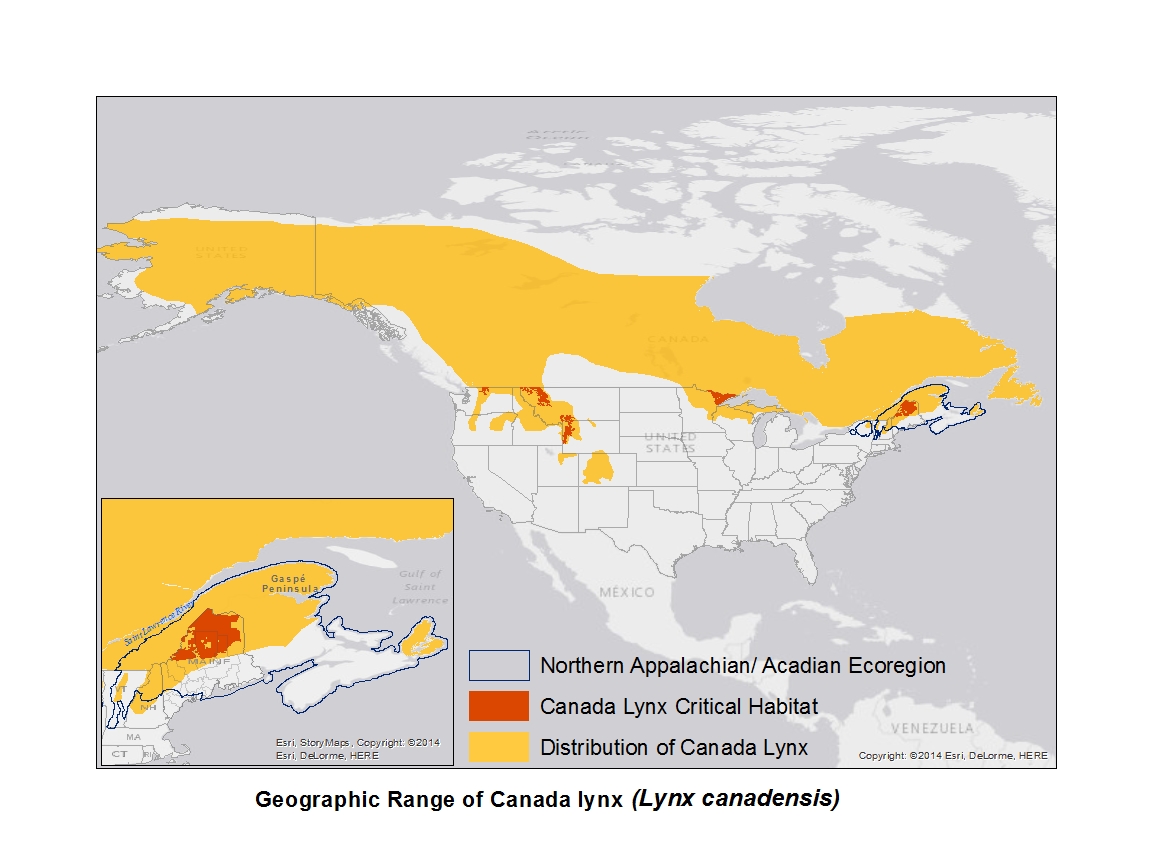
**PROJECT STATEMENT**

**Metapopulation Dynamics of Canada Lynx in the Northern Appalachian/Acadian Ecoregion**

**Background**

The Canada lynx (*Lynx canadensis*) is a wide ranging felid (Ward and Krebs, 1985; Slough and Mowat 1996) listed as threatened under the U.S. Endangered Species Act (ESA) in March 2000 (Federal Register, 2014). The listing decision pertained to the lower 48 contiguous States, and identified lynx occurring there as part of one Distinct Population Segment (DPS). A DPS is defined for listing purposes under the Endangered Species Act as a discrete population or group of populations that are biologically and ecologically significant, and may differ markedly from other populations in their genetic characteristics. Known populations of lynx within this DPS, however, are separated geographically within the United States, with no known lynx population occurring between the Northern Appalachian/Acadian Ecoregion (Figure 1) and the western Great lakes population. Advances in molecular genetic approaches are urgently needed in order to understand population dynamics within this DPS and develop recovery strategies

Figure 1. The complete distribution of Canada lynx in North America and designated areas of critical habitat. Distribution of critical habitat (Unit 1 – Maine) is shown in the lower left.

