, each of which has committed significant human and financial resources to this effort (SD, Part C: 4–10). Given the suspicion with which Indigenous peoples often view ethnobotanical research (Brown 2003, particularly chap. 4, "Ethnobotany Blues"; Hayden 2003; see also Laird 2002), this support is key to success. The collaboration of all three cooperatives will facilitate ethnographic fieldwork throughout the extensive area of study. The umbrella organization is Tosepan Titatanikske (see http://www.uniontosepan.org/), which comprises 22,000 members inhabiting 290 Nahuat and Totonac communities in 22 municipalities of the Sierra Nororiental. It has started construction of a two-story education, exhibition, and research center (SD, Part C: 4–6) that will serve as the local hub for research as well as the venue for the dissemination of results to communities in the Sierra Nororiental.

Articulating the efforts of the two above-mentioned groups is a uniquely qualified team, with skills in Nahuat and Totonac linguistics, ethnobotany, taxonomic and molecular botany, and historical linguistics.

**2. AREA OF STUDY |** Theresearch proposed here targets the Sierra Nororiental de Puebla, a region comprising 28 municipalities and 2,668 sq. km., ranging from an altitude of 100 to 3,100 meters. The Sierra includes the most north-eastern sections of endangered montane forests, considered high priority for research and conservation (CONABIO 2010). An extensive CONABIO (Consejo Nacional para el Conocimiento y Uso de la Biodiversidad) database of herbarium specimens from the Sierra Nororiental lists 1708 distinct species of angiosperms (flowering plants). With database errors and potential new state registers, an inventory of between 1750 and 1850 angiosperms is expected (about 8% of Mexico's flora [cf. Espejo-Serna et al., 2004] and about 33% of that of the state of Puebla). To this should be added 200–250 Pteridiphytes (ferns and allies; Daniel Tejero, p.c.). The potential for new information on biodiversity in the Sierra Nororiental was made apparent when Amith and Tejero surveyed Pteridophytes in the municipality of Cuetzalan and documented 161 species, of which 36 were new state registers (cf. Mickel and Smith's 2004 inventory of over 1000 Mexican Pteridophytes). In sum, from a botanical perspective, the Sierra Nororiental region is poorly studied, contains high priority ecosystems, and should yield new state registers and significant extensions of the geographic ranges of many plants.

Linguistically and culturally, the Sierra Nororiental is also important. It is home to Nahuat and Totonac communities that have undergone rapid deterioration of speaker competence in their native languages. Non-qualitative census data, however, is a poor indicator of language vitality as it registers passive speakers in the same category as monolinguals and fluent bilinguals. Amith has worked in the area for over five years. Over the last two generations the number of Nahuat and Totonac speakers may, in fact, have increased. But this reflects mostly demographic growth. It does not examine the internal composition of the "native speaker" census category, which has shifted heavily to passive semi-speakers. The nature of the expansion, however, portends a rapid "black hole–like" contraction of linguistic competence given that the present generation has mostly stopped transmitting even a passive knowledge to their offspring. Moreover, traditional ecological knowledge is one of the domains of linguistic and cultural expertise that has suffered extremely from "stylistic shrinkage" (see sect. 3, below).

This study will undertake comparative fieldwork in five Nahuat and two Totonac communities to analyze the nomenclature, classification, and use of plants. Study in one community is already significantly advanced: San Miguel Tzinacapan, a mid-altitude (860 m) Nahuat village in the Cuetzalan municipality. The other four Nahuat villages will be chosen from among the 28 municipalities based on: (1) ecosystem inhabited (at least one lowland and one highland community); and (2) degree language maintenance and expertise in TEK. Tosepan Titataniske will help to provide community contacts and to evaluate the best field sites and community collaborators. Local experts will be contracted to guide the collections and provide ethnographic and linguistic data. For ethnobotanical research, both fertile and sterile specimens will be collected and DNA barcoded. Fertile specimens will be identified by both taxonomic experts and through cross-referencing to the DNA barcode library. The level of agreement between the two methodologies (morphology vs. barcoding on the same specimen) will serve to evaluate and improve the library methodology's accuracy and indicate what species will need supplementary keys. For each collection a set of standardized linguistic and cultural data will be documented: name, classification (including number of related or similar plants), use, and expected habitat. In many cases native speaker experts will digitally record their knowledge of TEK in narrative or conversational format.

Research in Totonac ethnobotany will involve the first "blind" test of the methodology pioneered in this project. During the project's first two years, Gabriela Román and David Beck will compile a preliminary list of nomenclature in the two communities of their expertise: Huehuetla and Upper Necaxa, respectively (Beck is already well advanced, with a preliminary checklist of 300+ biotaxa terms, most still unidentified). Collecting in the Totonac communities will take place between project months 18 and 30. Detailed ethnographic and linguistic data information (including digital recordings in Totonac) about all collected flora will be gathered through consultation with three native experts from each of the two communities. Fieldwork will be carried out by Amith, Gorostiza, Beck, Román and a project botanist. Beck and Román will analyze the ethnographic and linguistic data. This one month of rapid fieldwork to collect fertile and sterile specimen named and classified in two Totonac communities will test the viability of a DNA barcode library to establish the basic taxonomic referents of Indigenous nomenclature. Success would constitute a transformative step in the use of molecular biology for ethnobotanical research.

**3. URGENCY |** The urgency of the proposed research is based on three factors: (1) the precarious state of endangered languages and the threatened status of ecological knowledge among Sierra Nororiental Nahuat and Totonac communities; (2) the rapid decrease in the number of botanical taxonomists able and willing to identify voucher specimens to species from morphological characteristics; and (3) the difficulty of ensuring that the multiple resources leveraged in this project will continue to be available.

***3.1 Linguistic and cultural knowledge***: We are at a unique moment in our ability to comprehend and preserve threatened domains of human knowledge and natural history, poised precariously on the cusp of waning knowledge and biodiversity and waxing possibilities to study and safeguard these cultural and natural resources. Indigenous languages are threatened as never before, with at least half of the approximately six thousand doomed to extinction in the present century. Yet within these languages a particular domain of cultural knowledge is even more at risk: the rich understanding of the natural environment (see Hill 2001, 2004; Zent 2001, and the case studies in section 2 of Maffi 2001). The threat is not simply from the loss of biodiversity. It is part of an inexorable process of life changes and population shifts, often mundane and simple. New building materials for houses and fences accelerates the loss of knowledge about the qualities and characteristics of local woods; plastic bags and ropes are not only ecological hazards but herald the loss of skills in the production of material culture from natural resources.

Languages disappear and cultures shift and die in many different ways. Nancy Dorian (1978) once stated that some languages die with their "morphological boots on," suggesting that the last speaker may be highly proficient in the moribund language. While this might be true, *it will not be true* that this last speaker has full knowledge of regional natural history. Even when language skills are relatively stable, the biosystematic lexicon may suffer significant loss (Hill 2001) as a result "stylistic shrinkage" (Palosaari and Campbell 2011:116). It is a domain of linguistic expression and cultural knowledge that is of high priority for documentation. Yet few projects are able to leverage the necessary resources to accomplish this.

