Subsample amino acid k-mers from each transcriptome

Species A transcriptomes

Species B transcriptomes

Input reads

- Nucleotide
- Protein

Protein-coding

Noncoding

Convert to reduced alphabet

Decompose to k-mers

Reduced amino acid alphabet k-mers

Cluster on *k*-mer similarity

Manual
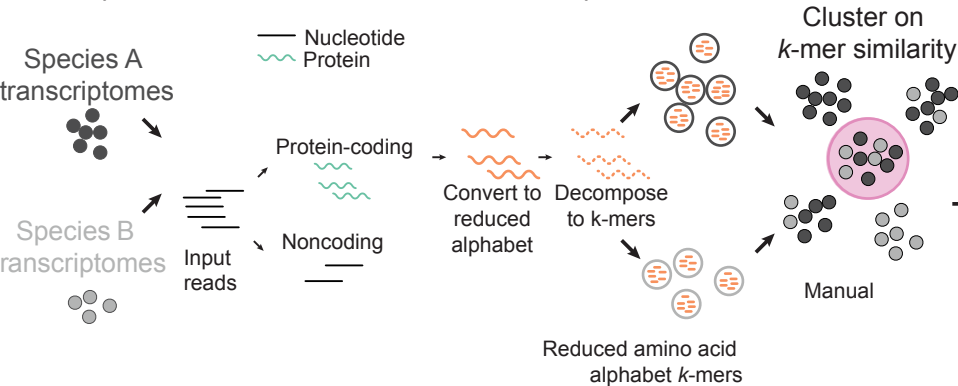
Perform differential k-mer expression and query for functional annotation

Differential k-mer expression within cluster, across species

k-mer abundance

Cluster   Rest

Cluster-enriched differential *k*-mers

In reference genome

Not in reference genome*

Not in an annotated gene*

In annotated gene

Unknown homolog*

Known homolog

RefSeq

Query mystery k-mers for functional annotation

*Putative homologs not identifiable by traditional methods