Conservation initiatives designed to sustain lynx in Maine must address their unique habitat requirements and life history strategies. Lynx co-evolution with snowshoe hare (*Lepus americanus*) resulted in a number of life history strategies, including long distance dispersal during periods of snowshoe hare decline in order to find suitable habitat and prey. Regenerating spruce-fir forest, principal habitat for snowshoe hares, was preferentially utilized by lynx in northern Maine over other forest types (Vashon et al. 2008). This habitat becomes suitable for hare and lynx around 10 years post-harvest and may lose its suitability around 40 years post-harvest (Scott et al. 2009; Olsen et al. 2015). The ephemeral nature of suitable lynx and hare habitat in Maine suggests a “shifting mosaic” pattern of local habitat occupancy and abandonment (Hagan et al., 2005). While lynx populations in Maine (and adjacent New England states) are not contiguous with the western Great Lakes population, they are contiguous with populations of bobcat (*Lynx rufus*) at the southern extent of their range. Bobcats are a distinct species from lynx (Werdelin 1981) and are widespread throughout the contiguous United States and reach their northern extent in Southern Canada. The two species will occasionally co-exist and are likely competitors (Aubry et al. 2000; Buskirk et al. 2000). Evidence of hybridization between bobcat and Canada lynx has been confirmed in the western Great Lakes population (Schwartz et al. 2004) and suspected in Northern Maine. Our proposed application of genomic approaches will provide valuable information as to the separation and speciation of lynx and bobcat, the factors that influenced it, and the significance relative to future management strategies for both species in Maine. Hybridization between these two taxonomically similar species may be a limiting factor to the distribution (Barton 2001) and recovery of Canada lynx. Management of hybrids will largely be dependent on the cause (natural or anthropogenic) of the hybridization, and the genetic consequences of hybridizations; two extremes of which may be (1) the widespread genetic introgression or complete admixture of taxa, and (2) hybridization without genetic introgression (Allendorf et al., 2001).The presence of confirmed hybrids (Schwartz et al. 2004) is a novel factor in the population management of both species with potential implications for hunting and trapping of bobcats. To the north, Maine lynx also meet neighboring populations in the Canadian provinces of Quebec and New Brunswick. The Gaspé Peninsula of Quebec has a lynx population that has sustained a legal fur harvest season for decades, and exhibits dynamics similar to populations in the heart of lynx range in northern Canada. Can the New Brunswick or the Gaspé lynx population be a source population for Maine? Do lynx within the Northern Appalachian/Acadian Ecoregion represent a metapopulation centered in Eastern Canada? If so, strategies focused on maintaining occupancy channels and gene flow across the border will likely be key to sustainability of lynx in Maine.

**Need**

The historic, current, and future status of Canada lynx in Maine has been a matter of speculation, resultant in lynx being listed as a single DPS under the Endangered Species Act (ESA) in 2000. Recent studies (Organ et al. 2008, Vashon et al. 2008a,b) regarding the current status of lynx in Maine have yet to address uncertainty over historic population levels and persistence, and population sustainability and trends in the future. Clarification of historic trends is needed to frame reasonable and appropriate recovery targets. Greater certainty on future trends is needed to establish pragmatic goals and assess the intensity of management that will be required. The conservation genomics approaches we propose will provide critical insights into historic population levels and trends, allow prediction of population resiliency as fueled by linkages to populations beyond Maine’s border. Furthermore, our approach will identify barriers – physical and otherwise – that can inhibit resiliency, and provide insights on gene flow pathways that may inform land conservation strategies beneficial to lynx.

In the last two decades, conservation genetics studies have extensively confirmed that declining and isolated populations lose genetic diversity, develop inbreeding depression and differentiate significantly from other populations (Frankham, 2010; Ouborg et al., 2010). Some studies have demonstrated concomitant fitness reductions in genetically compromised populations and lower potential for these populations to adapt to environmental change (Frankham, 2010; Hoffman & Sgro, 2011). Studies of this nature focused on Canada lynx populations have almost exclusively been based on the use of a few neutral molecular markers (e.g. nuclear microsatellites) and mitochondrial sequences (Carmichael et al., 2000; Schwartz et al., 2003). Extracted DNA has thus far been genotyped at 21 microsatellite loci or less – only 6 of which are characterized in lynx (Lc106, Lc109, Lc110, Lc111, Lc118, and Lc120; Carmichael et al. 2000) and 15 in the domestic cat (Menotti-Raymond et al. 1999). Utilization of such few markers fails to reflect genome-wide patterns of the functional variation upon which both adaptive potential and fitness depend (Ouborg et al., 2010). There is a shift towards conservation genomics as a necessary transition to fill this gap (Ouborg et al., 2010; Steiner et al., 2013). Genomic data is generated by sequencing DNA across all chromosomes of an organism (38 chromosomes for the domestic cat), providing a drastic increase in the number of potentially informative markers used in addressing fundamental and important questions, such as estimation of demographic parameters and viability of small or isolated populations. Furthermore, the promise of identifying adaptive loci offers a huge benefit in prioritizing the conservation of unique populations. Conservation genomics approaches employing whole-genome sequencing can ultimately provide more precise and unbiased estimates of effective population size, demographic history, levels of inbreeding, rates of gene flow, differentiation among populations and taxonomic status (Frankham, 2010; Luikart et al., 2003; Ouborg et al., 2010).

