Proposal Title: The UC Conservation Genomics Network
Host Campus: UCLA
Name of Applicant/Principal Investigator (PI): Wayne, Robert

Proposed Research Activities

Introduction: The potential of new cutting-edge molecular techniques to solve pressing environmental problems is yet unrealized. Key challenges hindering their application include matching appropriate molecular techniques to important conservation questions, modifying these techniques for use in non-model species, developing specialized computational and analytical resources, and engaging interest groups beyond the research community to collect and productively use new data and resources. Several recent reviews of conservation genomics have highlighted the considerable potential of the field (e.g. Allendorf et al., 2010; Ouborg et al., 2010a,b; Angeloni et al., 2011; Steiner et al., 2013; Shafer et al., 2015) or focused on a few specific approaches such as seed banking, restoration ecology or defining units of conservation (e.g. Williams et al., 2014; Funk et al., 2012; Shaffer et al., 2015). However, the central role of genomics in effective conservation has not been demonstrated empirically, nor have new computational approaches developed from basic molecular disciplines been applied to environmental questions. A primary problem is assembling a group of experts with the necessary expertise required for an integrative analysis of complex problems in the environment, including those with expertise in field biology, conservation, genomics, molecular and cell biology, and computational and analytical methods. Such consortia have been critical in genomic studies, and many leading papers in the analysis of genome data have dozens of authors (e.g. The 1000 Genomes Project Consortium, 2010). To address this multi-disciplinary challenge, a new form of training is needed that draws on the expertise of UC faculty in all of these disparate fields to create a new generation of researchers that can work effectively across many disciplines. Without such a revolution in graduate training, an effective synthesis that can be translated into plans and actions to better conserve natural resources, given ongoing and future environmental challenges, is not possible.

We propose a centralized and focused approach to collaboration and education that engages students and faculty in both training and doing research, as well as translates findings into recommendations that are actively pitched to conservation practitioners. Such an approach is catalytic by enabling researchers in traditional conservation fields to develop innovative genomic approaches to long-standing problems in the environment. Further, the application of cutting-edge genomic technology to an existing field when coupled with public outreach will help foster broader societal interest in science and attract young researchers to STEM fields. Our proposed consortium and the tools it produces will position the UC as a global leader in this emerging and vital field.

Our proposal has three broad aims: 1) to develop new molecular and analytical approaches; 2) to perform innovative and precedent setting research in conservation; and 3) to provide ready access of these approaches and results to the broader research community. We will also enhance public education and engagement through social media and a specific program utilizing public participation. Through these research aims, we will produce a new generation of broadly trained scientists that can function as independent researchers. To facilitate cross-disciplinary education and collaboration, we will support graduate student and faculty exchange across all UC campuses and associated institutions. The project will provide the integrative tools necessary for young scientists to thrive in the complex field of conservation science and to develop the "computational thinking" skills that are lacking in current educational programs, but essential for the mega-datasets of the future (Wing, 2006).

Products of our proposal will include: (1) a series of efficient computational and analytical tools that can be used by the wider communities in conservation and genomics, including online tutorials, as well as modules to explore the conservation implications for management and policy; (2) focused workshops in which a structured set of learning experiences convey appropriate questions, methodological and computational approaches and policy implications related to our specific aims (see below). These recommendations will be conveyed through our network of collaborators to State, Federal and NGOs involved in environmental monitoring and management; (3) educational tools, such as YouTube style videos, downloadable pre-recorded PowerPoint presentations and online line course materials that will provide content defining the field and methods of conservation genomics; and (4) a sampling toolkit for crowd-sourced data collection, which aims to actively engage the public in biodiversity research and conservation as participating members of the scientific team. This toolkit will be linked to a digital application that provides instructions for sampling and documentation, and that links participants to the broader research community through social media platforms, where they will be encouraged to continue to interact with the data and the UC-based team.

Campus	Species	Area	Status	Threat	Conservation Problem & Aim*	PI	Ref genome?/ Size/Tissue
1. UCM	Vernal pool fairy shrimp	Central Cal	Threatened	Habitat loss/grazing/ drought	Connectivity, local adaptation, drought (1)	Dawson	None/~3Gb/ Leg tissue
2. UCM	Vernal pool Colusa grass	Central Cal	Threatened	Habitat loss/grazing/ drought	Connectivity, local adaptation, drought (2)	Sexton	None/ unknown Leaf tissue
3. UCLA	Red Abalone	Cal coast	Threatened	Overharvesting	Population specific markers for law enforcement (3)	Wayne	None/250Mb Mantle Tissue
4. UCLA	California Tiger Salamander	Central Cal	Threatened/ endangered	Habitat loss	Hybridization with related invasive species (4)	Shaffer	None/>32Gb Tail Tissue
5. UCLA	Tricolored Blackbird	Central/SoCal	Declining	Urbanization/nest habitat loss	connectivity/ local adaptation (5)	Smith	None/~1.3Gb Blood
6. UCLA	Bobcat	Urban habitats in California	Declining	Rodenticide/leading cause of mortality	How toxicants affect immune function (6)	Wayne	None/~3Gb Blood
7. UCLA	Valley Oak	California	Declining	Habitat loss, drought	connectivity/ local adaptation (7)	Sork	None/~800 Mb Leaf tissue
8. UCB	California Black Rail	Central Cal Wetlands	Threatened	West Nile Virus/ drought	connectivity/ local adaptation (8)	Beissinger	None/~1.3Gb Blood
9. UCB	Yellow-legged frog	Sierra Nevada Mountains	Endangered	Disease, introduced predator/competitors	Connectivity/disease, local adaptation (9)	Rosenblum	In progress/ ~1Gb/ skin
10. UCSC	Mountain Lion	California	Threatened near urban areas	Habitat loss/ urbanization, human conflict/rodenticide	connectivity, gene expression & rodenticides (10)	Wilmers	None/~3Gb Blood
11. UCSC & UCLA	Sea Otter	Cal coast	Threatened	Habitat loss/pollution	connectivity, pollutants, bottleneck (11)	Estes/Wayne	None/~3Gb Blood
12. UCSC	Black Abalone	Cal coast	Critically endangered	Habitat loss	Local adaptations, connectivity (12)	Shapiro	None/250Mb Mantle Tissue
13. UCSC	Steelhead Trout	Cal Coast/ central valley	Threatened/ endangered	Dams/stream dewatering	Connectivity/local adaptation (13)	Palkovacs	Yes/2Gb fin tissue
14. UCSF	Northern Spotted Owl	West Coast/ Northern Cal	Threatened	Habitat loss, competition from invasive species	Hybridization with related invasive species (14)	Wall	Yes/~1.3Gb Gb /Blood

