**Quantifying and mitigating reference bias in conservation and population genomics**

The study of evolution and the conservation of species now depends on the sequencing of a diverse sample of species from across the tree of life. A common strategy in both population and evolutionary genomics, especially for non-model organisms, is to map the sequenced reads from new samples to a high-quality reference genome of a closely related species. However, mapping many samples with varying levels of divergence to a single reference can lead to reference bias, when the variation present in the reads prevents them from mapping. This means that the most divergent reads are not included in downstream analyses and subsequent measures of genomic variation such as heterozygosity, substitution rate, and positive selection may all be affected. Despite its wide-ranging influence on the conclusions drawn from genomic data, reference bias and its effects have not been exhaustively quantified. Here, we use simulated sequence reads of varying levels of divergence relative to a reference genome to show the extent of reference bias in modern sequencing data. We find that as divergence from the reference increases, the number of unmappable reads also increases linearly, from 10% of reads being unable to map at 2% divergence up to nearly 40% at 10% divergence. This has wide ranging implications for comparative analyses that aim to identify evolutionarily important mutations – much of this variation may be lost during mapping. We also explore various ways forward to capture the variation lost due to reference bias, such as with iterative mapping or assembly graphs.