The Canada lynx is a unique model on which to apply conservation genomics tools as lynx have historically gone through a documented decline that has affected genetic variation and fitness (McKelvey et al., 2000). Reference materials available for assembly of the Canada lynx genome include a quality annotated reference genome, gene expression data and variation data from the closely-related Iberian lynx (Godoy, 2010). Genomic approaches can only realize their full potential when combined with ecological, phenotypical, demographical and genealogical information, which is generally scarce for threatened and endangered species other than those that are emblematic or that have been intensively studied, monitored or managed. A broad collection of tissue, blood, hair and saliva samples has been made available for the purposes of this study by the Smithsonian Institute’s Frozen Collection and Maine Department of Inland Fisheries and Wildlife, providing representative geographical and temporal coverage of the North American lynx distribution with special focus on the Maine population. A wealth of genomic information, and an extensive set of population samples, provides replicates and comparison points for different demographic histories. The accumulated information enables us to evaluate the consequences of recent and historic decline and fragmentation on functional genomic variation in lynx, and to assess the possible role of natural selection in maintaining adaptive diversity (Frankham, 2010; Oleksyk et al., 2010). We will ultimately use genetic markers, primarily high density, whole-genome single nucleotide polymorphisms (SNPs) - a strong tool for researchers of quantitative and population genetics that are commonly used for estimation of historical effective population size. SNPs will be discovered through sequencing, assembly and annotation of the lynx and bobcat genomes and arranged in a custom capture array. Each SNP represents genetic variation involving a single nucleotide. We will develop assays, and subsequently genotype these SNP markers on many individuals (Ranz and Machado, 2006; Frankham, 2010) to compare variation patterns at assumed neutral and functional loci across individuals and populations over time. SNP assays will be developed and applied to assess: (1) Whether Maine has a genetically distinct population segment in relation to the contiguous North American DPS: (2) Whether Canada lynx in Maine are part of a metapopulation centered in eastern Canada, and to what extent population sustainability will necessitate transborder strategies versus sole focus on within-state efforts? We will identify populations, rates and sources or barriers to gene flow between Maine and populations in the North American DPS and in eastern Canada. At the most fundamental level, generation of lynx and bobcat genomic data will provide insight into the basic understanding of genome structure, function, evolution and variation, and demographic and evolutionary history of these two species. We aim to generate resources necessary for science based management decisions, informed recovery strategies, and resources and genomic approaches that will be broadly applicable to the contiguous North American lynx DPS and bobcat populations.

**Job 1: Construction and application of a custom array of SNPs and informative markers for Canada lynx and bobcat based on genomic data.**

***Objectives***

The objective of Job 1 is to inform management and recovery strategies for Maine lynx. We will meet this objective by generating genomic data for Canada lynx and bobcat through whole-genome sequencing. Raw genomic data will be assembled, analyzed, annotated and mapped onto reference genomes. Single nucleotide polymorphisms and new informative genetic markers discovered through this process will be compiled across all chromosomes of both species into a custom capture array used in the development of assays. Assays will be employed across individuals, populations and demographic histories to test specific conservation genetics hypotheses relative to Canada lynx populations in Northern Maine and their intersection with bobcat populations, lynx populations in Eastern Canada, and lynx populations across the greater lynx DPS.

Reaching our objective will require completion of these tasks:

1. We will determine the extent to which population recovery and sustainability of lynx in Maine will necessitate transborder strategies versus sole focus on within-state efforts. Our approach to reach this goal is to analyze genetic diversity and genetic difference between the Maine lynx population and metapopulations in Eastern Canada. Specifically, we will develop and employ SNP assays developed from a custom capture array of informative genetic marker and SNP’s discovered across all chromosomes of the Canada lynx genome. We will utilize these assays to test various hypotheses related to source populations (e.g. Canadian Gaspe peninsula, New Brunswick) and barriers to gene-flow between Maine and eastern Canada (e.g. Saint Lawrence River).
2. We will assess whether the Maine lynx population is a genetically distinct population segment in relation to other lynx populations in the currently defined DPS. Our approach is to develop and employ SNP assays of genetic differentiation using genetic sample material obtained from lynx within each subpopulation throughout the current DPS. The origin of genetic material outside of the Maine lynx population will either be provided by the Smithsonian Frozen Collection or partners yet to be identified.
3. We aim to identify the consequences of recent and historic lynx population decline and fragmentation and occurrence of Maine lynx and bobcat hybridization on functional genomic variation, occurrence of deleterious alleles and severely detrimental genetic variants, and potential to adapt to environmental change.
4. We will provide insights to genome structure, function, and common evolutionary relationships between Canada lynx and bobcat

**Approach**

The field of bioinformatics is rapidly evolving and consequently the methods and state-of the art sequencing technologies and analytical software change in regards to accessibility and opinion of the scientific community. Methodology at the time of data analyses may change accordingly, but will be very similar to what is proposed. The samples used for genome sequencing will be acquired from the Smithsonian Frozen Collection and from the Maine Department of Inland Fisheries and Wildlife. Data analyses and workflows will leverage an established partnership with the Smithsonian BioGenomics Initiative and their partners, which will provide bioinformatic support, training and pipelines for genome annotation, analysis, data management and visualization that benefit from high memory and highly scalable, parallel modules. These systems will speed parallel bioinformatics applications and, along with automated workflows, will enable efficient data analysis and large-scale knowledge discovery.

*Preparation of DNA for sequencing*

DNA will be extracted using standard extraction kits (Qiagen). DNA will be used to construct paired-end and mate-pair libraries of several insert sizes (e.g. 250bp, 500bp, 2KB) for sequencing. The whole genome cannot be sequenced all at once, so it will be subdivided, sequenced, and then reassembled in order to arrive at the sequence of the whole genome. DNA is prepared for sequencing by subdividing, copying, chemically modifying, and tagging portions of the genome corresponding to the four DNA bases (A, C, T, G). “Clone-by-clone” and “whole- genome shotgun” are two approaches to subdividing the genome and reassembling the sequenced pieces. “Clone by clone” involves breaking the genome up into relatively large chunks, called clones (~150,000 base pairs long), which are partitioned into smaller pieces with roughly 500 overlapping base pairs. These smaller pieces are sequenced and the overlaps are used to reconstruct the sequence of the whole clone. Genome mapping techniques are then used to figure out where in the genome each clone belongs. “Whole-genome shotgun,” involves breaking the genome up into small pieces, sequencing the pieces, and reassembling the pieces into the full genome sequence. Many genomes are assembled using both approaches.