* * Model for the following processes: 1. Maintenance of meta-populations through increasingly extreme climate cycles including drought and flood in vernal pools, Alpine ponds, and mosaics of managed and wilderness areas. 2. A species endemic to California vernal pool habitat, of which only ~10% remains and a large proportion of which are on private land; models of population structure and local adaptation will facilitate seed banking and recovery strategies. 3. Legal fishery in central California masks illegal take elsewhere; a genetic test for law enforcement is difficult because of high gene flow among populations, a genome wide study focused on local adaptation is needed. 4. Hybridization with an invasive species resulting in high hybrid fitness and the spread of non-native genes; models will predict future change and how best to reverse the spread of alien genes. 5. Habitat fragmentation and urbanization affecting North America's most colonial landbird. 6. Changes in patterns of expression in immunological genes due to exposure to anticoagulants in rodenticides. This is the primary source of mortality in SoCal bobcat. 7. Habitat fragmentation and urbanization affecting the most jeopardized oak in California, which serves as a model for all oaks -- foundational species for biodiversity of California ecosystems. 8. Genetic effects on gene expression and genomic diversity caused by disease (West Nile) and loss of wetland. 9. Connectivity and local adaptation (in particular disease resistance), in particular as influences translocation plans. 10. Inbreeding as a consequence to in proximity to cites, including genetic and immune effects of rodenticides. 11. Understanding the effect of population bottlenecks on genomic diversity and the impacts of environmental stressors on the gene expression response of a species emblematic of California kelp ecosystems whose protection has not led to population expansion. 12. Local adaptation and population fragmentation in response to warming water and other stressors. 13. Population fragmentation due to dams and dewatering and the evolution of freshwater resident populations. 14. Hybridization with Barred owls threatens persistence of this essential species for protection of old growth woodland. Mapping the spread of hybridization and understanding the genetic consequences is essential for predicting future changes and decreasing its effects.

Fundamental questions and specific aims: Our overall goals are defined by a series of challenging and fundamental questions in monitoring and managing natural environments and the species they contain:

- **Effects on population structure -** How is the population structure and demography of common, rare and endangered species affected by urbanization, habitat conversion, invasive species and climate change?
- **The landscape of adaptive and deleterious variation -** What is the landscape of adaptive and deleterious variation across disturbance levels from large pristine areas to isolated habitats at the urban-wildlands interface?
- **Response to biotic and environmental challenges -** How do individuals respond through changes in gene expression to immediate physical and biotic challenges, such as environmental degradation, pollution, disease and other stressors?

- **Baselines and long-term monitoring** - How can biodiversity be monitored across habitats and ecosystems to establish baselines and identify factors that maintain healthy ecosystems? How can this information be used to enhance management of economically valuable and rare and endangered species?

Given this set of questions, we have identified the following specific project outputs:

(1) Development of a suite of tools for the conservation genomics research community. Specifically, we will produce a framework for conservation genomic research and propose testable hypotheses; establish sampling designs for individuals and genes; and construct the experimental and bioinformatics pipelines that will be needed to address the above questions.
(2) Application of approaches developed in (1) to a demonstration panel of species of conservation importance to California and representing important ecosystems (Table 1).
(3) Education of a new generation of students with the skills necessary to use cutting-edge molecular approaches to address current and future environmental problems. Develop and convey recommendations to appropriate State, Federal and NGO organizations.
(4) Educate the public about the importance of research as a means to solve environmental and societal problems through direct participation in the scientific process.

Experimental Approach: Our genomic approaches will use instruments and laboratory techniques that have only recently been applied to natural populations. We will demonstrate these methodological and computational approaches on species and habitats of conservation concern to California and the US (Table 1).

Question 1. What are the effects of population structure on genetic diversity? We will characterize genome-wide patterns of variation in the target species using principally low coverage genome sequencing. We will use two different approaches for this effort depending on the nature of the species. In species with genomes <4 billion base pairs (the vast majority of birds and mammals), we will first develop a *de novo* assembly of the genomes using mate-pair libraries with variable insert sizes sequenced at high coverage (>60X). We will use genome-wide low coverage sequencing (1X) of a panel of 100-150 individuals per species sampled across their geographic range in California, encompassing disturbed and pristine areas that include a variety of threats (Table 1). To develop an approach applicable to species with large genomes, such as the California tiger salamander (Table 1), we will instead construct a reference transcriptome using RNA sequencing (see description in Question 3), and then on the basis of this transcriptome, construct an exome capture system and sequence a large panel of individuals at moderate coverage (e.g. Bi et al. 2013, Schweizer et al. b, in review). The Wayne lab has successfully applied this approach to a sample of 600 wolves using a capture array developed to enrich DNA samples for 1100 known genes, and 5000 1 kb neutral regions (for demographic analysis). We found that 25-50 individuals could be sequenced in a single Illumina lane, giving at total of 8Mb of sequence for each individual at an average coverage of 40X (Schweizer et al. b, in review). To analyze these data, we will use and develop analysis methods that can fully take advantage of all the information in low and medium coverage NGS data. We have previously developed statistical methods for quantifying population structure using low-coverage sequencing data including calculations of F_{ST} and principal component analyses (e.g., Fumagalli et al. 2013) and for assigning local ancestry (Wall, unpublished). Co-PIs Nielsen and collaborators Lohmueller, Green and Wall will extend these methods to enable advanced conservation genetic analyses based on such data, in particular Bayesian methods for population assignment of individuals, including Bayesian methods for detection of recent migrants (Wilson and Rannala, 2003), and methods for detecting recent changes in population structure and effective population size. A great advantage of genome-wide data is that it allows analyses of processes at recent, conservation relevant, timescales of 1-20 generations rather than more ancient, evolutionary timescales modeled previously. Further, using a capture array approach on species with large genome will allow a focus on sequencing transcribed genes whose functional properties can be understood, rather than random genomic segments (as in RADseq approaches).