From an endangerment perspective, the urgency of documenting traditional ecological knowledge is so great that the most common ethnobiological approach, single-sited and focused on describing a shared system of localized knowledge, is not adequate. A new generation of studies is needed that meets the urgency of documentation with a novel strategy capable of both breaking the bottleneck in data processing (the number of voucher specimens that can be correctly identified to species in a given time) and expanding the range of possible inputs (i.e., by creating a system capable of quickly processing and identifying not only fertile but also infertile specimens). This present project meets these challenges.[[1]](#footnote-1)

***3.2 Western taxonomic expertise:*** From the Western scientific perspective, there is a "dwindling pool of taxonomists" (Hebert et al. 2003:313) able and willing to determine voucher specimens to species, particularly specimens having no direct relation to their own research agenda. Indeed, NSF had recognized the problem of diminishing taxonomic expertise by creating the PEET (Partnerships for Enhancing Expertise in Taxonomy) program (now discontinued), which sought "to enhance taxonomic research and help prepare future generations of experts." The dearth of taxonomists has not, however, abated and from a practical perspective documenting the nomenclature, classification, and use of local flora (particularly in diverse neotropical environments[[2]](#footnote-2)) is fraught with difficulties because of this.

Unless a Western-trained (ethno)botanist is familiar enough with the regional flora to identify sterile collections to species, plant material collected in ethnobotanical research needs to be identified by collaborating botanists, who invariably want fertile specimens that are flowering or fruiting (sometimes both). Taxonomists involved in ethnobotanical projects, moreover, are often called upon to simply provide a service: determination to species, species that are often common and, from a botanist's perspective, uninteresting. The dynamics of interdisciplinary collaboration (linguistics-anthropology-biology) would benefit immensely if ethnobotanical projects also addressed issues pertinent to the research agendas of biologists (cf. sect. 4) and relieved expert taxonomists from the burden of identifying common species.

If indeed the (ethno)botanist does become highly familiar with local flora, this will occur after many years, often at the end of a project. Moreover, there is no convenient mechanism to communicate this expertise to benefit future research in the region. Botanical identification keys are not included in ethnobotanical projects and keys that are developed in floristic studies are usually too specialized for use by anyone but an expert. Practical field guides are rare, though this project will produce an illustrated guide. More important, however, this project, for the first time in ethnobotanical research, will develop a DNA barcode reference library covering the floristic inventory of a targeted region (the Sierra Nororiental). This will be an enduring, easily accessible reference that will facilitate future (ethno)botanical research in the region. The Sierra Nororiental DNA barcode reference library, a digital register of sequences, will also serve as a reliable foundation for incremental advances in building a reference library for an ever expanding region. Considering the high degree of floristic overlap with neighboring Indigenous areas in northern Veracruz to the east and the state of Hidalgo to the west, the added cost of extending this reference library to include the flora of neighboring Indigenous areas will be greatly reduced.

Thus taxonomic expertise, like Indigenous TEK, is "endangered." Many experts who have helped Amith in identifying specimens since he began ethnobiological research are either retired or close to retirement. Within a decade or two, it may no longer be feasible to obtain expert identifications of voucher specimens collected in ethnobotanical research. It will certainly not be possible to assemble a taxonomic team of the same caliber as that which supports the present effort.

***3.3 Synergistic teamwork:*** Finally,the PI of the present project, Amith, has after many years forged the unique level of institutional and individual collaboration represented in this proposal. While not "endangered," the opportunity for so many individuals to work together on a common research project will eventually respond to shifting career developments and goals. For example, the project's two experts in DNA barcoding, John Kress and Gerardo Salazar, are now well situated (as director of the Smithsonian's Consortium for Understanding and Sustaining a Biodiverse Planet and as chair of the Department of Botany, Instituto de Biología, respectively) to collaborate at both an individual and institutional level. Other team members are also committed to this project, but in five or six years might be focused on other areas of research not so linked to the present project. This is particularly true of the Indigenous collaborators.

**4. THE CREATION OF A DNA BARCODE REFERENCE LIBRARY |** Given that plant DNA barcode reference libraries have not been used to facilitate ethnolinguistic research in TEK, this section begins with a brief introduction of the concept. It is followed first by an account of how the necessary voucher specimens will be obtained and then by a more technical presentation of DNA extraction, amplification, and sequencing. As this proposal may be co-reviewed by Biology, section 4 concludes with a statement of the project's potential impact on studies of biodiversity and evolutionary phylogenetics.

***4.1 Floristics and the DNA barcode reference library:*** A DNA barcode is a short universal gene sequence taken from a standardized portion of the genome and used to identify specimens to species. The taxonomic foundation for each DNA barcode is a fertile herbarium voucher specimen identified to species by an expert taxonomist. The sequences generated from this voucher represent first of all the specimen itself. If the set of genetic markers is truly unique to a species, the same sequences will be found in all material (e.g., a small leaf sample) from any plant of the same species, and *only from* that species. The selection of the best DNA barcode markers for plants, however, has been debated (Kress et al., 2005, 2007; CBOL Plant Working Group, 2009). Different genetic markers possess pros and cons with respect to DNA barcoding: they may be easy to amplify and sequence, but too evolutionarily slow to distinguish species, working only at the generic level (e.g., *rbcL*). They may be more rapidly evolving and thus excellent species-specific markers but, precisely because they change so rapidly, they are difficult to amplify and sequence (e.g., *matK*). Or they may be of varying length (in numbers of base pairs) so that comparing the same region across widely different species is difficult (e.g., the intergenic spacer plastid *trnH-psbA* and nuclear ITS). Many of these problems have been solved. For example, advances in the development of primers have increased the success rate of *matK* amplification and sequencing. With any given specimen, however, failure to sequence one locus may be compensated for by success with others.

The four marker regions chosen for sequencing in this project—*rbcL, matK, ITS, trnH-psbA*—represent a somewhat "vigorous" expansion of the "core" *rbcL* and *matK* DNA barcode (for the historical development of DNA barcodes, see Hollingsworth, Graham, and Little, 2011). Yet given its liberal use of sequencing over four loci, it is perhaps the most accurate and successful option available. Hollingsworth (2011), for example, supports the use of these same four markers and writes that they distinguish over 80 percent of material to species. Given that this study is limited to a geographically delimited floristic region, DNA barcode accuracy will be increased: many congeneric species that may not be distinguished solely by a DNA barcode will not co-occur in the study region. Geographic specificity, therefore, will discriminate species that cannot be distinguished solely on the basis of the four-locus DNA barcode. In the few cases in which more discriminatory power is needed, this project will develop supplementary tools: (a) Kress will determine a clade-specific DNA locus to distinguish the congeneric species; (b) Amith, working with taxonomists and native natural historians, will identify discriminatory morphological features present in sterile specimens. The team is confident it can approach 100% accuracy in sterile species identification.

Two proponents of DNA barcoding state that this methodology would "free taxonomists from routine identifications," allowing them to concentrate on specimens of greater floristic or phylogenetic interest (Hebert and Gregory 2005:855). For those who rely on the expertise of taxonomists, such as ecologists or ethnobotanists, this outcome in itself is important, as it relieves them from direct dependency on the limited time of taxonomists for routine determinations. However, there is an even more important consequence that emerges when a DNA barcoding initiative creates a reference library for the flora of a given region: it allows the identification to species of sterile material collected from plants in the region. Academic researchers in the field with native speaking natural historians, as well as native speakers working alone in the field, will be able to take a small section of leaf or twig or bark or fruit from a plant, record ethnobotanical information for this plant (on a standardized survey form or through digital recordings) and be confident that the tissue specimen can be identified to species. This has immense implications for linguistic and anthropological TEK research. Ethnobotanists are often faced with situations in which a native speaker in the field will give an excellent, detailed account of use for a plant that is at that time unidentifiable, a simple stem with leaves. With a DNA barcode reference library this insignificant specimen can be accurately identified. In sum, a method that simplifies collection and often makes determination a technical exercise will greatly facilitate anthropological and linguistic research on Indigenous natural history.

