# MitoGeneExtractor: Efficient extraction of mitochondrial genes from next generation sequencing libraries

Marie Brasseur[1], Jonas Astrin[2], Matthias Geiger[3], and Christoph Mayer[2]

[1]Leibniz Institute for the Analysis of Biodiversity Change
[2]Zoological Research Museum Alexander Koenig
[3]Zoologisches Forschungsinstitut und Museum Alexander Koenig

April 4, 2022

## Abstract

Mitochondrial DNA sequences (mtDNA) are often found as byproduct in hybrid enrichment data sets originally created to capture anchored hybrid enrichment (AHE) or ultra-conserved element (UCE) nuclear loci. The mtDNA sequences in these data sets are currently rarely used, even though mitochondrial genes such as COI, ND5, CytB, and 16S are of general interest and often not yet known and deposited in public databases. We developed MitoGeneExtractor to extract mitochondrial genes of interest from genomic libraries. Gene sequences are reconstructed through multiple sequence alignments of sequencing reads to an amino acid reference. We applied MitoGeneExtractor to recently published data created for UCE enrichment and were able to extract complete or nearly complete COI and ND5 sequences for a large proportion of the sequencing libraries. MitoGeneExtractor can be used to extract mitochondrial protein coding genes from a wide range of next generation sequencing data sets.

## Hosted file

MitoGeneExtractor_MB_MFG_JA_CM.pdf available at https://authorea.com/users/473852/articles/563896-mitogeneextractor-efficient-extraction-of-mitochondrial-genes-from-next-generation-sequencing-libraries