Using these methods, we will generate a map of connectivity across the geographic range of each species and an assessment of how fragmentation and small population size has impacted levels of genetic variability. Our analysis will identify the populations that are the most isolated, have lost the greatest amount of variability, and are potentially at the greatest risk of extinction. This analysis performed across a wide range of species will provide some estimate of the habitat disturbances that result in population isolation and endangerment. We will utilize geographic information system (GIS) techniques to better understand the specific environmental variables that correlate with genetic measures of variation (e.g. Sork and Waits et al., 2010; Thomassen et al., 2010, 2011; Schoville et al., 2012) and thus identify life history and environmental traits most predictive of genetic loss associated with disturbance. Further, we will use and develop new population viability analysis (PVA) methods to predict population change into the future under different environmental scenarios.

Question 2. What is the landscape of adaptive and deleterious variation? Preservation of adaptive genetic variation is a principle concern for conservation as it maximizes the potential that populations can respond to changing environmental conditions such as will occur with global climate change (Schoville et al., 2012). Adaptive variation can be identified through analysis of low coverage genomic sequences by diagnostic patterns of genetic variation and differentiation. Specifically, genes influencing local adaptation will tend to have low genetic diversity in flanking regions and will be more highly differentiated between populations due to divergent natural selection (Nielsen et al., 2009; vonHoldt et al. 2010; Zhang et al. 2014; Pilot et al. 2014; Schweizer et al. a, in press). We have previously developed methods for quantifying genetic variability and detecting selection from low- and moderate coverage sequencing data (see e.g., Korneliussen 2013). Further, we predict that small isolated populations will have high levels of deleterious variants (i.e. those mutations which reduce reproductive fitness; e.g. Li et al., 2014; Xue et al., 2015), and that populations inhabiting unique habitats will show evidence of local adaptation (e.g. Zhang et al. 2014; Schweizer et al. b, in review), highlighting the need for their conservation. Using either full genome sequencing, or exome sequencing, we can test these predictions. Further, we can identify genes within regions under selection that are likely candidates for adaptation as well as putatively deleterious variants. For example, we will study SNPs within genes to identify those which may be damaging using bioinformatics programs such as SIFT, PolyPhen and GERP (e.g. Flanagan et al., 2010; Schubert et al. 2014; Henn et al., 2015; Marsden et al. in review) which predict the functional consequences of mutations. Together this will provide a map of adaptive and deleterious variation within populations that will identify populations with novel local adaptations at the genome level and those which are most under threat from the accumulation of deleterious variation. We will use population genetic models (see Lohmueller 2014, Marsden et al. in review, for a more details of the modeling approach) to assess the overall impact of the deleterious variation on population fitness. Additionally, these models will be used to predict future viability of populations in response to changes in mating structure and population size. The inference methods used in this and other questions will be assembled in distinct analysis modules encoded in software made available to the research community.

Question 3. What is the range of responses to biotic and environmental challenges? Genetic variation represents a well-accepted source for medium and long-term responses to environmental change across generations. However, the potential for immediate response at the individual level through changes in gene expression is relatively unexplored. Recently, RNA-seq analyses have shown that short-term response to environmental challenges such as higher temperature and algal blooms can be mediated by gene expression changes (Barshsis et al., 2013; Taylor-Knowles and Palumbi, 2014; Palumbi et al., 2014; De Wit et al., 2014; Dixon et al. 2015). RNA-seq involves high-throughput sequencing of RNA transcripts with the goal of estimating relative expression based on the abundance of individual transcripts of all genes within a tissue (Wang et al., 2009; Charruau et al., *in preparation*). Gene expression response represents the first line response to environmental challenges, which as mentioned above, is poorly researched relative to genomic DNA sequence-based variation. To assess gene expression response in natural populations, we will perform RNA-seq analysis of several targeted species through blood sampling across a range of populations (Table 1). For a few species, we will also study tissue-specific patterns of differentially expressed genes. Although extraction and sequencing protocols are well developed (Wang et al., 2009), computation and analytical challenges exist for natural populations, especially those without a reference genome (e.g. Konczal et al., 2014). We will develop new methods to align reads to low coverage genomes and transcriptomes that will enhance measures of differential expression and SNP genotype calling and will be analyzed in a new statistical framework. Our goals are to determine if there are population specific patterns of gene expression that vary with habitat composition, size, quality and disturbance and to determine effects of specific stressors such as pollutants and toxicants (e.g. rodenticides in bobcats, Table 1). Our experimental design will use 3-5 replicates of each condition (e.g. from small isolated populations or disturbed areas vs. large connected populations in similar or pristine habitats). We predict that environmental stressors, such as disturbance associated with urbanization, profoundly affect gene expression and survival (e.g. Veldhoen et al., 2012).

Question 4. How should we establish baselines for long-term monitoring? To gauge the scope and rate of loss, baselines of biodiversity need to be established for the full range of habitats and taxa worldwide. Given the expense and lack of taxonomic expertise available at the global scale, we instead propose to design, develop and test a comprehensive "non-invasive" biodiversity sampler kits that can track the diversity of vertebrate, invertebrate and plant taxa in a systematic fashion and that can be used by relatively untrained individuals over a short time period. The approach is conceptually supported by the emerging literature on "Environmental Sampling" that takes advantage of the natural and extensive preservation of DNA sequences in every environment through a variety of processes (see review in Shokralla et al., 2012). We propose to design a universal sampler kit to capture DNA from these varied sources and preserve it for shipment to a centralized facility for analysis by the full community of researchers interested in biotic change from microbes to mammals and for archiving for future research. We will develop, verify and produce biodiversity sampling kits to be used in the same habitats containing our target species across California for which we often have long-term data (Table 1). The target species should be detected along with others known to occur at each location, proving a verification of our approach. This analysis will provide a genetic baseline at our localities and, as the kit is distributed to other researchers and the public (see below), a wide diversity of localities across the state. Data produced by this environmental sampling can be used to address a wide variety of environmental questions and to provide a resource for assessing genetic and biodiversity changes in the future. This component of our research design will specifically catalyse engagement with young researchers and the public and will provide a forum for demonstrating the value and process of science. It will provide a resource that will form the basis of new collaborations and research across the UC system and elsewhere and provide a critical precedent for biodiversity research worldwide.

