

TITLE: TreeParser-aided Klee Diagrams Display Taxonomic Clusters in DNA Barcode and Nuclear Gene Datasets

AUTHORS: Mark Y. Stockle and Cameron Coffran

SUPPLEMENTARY INFORMATION

Fig. S1. Comparison of TreeParser output with template NJ tree for *A. fulgerator* dataset.

NJ tree is shown at left and labeled columns list sequences as they appear in NJ tree and TreeParser output file. Sequences are colorized by species; those that differ in order between NJ tree and TreeParser output file are marked at far right. Lines between columns connect sequences with non-identical positions. All cases reflect topology-equivalent branch rotations within species.

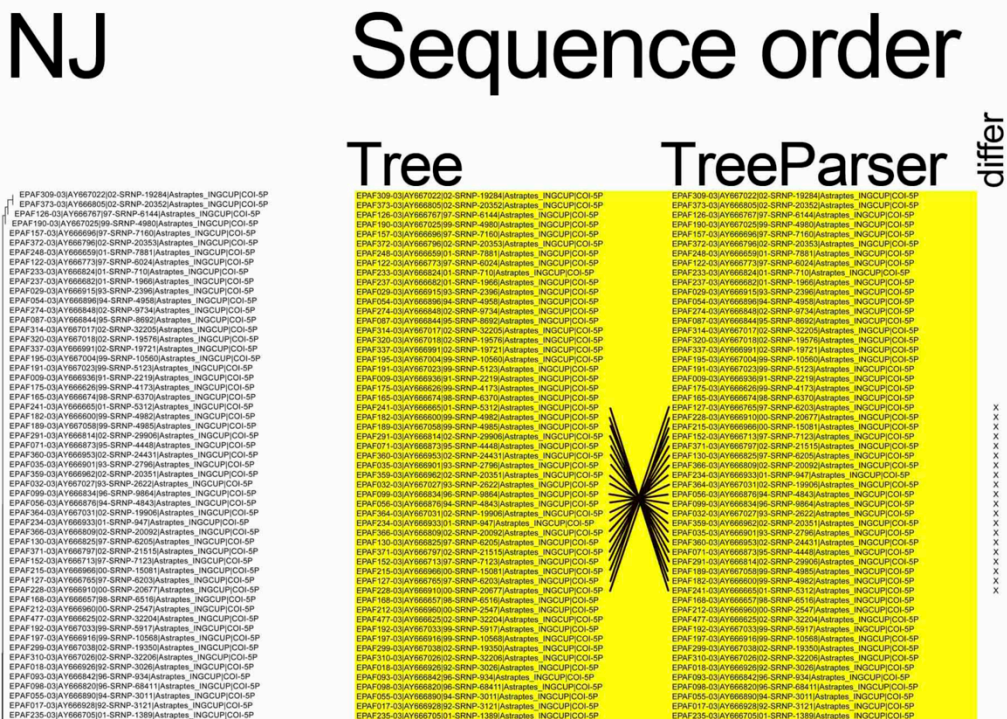
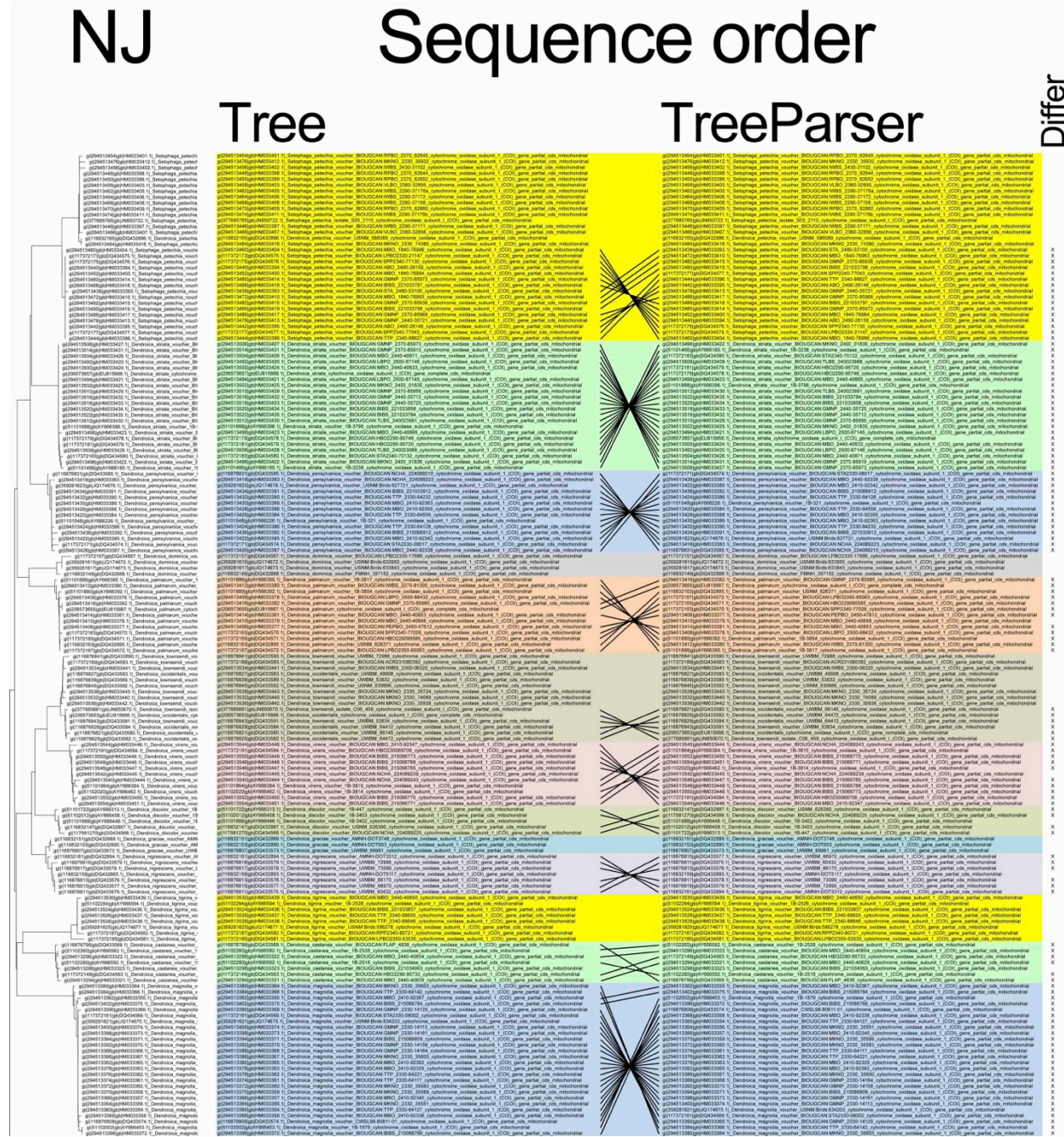