***4.2 Building a floristic inventory of voucher specimens:*** The voucher specimen inventory for the DNA barcode library will be worked on intensively in years 1 and 2 (during this same period Amith will work in a parallel manner on ethnobotanical fieldwork). There are three sources of specimens for the DNA barcode library: (1) fertile material that Amith has already collected (approximately 450 species of angiosperms and 150 Pteridiphytes) or will collect as part of his ethnobotanical research; (2) existing herbarium vouchers of specimens collected in the Sierra Nororiental; and, (3) new intensive fieldwork that targets geographical areas and families of flora deficiently covered by the preceding sources.

Sequencing will begin with Amith's material, identified to species by the same taxonomists supporting this project (SD: Part D). Herbarium vouchers will be obtained at the six participating herbaria. At each herbarium botanists supporting this project will be asked to verify the species determination of the voucher specimen before a vegetative section is sent to a molecular lab (Smithsonian or Instituto de Biología) for DNA barcoding. Given that vouchers from Amith or herbaria were heat dried, some specimens will not yield material that can be sequenced. Nevertheless, this should be a small minority of cases. The target is to obtain through the first two sources DNA barcodes for at least fifty percent of the floristic inventory.

Parallel to the sequencing of Amith's material and herbaria specimens, two highly experienced Mexican field botanists will carry out floristic fieldwork designed to complement the DNA barcode data obtained from the first two sources. They are budgeted for 288 totals days in the field over the first two years (72 days each/year). Each collection will comprise 3 vouchers of fertile material to be deposited in MEXU, US, and the institutional herbaria of collaborating taxonomists. MEXU will fast-track for digital photography all vouchers as part of a CONABIO-supported project to photograph MEXU holdings.

***4.3.*** ***Generating DNA barcodes Within and Across Forests: Field and Lab Protocols:*** The field and lab protocols for generating the four-locus DNA barcode (*rbcL, matK*, *trnH-psbA*,and *ITS* regions) will follow those outlined by Kress (Kress et al. 2009, 2010) and Fazekas (Fazekas et al. 2012), which have been implemented successfully in multiple plant groups and floristic regions in the tropics and temperate zone. For new collections, tissue samples will be field collected and preserved by silica gel desiccation. Approximately 0.1 g of material will be collected from each individual sampled. Of the multiple accessions per species one must have a fertile voucher associated with it (either from project fieldwork or from a herbarium) while the ethnobotanical collection may be of fertile or sterile vouchers. Tissues will be collected only from photosynthetic material, primarily young leaves. DNA will be extracted from this material using a GUSCN extraction method and stored at -20°C. Each sample will be assigned a unique DNA extraction number for tracking before, during, and after the PCR reaction. Routine PCR will be used, in conjunction with Sanger sequencing to collect the sequence data. Recovered trace files will be imported into Geneious Pro 4.6 (Biomatters Ltd. Auckland, New Zealand) edited and assembled into contigs. All sequence data will be submitted to GenBank as it is developed. The two labs that will handle the DNA processing (Smithsonian and Instituto de Biología) are both highly experienced, having processed thousands of plant samples for DNA barcoding. Thus the project is entirely feasible.

***4.4 Supplementary keys***: When discrimination of congeneric species is not possible, two supplementary tools will be developed: (1) in consultation with the collaborating taxonomists, Kress, Salazar, and Erickson will identity supplementary DNA barcode loci to resolve problematic clades; (2) Amith, in consultation with native speakers and collaborating taxonomists, will develop supplementary morphological keys that are valid for sterile specimens (such morphological distinctions were used to distinguish congeneric species in *Aspalathus* [Fabaceae] that DNA did not discriminate; Edwards et al. 2008). Provisions are in place to evaluate the accuracy of the DNA barcode by testing agreement in fertile ethnobotanical collections between specimens identified on the basis of both morphology and DNA barcodes.

***4.5 Regional biodiversity investigations*:** Though not a targeted research goal, this project will provide material important to understanding both species and phylogenetic diversity in the Sierra Nororiental. A floristic inventory accompanied by DNA barcodes for each species will permit the generation of regional and local floristic phylogenies valuable to future ecological and conservation investigations. Once the DNA barcode reference library is complete, a specimen sequence that does not match any DNA barcode in the library will be considered a potential new register or species in need of additional study (although it might also represent intraspecific genetic diversity within an already sequenced species). If the sequence was taken from a sterile specimen, this specimen will be targeted for recollection when fertile.

Given the collaboration with Tosepan, this project will benefit from immense "boots-on-the-ground" support. This is important not only for the linguistic and anthropological research that is the primary focus of this grant (see sect. 5) but also for building and completing the floristic inventory. For example, abundance and ubiquity are factors that contribute to species diversity and function as a warning sign of potential endangerment. This project is not equipped to carry out quantitative research on these topics. However, it will be able to ask native speaker collaborators, as part of the interview data for each plant collection, for an impressionistic observation as to both the relative abundance of the species and for information as to its geographic distribution. Speakers are often very aware of each factor. During Amith's previous work in the region collaborators were often able to give information on the distribution (by altitude and habitat) and relative abundance of many species identified by an Indigenous name.

Finally, Amith will develop a new strategy during his ethnobotanical research: asking Indigenous collaborators to collect specimens (both fertile and sterile) that they have never seen. This "negative cognitive imaging" of the natural environment is the flip side of most ethnobotanical research, which generally focuses on plants that are named, classified, and used by native speakers. Negative cognitive imaging is possible given the utilization of a DNA barcode reference library, which will greatly reduce the cost of determining vouchers to species, regardless of the presence of flowers or fruits. Thus the study of biodiversity or development of a floristic inventory will benefit from collaboration with Indigenous natural historians who will be asked to collect specimens they consider rare or unknown.

***4.6 Evolutionary ethnobotany*:** To date most studies of ethnobotany and TEK combine basic plant taxonomic knowledge with ethnographic information. With recent advances in applying DNA barcode sequence data to reconstructing phylogenetic relationships among plant species in an ecological community (e.g., Kress et al. 2009, 2010), the opportunity now exists to place the traditional ecological knowledge of Indigenous communities into both a taxonomic framework and an evolutionary context. The ability to construct a robust and well-resolved phylogeny using DNA barcodes will enable specific questions to be addressed regarding changes in ethnobotanical use and local nomenclature of plants in Indigenous communities in both space and time relative to the actual evolutionary relationships of the plant species themselves. For example, are multiple plant species used to treat a specific ailment in a community or between communities more closely related phylogenetically than would be expected by chance alone? In other words, is there an evolutionary signal in the species being utilized that may indicate the inheritance from a common ancestor of a specific medicinal property?

We believe that the ability to quantify the phylogenetic relationships among species used by local peoples will provide insight into how different species or parts of species are chosen and used in different ethnobotanical contexts. We will also be able to quantify evolutionary differences in regional species diversity and relate how changes in this diversity, both species richness and phylogenetic diversity, affect usage of individual species for local specific purposes. In an ecological context it has been shown that species that share the same habitats or functional traits in a community may be phylogenetically clustered (i.e., more closely related than expected at random), phylogenetically overdispersed (i.e., more distantly related than expected), or phylogenetically neutral (Kemble and Hubbell, 2006; Kress et al., 2009; Swenson et al., 2013). These studies were based upon well-supported data on phylogenetic relationships among the taxa under study. We will utilize the same types of analyses, substituting ethnobotanical uses of plant species for habitats and functional traits. Studies such as these focused on plant uses and TEK have not yet been conducted. However, we believe that the phylogenetic framework provided by the DNA barcode sequence data will allow such investigations in the future.

***4.7 Substantive products:*** Botanical fieldwork and sequencing of vouchers will yield: (1) a floristic specimen inventory of the Sierra Nororiental comprising one to two voucher specimens for each of the estimated 1750 to 1850 angiosperms and 200 to 250 ferns and allies; (2) connected to each voucher, a four-locus DNA barcode (the entire reference will comprise some 3250 angiosperm and 350 Pteridophyte DNA barcodes; and (3) digital photos of the vouchers with collection coordinates and altitude, available through GenBank and UNIBIO, the Instituto de Biología website.

