Supporting Information for Online Publication for:

Similarity thresholds used in short read assembly reduce the comparability of population histories across species

Michael G. Harvey, Caroline Duffie Judy, Glenn F. Seeholzer, James M. Maley, Gary R. Graves, and Robb T. Brumfield

**Table S1.** Sample information for individuals used in this study.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Population** | **Genus** | **Species** | **Subspecies** | **Museum1** | **Tissue #** |
| 1 | 1 | *Cranioleuca* | *antisiensis* | *antisiensis* | LSUMZ | 67126 |
| 2 | 1 | *Cranioleuca* | *antisiensis* | *antisiensis* | LSUMZ | 67136 |
| 3 | 1 | *Cranioleuca* | *antisiensis* | *antisiensis* | LSUMZ | 67177 |
| 4 | 1 | *Cranioleuca* | *antisiensis* | *antisiensis* | LSUMZ | 67243 |
| 5 | 2 | *Cranioleuca* | *baroni* | *capitalis* | LSUMZ | 72753 |
| 6 | 2 | *Cranioleuca* | *baroni* | *capitalis* | LSUMZ | 72754 |
| 7 | 2 | *Cranioleuca* | *baroni* | *capitalis* | LSUMZ | 72756 |
| 8 | 2 | *Cranioleuca* | *baroni* | *capitalis* | LSUMZ | 72759 |
| 9 | 1 | *Rallus* | *crepitans* | *saturatus* | LSUMZ | 63400 |
| 10 | 1 | *Rallus* | *crepitans* | *saturatus* | LSUMZ | 63404 |
| 11 | 1 | *Rallus* | *crepitans* | *saturatus* | LSUMZ | 63475 |
| 12 | 1 | *Rallus* | *crepitans* | *saturatus* | LSUMZ | 63477 |
| 13 | 2 | *Rallus* | *elegans* | *elegans* | LSUMZ | 63464 |
| 14 | 2 | *Rallus* | *elegans* | *elegans* | LSUMZ | 63467 |
| 15 | 2 | *Rallus* | *elegans* | *elegans* | LSUMZ | 63471 |
| 16 | 2 | *Rallus* | *elegans* | *elegans* | LSUMZ | 63473 |
| 17 | 1 | *Trochilus* | *polytmus* | *polytmus* | USNM | 4134 |
| 18 | 1 | *Trochilus* | *polytmus* | *polytmus* | USNM | 4136 |
| 19 | 1 | *Trochilus* | *polytmus* | *polytmus* | USNM | 4147 |
| 20 | 1 | *Trochilus* | *polytmus* | *polytmus* | USNM | 4160 |
| 21 | 2 | *Trochilus* | *polytmus* | *scitulus* | USNM | 3914 |
| 22 | 2 | *Trochilus* | *polytmus* | *scitulus* | USNM | 3915 |
| 23 | 2 | *Trochilus* | *polytmus* | *scitulus* | USNM | 3931 |
| 24 | 2 | *Trochilus* | *polytmus* | *scitulus* | USNM | 3935 |
| 25 | 1 | *Xenops* | *minutus* | *mexicanus* | KUMNH | 2044 |
| 26 | 1 | *Xenops* | *minutus* | *mexicanus* | LSUMZ | 60935 |
| 27 | 1 | *Xenops* | *minutus* | *littoralis* | LSUMZ | 2209 |
| 28 | 1 | *Xenops* | *minutus* | *littoralis* | LSUMZ | 11948 |
| 29 | 2 | *Xenops* | *minutus* | *obsoletus* | LSUMZ | 4244 |
| 30 | 2 | *Xenops* | *minutus* | *obsoletus* | LSUMZ | 6862 |
| 31 | 2 | *Xenops* | *minutus* | *obsoletus* | LSUMZ | 9026 |
| 32 | 2 | *Xenops* | *minutus* | *obsoletus* | FMNH | 433364 |
| *1 Museums are Field Museum of Natural History (FMNH), University of Kansas Natural History Museum (KUNHM), Louisiana State University Museum of Natural Science (LSUMZ), and Smithsonian Institution National Museum of Natural History (USNM)* |

**Table S2.** Results of Kruskal Wallis test for differences in number of alleles across different similarity thresholds.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **chi-squared value** | **df** | **p-value** |
| *Cranioleuca* | 2682.455 | 6 | 2.20E-16 |
| *Rallus* | 1007.52 | 6 | 2.20E-16 |
| *Trochilus* | 2301.125 | 6 | 2.20E-16 |
| *Xenops* | 15267.58 | 6 | 2.20E-16 |

**Table S3.** Proportion of loci containing putative under-split alleles (containing individuals with three or more alleles).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Similarity Threshold** | ***Cranioleuca*** | ***Rallus*** | ***Trochilus*** | ***Xenops*** |
| 99 | 0.0022 | 0.0020 | 0.0011 | 0.0023 |
| 98 | 0.0028 | 0.0028 | 0.0016 | 0.0031 |
| 97 | 0.0030 | 0.0028 | 0.0016 | 0.0033 |
| 96 | 0.0029 | 0.0029 | 0.0015 | 0.0033 |
| 95 | 0.0031 | 0.0029 | 0.0016 | 0.0033 |
| 94 | 0.0031 | 0.0030 | 0.0017 | 0.0034 |
| 93 | 0.0031 | 0.0033 | 0.0018 | 0.0035 |

**Fig. S1.** Adjusting minimum read depth used to form alleles in Stacks affects both the total number of bases in a dataset and the number of singleton alleles, the latter serving as a proxy for the frequency of errors. We selected a read depth of 7x for further analyses because, among the settings examined, it appeared to balance the inclusion of potential errors with the total size of the data matrix.

****

**Fig. S2.** Variance across the four study species in mean number of alleles and mean estimates of phylogeographic and population genetic parameters (aside from Fst) varies depending on the similarity threshold applied.



**Fig. S3.** Mean values of mutation-scaled effective population size (theta) of the two extant populations did not show an association with similarity threshold for any species. Circles and triangles are used to distinguish the two daughter populations from each species.

****

**Fig. S4.** Mean values of population divergence time (tau) between the two extant populations did not show an association with similarity threshold for any species.

