A new approach to identify genetic boundaries of species could also impact policy

May 13 2021

The California gnatcatcher was at the center of a decades long fight because it was designated as a threatened species. A new approach to genomic species delineation put forth by evolutionary biologists led by Jeet Sukumaran at SDSU could impact policy and lend clarity to legislation for designating a species as endangered. Credit: Tom Benson, CC BY-NC-ND 2.0, https://www.flickr.com/photos/40928097@N07/48246028557

A new approach to genomic species delineation could impact policy and lend clarity to legislation for designating a species as endangered or at risk.

The coastal California gnatcatcher is an unassuming little gray songbird that's been at the epicenter of a legal brawl for nearly 28 years, ever since U.S. Fish and Wildlife Service listed it as threatened under the Endangered Species Act.

Found along the Baja California coast, from down south in El Rosario, Mexico to Long Beach, Calif., its natural habitat is the rapidly declining coastal sagebrush that occupies prime, pristine real estate along the West Coast. When this particular gnatcatcher, Polioptila californica, was granted protection, the region's real estate developers went to court to get it delisted.

Central to their argument, which was dismissed in a federal court, was whether it was an independent species or just another population of a more widely found gnatcatcher. This distinction would dictate its threatened status. Evolutionary biologists have developed a new approach to genomic species delineation that improves upon current methods and could impact similar policy in the future.

This approach is based on the fact that in many groups of organisms it can be problematic to decide where one species begins and another ends.

"In the past, when it was challenging to distinguish species based on external characters, scientists relied on approaches that diagnosed signatures in the genome to identify 'breaks' or 'structure' in gene flow indicative of population separation. The problem is this method doesn't distinguish between two populations separated geographically versus two

populations being two different species," said Jeet Sukumaran, computational evolutionary biologist at San Diego State University and lead author of a study published May 13 in *PlOS Computational Biology*.

"Our method, DELINEATE, introduces a way to distinguish between these two factors, which is important because most of the natural resources management policy and legislature in our society rests on clearly defined and named species units."

Typically, scientists will use a range of different methods to identify boundaries between different species, including statistical analysis and qualitative data to distinguish between population-level variation and species-level variation in their samples, to complete the classification of an organism. In cases where it is difficult to sort the variation between individuals into differences due to variation within a species as opposed to between two species, they often turn to genomic data based approaches for the answer. This is when scientists often use a model that generates a population phylogeny, or an evolutionary tree relating different populations.

Sukumaran and co-authors evolutionary biologists L. Lacey Knowles with the University of Michigan, Ann Arbor and Mark Holder with the University of Kansas, Lawrence add a second layer of information to the population phylogeny, to explicitly model the actual speciation process. This allows them to understand how these separate populations sometimes evolve into distinct species, which is the basis for distinguishing between populations and species in the data.

Whether some of the population lineages in the sample are assigned to existing species or classified as entirely new species depends on two factors. One is the age of the population isolation events such as the splitting of an ancestral population into multiple daughter populations, which is how species are "born" in an extended process of speciation.

The other is the rate of speciation completion, which is the rate at which the nascent or incipient species "born" from population splitting events develop into true full species.

"We're coming to realize now that many organisms are cryptic species," Sukumaran said. "Many of them are similar looking even though they are actually distinct species separated by many tens or hundreds of thousands or even millions of years of evolution."

This is either due to strong selection pressures to maintain the same morphology, or, more typically, due to very recent speciation resulting in insufficient time for external differences to develop.

"When rivers change course, when terrain changes, previously cohesive populations get fragmented, and the genetic makeup of the two separate populations, now each a population in their own right, can diverge," Sukumaran said. "Eventually, one or both these populations may evolve into separate species, and may (or may not) already have reached this status by the time we look at them.

"Yet individuals of these two populations may look identical to us based on their external appearances, as differences in these may not have had time to 'fix' in either population. This is when we turn to genomic data to help guide us toward deciding whether we are looking at two populations of the same species, or two separate species."

Currently, scientists apply a model based on multispecies coalescent theory to genomic data to identify the disruption of gene flow between different groups of organisms. This disruption is fundamental to species formation, but it can also occur between two different populations as well as two different species.

While scientists agree that it is critical to distinguish between

populations and species boundaries in genomic data, there is not always a lot of agreement on how to go about doing it. "If you ask ten biologists, you will get twelve different answers," Sukumaran said.

By modeling the dynamics of speciation itself in the species delimitation analysis, something previous methods did not do, the researchers' approach allows for distinguishing between interpopulation boundaries to gene flow versus interspecies boundaries, based on the predicted pace of the speciation events.

With this framework, scientists can have a better understanding of the status of any species, but especially of species that are members of a species complex—multiple independent species that all look alike.

Many fields of science and medicine depend on the accurate demarcation and identification of species, including ecology, evolution, conservation and wildlife management, agriculture and pest management, epidemiology and vector-borne disease management etc. These fields also intersect government, legislature and policy, with major implications for the day-to-day lives of broader human society.

The DELINEATE model is a first step in a process that will need to be further refined. Funding for this research came from the National Science Foundation.

More information: *PlOS Computational Biology* (2021). journals.plos.org/ploscompbiol … journal.pcbi.1008924

Provided by San Diego State University

Citation: A new approach to identify genetic boundaries of species could also impact policy