*Genome sequencing*

Prepared DNA libraries will be sequenced on an Illumina HiSeq 2000 or similar Sequencing System. Regardless of how you subdivide and then assemble sequences, the actual process of sequencing DNA in the genome is the same and employs electrophoresis. Pieces of DNA are marked with fluorescent tagging and arranged into batches by the last nucleotide in each piece (A,C,T,G). DNA to be sequenced is placed at one end of a gel, an electrical current is applied and DNA molecules move through the gel. Smaller molecules move more rapidly and DNA molecules become separated into different bands based on their size. This method can only separate ~500base pairs and hence the need for subdividing the genome before sequencing. As DNA molecules move through the gel, a sequencing machine reads the order of the DNA bases by their fluorescent tagging. The “raw sequence” grows base by base and reads are hooked together in the proper order. Corrections for breaks and errors are made during “finishing”.

*Assembly*

Genomes can comprise of a lot of repetitive DNA and can be difficult to assemble. Computer programs, “assemblers” put together sections of the sequenced genome by finding and analyzing overlapping sequences or identical sequences at either end of two different reads. The assembler compares each read to every other, and puts all the reads in the proper order based on how they overlap. Due to the overwhelming number of comparisons the assembler must make and keep track of, a huge amount of memory is required. A combination of redundancy and quality control ensures that errors in the genome sequencing are kept to a minimum. The genome may be copied multiple times and partitioned so that each base is sequenced 6 – 10 times on average. The assembler software will determine the “consensus” sequence of a base using the compilation of various reads. If the sequencing machine gets the base wrong or if a piece of DNA doesn’t get sequenced, there are likely to be other correct reads to correct for errors and fill gaps. Error probabilities for all of the bases read by the sequencing machine are added together for an estimate of the number of errors in the sequence. Bad reads or partial reads are weeded out before the assembly stage. Once the sequence is assembled, it may be checked against small parts of the genome that have been “finished” or against various landmarks on genome maps. The final polishing involved identifying and filling information gaps of the genome is completed by hand. For the high coverage genomes, raw reads will be filtered for quality and assembled into contigs and scaffolds (Oleksyk et al., 2012) that will then be used to complete de novo genome assemblies using SOAPdenovo or a similar assembly program. The de novo assembly will be complemented by an assisted assembly built from the domestic cat (Pontius et al. 2007) and the Iberian lynx genomes. Genes will be predicted de novo using AUGUSTUS and GENSCAN or similar programs. Homologous proteins will be mapped with tBLASTn using a reciprocal-best-match strategy and the aligned sequence and its query protein will be filtered and passed to GeneWise to search for accurately spliced alignments. Domestic cat EST and full-length cDNA sequences will be aligned to the genome using BLAT to generate spliced alignments. A consensus gene set will be obtained with GLEAN. The genome sequences will be aligned to well-assembled genomes (e.g. human, domestic cat, domestic dog) using LASTZ.

*Genome mapping and annotation*

A genome map includes landmarks such as short DNA sequences, regulatory sites that turn genes on and off, and genes themselves. Genome mapping is often used to find new genes and navigate around the genome for annotation and interpretative purposes. Creating the map and sequencing the genome go hand in hand. The map will determine where each sequenced piece of the genome belongs in relation to the other pieces. A more detailed map will make the assembly process easier and more accurate. Genome maps serve to illuminate the overall structure of the genome and identify where related genes are clustered together or where unusually rich concentrations of genes reside. Genome maps also enable researchers to compare the genomes of different species. Protein-coding genes and other genome features will be annotated on the genome map using algorithmic (Burge & Karlin, 1997) and homology-based gene prediction methods (Altschul et al., 1990). Using the high coverage and low coverage genomes of bobcat, European lynx, Canada lynx and Iberian lynx, SNPs will be identified using established programs (Li et al., 2009) and pipelines (Goecks et al.,2010). We will design a custom capture array of around 6,000 SNPs distributed across all chromosomes and located in annotated protein-coding genes and known gene pathways. This array will be used to assay genomic diversity in existing populations as well as historical museum specimens. The Genome Diversity toolkit in Galaxy (Bedoya-Reina et al., 2012) and other methods will be used to analyze genomic diversity and historical demography.

To more efficiently identify genes within the genome and compare patterns of gene expressions across individuals, blood samples for RNA have been and will continue to be collected during routine response to incidental capture of lynx in Maine. Blood is an ideal first source of RNA because it contains multiple organs and has a crucial role in immune response. All samples will be preserved in buffer to prevent degradation of the intracellular RNA and minimize gene induction (PAXgene Blood RNA kit). Total RNA will be extracted and the globin mRNA and rRNA removed. RNAseq libraries will be pair-end sequenced on an Illumina HiSeq2000, with uniquely indexed libraries pooled in sets of 6 per lane to conservatively ~40Gb of sequence. Since the approximate size of the felid exome is 30Mb, this is an average coverage of at least 200X for each sample.

