XSEDE New Research Allocation Renewal:

Harnessing cloud computing and genomics to facilitate adaptive species conservation of non-model organisms

Abstract

Conservation of endangered species relies on accurate partitioning of biodiversity into units which are demographically or evolutionarily independent from one another. Utilizing genetic data gives conservation scientists unprecedented resolution in examining how the evolutionary process is realized at various spatial scales, allowing the delimitation of boundaries between species, and among units within species. However, as our ability to assay increasingly large portions of the genome allows greater resolution, applications is simultaneously hampered by the lack of an appropriate analytical framework, as well as difficulties in bridging the gap between traditional conservation genetics − where previously workloads weighted heavily towards field and bench work − to the computationally intensive landscape of today, driven by tectonic developments in next-generation sequencing and genomics. Traditional High-Performance Computing (HPC) environments, being designed for highly scalable applications, don’t provide an optimal space for our analytical needs, which tend to have high memory requirements and runtimes, with limited scalability. Additionally, unfamiliarity with both developers and practitioners in the field limits adoption. In both regards, the utility of the Jetstream cloud system is particularly evident, in that it meets our hardware needs within an environment which is user-friendly and flexible, whereas traditional HPC is often static. In this allocation renewal request, we hope to continue our earlier work in fleshing out computational approaches to big data in conservation genetics, as well as extend these methods to numerous case studies, with large collaborations including national and state wildlife and natural resource departments, as well as international conservation organizations and management agencies. Our specific research objectives are to: 1) Evaluate genetic delimitations and proposed taxonomies within species of management concern, with a case study in Speckled dace (*Rhinichthys osculus*) in the desert American Southwest; 2) Investigate ‘blurred’ boundaries between biodiversity units in association with anthropogenic habitat degredation in endangered Humpback Chub (*Gila cypha*) of the Grand Canyon and Colorado River; 3) Apply genetic data to model connectivity and enhance management action to control spread of Chronic Wasting Disease in Whitetail deer (*Odocoileus virginianus*); 4) Inform responsive economic development in Bhutan by assessing migratory habits of a recreational and local spiritual icon, the Golden Mahseer (*Tor putitora*); 5) Understand environmental and clade-level traits driving species diversification in the Ozarks, a region harboring many endangered and locally endemic aquatic species; and 6) Understand adaptive evolution in invasive species, particularly with regard to the potential for colonization of parasites into native species. As a cumulative effort across projects, we aim to continue efforts to advance the field of conservation genetics and encourage adoption of cloud-computing resources by continued software development porting established methodology to large datasets and scalable environments, fostering new academic and governmental collaborations, and in training the next generation of conservation scientists with the many dissertations this allocation supports.