Faculty development and graduate training. We will offer a multi-layer educational resource aimed at positioning the UC as the leading source of expertise and education in genomic approaches to conservation problems. We will also reach out to the public sector through workshops and the presentation of online programs. Below we divide the program into components offering differing education venues under the premise that diverse sources of instruction provide the best system to reach a wide audience and more effectively educate.

A. Workshops. We will organize a series of one week workshops focused on the different components of conservation genomics and held at each of the five partner institutions. For example, PI Wayne will organize a focused workshop on field and laboratory techniques, such as collection methods and strategy, DNA and RNA preservation and extraction, library construction for genome sequencing and RNA-seq to be taught at UCLA with collaborating faculty. Similarly, co-PI Nielsen would host a workshop on computational and analytical methods featuring software he has developed such as ANGSD (http://cteg.berkeley.edu/software.html) and involving other analytically oriented collaborators including Ed Green (UCSC), Jeff Wall (UCSF), co-PI Beth Shapiro (UCSC) and Kirk Lohmueller (UCLA). Each of the five campuses will develop their own workshops, and minimally, we will offer 3-5 such workshops each year for graduate students, post-docs and faculty who wish to improve skills in this developing field. We may also offer specialized two-day "bootcamps" aimed at developing expertise in critical resources, such as the statistically oriented programing platform "R" (http://www.r-project.org/). We can improve on successful models of workshop and bootcamps existing at UCLA and the other campuses, for example, the UCLA Computational Biology Institute currently offers bioinformatic workshops every few months.

B. Inter-campus exchange. To provide expertise necessary for specific projects, we will support 12 one-quarter fellowships each year for graduate student exchange among the 6 campuses. These fellowships will be awarded to trainees who have reached a point where specialized expertise would greatly accelerate research progress, for example, to develop and troubleshoot specific experimental or laboratory techniques or create an analysis pipeline. At the end of their fellowship, the trainee will document advances in protocols and pipelines, and this documentation will be posted on the project website for use by the community. These fellowships would be available to any students in the UC system and sponsor travel to any lab with the necessary expertise. However, we will give priority to labs associated with our program to enhance interaction and collaboration among partners.

C. Online materials. We will develop online modules aimed at the questions addressed in the proposal (see above). These modules will include discussion of questions, sampling design, appropriate experimental and methodological approaches, computational tools, analytical approaches and tutorials associated with commonly used software packages. We will provide YouTube-like instructional videos, and seek out expertise from the digital media department at UCLA in which the PI has participated in the development of educational media (e.g. http://online2.tft.ucla.edu/X9ED5PSQ3p). Each video will be accompanied by notes and online heuristic exercises.

D. Software. This proposal will develop a series of software tools aimed at research in conservation genomics. Specifically, we will develop pipelines associated with each of the primary molecular techniques (low and moderate coverage genomic sequencing, RNA-seq and transcriptome assembly, filtering, read mapping and SNP calling), as well as analysis tools, e.g. embedded into co-PI Nielsen's *ANGSD*. All tools will be available on the project website with full documentation, tutorials, example datasets and troubleshooting advice. We will also develop help facilities that connect project personal and students. The results of all these efforts will be a leading facility that will accelerate research in conservation genomics at the UC and elsewhere. It will define the UC as a leader in applying new genomic techniques to problems in the environment.

E. Public engagement and the process of science. The species and questions addressed in this proposal are of general societal interest and importance. The taxonomic (oaks, grasses, abalone, fish, salamander, carnivores) and ecosystem (marine, terrestrial, vernal pools, urban-wildlife interface) ranges are considerable (Table 1), offering the public and the California conservation community a wide diversity of heuristic examples to educate and motivate interest in the environment and science. Charismatic species such as sea otters, mountain lions, bobcats and Northern spotted owls (Table 1), provide additional opportunities for engagement. For example, due to declining whale stocks, killer whales in the Aleutian Islands switched to decimating sea lions, then harbour seals, and lastly sea otters. With the loss of the latter, kelp forests disappeared from large areas due to overgrazing by urchins, a prime prey item of otters (Estes et al., 1998). As otter populations declined, important genetic variation may have been lost, which may partly explain why recovery has not occurred in California. Educating the public about such linkages provides critical lessons in ecosystem recovery and management, and PI Wayne is in discussion with the Monterey Bay Aquarium about an exhibit on the subject. To reach as broad an audience as possible, our results will be featured in public workshops, social media, and in classes taught by the PIs. However, the most consequential public interface of this project will concern the biodiversity sampling (question 4). Here we will make innovative use of social media to recruit volunteers to visit and sample over 500 localities chosen by the PIs. Volunteers will receive instructions online from students and faculty involved in the project, will be sent a biodiversity sampling kit, and then be guided to specific localities where will they will document habitat parameters and structure using photos and notes entered into a digital phone app. All participants will have access to a special section of the website where habitat and DNA information as well as analytical tools will be available for the participants to use and discuss results. This interface will both educate and energize public interest across a wide citizen base. Finally, student and faculty will directly interact with this public effort through regional meetings and presentations to volunteers as well as online forums. If this ambitious project to involve the public is successful, we expect to open the website information to the general public and the conservation community. We are already working with the appropriate State and Federal authorities to obtain the necessary permits for such a citizen science based effort. We hope this will provide a new paradigm for biodiversity baseline sampling worldwide in a crowd-sourcing format.

