Solution

-

-

-

-Move empirical data

Carry

-Combine fingerprints

Repeat for multiple fingerprints

Current activity

-Find efficient 2D mapping: nonmetric

multidimensional scaling (NMDS)

-Develop fast distance computation algorithms: pre clustering plus hasing

-develop accuracy statistics: perturbation analysis

-Determine how many “registration sequences” are best for 20 million

empirical points

-Precompute registration libraries for different sample types

-Determine accuracy with simulation and known sequences

Future work

-parallelize: use reference triangles

-What to do with outliers?

-What do cluster shapes/density say about fingerprints, species,

ecology

\*which fingerprints are good for which species?

\*which fingerprints are most misleading in given environments?

-How are clusters and evolution related?

-Application to empirical data: milk project

-…