**Chapter 3: Germline and somatic mutational processes across the Darwin Tree of Life**

**Introduction**

The Tree of Life encapsulates biological entities with 5 billion years of history on Earth, extinct species, survivors, and their descendants. The genomes of a select number of species, deemed biologically important, have been sequenced and assembled [ref,ref,ref, Drosophila melanogaster, C.elegans, Zebrafish, Mouse]. Homo sapiens, as a matter of fact, is one leaf in the Tree of Life and an unknown number of leaves remains to be studied. The completion of the human genome project and ramifications of the human genome project is undoubtedly a monumental moment in human genomics, but we far from studying and understanding the question “What is Life?” “What constitutes Life on Planet Earth”.

Contracts, expands, fuses, inverts, rearranges, inserts, deletes, substitutes and copies and pastes, recombines, and the combination of all the above mechanisms to change the genome.

A number of factors has thwarted our efforts to understand species across the Tree of Life. These factors include the sequencing cost, read length, base accuracy, and computational costs, genome sequence complexity and ploidy.

The human genome project cost approximately 3 billion dollars, equivalent to dollar per base pair and required colossal effort requiring international collaboration across major sequencing institutions. Despite the gargantuan effort to physically map and assemble individual BAC clones, the human reference genome had missing sequences, unplaced and unlocalized scaffolds with unknown locations on the human reference genome. The p-arm of acrocentric chromosomes and centromeric sequences of every chromosome, for example, remains unassembled because of their highly repetitive sequence content. The centromeric sequence in the latest human reference genome grch38, hence, is modelled and is not a true representative of the underlying sequence. In addition, the palindromic sequences in chromosome Y makes chromosome Y particularly difficult to assemble and the high degree of similarity between chromosome X and chromosome Y because of X-degenerate and X-transposed sequences. The human reference genome required the advent of new sequencing technologies with higher base accuracy and longer read length to correct misassemblies and minor errors and a new generation of human reference genome [ref, ref].

Segmental duplications defined as non-repetitive sequences with >90% sequence homology between multiple copies makes de novo assembly an extremely difficult problem.

The whole-genome shotgun sequencing approach and assembly approach with Illumina short reads might be scalable

The human reference genome is undoubtedly the most accurate mammalian reference genome and required a colossal effort to generate BAC clones, to determine the location of BAC clones through physical mapping and determining the BAC clones for minimal tiling path generation.

There were initially two competing approaches for human genome construction: whole-genome shotgun sequencing by JCVI and minimum tiling path by the NCBI?

The advent of PacBio CCS and ONT sequencing has been a game-changer/monumental/pivotal moment

The use of Hi-C reads

thwarted

Inability to culture microbiome\*

The human genome project

Relatives in the tree of life

A subset of the branches in the trees of life

Select number of leaves on the tree of life has been studied

The study of other species has been limited by the lack of high quality assemblies, which in turn was limited by the sequencing cost and the technical limitations of the next-generation sequencing platform.

**Materials and Methods**

**CCS library preparation and sequencing**

**Phorcus lineatus preparation**

**CCS read alignment and germline and somatic mutation detection**

**HDP mutational signature extraction**

**Mutation signature analysis**

**Results**

**Germline and somatic mutations**

**TiTv ratio**

**Mutational signature analysis**

**Conclusion**

We discover XX number of mutational signatures previously undiscovered in previous studies.

**Discussion**

%% Samples without somatic mutations

%% Somatic theory of aging

%% Life cycle of Insects

%% Environmental mutagenesis

%% number of eukaryotic species from the DToL project

%% Peto's paradox

%% limited study of somatic mutational processes in other non-human samples: C-elegans and Drosophila melanogaster

%% limited study of germline mutational processes

%% TiTv ratio

%% SBS6 classification

%% Life cycle of animals

%% recipe for evolution: mutation, natural selection, speciation,

%% cold blood

%% environment, habitat

%% endogenous, exogenous processes

%% somatic mutation: DNA damage, repair and fixation %% clonality

%% germline and somatic mutational processes could be shared

%% Darwin Tree of Life project aims to sequence and assemble 66,000 species in Britain and Island

%% somatic theory of aging: somatic mutations increase with age and the accumulation of somatic mutations impairs cellular functions

Genetic components/modules for synthetic biology purposes and applications