Translating results into management recommendations: A critical deficit in conservation science is the translation of results into management and policy recommendations. We will emphasize the translation aspect of our research by specifically providing guidance through semi-annual workshops for resource and policy personnel and website communications including a website blog about molecular approaches in conservation. We will also engage UC faculty working on the forefront of science and society including faculty at UCLA's Institute of the Environment and Sustainability (e.g. colleague Stephanie Pincetl; Director, Peter Kareiva). Collaborator Brad Shaffer has worked on the interface of science and policy as head of the UCLA/La Kretz Center for California Conservation Science. Similarly, collaborator Thomas Smith as head of the Center for Tropical Research at UCLA, has worked closely with State, Federal and international conservation organizations for many decades. Collaborator Steve Beissinger (Berkeley) has trained agency personnel in population viability analysis at the National Conservation Training Center for the U.S. Fish and Wildlife Service for a decade.

Collaboration, Mentorship and Training Opportunities

Intellectual contributions. The novel and innovative intellectual contribution of our project centers on the application of cutting edge molecular, computational and analytical techniques principally used in human genome biology, to natural systems. We will translate our findings into recommendations for management of threatened and endangered species and monitoring of environmental resources, all critical concerns in California given population growth and climate change. Achieving these ambitious goals will require the assembly and coordination of expertise across disciplines, interests, and career stages.

Project leadership. Lead PI Wayne (UCLA) has extensive experience in leading successful, cross-disciplinary international consortia and in training and mentoring undergraduate and graduate students, postdoctoral scholars, and junior faculty. He is a founding member of the CanMap project which, like the human Hapmap project, aimed to characterize genome-wide variation in dogs and wolves and has involved more than a thousand genotyped samples and included a consortium of over 50 researchers (e.g. vonHoldt et al. 2010, 2011). Wayne will coordinate joint research activities via quarterly conference calls and annual on-site meetings. Our management team also includes one PI from each institution who will coordinate the activities within their own institution including tracking the progress of postdoctoral, graduate and undergraduate student members. Each targeted field project (Table 1) will be coordinated by a designated faculty member, who will report progress to the leadership team during quarterly calls. All project staff will be involved in maintaining and updating a project website that will contain a full description of all supported projects, details of and links to results and data, as well as documentation of laboratory and field approaches and computational and inference tools. The website will have a tiered access structure, protecting unpublished data and limiting editing of scripts, programs and protocols to specific users and collaborators.

Mentorship and career development. Our project takes advantage of the diverse and complementary expertise in genomics and conservation across the UC and thus will facilitate the progress of junior and mid-career faculty struggling with limited technical and intellectual resources. The field of molecular genetics is becoming a *have and have not* endeavor whereby major universities which have the extensive instrumentation, bioinformatics and scientific expertise needed for big data projects are the only ones that can engage in cutting-edge research and address consequential questions. This is exemplified in human genomics where many papers have dozens of authors and represent international collaborations, which are often exclusive and inhibit the engagement of young researchers. Our catalytic enterprise is aimed specifically at this problem by assembling a UC wide panel of expertise that will provide the necessary intellectual and analytical tools so that young researchers in any UC can have entrée into advancing population genomic approaches as well as ecological and evolutionary concepts necessary for creative synthesis. Specifically, we have field and laboratory expertise centered at UCLA and UCSC, with 8 UCLA and 5 UCSC faculty conducting field sampling coupled with genomic analysis (UCLA: PI Robert Wayne and project collaborators Victoria Sork, Brad Shaffer and Thomas Smith, with aligned faculty including Mike Alfaro, David Jacobs, Paul Barber and Pamela Yeh; UCSC: co-PI Beth Shapiro and project collaborators Chris Wilmers, Eric Palkovacs, and Jim Estes with aligned faculty including Pete Raimondi), and at UCB (project collaborators Michael Nachman, Steve Beissinger and Erica Rosenblum) and UCM (co-PI Mike Dawson and Jason Sexton). This experimental arm is complemented by computational expertise at UCSC (Shapiro and collaborator Ed Green, with aligned faculty including David Haussler) and UCB (co-PI Rasmus Nielsen with aligned faculty including Monty Slatkin). Development of new analytical methods will be centered at UCB (Nielsen), UCSF (co-PI Wall), and UCLA (Lohmueller). These computational and analytical efforts will be paired with campus targeted field projects (Table 1) and the interaction between empirical and computational biologists will foster and sustain a long-lasting liaison which promotes further research activity on multidisciplinary, precedent-setting projects. As our team comprises faculty across all professorial ranks, these collaborations also provide unique opportunities for mentorship within and between institutions and the establishment of novel long-term research projects. Specially, we will engage faculty, at all levels including 4 assistant professors (Lohmueller, Green, Saxton, Rosenblum), 4 associate professors (Shapiro, Dawson, Wilmers, Wall) and 9 full professors (Wayne, Nielsen, Smith, Sork, Shaffer, Estes, Wall, Nachman). The collaborative research and workshop setting provided by the project will facilitate the transfer of academic and career advancement approaches from senior to more junior faculty, thereby enabling more rapid advancement.

Faculty, post-doctoral and researcher engagement and participation across the campuses. The intercampus interaction will be actively sustained by a vigorous training and education component. Specifically, project fellowships for graduate and post-doctoral scholars will move laboratory and computational expertise among campuses and ensure a collaborative atmosphere where project participants are actively involved and informed of research activities. Project-specific meetings will encourage communication and collaboration between computational and experimental researchers and between partners at different career stages. To ensure that results will additionally inform conservation management, project researchers will participate in meetings with NGOs, and State and Federal scientists involved in management. This effort will be facilitated by project researchers extensive connections with practitioners including collaborators Wayne, Shaffer, Smith, UCLA Life Sciences Dean, Sork, and Beissinger. Our intense public engagement through the biodiversity baseline project will translate our results to further to stockholders and students. The proposed series of workshops and bootcamps will additionally insure intellectual interchange and movement of project participants, including affiliated faculty, several times a year among campuses. Graduate students and postdocs will be encouraged to develop and teach components of these workshops, including but not limited to those designed to stimulate interaction among the UC-based research team as well as the public.