**5. ETHNOBOTANICAL RESEARCH: SYNCHRONIC AND DIACHRONIC PERSPECTIVE |** Project organization will be divided into two interlocking tracks: (a) botanical research oriented toward developing a floristic inventory; and (b) ethnobotanical research headed by Amith for the five Nahuat communities and by Beck and Román for each of the two Totonac communities.

***5.1 Synergy with botanical fieldwork for the DNA barcode library:*** The major strategies for developing the DNA barcode library are described in the previous section. One source of vouchers will be extensive collecting (288 person days over two years) by two field botanists. As often as possible Allen Coombes, curator of scientific collections at the Herbarium y Jardín Botánico of Puebla, who is partnering in the development of this floristic inventory, will be present on botanical and ethnobotanical collection trips. Coombes is the botanist most familiar with state of Puebla floristics as he has an long-term project to develop an inventory of the entire state. Collaboration with him (SD: Part D:34) will be highly beneficial, as will be the full support of the Puebla state herbarium and botanical garden (SD: Part C:13).

The extensive botanical activity associated with this project will greatly enhance ethnobotanical research in the Sierra Nororiental. Botanically driven collection of fertile voucher specimens for the DNA barcode library will not be limited to the seven Indigenous communities (5 Nahuat, 2 Totonac) targeted by Amith, Beck, and Román but rather attend to the needs of documenting the Sierra Nororiental flora (see p. 6). When the field botanists collect in Indigenous communities, however, they will be accompanied by a local guide who will provide Indigenous names and uses for any recognized plant. For these plants, the botanist will fill in a survey sheet and digitally record basic information (Indigenous name, classification, and use) on a handheld Zoom H4n for later review by language experts. Although botanical fieldwork is designed to create an inventory of *fertile* specimens for the reference library, for plants with Indigenous names, even is the specimen is sterile, a branch will be collected and dried and a small leaf sample placed in silica gel for future sequencing. The result will be an extensive corpus of highly valuable research material: (a) a plant specimen (branch and silica-gel dried vegetative material); (b) ethnobotanical information written on a standardized survey questionnaire accompanied by a short digital recording; (c) full collection data (coordinates and altitude, habitat). Although most of this material will not be DNA barcoded as part of the present project, the ethnographic data obtained by botanical fieldworkers will be entered into the project database and the specimen kept for future reference and sequencing.

***5.2 Nahuat and Totonac ethnobotany: Fieldwork methodology:*** Amith will target 5 Nahuat communities in the Sierra Nororiental (see p. 3, top paragraph for details on how these will be chosen) and has budgeted 48 days/year for ethnobotanical fieldwork (total = 144 days). Fieldwork will be carried out by Amith with two native speaker team members: Eleuterio Gorostiza and Amelia Domínguez (see sect. 9). Again, as with strictly botanical fieldwork, Allen Coombes or his assistant will accompany the ethnobotanical research team when possible. For fieldwork in each Nahuat or Totonac village, three local native speaking consultants will accompany the team each day of collection. Two types of data will be gathered: (1) ethnographic information (nomenclature, classification, and symbolic and economic use) will be written down in field notebooks with standardized fields to ensure consistent ethnobotanical information across specimens and consultants; and (2) the opinions of Indigenous consultants about relative abundance and common habitat or geographic location of the collected specimen (see sect 4.5). The results will be entered into a database of botanical and ethnobotanical information. Nomenclature and classification will be added to the Nahuat dictionary that Amith is developing (presently comprising 3,509 manuscript pages: 7,874 headwords/21,208 senses).

In addition to the aforementioned data, digital recordings will be made of native speaker natural historians discussing some of the collected plants (particularly those with utilitarian value). Project commitment is to transcribe 20 hours of Nahuat from each of the five communities, 8 hours of Totonac from a village speaking the variant of Gabriela Román, and 12 hours of Totonac from a village speaking the Totonac studied by Beck. Domínguez, Gorostiza, and Román are all extremely literate in their respective language. Beck will work with a native speaker for 6 months to produce Upper Necaxa transcriptions. A one hour recording takes about 40 hours to transcribe (time-coded format) and review.

***5.3 Substantive contributions***: This project will yield extensive ethnobotanical data from Nahuat and Totonac villages in the Sierra Nororiental: (1) a database of nomenclature, classification, and use from each community; (2) an extensive corpus of recordings (48KHz, 16-bit) on TEK with a minimum 120 hours (100 Nahuat/20 Totonac) transcribed in time-coded format (Transcriber or ELAN); (3) TEK data resulting from botanical fieldwork in Indigenous communities; (4) enhancement of Amith's dictionary with ethnobotanical nomenclature and classification from Sierra Nororiental de Puebla Nahuat; (5) illustrated field guide of Sierra Nororiental Nahuatl TEK for community and school use; (6) metadata database that meets the standards of AILLA and DobeS, where Amith has already deposited much material; (7) open access archiving of the preceding at AILLA and DobeS.

***5.4 Theoretical contributions to linguistics and anthropology:*** Comparison of the nomenclature, classification, and use of biological taxa, as well as of culturally specific beliefs (such as black cats being bad luck), both within genetically related languages and among unrelated languages that have been in contact, can shed important light on diachronic lexicosemantics and cultural history.[[3]](#footnote-3) Biosemantics is a particularly interesting area of study in part because of the very nature of cognitive categories in this domain. One notable feature is the constraint on arbitrariness. Thus even though "the categorization function of language imposes a structure on the world rather than just mirroring objective reality" (Geeraerts 1997:9), flora and fauna are recalcitrant objects of contemplation, having their own "thinginess" (Ellen 2004) and "discontinuities" (Hunn 1977) that constrain the freedom of human societies to classify nature according to arbitrary and artificial cultural criteria. Moreover, as studies of the classification of biotaxa have shown (see particularly works by Berlin and Ellen in References), categorization in the biological domain can best be analyzed in the framework of prototype theory as (a) internally structured, with some members more representative than others; (b) non-discrete and blurred at the edges; and (c) comprising "internally coherent classes" of referents or meanings linked by a structure of family resemblance (cf. Rosch and Mervis 1975; Geeraerts 1997; Sweetser 1990; Lakoff 1987). Significantly, the prototypical nature of biosemantic categories is not only important in synchronic studies: the nature of and variation in biotaxa nomenclature and classification within a community. Prototype theory, as Geeraerts (1997) has demonstrated, is also relevant to understanding the nature of diachronic lexicosemantic shift within both sedentary and migrating linguistic communities. That is, the internal structures of categories is one factor that affects that nature and directionality of semantic shift.

Migrating populations, in particular, probably manifest higher rates of change in biosemantic terms than in other areas of the lexicon as well as higher rates of change in biosemantic terms than occurs in stable populations. Thus as human populations move into ecosystems with biotaxa distinct from that of their place of origin, the new species encountered may be named by (1) semantic shift of extant terms to new denotata, (2) borrowing from languages encountered in the new habitat, or (3) neologisms, most likely descriptive, semantically transparent terminology. Andersen (2003:1) defines linguistic stratigraphy, a term that well describes one of the research goals of this project, as "the systematic investigation of the layering of grammatical and lexical material in a language or dialect which reflects its historical development and past contacts between its speakers and bearer of other linguistic and cultural traditions." This leads to a focus on loanwords and loan translations (calques), on borrowings and intrusions, for historical reconstructions of convergence and contact. This reliance on borrowings and intrusions is in effect the flip side of a methodology that utilizes lists of basic vocabulary (Swadesh or Leipzig-Jakarta; for a discussion see Haspelmath and Tadmor 2009: chaps. 1–3) that comprise words that are universal, relatively culture-free, and "less subject to replacement than other kinds of vocabulary" (Campbell, 2004:201–7). Again, Andersen (p. 4) provides a concise summary of the distinction: "Regular correspondences form the basis for the determination of genetic relations among languages. Irregular ones bear witness to the affinal relations of individuals languages during their history." Meroz (2013) makes a similar observation, and points to the importance of biosemantic nomenclature: "Words in this domain [plants and animals] are typically more prone to borrowing than basic vocabulary, especially when speakers of a language move and encounter different species. A survey of such vocabulary is especially suited to identifying and highlighting old language contact" (p. 2).