The transcriptomes will be directly mapped onto their reference genome using a combination of currently available software such as TopHat, Bowtie and BWA. For the downstream analyses, the Tuxedo Suite package contains tools to complete the RNAseq data analysis workflow, including Cufflinks to generate the transcriptome assembly, quantify gene expression, and assess differential expression among populations and species. The software Cufflinks is currently the program of choice to estimate expression levels and test for differential expression between samples after normalization. To explore the biological foundations of genes that are differentially expressed, we will perform categorical enrichment analysis (Gene Ontology; GO) based on publicly available gene annotation sets and the Database for Annotation, Visualization and Integrated Discovery (DAVID). To go beyond a list of differentially expressed genes, we will perform a weighted gene co-expression network analysis to identify clusters of genes whose expression levels covary across individuals and which measures every possible pairwise correlation between genes and identifies groups of genes with significant correlations to each other. If necessary, we will implement and visualize simplified pathways using open source software such as Cytoscape. This will allow us to choose which nodes should be used as targets in order to form a more informative gene protein network. We can also perform pathways analysis in which data available on human transcripts can be used as a reference to analyze the transcripts from our study. Finally, the correlations between phenotype and gene modules can be inferred to identify genes or gene networks associated with phenotype (such as derived from blood profiles). Using algorithms like Random Forest, which are more robust and reliable than hypothesis testing approaches, we will report and visualize significant groups of biologically defined gene sets and their contribution to the phenotypes of interest.

*Gene evolution*

We will identify Canada lynx and bobcat lineage-specific amino acid changes by comparison to human, dog, cat, and mouse from Ensembl. To detect genes evolving under positive selection, we will use conserved genome synteny methodology to establish a high-confidence orthologous gene set based on large-scale synteny of high quality alignments and conserved exon-intron structure. We will align ortholog genes by PRANK and use the optimized branch-site model of PAML and likelihood ratio tests (LRTs) (P<0.05). We will test for over-represented functional categories and identify GO categories significantly above or below average in the target genome. The historic population size of each species will be inferred using a pairwise sequentially Markovian coalescent model (PSMC) with consensus sequences of each species by estimating X Chr and Y Chr separately.

*Development of new genetic markers*

Gene markers will be assessed for their utility to identify current and past levels of introgression between bobcat and Canada lynx in developed assays. Judicious selection of SNPs will easily discriminate among various permutations of first- and second-generation crosses and the sex of individuals involved. A reference ITMImap consisting of anchor markers will be prepared using MAPMAKER/Exp Version 3.0b and marker data will be used for mapping each of the new single nucleotide polymorphisms (SNPs). Markers with minimal missing data or segregation distortion will be selected to build a skeleton map for each chromosome. Other markers will be assigned to intervals between the anchor markers and the loci on the skeleton map will be checked using MAPMAKER. Ultimately, the custom capture array will be distributed across all chromosomes of the bobcat and lynx genomes.

*Genetic diversity and differentiation*

Using SNPs from annotated protein-coding genes, we will examine how allelic diversity profiles have changed across time. If indeed we find significant genetic differentiation in the

contemporary Maine lynx population with respect to the rest of the North American DPS and the Canadian Gaspe Peninsula population, then it will be possible to determine specifically if and how dispersal events between Maine and Canada are impacting genetic variation or function of key genes and genetic pathways related to fitness, such as innate immunity and reproductive effectiveness (Paige 2010; Zhang et al., 2013).

Our samples will be divided into subsets based on geographic distribution (e.g. Maine, Parc de la Gaspesie, Quebec, Bas-Saint-Laurent). Samples outside of the sampling regions or outside proposed boundaries will be omitted. Samples in our most intensively sampled region, Northern Maine’s Aroostook County, may be regrouped to test putative dispersal barriers at a finer scale. DNA will be extracted from samples using Quiagen Blood and Tissue Extraction Kits (Quiagen, Hilden, Germany). Gene fragments from the Canada lynx genome sequence will be selected and primers designed to amplify corresponding sequences from DNA samples on either side of proposed barriers to gene flow (e.g. the St. Lawrence River, climatic barriers, isolation by distance hypotheses) or; alternatively from separate proposed lynx populations (e.g. Maine population and the Gaspe Peninsula population).

PCR amplification will be performed in GeneAmp PCR System (or similar). PCR products will then be sequenced using the original primers and electrophoresed on a 3130XL Genetic Analyzer Applied Biosystems). The final sequences will then be eye checked and aligned using MEGA, from which SNPs will be ascertained. Nucleotide diversity will be calculated for each identified genetic population using MEGA. We will test linkage disequilibrium between any two SNP loci within each inferred genetic population. We will use GenALEX to calculate observed and expected heterozygosity and to test for Hardy Weinberg equilibrium (HWE) within each group for each locus. To assess differentiation between geographic populations, we will calculate pairwise D-values using the package DEMEtics (Gerlach et al., 2010) in R. We will also test for isolation by distance between regions (using centroids as locations) with the ecodist 1.2.2 package (Goslee and Urban 2007) in R (R Development Core Team 2009).

**Expected Results/Benefits**

The present proposal will become one of the first genomic studies of Canada lynx and bobcat to date, expected to yield novel and valuable insights on the demographic history and occurrence of hybridization between Canada lynx and bobcat, inbreeding, rates and sources of gene flow (Ouborg et al., 2010), differentiation among populations (Row et al., 2012) and taxonomic status of both species. Assessment of Maine lynx as a DPS will provide information needed by the Maine Department of Inland Fisheries and Wildlife and the U.S. Fish and Wildlife Service to inform recovery strategies and present evidence for management approaches toward long term sustainability and connectivity in the Northern Appalachian/Acadian ecoregion. If long-term persistence of lynx in Maine is dependent upon metapopulation dynamics in eastern Canada, appropriate policies and actions can be designed. The pioneering aspects of the project, together with the iconic nature of both species, will enhance the local visibility and interest in lynx conservation and bobcat management. Inferences for the study will have broad implications for management of Canada lynx and bobcat, and many of the genetic markers and tools developed will serve colleagues and collaborators doing research on these species.