Postdoctoral and graduate training opportunities. Graduate students and postdoctoral scholars supported through this project will be formally based at one UC campus, but will be required to spend one third or more of their supported time working in at least one or more other UC institutions. This rotational system will ensure that the student/postdoc will receive cross-disciplinary training and mentorship, and foster the development of collaborative research across the UC. Through their involvement with the planning and coordination of the workshops, graduate students and postdocs will develop effective leadership and presentation skills and enhance their communication to fellow researchers. Further, because they will also be involved in the biodiversity baseline project, they will develop skills to educate the public in the value and application of the scientific method. Finally, intercampus exchange will provide post-doc and graduate students with different intellectual frameworks enhancing synthesis, innovation and novelty.

Undergraduate training opportunities. Undergraduate students will be actively involved in all aspects of this project from design to data collection to computational analysis. Undergraduate students will be recruited from all departments represented by our affiliated faculty, and will include students from computer science, engineering and statistics as well as molecular and evolutionary biology. For example, the web and phone-based app will be designed as part of a targeted undergraduate experience program at UCSC. Undergraduates will be encouraged to attend and participate in workshops and bootcamps and to participate in the publication and dissemination of results. Graduate students and postdocs will be expected to train and mentor undergraduates through established research internship programs providing college credit during the academic year and summer sessions. The PI has had extensive experience through his board membership of the PEERs program at UCLA (http://www.ugresearchsci.ucla.edu/progpeers.htm), as well as offering research opportunities to student in his conservation biology course (EEB116, with ~250 students). As with many project participants, he generally has 3-5 undergraduates working in his lab each quarter.

Systemwide Engagement and Benefits

Our ground-breaking research program will be a catalyst for defining the emerging field of conservation genomics for the UC system and help to establish the UC as a world leader in the application of molecular techniques to environmental problems. Our multi-campus program of student education and public involvement in research will establish a new educational model for the UC that will be accompanied by a set of heuristic tools and guidelines to facilitate wider application in future. Faculty collaboration across disciplines will be formalized through workshops and bootcamps, and by the integrative nature of the work outlined in the proposal, which necessarily fosters collaborations among faculty with diverse resources and expertise spanning field biology, conservation, population genetics and molecular biology. We believe this collaboration system provides a new paradigm for solving environmental problems that involves not only science but also education, politics and management, and will lead to successful future funding for inter-campus collaborations. Our proposal creates a UC-based alliance of experts that, in collaboration, will be ideally positioned to define and develop the emerging field of conservation genomics and raise the international stature of UC research.

We aim to achieve several distinct and important goals. First, we will demonstrate the utility of our applications, techniques and tools that arise from our efforts to Californian species of special conservation importance. We will translate our findings into recommendations for management and monitoring of environmental resources, a critical concern in our state given population growth and climate change. Specifically, we will develop methods to assess the accumulation of adaptive and deleterious variation in populations and identify populations that may hold critical adaptations for the future or are under the greatest threat of genetic decline and extinction. Our research will assess the potential of populations for short-term response to environmental change through the analysis of gene expression and make predictions about the response of populations to specific environmental stressors. In general, the proposed research addresses major unanswered questions in conservation biology such as how the adaptive potential of populations can be preserved as a hedge against future climate change and the specific management practices that best foster the preservation of this potential. Finally, we will develop new ways of assessing biodiversity and monitoring biodiversity change. We will establish baselines of biodiversity in a wide group of taxa and habitats across California. These specific data will be invaluable in assessing the response of biodiversity to future habitat loss and fragmentation, invasive species and climate change. Finally, we will formally initiate heuristic interactions between the UC and the public by engaging them as active members of the research effort.

Multi-campus engagement will be actively sustained by three primary mechanisms. First, quarterly team-wide conference calls organized by lead PI Wayne will keep all campuses and project members up-to-date on research progress and coordinate future actions. Similarly, project-specific calls by project leaders will integrate the components of each specific project across campuses. Second, the training and educational component of the project requires that postdoctoral scholars and graduate students actively move between campuses throughout the duration of the program. This continuous movement of personnel will restructure the relationships between campuses, and foster an atmosphere of collaboration and communication that benefits both the project and the UC community of scientists. Finally, the planned series of week-long workshops and research “bootcamps” for students will bring together participants from all campus locations to interact with each other and, through regional meetings, with members of the public.

Public Engagement and Research Benefit

We will develop a novel and integrative interface for the involvement of the public in the process of science through our biodiversity baseline project (question 4). This process will involve four essential steps. First, we will develop a sampling kit appropriate for the preservation of environmental DNA from several sources including soil, aquatic and terrestrial environments. The science of such sampling is fairly well advanced and was first applied to ancient permafrost soils, lakes and rivers, and trace remains such as saliva, feces and urine (e.g. Thomsen et al. 2002; Epp et al. 2012). Our kit will aim at sampling soils, aquatic environments and feces as they have DNA from a wide range of communities and species. The kit will contain various (non-toxic) buffers, sampling instructions and a mailer. Second, we will develop a digital smart phone app with a dynamic and interactive interface that will provide a step-by-step standardized sampling procedure complete with interactive questions which will allow the public to act as scientists making critical observations about their environment and educate them in the scientific process through making observations. The volunteers will also be prompted to take a standardized series of pictures, and GPS locations will be noted by the app as well as observer information. This app will provide standardized information the will be useful to researchers although it will be gathered by the public and will also in the process educate the public about the process of science, enhance their observation skills and encourage conservation values. Notably, this application will be developed by undergraduates at UCSC as part of an established course. Third, we will use social media to target interested public and recruit volunteers. This media program will be developed by undergraduates as well (who are experts at social media), as part of a weekly class section in the undergraduate conservation biology course taught by the PI and in a computer science course taught by co-PI Shapiro and collaborator Green at UCSC. The media used will include the full range of common platforms such as tumblr, twitter and facebook. Finally, graduate students, post-docs and faculty will directly engage the volunteer public through a website interface blog format, where questions will be answered and project information will be provided. We will provide travel support for students and faculty to have regional meetings with volunteers at UC partners' institutions where results and conclusions are presented so that the public can see the fruits of their efforts and complete their education about the scientific process. A section of the project website will be open to volunteers allowing them to see the assembled data, run some preliminary analyses with provided analytical tools and make their own observations and conclusions that will be added to the scientist-public interactive blog.