For Mesoamerica, as Dakin (2003:259) notes, loanwords are "one of the few kinds of evidence available to help unravel the history of the region." Smith-Stark (1982, 1994) and Campbell, Kaufman, and Smith-Stark (1986), in defining Mesoamerica as a cultural area, included calques (loan translations) as important evidence. Loan words and loan translations, as well as semantic shift in plant nomenclature, precisely because of its sensitivity to change through migration to new ecosystems, has been selected for use in stratigraphic studies of cultural history in Zambia (Bostoen 2007), Northern Australia (Bowern, 2007, 2011), and among native languages of California (Meroz 2013).

A few scholars have also explored biological nomenclature to determine the factors (effectively independent variables) that affect the relative stability or instability of lexical terms in this semantic domain. Berlin et al. (1969, 1973) limit their pioneering study of retention and loss to two closely related Mayan languages (Tzeltal and Tzotzil) and suggest that retention of nomenclature is directly proportional to the “cultural significance” of the biota, defined as the “practical value of biological knowledge for a given culture” (1973: 273). Balée and Moore (1991) further explore this topic, increasing the number of languages studied (to five Tupi-Guarani Amazonian languages in four different subgroupings) and considering the level of human intervention in plant use— dividing the domain into domesticates, semi-domesticates, and non-domesticates—as the independent variable affecting retention. They also add an independent linguistic variable: the semantic nature of the plant term, either “literal” (a term whose sole referent is a specific plant) or metaphorical (a name that contains a metaphorical use of a term or terms that do not refer to a specific plant). Both Berlin et al. and Balée and Moore find that terminological retention is strongly correlated to what may be called the level of human intervention. For example, considering all ten possible pairs of the five languages, Balée and Moore demonstrate that lexical similarity is 80.3% for domesticates, 59% for semi-domesticates, and only 30.8% for non-domesticates. Balée and Moore also show that the level of terminological retention is directly related to the “literal” nature of the term (e.g., 97.3% for domesticates the terms for which are "literal").

Several scholars (most notably Siebert 1967; Friedrich 1970; Fowler 1972a, 1972b, 1983; Whistler 1977) have explored the methodology of "matching reconstructible plant and animal vocabulary to reconstructed natural environments" (Whistler 1977:160). Fowler (1983:224) expanded her original focus on Numic, applying the same search for "the location of potential homelands for protolanguages" to Uto-Aztecan. Such a deep level of cultural historical reconstruction is beyond the scope of this present project, although the biosemantic database that will be developed for Sierra Nororiental Nahuat and Totonac communities might produce material relevant to such ancient history.

The main theoretical focus in this project, however, is on more local patterns of contact and lexicosemantic change. Nahuat nomenclature for biotaxa will be gathered from five communities in mostly distinct ecosystems. It will thus be possible to consider a question that Fowler broached in her thesis (1972a:12–13): the role that ecology plays "in semantic development in ethnobotany" such as the emergence of new terms and classificatory schemes, or the shift in meaning of cognate lexemes. Exploration of nomenclature in five communities will also allow the project team to explore any linguistic and cultural factors, such as those suggested by Berlin et al. (1969, 1973) and Balée and Moore (1991) that might influence retention, shift, or loss in biotaxa nomenclature and classification. Work in both Nahuat and Totonac communities will facilitate documentation of any possible loanwords and calques that have occurred (preliminary study by Amith shows borrowing of biotaxa terminology from Totonac into Nahuat, including the term *xopepe*, 'cockroach', a word that seems an unusual one for borrowing).

Finally, an attempt will be made to situate Sierra Nororiental Nahuat traditional ecological knowledge within the context of Nahuatl in general. To accomplish this, a database will be developed of all historical and modern Nahuatl material on the nomenclature, classification, symbolic and economic use of plants (see references, section 2). There is enough comparative data both from secondary sources and from Amith's own ethnobotanical research in Nahuatl-speaking communities of the Balsas Valley in central Guerrero to provide a firm foundation from which to evaluate any particular changes in nomenclature and classification (shifts, neologisms, borrowings) that have affected Sierra Nororiental Nahuat communities.

In sum, this project will not only develop an innovative methodology and a large amount of substantive materials on Nahuat and Totonac traditional ecological knowledge. It will also situate the material in the context of debates on cultural history and diachronic lexicosemantics in reference to the domain of ethnobotanical knowledge.

**6. INTELLECTUAL MERIT ACROSS THREE DISCIPLINES |** This project will have a major impact, both theoretically and methodologically, in three areas of research: linguistics, anthropology, and botany.

**Linguistics**: A particularly challenging field of linguistic documentation is Indigenous natural history, including the nomenclature and classification of regional flora. About 10 percent of a language's lexicon references the natural environment (including nomenclature for flora and fauna). Inadequate attention to this domain leaves a significant lacuna in corpora, lexicons, and morphosyntactic research (e.g., compounding). Moreover, the lexical richness of a language is revealed not only through an extensive terminology for biotaxa but in extended, metaphoric usage of such terms and in the rich vocabulary that references culturally specific technologies that transform local flora into material culture. Often, the semantics of Indigenous terms for biotaxa may be opaque unless the referent is identified to scientific species, a goal attainable only through extensive field research and close collaboration with biologists.

Among the Sierra Nororiental Nahua, the project team expects to find significant regional variation in ethnobotanical nomenclature, classification, and symbolic and economic use. However, intercommunity communication about specific plant species may not be significantly hindered as contextual factors and shared knowledge of use may aid in communication despite absence of a shared lexicon. The project will digitally record conversations between Nahuat-speaking individuals (particularly herbal curers) from different villages to explore how knowledge is communicated in cases of lexical divergence in plant nomenclature, i.e., how speakers "negotiate" meaning in the context of natural discourse.

This project will also address issues of relative rate of loans in different semantic domains. Bowern (2011) suggests a varying though not particularly high level of loans in biotaxa nomenclature in Australian languages although there is a lower level of inheritance as compared to basic vocabulary. The present project will add to the comparative data on loan levels of ethnobiological terminology by exploring this issue in Nahuatl languages. Also related to Bowern's research (Bowern and Atkinson 2012) is the degree to which computational phylogenetics can be applied to ethnobotanical nomenclature and how the results compare to other reconstructions of language history. This will involve the encoding of the lexical terms from different communities in a way that expresses the relationship among the terms. To accomplish this a set of basic terms will be selected after an evaluation of the entire dataset.

The contribution of this project to linguistic research, particularly among endangered languages, is both substantive and transformative: it develops new strategies and techniques implemented by a multidisciplinary academic team collaborating with Indigenous cooperatives. The innovative methodology will facilitate study of important areas of cultural and linguistic endangerment (natural history and material culture) either insufficiently targeted or incompletely resolved in present language and cultural studies. The extensibility of this interdisciplinary approach will be tested in collaboration with Gabriela Román and David Beck, two linguists (the first a native speaker in a doctoral program) working on Sierra Nororiental Totonac. Beck has a list of over 300 Totonac plant terms, the majority not identified. The test will be to identify these to species in one month of fieldwork, utilizing the reference tool developed for identifying sterile specimens. Success will represent a significant advancement in language documentation.