The application of genomic approaches to threatened species promises a boost in analytical and statistical power for demographic inference and relatedness estimates. The genomic data, annotations and developed assays will be significant scientific achievements that will lead to new research and resources linking phenotypic adaptations with gene function and providing insights on the processes of speciation and maintenance of genetic diversity (Frankham, 2010). The custom capture array will encompass single nucleotide polymorphism (SNPs) and informative markers across all chromosomes of the lynx and bobcat genomes used for more precise and unbiased estimates of effective population size, demographic history, levels of inbreeding, prevalence of hybridization, rates of gene flow, and differentiation among populations. Application of a more comprehensive suite of neutral markers will enable better understanding of genetic differentiation and metapopulation dynamics at course and fine geographic scales (isolation by distance barriers and landscape variables) which may provide insight into the effects of landscape and climate on patterns of genetic differentiation and dispersal (Cushman et al. 2006; Row et al., 2012). This information will assist with identifying and maintaining current patterns of connectivity across the species range, particularly for southern populations such as Maine’s which are likely more at risk and more fragmented (Murray et al. 2008).

**Timeframe**

Sample preparation: starting August 2015 and continuing as necessary

Sequencing: August 2015 – December 2015

Sequence analyses including genomic assembly: September 2015 – August 2016

Comparative genomic analyses: January 2015 – December 2016

Additional lab work as needed: January 2015 – December 2016

Gene Mapping; Development of markers: January 2015 – July 2015

Application of markers in genetic diversity and differentiation: January 2016 – May 2017

Data summary, analyses, and manuscript preparations: July 2016 – December 2018

**Cost**

**Total Direct Cost: $214,727.48**

Federal:

State Matching: Non- federal match will be provided by in-kind contributions from the Smithsonian (non-federal resources only).

Pending Support: There is no other pending support at this time, but additional proposals will be prepared to solicit funds from other agencies (including the Smithsonian Institution and University of Massachusetts Amherst) to extend this work to other Canada lynx and bobcat populations in North America.

***Facilities, Space and Other resource needs:***

Our primary needs to conduct this work are for materials, supplies and modest funding for travel and at least three publications in peer reviewed journals. Samples for the research are (1) archived in frozen collections stored at the Smithsonian; (2) obtainable from archived samples from Maine Department of Inland Fisheries and Wildlife; or (3) will be obtained from incidentally captured individuals during the 2016 coyote fur-trapping season. Genome sequencing will be done at an academic, government or private sequencing center that will be identified at the appropriate time based on costs and possible collaborative agreements. Needed laboratory work at Smithsonian will be done at the National Museum of Natural History (NMNH) Laboratories of Analytical Biology (LAB), which consists of multiple laboratory spaces, computer rooms and office spaces in the National Museum of Natural History.

1. We are requesting a stipend for 3 years minimum ($115,485) for a 20 hour graduate student appointment The project will be coordinated by doctoral student, Tanya Lama, over the course of 3 years, including collection of samples (archival and field), conducting data analyses, and summarizing research findings for publication in peer-reviewed journals under the guidance of senior investigators.
2. We are requesting funds as well as contracting of next generation sequencing services for whole genome sequencing of one Canada lynx and one bobcat at high coverage (>30x), along with re-sequencing of eight additional individual Canada lynx at low coverage (<10x) ($27,200), including preparation of genomic libraries, sequencing and assembly of the high coverage genomes, to be done through Axeq Technologies/Macrogen or a similar company depending upon lowest costs and highest efficiencies at the time, and transcriptome sequencing of individuals of each species ($20,000); and consumable materials and supplies ($11,110). Any reduction in these estimates of anticipated costs will be used for additional data collection.
3. Support is requested ($40,695.48) for the graduate student, supervising senior investigators and associated staff to travel to UMass, Maine and Washington DC for joint meetings. Travel funds are also requested for the graduate student to gain necessary training and guidance, collect additional samples (archival and field), conduct data analyses and summarize findings for publication.
4. Funding is requested for travel ($1,800) for the graduate student to attend a scientific conference in Year 3 to present the data and/or to share findings at an annual conference.
5. Support is requested in Year 3 to cover the cost of publishing at least three peer reviewed articles ($1,500) in a high impact journal.

