Pipeline Proposal

Overview
1. Find potential dietary hits
2. Of those hits, remove any which may align to microbes (contamination)
3. Of remaining hits, get aDNA authentication
4. Phylogenetic precision score per taxon identification (missing in DB?)
   - Similar to PIA but don't discard just penalise

Output

Sample	Taxon	Reads	Damage	Length	Phylo. Preci. Score

Run alongside
Negative controls for
comparison?

e.g. Chordata + Mollusca

Dietary Target Database (e.g.)

Input Database(s) should be cleaned against Steinegger etc and Salzberg

Already clipped/ merged

fastq

Off-Target Taxa Screening
MALT
MiniMap2 + sam2lca

bam

fastq

Remove reads with possible microbial sources
samtools

bam

Generate NCBI Accession to Taxonomy Map

text

aDNA Authentication
DamageProfiler

Off-Target Taxa Screening
MALT
MiniMap2 + sam2lca

Microbial Database (e.g.)

text

Extract Reads IDs
samtools / unix

Split BAM per Taxon Node

bam

Identification Precision
Phylogenetic Diversity of subhits

aDNA Authentication
pydamage