**Anthropology**: This project will address important issues in cognitive anthropology and cultural history. The first issue concerns the cognitive structure of classificatory systems of plants. Much ethnobotanical research relies on a few local consultants to construct models of nomenclature and classification considered representative of community ethnobotanical knowledge. The internal structure of the categories, the absence of clearly delimited category boundaries, the mechanisms used to extend classifications to new members, and variation in nomenclature and classification among different groups (men vs. women; neighboring communities of the same or different cultural-linguistic groups) are issues less commonly explored (but cf. Hays 1974, 1976 who gives a very complete account of intracommunity variation). Wide-ranging ethnobotanical work among a large set of native speaker consultants is a better way to research these issues. Rather than a consensus model, the goal is to explore variation and tension in nomenclature and classification. A resource, such as the DNA barcode reference library, that enables scientific determination of plant specimens in the absence of the flowering parts upon which traditional keys are based, will significantly reduce the time required to complete such wide-ranging, intercommunity anthropological (and linguistic) research on local flora and significantly impact the facility with which theoretical issues that depend on extensive, multisited research can be addressed.

The second issue involves comparative ethnobiological research among cultural groups that share historical and linguistic roots to gain insights into their cultural history and the factors that affect retention or loss of biological nomenclature and classificatory systems over time. A comparison of similarities and variation in the nomenclature and classificatory systems of genetically related languages can reveal the biological and linguistic variables most closely correlated with retention, loss, or extension of nomenclature and classification in the ethnobiological system. As a semantic field particularly sensitive to borrowings and intrusion, a large scale study of biosemantic nomenclature in an area of contact between two distinct linguistic groups (Nahuat and Totonac) might shed light on the cultural history of each group.

**Botany**: The project will develop a specimen-based floristic inventory, determined to species by taxonomists committed to this effort. The inventory will be photographed (in situ and as mounted vouchers); documented as to habitat and coordinates of collection; associated with Indigenous nomenclature, classification, and use; and used to create a DNA-based reference library, a significant contribution to the study of Mexican flora covering approximately 1,750–1,850 angiosperms and 200–250 ferns and allies. Work will be carried out with Indigenous natural historians and Western taxonomists to develop a supplementary morphological key to separate congeneric species that the 4-locus DNA barcode does not distinguish to species. Kress will head the effort to document the use of any clade-specific regions that could also be used to separate these congeneric species.

The project budget includes funds to initiate a bilingual (English/Spanish) website for a "Next Generation Flora of the Sierra Nororiental de Puebla". Site architecture will allow users multiple paths (scientific name, geographic location; Indigenous nomenclature, classification, and use; thumbnails of flowers and fruit, organized grosso modo by morphological characteristics) to access at one site a complex of information unlike previous published and online Floras: (1) voucher specimen with photos of the plant in situ and mounted; (2) scientific names, including synonyms; (3) biogeographical data on species collection points in the targeted research area; (4) Indigenous nomenclature, classification, and symbolic and economic use; (5) selected digital recordings (transcribed and translated into English and Spanish) of native experts speaking in Nahuat and Totonac about local flora; (6) presentation of DNA barcode data.

The project will establish close collaboration at the international (Smithsonian and UNAM) and interethnic (Western scientists and Indigenous natural historians) levels. It has a strong education and outreach component, creating material for local use in Indigenous communities and schools and by native specialists such as herbal curers.

**Transformative Nature of the Project**:This project is transformative. It qualitatively changes the nature of ethnobotanical fieldwork. It has profound implications for anthropological and linguistic research involving ethnobotany. It empowers Indigenous peoples to participate in documenting theirTEK, as the three partnering collectives have elected to do. Present ethnobotanical methodology involves the collection of fertile specimens linked to ethnographic information. The specimens are then determined to scientific species, an "etic grid" that provides a universally shared referent. Even a highly skilled ethnobotanist familiar with the flora of one given community will find it a challenge to move into a new community, with different nomenclature, and identify sterile specimens. By using a DNA barcode reference library, collection becomes "routine" (a simple leaf dried in silica gel to accompany one pressed voucher for reference) and identification to species more rapid and objective, a "technical" laboratory procedure. This will allow greater participation in ethnobotanical research by social scientists and Indigenous communities. Extensive community surveys become feasible, enabling research to address the theoretical questions posed in section 5. Finally, the DNA barcode library is reliable, accurate (based on expert determinations and state-of-the-art sequencing), and extensible. It will cover about 8 percent of Mexican angiosperms, a significant level of resource development for a project in the social sciences.

**7. BROADER IMPACTS |** This project builds collaboration among a liberal arts college (Gettysburg), three Indigenous collectives that offer a remarkable degree of support, Mexico's National University, and the Smithsonian. Three full-time Indigenous collaborators will be trained in botanical, linguistic, and anthropological research; many others will work part-time on individual projects. A Mexican advanced degree student will work closely with researchers at the Smithsonian, the Instituto de Biologia, and the Indigenous collectives and help carry out DNA extraction, amplification, and sequencing. Tosepan Titataniske is building an educational, exhibition, and research center that will function as a hub of project activities and create exhibitions of project results for Indigenous people, visiting students, and the general public. The project team and Tosepan's Montessori school teachers will develop pedagogical material, which will be placed, along with the primary linguistic documentation, on an open access project website. Collaboration with an Indigenous health collective, Tosepan Pajti, includes training of Indigenous herbal curers and the creation of an anthology of medicinal plant knowledge to be shared within the collective and, pending the IPR approval by the curers and the collective, with the general public.

In regard to public dissemination, all primary project results will be made available quickly: (1) the voucher specimens will be accessioned at MEXU, US, and the taxonomic expert's herbarium, if different; (2) a digital photograph of the MEXU voucher will be uploaded to the Instituto website UNIBIO; (3) the DNA barcodes will be submitted to GenBank; (4) linguistic material (digital recordings, transcriptions, translations; a database of all plant species with Indigenous nomenclature, classification, and use) will be deposited at the Archive of Indigenous Languages of Latin America (U. Texas). The project will construct its own bilingual (English/Spanish) website for an online "Next Generation Flora of the Sierra Nororiental de Puebla". The site will be developed toward project end and hosted either on the National Museum of Natural History server or on the Endangered Language Fund 4-terabyte server (see support letter of ELF president, Douglas Whalen, SD, C-14) both of which have offered support.

**8. SUPPORT: INSTITUTIONAL AND INDIVIDUAL**

***8.1 Institutional support:*** Institutional support is international and interethnic. Botanical research will be based at the Smithsonian and the Instituto de Biología (each with molecular labs and associated with the national herbariums of the US and Mexico) with added botanical support from four major herbaria (New York, Missouri, Texas, California Academy of Sciences) and the Puebla state herbarium and botanical garden, an institution strongly interested in and committed to the floristic study of the Sierra Nororiental. As part of a CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad) grant to digitally photograph all MEXU holdings, Salazar and Gernandt (director of the national herbarium) will place all Sierra Nororiental voucher specimens used for barcoding on a fast track to be photographed and uploaded to UNIBIO, an Instituto de Biología website (http://unibio.ibiologia. unam.mx/). Salazar, Kress, Erickson, and Gernandt will also oversee a Mexican researcher who will provide day-to-day oversight of the biological facet of the project, coordinate sequencing between the Smithsonian and the Instituto de Biología, and lead the effort to utilize project results to address issues in community phylogenies, biogeography and biodiversity, while analyzing the genetic data for indications of new species or new state registers. Finally, Dr. José Saruhkán, director of Mexico's Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, has extended a letter of support for this project (SD, Part C:11), stating that it is of "great interest" to the government agency that he directs. Considering the molecular work that this project entails, this support and interest is extremely valuable. CONABIO has already facilitated access to CONABIO's floristic database resources, to which this project will add significant data.