|  |
| --- |
| Stipend (**$115, 485**) *20 hour appointment (Fall and Spring)*  Stipend $23,670  Fringe $5,781  Curriculum Fee (if applicable) $9,044  subtotal per year $38,495  **total for 3 YR:**  $115,485 |
| Supplies (**$11,110**)  DNA extraction kits for prep. of genomic DNA $800.00  Kapa HiFi Hoststart Uracil+ (KK2802) $850.00  Qubit dsDNA HS Assay (Q32851) $240.00  Bioanalyzer Agilent High Sensitivity DNA Kit (5067-4626) $440.00  Kapa qPCR Complete Kit (KK4824) $920.00  Kapa Library Preparation Kit for SNP array (KK8232) $2,100.00  MyBaits (Custom Bait Libraries, 48 captures, MYcroarray) $5,760.00 |
| Sequencing and analysis (**$47,200**)  Whole genome sequencing budget (based on prices from Axeq Technologies):  Paired-end libraries (200-500bp insert size) construction: $300.00 each x 3 = $900.00  5kb mate pair library construction: $600.00 each x 1 = $600.00  Illumina HiSeq2000 sequencing: $1980.00 per lane x 3 = $5,940.00  Genome assembly using ALLPATHS-LG: $1000.00 x 1 = $1,000.00  Subtotal total per genome: $8,440.00  **subtotal for both *Lynx* genomes: *$16,880.00***  Transcriptome sequencing, both species**:  *$20,000.00***  Genome re-sequencing budget (based on prices from Axeq Technologies):  Paired-end libraries (200-500bp insert size) construction: $300.00 each x 8 = $2,400.00  Illumina HiSeq2000 sequencing: $1980.00 per lane x 4 = $7,920.00  **subtotal: *$10,320.00*** |
| Travel (**$37,632.48**)  **Tanya to Maine $8406.94**  1 week x3 per year x 3 years =  Lodging Bangor, ME $83 x 7 = *$581*  Travel Hadley MA to Bangor, ME (POV) $0.57/mile x 309.6mi = $176.47  x 2 = *$352.94*  Subtotal per visit: *$933.94*  TOTAL: $933.94 x 3 visits x 3YRs = **$8406**  **Tanya to TBD - lab work e.g. to Tucson, AZ $8402.96**  5 weeks (YR1) + 5 weeks (YR2)  Travel Hadley MA to BDL (POV) $0.57/mile x 39.9 = $22.74  x 2 = *$45.48*  Travel BDL to e.g. Tucson AZ maximum round trip = *$656*  Transportation in Tucson  Lodging Tucson maximum lodging $100 x 7 = $700/ week  x 5 weeks = *$3500*  Subtotal per visit: *$4201.48*  TOTAL: $4201.48 x 2 = **$8402.96**  **Tanya to DC - meetings with Warren & John $11900.88**  1 week x 2 per year x 3 year =  Travel Hadley MA to BDL (POV) $0.57/mile x 39.9 = $22.74 \*2 = *$45.48*  Travel BDL to Reagan (DCA) maximum round trip = *$335*  Lodging DC maximum lodging $229 x 7 = *$1603*  Subtotal per visit: *$1983.48*  Subtotal per year: $1983.48 \* 2 = *$3966.96*  TOTAL: $3966.96 X 3 YR = **$11900.88**  **John, Warren and Jen to UMass $6065.82**  1 per year x 3 days x 3 years  Rate 106/56  **JEN to UMass**  Travel Bangor, ME to Hadley, MA (POV) 0.57/mile x 309.6mi = $176.47 \*2 = *$352.94*  Lodging Hadley, MA $106 x 3 = *$480*  Per diem $56 x 3 = *$168*  Subtotal Jen per visit: *$1000.94*  TOTAL JEN: $1000.94 x 3YR = **$3002.82**  **WARREN to UMass**  Travel Dulles to BDL maximum round trip *$373*  Transportation  Lodging Hadley, MA $106 x 3 = *$480*  Per diem $56 x 3 = *$168*  Subtotal Warren per visit: *$1021*  TOTAL WARREN: $1021 x 3 = **$3063**  **JOHN to UMass**  John USGS funded travel    **Tanya, Steve, Todd, John &Warren to Maine $2855.88**  3 days x YR1  Rate 83/46  **TANYA to Maine**  Lodging Bangor, ME $83 x 3 = *$249*  Travel Hadley MA to Bangor, ME (POV) $0.57/mile x 309.6mi = $176.47  x 2 = *$352.94*  TOTAL TANYA: **$601.94**  **STEVE to Maine**  Lodging Bangor, ME $83 x 3 = *$249*  Travel Hadley MA to Bangor, ME (POV) $0.57/mi x 309.6mi = $176.47  x 2 = *$352.94*  Per diem $46 x 3 = *$138*  TOTAL STEVE: x YR1 = **$739.94**  **TODD to Maine**  Lodging Bangor, ME $83 x 3 = *$249*  Travel Hadley MA to Bangor, ME (POV) $0.57/mi x 309.6mi = $176.47  x 2 = *$352.94*  Per diem $46 x 3 = *$138*  TOTAL TODD: x YR1 = **$739.94**  **JOHN to Maine**  John USGS funded travel  **WARREN to Maine**  Travel Dulles to BGR maximum round trip *$370*  Transportation in BGR  Lodging Bangor, ME $83 x 3 = *$249*  Per diem $46 x 3 = *$138*  TOTAL WARREN: x YR1 = **$757** |
| Conference Travel (**$1,800**) |
| Publishing (**$1,500**) |
| **Total Direct Cost: $214,727.48** |

**References**

Allendorf FW, Leary RFm Spruell P, Wenburg JK (2001) The problems with hybrids: Setting conservation guidelines. *Trends Ecol. Evol*., **16**, 613-622.

Altschul SF, Gish W, Miller W, Myers EW,Lipman DJ. 1990. Basic local alignment search

tool. J. Mol. Biol.215: 403-410.

Aubry KB, Koehler GM, Squires JR (2000) Ecology of Canada lynx in Southern Boreal Forests. In: *Ecology and Conservation of Lynx in the United States* (eds. Ruggiero LF et al.) pp. 83-100. University Press of Colorado, Coulder, Colorado, USA.