Fig. S2. Comparison of TreeParser output with template tree for *Setophaga* warbler dataset. Figure is organized as described in Fig. S2 legend. Lines between columns connect sequences with non-identical positions. All cases reflect topology-equivalent branch rotations within species.



Differ

Fig. S3. Large dataset avian COI Klee diagram. A published alignment of avian COI BARCODEs [1] was downloaded from the journal website. DNAdist and Neighbor programs in Phylip software package [2] were applied to generate a K2P NJ text file using a high performance computer (HP ProLiant DL380/G5 64-bit Linux server with dual-quad core Intel/Xeon 3.16 GHz processors, 64 GB RAM). The alignment and tree files were uploaded to TreeParser, and the resulting rearranged alignment was filtered to exclude short sequences (minimum length positions 100 to 600). The final file contained 11,249 sequences representing 2,679 species. A Klee diagram was generated with parameters $n = 1$ sequence/vector and bp window = 100-600 using high performance machine. Larger blocks of higher correlation corresponding to taxonomic divisions are labeled.

REFERENCES

1. Stoeckle, M.Y. & Kerr, K.C.R. Frequency matrix approach demonstrates high sequence quality in avian BARCODEs and highlights cryptic pseudogenes. *PLoS ONE* **78**, e43992 (2012).
2. Felsenstein, J. PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166 (1989).