As discussed in the proposal, the species and questions we address are of general societal interest and importance. The range of taxa and ecosystem samples are considerable (Table 1), offering broad set of questions of general interest to the public and the California conservation community. We included, keystone and charismatic species such as sea otters, mountain lions, bobcats and Northern spotted owls (Table 1) that will enhance our ability to attract public interest and will act as flagship species for conservation of specific environments and ecosystems. The PI was co-curator of a multi-million dollar, NSF sponsored exhibit at the Los Angeles County Museum that used dog domestication to teach evolutionary biology, and which travelled for several years to other venues. He would like to parlay that experience to help create an exhibit at the Monterey Bay Aquarium featuring sea otters as a tool to educate the public about the use of genomic approaches in conservation at the UC.

Translating results into management recommendations: As discussed in the proposal, a critical deficit in conservation science is the effective translation of results into management and policy recommendations and practice. We will convey management recommendations specifically through semi-annual workshops for resource and policy personnel as well as website communications including a research blog specifically aimed at conservation practitioners (distinct from the public blog) which will require membership and enhance direct communication between faculty and scientists needing to solve pressing environmental problems. This blog may enhance the probability of additional funding to the UC and project participants.

Timeframe, Milestones and Evaluation Metrics

This project requires three years of support. In the first year, we will complete sampling of target taxa across the State. Given that all or the majority of samples for target taxa are already in place (Table 1) and the existence of long-standing field programs for the target species, we are confident we can meet this deadline. The computational and analytical groups will spend the first year developing sequencing and analysis pipelines that will be applied to data in the second year when the majority of sequencing will take place. The computationally oriented PIs have considerable experience with developing these resources; co-PI Nielsen has developed several software analytical pipelines, most notably ANGSD, which has become a leading tool in evolutionary genetics. The third year will involve data and analysis and write-up of results. Additionally, we will focus on translating results into management recommendations and meeting with State and Federal biologists (see educational plan). Academic interchange will begin in the second half of year one whereas other activities will begin in the second year (see educational components above). The requested budget is sufficient for these activities. Finally, fellowship and workshop support are recurrent costs that are straightforward to estimate.

PI-Wayne will be charged with monitoring progress and compliance with the timelines above. He has been involved in numerous multi-PI projects (e.g. Freedman et al., 2014; Thalmann et al., 2013) and has published over 250 articles with hundreds of collaborators. He will actively work on project bottlenecks and deal with UC and State bureaucracy when necessary. As the proposed collaboration is chiefly aimed at research, the primary milestone will be the submission and publication of papers in the third year. Minimally, we can expect one or two submitted papers for each of the target taxa, and in addition, a similar number from each of the computational PIs as senior or first author. Consequently, we expect at least 10-20 submitted papers by the end of the project. All heuristic activities will have project evaluations during and after the activity. For these evaluations we will draw on the office of instructional development at UCLA (http://www.oid.ucla.edu/), which has standard approaches and expertise for evaluation using tools such as targeted questionnaires and interviews. The usual metrics will be used to keep track of project website activities, and the use of all developed tools and concepts will be tabulated through web searches. Finally, we will collate comments from our web-blog in a report. Review articles will be written outlining general project achievements and novel insights from the research.

Conservation genomics is a young field that has direct relevance to both state and federal agencies. Particularly though the California Department of Fish and Wildlife (CDFW), we anticipate developing an MOU that would allow for long-term funding for our UC Conservation Genomics initiative after this initial funding expires. There are many examples of this kind of arrangement, often accomplished through Cooperative Extension or (at the federal level) the CESU initiative, and we anticipate that such support should be a realistic goal after three years. Project collaborator Shaffer (UCLA) has already approached the CDFW about their interest in such a program, and has a written statement of enthusiastic support from Chuck Bonham, Director of the CDFW. Other agencies tasked with protecting California at-risk species, including the Department of Transportation and the US military have supported similar projects in the past, and we anticipate working across agencies to develop a viable, long-term financial model for our initiative. There are several other avenues for long term funding. First, traditional mechanisms such as support at NSF could be generated in permanent panels in the Division of Environmental Biology, including the Clusters of Ecosystem Science, Evolutionary Processes, Population and Community Ecology and Systematics and Biodiversity Science. Collectively the PIs have had funding experience for all of these clusters. Secondly, special NSF programs such as the dynamics of coupled natural and human systems which focus on the human natural environment interface would be appropriate for long term support. Other government agencies would have interest in this research including the USDA (e.g. http://www.usda.gov/oce/climate_change/regional_hubs.htm), EPA, NOAA, the USFWS and the National Park Service. Lastly, a long list of NGOs including the Pew Foundation, the Keck foundation, the Packard Foundation, the MacArthur Foundation, The Nature Conservancy, The Environmental Defense Fund, the Center for Biological Diversity, The Wilburforce Foundation, the National Resource Defense Council, and the Animal Morris Foundation would have interest in this research. For example, PI Wayne has had recent support from the last four NGOs for conservation related projects.

Literature Cited

1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061-1073.

Allendorf FW et al. 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697–709.

Angeloni F. et al. 2011. *De novo* transcriptome characterization and development of genomic tools for Scabiosa columbaria L. using next- generation sequencing techniques. Mol. Ecol. Resour. 11, 662–674

Barshis, DJ et al. 2014. Lineage specific transcriptional profiles of Symbiodinium spp. unaltered by heat stress in a coral host. Molecular Biology and Evolution. doi: 10.1093/molbev/msu107

Bi, K. et al. 2013. Unlocking the vault: next-generation museum population genomics. Molecular Ecology 22, 6018–6032. doi: 10.1111/mec.12516.

Charruau, P. et al. Influences of Demographic and Environmental Factors on Gene Expression Patterns in a Wild Wolf Population. *In preparation for submission to Cell.*

De Wit, P. et al. 2014. Forensic genomics as a novel tool for identifying the causes of mass mortality events. *Nature communications*, *5*.

