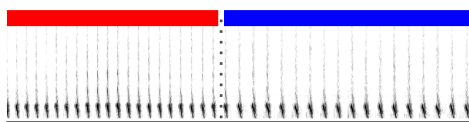


Figure 1 | Annotating large datasets of birdsong is a hard task

- The zebra finch repeating motif allows annotation by matching its template sonogram.
- Bengalese finch songs segmented to syllables.
- A third of one domestic canary song of median duration segmented to syllables reveals repetitions (phrase) structure.
- The median, 0.25 and 0.75 quantiles of song durations (x-axis) and # of syllables per song (y-axis) for 2 canary strains, Bengalese finches and Zebra finches (color coded)
- Variable songs are not suited for template matching. Songs contain repeating sequences of syllables but because of sequence variability songs with more syllables (x-axis) share smaller sequence fractions (y-axis)
- Distributions of syllable durations for one domestic canary. The bird had 20 different syllable types (x-axis, ordered by mean syllable duration). Box plot shows median, 0.25 and 0.75 quantiles of syllable durations. Whiskers show the entire range



Supplementary Panels:

Example of two consecutive canary phrases that differ mostly in inter-syllable gaps. These two elements necessitate combining the segmentation and annotation steps