

Humpback Chub Genetics

Project Description:

Life history of *G. cypha* in the Colorado River of Grand Canyon (GC) is mostly enigmatic and interrelationships among subpopulations are virtually unknown. Lack of an historic baseline further complicates understanding of present-day patterns, and causal relationships between physical and biological parameters are merely the source of speculation. The most pressing questions pertain to genetic distinctiveness of aggregations in the mainstem Colorado River (MCR), the interrelationships among these and tributary populations, and how the sum can be adaptively managed in a dam-perturbed environment.

Objectives of the study are to infer interrelationships among populations of *G. cypha*, to identify (if possible) genetically distinct units, and to derive a management strategy for this endangered species. Primary focus is on genetic interrelationships among aggregates (populations) of *G. cypha* within GC. To gain perspective on basin-wide intra-specific relationships of the species, four populations from the Upper Colorado River basin are also included in the study.

A combination of molecular markers is employed to investigate genetic relationships within- and among-populations of *G. cypha*. Mitochondrial (mt) DNA sequence data from two regions of that molecule allow identification of phylogenetic lineages within GC and elsewhere in the basin. Genetic variation is also analyzed on a finer scale within- and among-populations by evaluating 20 faster evolving microsatellite DNA loci to assess genetic structure and levels of gene flow among populations. By contrasting results from different molecular markers, both recent and historic population events can be inferred. Further, estimates for population parameters such as N_e (effective population size), and N_m (number of migrants) can be explored.

Preliminary Results & Status:

Mitochondrial DNA sequence data have been generated for GC aggregates and upper basin populations and are currently being analyzed. While sample sizes for the LCR and Randy's Rock (both Grand Canyon) as well as for Desolation, Black Rocks and Westwater canyons are sufficient, small and varying sample sizes for most GC aggregates and the Yampa River population complicate analyzes and require appropriate procedures so as to avoid spurious results caused by unequal sample sizes. Preliminary analyses of mtDNA sequences reveal that the majority of Humpback Chub haplotypes are basin- or population-specific, supporting an hypothesis of reduced gene flow among basins. However, shared ancestral polymorphism also indicates that lineage sorting is incomplete within these populations. More detailed analyses are needed to assess historic and contemporary levels of gene flow.

Microsatellite DNA data have been generated for Grand Canyon and upper basin populations across 20 nuclear loci. Preliminary analyses reveal high allelic diversity and considerable heterozygosity (a surprising, but very positive finding). However, high levels of microsatellite polymorphism underscores the need for large (i.e., 50-100 individuals /

population) and balanced sample sizes. Since the latter cannot be accommodated with a reasonable sampling effort, statistical protocols must again be employed to avoid spurious results. Again, this complicates analyses and requires additional time.

Further, allele patterns are also very complex and require substantial time to be scored consistently and correctly, using specific analytical protocols designed to deal with that kind of variation. To facilitate allele identification and reduce ambiguity in their allocation, microsatellite loci with tetranucleotide repeats (i.e., the microsatellite motif consist of four repeated nucleotides) were selected rather than commonly used dinucleotide repeats. Due to the larger differences between alleles (four instead of two base pairs), tetranucleotide repeats are easier (and thus more reliable) to score. However, these tetranucleotide repeats do not always mutate in a consistent manner (i.e., by adding or losing four base-pairs), but instead may vary in three, two or even one base-pair. This is problematic in that many algorithms that analyze microsatellite data assume a step-wise mutation model, where the same number of base pairs is added or deleted in mutational steps. This confusion adds additional complexity to the analysis and often requires sequencing of these alleles to determine the order of base pairs.

Reasons for Extension & Timelines:

Various technical problems hampered progress over the project duration (e.g., delay in laboratory renovations, equipment failure, optimization of protocols for specific populations, etc.). These, in combination with complicated allele patterns, difficulties in scoring genotypes consistently, coupled with complexity of statistical analyses have increased time needed to complete particular steps of the project. In addition, molecular genetic data differ from other data types insofar as conclusions should only be drawn from complete data sets. Preliminary analyses might reveal tendencies, but often these 'tendencies' must be revised once complete datasets have been analyzed. For example, to assess whether mainstem populations differ in allele frequencies, sufficient numbers of individuals must be examined for their statistical robustness. Findings based on too few individuals or loci might reveal differences that are not substantiated once complete datasets are analyzed.

Given the relevance of these genetic analyses for management decision on Humpback Chub populations, PIs elected not to reveal preliminary findings to avoid premature conclusions and precipitation of inappropriate management actions. PIs thus requested an extension for the final report on 'Humpback Chub Genetics.' Analyses and report compilation are expected to be completed within one year from original project end (1 January 2005). Thus findings should be available no later than 1 January 2006.

Respectfully submitted,

Marlis R. & Michael E. Douglas
Colorado State University