Problem Statement

Using non-invasive genetic samples for population estimation is a promising methodology. Not only is it often easier to obtain the necessary samples, but the use of genetic tags can remove many of the potential biases associated with trapping organisms. We believe that these advantages will lead to the dominance of genetically based capture/mark/recapture (CMR) population estimation for many species. Genotyping errors, however, can lead to mistakes in identification, and these mistakes can lead to large biases in abundance estimates. Genotyping errors cause abundance to be biased high, as much as 5.5 fold, based on published data. While there are several methods available for removing genotyping errors, they must be removed efficiently, and the resulting data must be shown to be error free.

Research Approach

Our program goal was to explore the probabilistic nature of the errors, and look for characteristic patterns associated with errors. These patterns, once identified, could be both targeted and eliminated, providing both an efficient approach to error removal and a strong test to determine error elimination.

Results

We determined that with a sufficiently large tag (enough loci) the genetic differences between individuals differed formed single mode, and that few organisms were either identical at most loci, or differed at all loci. An error free sample would have a minimum distance distribution that was composed of organisms that differed at an intermediate number of loci, and organisms that were identical at all loci (recaptures). With error, however, many organisms appear differ at only a few loci. This leads to a bimodal distribution that is a signature of a sample containing genotyping errors. Those samples that comprise the lower mode likely contain errors, and these errors can therefore be efficiently targeted and removed. Secondly, with a sufficiently large tag, either altering the tag size and composition should have no effect on the capture history. However, with genotyping errors, changing the tag size and composition will produce different numbers of individuals, depending on whether the errors are included in the tag. If including a particular locus in the tag results in a large number of new individuals, then this locus likely has a high error rate and the locus can be removed. This second understanding, that without errors organism identification should be insensitive to tag composition, provides a strong test demonstrating error removal. Together, these two tests facilitate the production of defensible and cost effective abundance estimates using non-invasive genetic tags.

Partners

This project is a partnership between the USDA FS - Rocky Mountain Research Station and the Idaho Department of Fish and Game.