Bean MJ, Principal Deputy Assistant Secretary for Fish and Wildlife and Parks. *Federal Register*, Dated: August 27, 2014. [FR Doc. 2014–21013 Filed 9–11–14; 8:45 am] BILLING CODE 4310-55-C

Bedoya-Reina OC, Ratan A, Burhans R, Kim HL, Giardine B, Riemer C, Li Q, Olson TL,

Loughran TP, vonHoldt BM, Perry GH, Schuster SC, Miller W. 2013. Galaxy tool to study

genome diversity. GigaScience2: 17.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. J.

Mol. Biol.268: 78-94.

Buskirk SW, Ruggiero LJ, Krebs CJ (2000) Habitat fragmentation and interspecific competition: Implications for lynx conservation. In: *Ecology and Conservation of Lynx in the United States* (eds. Ruggiero LF et al.) pp. 83-100. University Press of Colorado, Coulder, Colorado, USA.

Carmichael L, Clark W, Strobeck C (2000) Development and characterization of microsatellite loci from lynx (*Lynx canadensis*), and their use in other felids. *Mol. Ecol*., **9,** 2197 – 2199.

Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene Flow in Complex Landscapes: Testing Multiple Hypotheses with Causal Modeling. *The Amer Naturalist*, **168**, 486-499.

Frankham, R (2010) Challenges and opportunities of genetic approaches to biological conservation. *Biol. Conserv*., **143**,1919-1927.

Godoy JA (2010) La genomica y la conservacion del lince iberico. *Lychnos*: pp.14-8.

Goecks, J, Nekrutenko, A, Taylor, Jand The Galaxy Team. 2010. Galaxy: a comprehensive

approach for supporting accessible, reproducible and transparent computational research in

the life sciences. Genome Biol 11:R86.

Goudet J, Raymond M, de Meeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetic Soc of America*, **144**, 1933 – 1940.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Stat Software,* **22**, 1 – 19.

Hagan JM, Irland LC, Whitman AA (2005) Changing Timberland Ownership in the Northern Forest and Implications for Biodiversity. Manomet Center for Conservation Sciences, Report #MCCS-FCP-2005-1, Brunswick, ME.

Hoffman AA, Sgro CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479-485.

Oleksyk TK, Smith MW, O’Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philos T Roy Soc B*. **365**, 185-205.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. *Trends Genet.,* **26**, 177-187.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin

R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map

(SAM) format and SAMtools. Bioinformatics25: 2078-9.

Luikart G, England PR, Tallmon DA, Jordan S, Taverlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.,* **4**, 981-994.

McKelvey KS, Aubry KB, Ortega YK (2000) History and distribution of lynx in the contiguous United States. In: *Ecology and Conservation of Lynx in the United States* (eds. Ruggiero LF et al.) pp. 207-264. University Press of Colorado, Boulder, Colorado, USA.

Menotti-Raymond M, David VA, Lyons LA, Shaffer AA, Tomlin JF, Hutton MK, O’Brien SJ, Barton N (2001) A Genetic Linkage Map of Microsatellites in the Domestic Cat (*Felis catus*) The role of hybridization in evolution. *Mol. Ecol*., **10**, 551-568

Murray DL, Steury TD, Roth JD (2010) Assessment of Canada Lynx Research and Conservation Needs in the Southern Range: Another Kick at the Cat. *Jour Wildlife Mgmt*, **72**, 1463 – 1472

Oleksyk, TK, Pombert J-F, Siu D, Mazo-Vargas A, Ramos B, Guiblet W, Afanador Y, Ruiz Rodriguez CT, Nickerson ML, Logue DM, Dean M, Figueroa L, Valentin R, Martinez Cruzado J-C. 2012. A locally funded Puerto Rican parrot (Amazona vittata) genome sequencing project increases avian data and advances young researcher education. *GigaScience,* 1: 14.

Ranz JM, Machado CA (2006) Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol*., **21**, 29-37.

Row JR, Gomez C, Koen EL, Bowman J, Murray DL, Wilson PJ (2012) Dispersal promotes high gene flow among Canada lynx populations across mainland North America. *Conserv. Genet*.,**13**, 1259-1268.

Schwartz MK, Mills LS, McKelvey KS, Ruggiero LF, Allendorf FW (2002) DNA reveals high dispersal synchronizing the population dynamics of Canada lynx. *Nature***, 415**, 520-522.

Schwartz MK, Pilgrim KL, McKelvey KS, Lindquist EL, Claar JJ, Loch S, Ruggiero LF (2004) Hybridization between Canada lynx and bobcats: Genetic results and management implications. *Conserv. Genet*., **5**, 349-355.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics Soc of America*, **139**, 457 – 462.

Slough BG, Mowat G (1996) Population dynamics of lynx in a refuge and interactions between harvested and unharvested populations. *J. Wildlife Manage*., **60**, 946- 961.

Steiner CC, Punam AS, Hoeck PEA, Ryder OA (2013) Conservation Genomics of Threatened Animal Species. *Annual review of Animal Biosciences*, **1**, 261-281.

Van Oosterhout C, Hutchinson WF, Willd DPM, Shipley P (2004) Micro- Checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes,* **4**, 535 – 538.

Ward RMP, Krebs CJ (1985) Behavioral responses of lynx to declining snowshoe hare abundance. *Can. J. Zool*., **63**, 2817 – 2824.

Weir BS, Cockerham CC (1984) Estimating F- Statistics for the Analysis of Population Structure. *Evolution,* **38**, 1358 – 1370.

Werdelin L (1981) Evolution of lynxes. *Ann. Zool. Fenn*., **18**, 37-71.