Tosepan Titataniske, the Nahua and Totonac cooperative (p. 2, 2nd paragraph), is building a 4,300 sq. foot education, exhibition, and research center (SD, Part C:4–6) to serve as a hub for investigative, pedagogical, and outreach activities associated with this project. Tosepan Kali will lodge all project participants who travel to the Sierra Nororiental (SD: Part C:7–8). Finally, Amith and Tosepan Pajti (dedicated to health issues and traditional medicine, with 3000 members in 90 communities) have signed a collaborative agreement (SD: Part C:9–10): (1) to train traditional herbal curers in plant collecting; (2) to identify (in Indigenous and scientific terminology) medicinal plants known to Tosepan Pajti members; (3) to create a cross-referenced database of nomenclature correspondences among varying local Indigenous names for single denotatum; (4) to register Indigenous medicinal knowledge through digital recordings; (5) to create specialized illustrated guides of medicinal plants (with accompanying CDs of interviews with traditional curers) to facilitate knowledge sharing among Tosepan Pajti members and others. This level of collaboration to document Indigenous medicinal plant knowledge and create educational materials presents a unique opportunity to develop a mutually beneficial research and educational initiative linking Western academics and Indigenous natural historians.

***8.2 Key project personnel:*** The project team is uniquely qualified, to cover the following research areas: (1) Nahuat(l) ethnobiology and historical linguistics; (2) Totonac ethnobiology for comparative studies; (3) regional floristics, DNA analysis for a DNA barcode library, community phylogenies.

**Jonathan D. Amith,** **PI/Project director** (Ph.D. Anthropology; Yale, 2000). Amith, an independent scholar and has researched Nahuatl language and culture for over two decades. He has received two three-year NSF DEL grants to study modern Nahuatl, first (2005–9) in the central Balsas Valley, Guerrero, and then (2008–present) in the Sierra Nororiental. He recently was awarded a third NSF grant to study the Mixtec language spoken in Coastal Guerrero. Amith has published in linguistics, history, and anthropology. He has curated an exhibit of Indigenous protest art, produced and codirected an award-winning documentary film, and edited an anthology of Nahuatl-language texts (book and 6-CD set) distributed to Nahuatl-speaking communities by Mexico’s Instituto Nacional de Lenguas Indígenas. He has managed ambitious projects comprising large groups of diversely trained collaborators including computational linguists, phoneticians, biologists, native language speakers (whom as part of his grants he has trained in the skills necessary for language and cultural documentation), and video teams.

**W. John Kress, Co-PI** (Ph.D. Botany, Duke University, 1981). Research Scientist and Curator of Botany, National Museum of Natural History and director of the Consortium for Understanding and Sustaining a Biodiverse Planet, Smithsonian Institution. Kress is a leader in the study of tropical plants, including plant biodiversity, systematics, pollination biology, and conservation. He has spearheaded the application of DNA barcodes for the understanding of plant diversity and evolution, especially characterizing community-level phylogenies and designing the DNA barcoding protocols for the CTFS/SIGEO forest dynamics global plot network. The Smithsonian Plant DNA Barcoding Lab has generated DNA barcode sequences for thousands of plant species. He will oversee the collection and curation of the plant material to be sequenced as well as supervise the laboratory DNA barcode procedures for building the floristic DNA barcode library for the project. With David Erickson (see below) he has edited a recent book on barcoding: *DNA barcodes: Methods and Protocols* (New York: Springer Verlag, 2012).

**Gerardo Salazar, Senior project personnel** (Ph.D. Botany; University of London, 2003). Salazar is a Mexican botanist with over 25 years of experience in collecting, curating, and identifying flowering plants in Mexico, with a specialty in the floristics and systematics of Orchidaceae. Starting with his doctoral studies, he has been involved in plant DNA sequencing. He has been involved in the international DNA barcoding initiative since 2004 and has participated in several international research teams set to assess the potential of different portions of the plant genome as DNA barcodes. Salazar is presently chair of the Department of Botany at the Instituto de Biología, UNAM.

**Karen Dakin, Senior project personnel** (Ph.D. Linguistics; University of Wisconsin-Madison, 1972). Dakin is Full Professor/Researcher in the Seminario de Lenguas Indígenas, Instituto de Investigaciones Filológicas, UNAM. A specialist on historical/comparative Uto-Aztecan linguistics, with special reference to Nahuatl, Dakin has published papers identifying major isoglosses that provide chronological evidence for the diversification of the language family and for Nahuatl dialect splits.

**Daniel Austin: Senior project personnel** (Ph.D. Botany; Washington University, St. Louis, 1970). Austin, an Emeritus Professor at Florida Atlantic University and Research Associate at the Arizona-Sonora Desert Museum, has specialized in ethnobiology and systematics for over 40 years. An expert on the Morning Glory Family (Convolvulaceae), he has also researched and published extensively on plants used by people in the Southeastern United States (*Florida Ethnobotany*, 2004) and in the Southwest, among the Tohono O'odham, a Uto-Aztecan language community (*Baboquivari Mountain Plants*, 2010). Both books won the Mary W. Klinger Book Award given by the Society for Economic Botany.

**David Beck, Senior project personnel** (PhD Linguistics, University of Toronto, 1999). Beck, Professor of Linguistics, University of Alberta (since 2000), is a specialist in Totonacan languages, particularly Upper Necaxa Totonac, from the Sierra Norte of Puebla. Beck’s current research interests include comparative Totonacan morphosyntax and lexicon, the computational-historical reconstruction of the Totonacan family, and possible deep genetic links to other language families of Mesoamerica and the southern United States.

**Claire Bowern, Senior project personnel** (PhD Linguistics, Harvard University, 2004). Bowern, Associate Professor of Linguistics at Yale University, is a specialist in historical linguistics and language documentation, with particular reference to the languages of Australia, where she carried out fieldwork on Yolŋu and Nyulnyulan languages. Her current research involves the possible differences between languages spoken by hunter-gatherer groups and agriculturalists. To this effect she has led an interdisciplinary initiative funded by the NSF’s Human Social Dynamics program to investigate language change in hunter-gatherer societies, particularly differences in loan rates and in the nomenclature of biotaxa and material culture. Her recent work has explored the possibilities of computational phylogenetics for determining the nature of historical relationships among genetically related languages.

**David Erickson, Senior Project Personnel** (Ph.D. Botany, Univ. of Georgia, 2000). Biologist, Department of Botany, Smithsonian. Erickson has been responsible for the molecular component of the Smithsonian's plant DNA barcoding project for the past five years. He has overseen the development of increasingly automated methods for sample processing, DNA sequencing, and implementation of digital Laboratory Information Management Systems for tracking samples and he has supervised a set of graduate students and laboratory technicians who have contributed to that project. Erickson's research background in evolution and genetics has enabled him to both collect the basic data associated with DNA barcode library construction and apply the DNA barcode data in experimental analysis of community phylogenetics and ecological forensics, activities directly applicable to the present proposed research.

**Amelia Domínguez, Eleuterio Gorostiza, Collaborators:** Both individuals are native Nahuat speakers from the Sierra Nororiental de Puebla and fully bilingual in Spanish. They have worked continuously with Amith on NSF and Ford Foundation language documentation and education projects since 2008 and possess the skills (such as accurate transcriptions) necessary to support the development of the Nahuatl material that underlies the present proposal. Gorostiza has a bachelor’s degree in biology and will be a particularly valuable contributor for regional floristics and Nahuat ethnobiology.

