**Germline and somatic mutational process across the Tree of Life**

The Darwin Tree of Life (DToL) project is an ambitious project that aspires to sequence and assemble high-quality reference genomes for 70,00 eukaryotic species in Britain and Island. The advent of high-throughput long-read sequencing and genome mapping technologies, improvements in the base accuracy of long reads and algorithms that leverages the longer read length and long-range genomic interactions enables the construction of complete and contiguous chromosome-length scaffolds at unprecedented cost and speed.

To date, the germline and somatic mutational processes in *Homo sapiens* has been studied to measure the germline and somatic mutation rate, to understand the clonal structure of normal tissues, dissect the driver mutation landscape of tumors and to lineage trace tumor and normal tissue development. The high cost of hierarchical shotgun sequencing and technical limitations of the next-generation sequencing platform for generation of high-quality reference genomes has prevented similar analysis in non-human samples, except for a small number of model organisms. Here, we leverage the extraordinary public resource from the DToL project to investigate the germline and somatic mutational processes of a number of eukaryotic species for the first time.

Somatic mutation detection from normal tissues either requires 1) amplification of mutagenic DNA through single-cell clone expansion or whole-genome amplification 2) or increase in base accuracy through combination of approaches that uses a unique molecular identifier, redundant sequencing of the template molecule and generation of a consensus sequence. We hypothesized that circular consensus sequence (CCS) reads from Pacific Biosciences (PacBio) might be as accurate or more accurate than duplex reads and might be suitable for single molecule somatic mutation detection.

To enable somatic mutation detection from normal tissues agnostic of clonality and species, we measure the CCS error rate, demonstrate that CCS base accuracy is sufficient to enable single molecule somatic mutation detection where a single read alignment supports the mismatch between the sample and reference genome, and develop a method to detect single molecule somatic mutation detection with high confidence using CCS reads. In addition, we show that our method is applicable in non-human samples with higher heterozygosity through the measurement of somatic mutation rate in *Phorcus lineatus* samples of different ages. We apply our method to approximately 500 eukaryotic species from the DToL project to identify somatic mutations from normal bulk tissue and to *de novo* extract mutational signatures unique to a species and common across species. We time the emergence of these mutational processes through phylogenetic tree analysis and use the mutational signatures to better understand the germline mutational process in each species.

The ability to detect somatic mutations from bulk normal tissue without arduous experiments should accelerate our understanding of somatic mutagenesis across the Tree of Life and enable early detection of tumors and monitoring of tumor development during patient treatment.