Dixon, G. B., et al. 2015. Genomic determinants of coral heat tolerance across latitudes. *Science*, *348*(6242), 1460-1462.

Estes, J.A. et al. 1998. Killer whale predation on sea otters linking coastal with oceanic ecosystems. Science 282:473-476.

Epp, LS et al. 2012. New Environmental Metabarcodes for Analysing Soil DNA: Potential for Studying Past and Present Ecosystems. *Molecular Ecology* 21 (8). Wiley Online Library: 1821–33.

Flanagan SE, Patch AM, Ellard S. 2010. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet Test Mol Biomarkers 14:533-537

Fumagalli M. et al. 2013. Quantifying Population Genetic Differentiation from Next-Generation Sequencing Data. Genetics doi: 10.1534.

Funk WC et al. 2012. Harnessing genomics for delineating conservation units. Trends in Ecology and Evolutionary Biology 27, 489-496.

Henn, BM et al. 2015. Estimating the Mutation Load in Human Genomes. *Nature Publishing Group* 16 (6). Nature Publishing Group: 333–43. doi:10.1038/nrg3931.

Konczal M et al. 2014. Accuracy of allele frequency estimation using pooled RNA-Seq. Mol Ecol Resour 14,381–392.

Korneliussen TS, et al. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics. 14289, doi 10.11861471-2105-14-289.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK. 2006. Genomics and conservation genetics. 21:629-637.

Li, S. et al. 2014. Genomic Signatures of Near-Extinction and Rebirth of the Crested Ibis and Other Endangered Bird Species. *Genome Biology* 15 (12). BioMed Central Ltd: 557.

Lohmueller KE, 2014. The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits. 2104. PLoS Genet. doi:10.1371/journal.pgen.1004379.

Marsden, C. et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *PNAS*, *in review*

Nielsen R, et al. 2009. Darwinian and demographic forces affecting human protein coding genes. Genome Res. 19: 838-849.

Ouborg NJ, et al. 2010. Conservation genomics in transition to conservation genomics. Trends in Genetics 26, 177-187.

Ouborg NJ, Angeloni F, Vergeer P. 2010. An essay on the necessity and feasibility of conservation genomics. Conservation Genetics 11, 643-653.

Palumbi SR et al. 2104. Mechanisms of reef coral resistance to future climate change. Science doi: 10.1126/science.1251336

Pilot, M. et al. 2014. Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity*, *112*(4), 428-442.

Steiner CC, et al. 2013. Conservation genomics of threatened animal species. Annu. Rev. Anim. Biosci. 1, 261-281.

Schoville SD, et al. 2012. Adaptive Genetic Variation on the Landscape: Methods and Cases. Ann. Rev. Ecol. Evol. Syst. 43:23-43.

Schubert, M. et al. 2014. Prehistoric Genomes Reveal the Genetic Foundation and Cost of Horse Domestication. *Proceedings of the National Academy of Sciences of the United States of America* 111 (52). National Acad Sciences: E5661–69.

Schweizer, R. et al. Genetic subdivision and candidate genes under selection in North American gray wolves. (a) *in review*, Molecular Ecology, Special Issue.

Schweizer, R. et al. Targeted capture and resequencing of 1040 genes reveals environmentally driven functional variation in gray wolves.(b) *in review*, Molecular Ecology, Special Issue.

Shafer, A. et al. 2015. Genomics and the Challenging Translation Into Conservation Practice. *Trends in Ecology & Evolution* 30 (2). Elsevier: 78–87.

Shaffer, H. B. et al. 2015. Conservation Genetics and Genomics of Amphibians and Reptiles. *Annu. Rev. Anim. Biosci.*, *3*(ch_1), 113-138.

Shokralla S, et al. 2012. Next generation sequencing technologies for environmental DNA research. Molecular Ecology 21, 1951-1965.

Sork VL and Waits L. 2010. Landscape genetic contributions to molecular ecology – approaches, insights, and future potential. Molecular Ecology 19, 3489–3495.

Thomsen, P. et al. 2012. Monitoring Endangered Freshwater Biodiversity Using Environmental DNA. *Molecular Ecology* 21 (11). Wiley Online Library: 2565–73.

Thomassen HA et al., 2010. Spatial modelling and landscape-level approaches for visualizing intra-specific variation. Molecular Ecology 19, 3532-3548.

Thomassen HA et al., 2011. Mapping evolutionary process: a multi-taxa approach to conservation prioritization. Evolutionary Applications 4(2), 397-413.

Traylor-Knowles N, and Palumbi SR. 2014. Translational environmental biology: cell biology informing conservation. Trends in Cell Biology doi: 10.1016/j.tcb.2014.03.001

Veldhoen N et al., 2012. Molecular profiling of marine fauna: Integration of omics with environmental assessment of the world's oceans. Ecotox. and Environmental Safety 76: 23-38.

vonHoldt, BM et al. 2010. Genome-Wide SNP and Haplotype Analyses Reveal a Rich History Underlying Dog Domestication. *Nature* 464 (7290). Nature Publishing Group: 898–902.

vonHoldt, B M et al. 2011. A Genome-Wide Perspective on the Evolutionary History of Enigmatic Wolf-Like Canids. *Genome Research* 21 (8): 1294–1305.

Wang Z, Gerstein M, and Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10(1), 57-63.

Williams AV, Nevill PR, Kraus SG. 2014. Next generation restoration genetics: Applications and opportunities. Trends in Plant Science http:// dx.doi.org/10.1016/j.tplants.2014.03.011

Wilson GA and Rannala B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163: 1177-1191.

Wing JM, 2006. Computational Thinking. Communications of the ACM 49 (3):33-35.

Xue, Y. et al. 2015. Mountain Gorilla Genomes Reveal the Impact of Long-Term Population Decline and Inbreeding. *Science* 348 (6231): 242–45.

Zhang, W. et al. 2014. Hypoxia Adaptations in the Grey Wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. *PLoS Genetics*, *10*(7).