**Gabriela Román Lobato, Collaborator:** Gabriela Román (PhD candidate, Universidad Nacional Autónoma de México). Román is a native speaker of Totonac spoken in San Juan Ozelonacaxtla (municipality of Huehuetla), bordering immediately to the west of Cuetzalan in the Sierra Nororiental. She is versed in the phonology and orthography of her language and possesses the necessary skills for the ethnographic and linguistic fieldwork and analysis required in this project. With Beck, she will be responsible for the comparative Totonac ethnobiological component of this project.

**Taxonomists** (for complete list see letters of commitment in supplementary documentation, section D). The key to scientifically sound ethnobotanical research and to an accurate DNA barcode reference library is accurate scientific determinations. This is guaranteed by the collaboration of over 74 expert taxonomists who will provide accurate determinations of fertile specimens used for the DNA barcode library. Together they have expertise in families that constitute 90 percent of the regional flora.

**9. RESULTS FROM PRIOR NSF SUPPORT**

***9. a Amith****:* Amith has been awarded three NSF DEL grants for primary language documentation. In all projects Amith has collected ethnobiological data: 2,035 plant vouchers and 2,046 arthropods, documented with Indigenous nomenclature, classification, and use, and identified to the level possible through support of over 150 taxonomists. Selected products from all awards are in Amith's biosketch.

**Award #0756536** ($291,798) 2008–present (with supplementary funding): "Nahuatl Language Documentation Project: Sierra Norte de Puebla," one of three NSF DEL grants to develop primary language documentation materials: (a) corpus of digital recordings and time-coded transcriptions; (b) extensive dictionary; (c) reference grammar. The Puebla grant is part of a long-term commitment to collaborative work with the Indigenous collective Tosepan Titataniske.

***Intellectual merit*: Primary documentation material: *Corpus*:** To date Amith has recorded 893 separate items totaling over 180 hours; 680 recordings have been transcribed in time-coded format. About 25% of the material comprises discussions in Nahuat about local flora, fauna, and material culture.The ***dictionary*** presently comprises 3,509 ms. pages: 7,874 headwords/ 21,208 senses, including lexicalized phrases and collocations. About 750 terms refer to flora and fauna. All entries have multiple illustrative examples. New lemmas and senses from a review of the corpus are pending incorporation.The ***grammar*** comprises anoutline of a 51-chapter grammar; complete drafts have been written for the first 17 chapters, which cover the basics of phonology and morphology. Extensive notes for the remaining chapters have been compiled from the material in the digital corpus of recordings/transcriptions.

**Collaborations**: Amith's work with Michael Maxwell on a Nahuatl transducer merited mention in NSF Highlights for its innovative use of technology a technology that is the basis for another proposal submitted to DEL 2013. Collaboration with Douglas Whalen to utilize computational tools to process large amounts of documentary material is supported by NSF grant #0966462. Whalen's work on forced alignment (for Mixtec) is potentially transformative in its ability to segment large amounts of primary data (Amith's data has provided the primary test case) from under-described languages.

***Broader impacts:*** Through this award Amith has (1) Enhanced infrastructure for research and education; (2) Increased participation of under-represented groups; (3) Disseminated results to the general public.

Amith has forged a series of inter-institutional, international, and multiethnic collaborations including: (1) the National Museum of Natural History (Smithsonian), where he is a research affiliate; (2) the Instituto de Biología (UNAM), where he processes his ethnobotanical collection; and (3) Tosepan Titataniske, the Indigenous cooperative that facilitates his field research in the Sierra Nororiental. Recently, to increase the impact of its collaboration with Amith, Tosepan Titataniske contracted an architect to design and build a large (400 sq. m) education, exhibition, and research center. The floor plans (SD: C:6) are complete and construction has commenced. This building will be a permanent home to the materials produced through Award #0756536 and will have a continuing impact on the education of the area's Indigenous youth. The center will provide the facilities for this proposal's research. A climate control room will store project voucher specimens.

Finally, with grant #0756536 Amith produced and co-directed, with Roberto Olivares (director), an award-winning 65-minute Nahuatl-language documentary: *Silvestre Pantaleón*. It has won 5 awards, including Best Documentary at the prestigious Morelia International Film Festival and is being handled by Icarus Films, a leading distributor of educational videos to US institutions of higher learning. It has been broadcast three times by the major Mexican government cultural channel and 3,000 copies were acquired by a Mexican government institution for free distribution to Nahua communities and schools.

**9.b** **Kress: *PIs: E. Temeles and W. J. Kress Award: 0614218 Amount: $240,000 Period: 2006–11 Title: RUI: Coevolutionary Convergence and Displacement Across a Geographic Mosaic: Hummingbirds and Heliconias of the Lesser Antilles.***

***Intellectual Merit:*** (1) Field work in Dominica, W.I. on hummingbird mating behavior, plant fitness components, floral trait heritability, and flowering phenology in December 2006, March–June 2007, March–June 2008, and May–June 2009, and March–June 2010; (2) comparative fieldwork collecting data on floral structure and visitors on other islands in the Eastern Caribbean, including Hispaniola, Puerto Rico, St. Kitts, Guadeloupe, Martinique, St. Vincent, Grenada, and Trinidad and Tobago; (3) lab work at the Smithsonian’s Laboratories of Analytical Biology on microsatellite analysis of plant population structure in Dominican Republic, Puerto Rico, Guadeloupe, Dominica, St. Vincent, Grenada, and Trinidad and Tobago for two species of *Heliconia*; (4) completed drafts on eight manuscripts: seven published in journals such as *Ecology*, *Proceedings of the Royal Society B*., *Journal of Evolutionary Biology*, *Biotropica*; the remaining submitted and under review in peer-reviewed scientific journals; (5) fourteen invited presentations and four symposia organized on research work.

***Broader Impacts:*** (1) trained and advised one PhD student, one MS student, two post-doctoral students, eight undergraduates, two visiting scientists, and three students/faculty from Dominica State College; (2) developed and launched new permanent exhibit on coevolution (entitled *Partners in Evolution*) at the Smithsonian’s National Museum of Natural History; and (3) developed and launched a new temporary exhibit on Darwin and evolution (*Since Darwin: The Evolution of Evolution*) at the Smithsonian’s National Museum of Natural History, both of which included sections on the heliconia-hummingbird research; (4) worked with the film crews from BBC and PBS to develop public education documentary on hummingbird-plant ecology and evolution; and (5) interviewed on ABC News, AAAS podcast, Smithsonian Magazine, and *K En Tu Vida* magazine (Panama) about hummingbird/heliconia research. Articles that have resulted from this award are listed in the biographical sketch under Five Other Significant Products and marked \*.

1. An important ethnobotanical work (Taller, 1988) with bilingual texts (originally from recordings) exists of 150 useful plants in one village of Cuetzalan, one of the 28 Sierra municipalities. Though valuable, from a documentation perspective it suffers shortcomings: (a) no audio or archival component; original cassette recordings have deteriorated beyond recovery; (b) vowel length is not written; (c) the plants are mostly not identified in Western binomial nomenclature; (d) it covers less than 30% (medicinal plants) of species documented by Amith to date; (d) no vouchers have been deposited in a herbarium. [↑](#footnote-ref-1)
2. In neotropical areas a single Indigenous community might easily have 1500+ species on its lands alone. [↑](#footnote-ref-2)
3. Comparison among unrelated languages never in contact can also shed light on cognitive universals of biosemantics (cf. Berlin, 1992). Martin (1996) is an excellent example of this comparative approach. [↑](#footnote-ref-3)