South China Tiger Genome Review Paper

1. Study species:
   South China Tiger (*Panthera tigris amoyensis*)

2. Paper Title: "Chromosome-scale genomes reveal genomic consequences of inbreeding in the South China tiger: A comparative study with the Amur tiger"
   Journal: Molecular Ecology Resources
   Publication Date: February 2023
   Author List: Zhang L, Lan T, Lin C, Fu W, Yuan Y, Lin K, Li H, Sahu SK, Liu Z, Chen D, Liu Q, Wang A, Wang X, Ma Y, Li S, Zhu Y, Wang X, Ren X, Lu H, Huang Y, Yu J, Liu B, Wang Q, Zhang S, Xu X, Yang H, Liu D, Liu H, Xu Y.

3. Size of the haploid genome:
   Haploid genome size of the South China tiger is 2.5 Gb.

4. Type of genome assembly:
   De novo assembly was done using three types of genome assembly: PacBio long reads sequencing, Hi-C technology and Paired-ends short reads sequencing.

5. Summary of the genome assembly:

Genome Size	2.5 Gb
Number of scaffolds	2107
Longest Scaffold Length	238.2 Mb
Scaffold N50	145.6 Mb
Contig N50	9.47 Mb
GC Content	41.67%

6. Describe the strategy used to sequence the genome. Critically evaluate the genome sequencing strategy:

- A sample of a male South China Tiger (SCT) was used for genome assembly.
  Steps of Assembly:

- Sequencing Methods Used:

Three types of sequencing data were used: PacBio long reads, DNBSEQ short reads, and Hi-C sequencing data.

- PacBio long reads:
  Library preparation and sequencing was done, to generate long reads (average insert size - 20 kb). The short PacBio reads (<1000 bp) were filtered out and regions of overlap were identified in the remaining long reads. Read correction and trimming was done using the Canu software to obtain high quality sequences and contigs were generated along with draft genomes.

- Base Correction with Short Reads:
  Short reads of 100 bp were obtained from paired-end sequencing using DNBSEQ platform. They were used for base correction using the NEXTPOLISH software on the draft genomes obtained from PacBio sequencing. The draft genomes may have some errors or inaccuracies in individual bases, and base correction improves their accuracy and reliability. The paired end sequencing method is especially useful, as both the forward and reverse strands could be used to verify the accuracy of the draft genome.

- Hi-C Sequencing Data Integration:
  Hi-C sequencing data were used to create chromosomal scaffolds, where the contigs' orientation and location could be verified. Hi-C reads were mapped to the draft genomes using the Burrows–Wheeler Aligner software. The chromosome-level genome was generated using the 3D-DNA pipeline, which helped to give insights about the spatial organization of the genome. The final scaffold N50 was 146.5 Mb, indicating that this method generated many long contiguous fragments. Thus, the assembly was very efficient.

- Quality Evaluation:
  The quality of the assembled genome was analyzed using BUSCO analysis. It was found that the assembly covered 95.7% of the complete BUSCO genes, meaning it was highly complete.

  Thus, this method of using PacBio to get an idea about the larger picture of the genome (ex: structural variants, repetitive sequences), short read sequencing to correct errors and HiC to get a 3D structure was very effective at removing inaccuracies and providing a chromosomal level assembly of the genome.

7. Discuss five important findings of the paper:

    1. High-quality genomes:

This paper performs the first genome assembly of the South China subspecies of tiger. Additionally, the assembled genomes had more scaffold continuity and completeness, fewer gaps and more BUSCO genes annotated than previous assemblies of other subspecies. By using homology-based protein alignment, ab initio predictions and transcriptome mapping for gene annotation, 18,751 protein coding genes were predicted. These findings can help studies on tiger biology and conservation because of the insights they can provide on genetic diversity, population history and population structure.

2. Comparative Analysis of SCT and AT:
   The genome of the South China Tiger(SCT) along with another subspecies(the Amur Tiger(AT), was sequenced, and the two were compared using collinearity analysis. Sequencing of 28 other individuals(14 AT and 18 SCT) belonging to these two subspecies was also done, and PCA and phylogenetic analysis of the two populations were performed to evaluate the genetic relationships between them. This data revealed that the two species were two distinct groups which had diverged ~7000 years ago, but still 97.8% of the SC Tiger's genome matched almost identically with 97.9% of the Amur Tiger's genome.

3. Population Decline Phases in SCT identified:
   The dynamics of effective population size(Ne) was found by genomic analysis of the 18 SC tiger sequences. MSMC and SMC++ methods were used to find Ne, which represents the number of individuals in an ideal population that would retain the genetic diversity of the actual population (can be smaller than the census population size).
   - Three phases of decline of SC tigers were found to be:
   a. Around 5000 to 1000 years ago when Ne reduced from 14700 to 200
   b. Around 100 years ago, when, due to human activity (habitat loss and hunting) the census population size was affected: it fell from 4000 to functionally extinct numbers.
   c. About 66 years (1963) ago, when inbreeding occurred in captivity which led to further loss of Ne due to loss in genetic diversity.

4. Inbreeding Depression on SCT:
   The genomes of the 18 SC tigers were used to find Runs of Heterozygosity. The chromosome level assembly of this genome helped to find that the ROH in these tigers were long, continuous and covered a large portion of the chromosome, indicating that it occurred as a result of recent inbreeding. It was inferred that this occurred in the past 66 years, when these tigers were being bred in captivity.

5. Genetic Load Identification:
   Using the annotated genes, two kinds of deleterious mutations were identified in SC tigers - non-synonymous Single Nucleotide Variations (nsSNPs) and Loss of Function (LOF) mutations.
   - It was found that SCT has a high proportion of deleterious nsSNPs when compared to its total genetic variation, and these mutations occurred in genes related to reproduction, growth/development, disease.
   - SCTs had a large number of LOF mutations, and some of these mutations led to a reduction in sperm genesis, maturation and activity, and affected fetal growth and survival.

   This was in line with the inbreeding depression effects observed in the captive population(low reproductive rate and high juvenile mortality).

Since loss of genetic diversity and inbreeding depression are common threats to captive populations and are the main reason that such captive endangered species end up going extinct, good management of these populations in order to improve the gene pool (by performing outbreeding by distant relative pairing to minimize ROHs and remove deleterious mutations) is needed.

References:

1. Zhang L, Lan T, Lin C, Fu W, Yuan Y, Lin K, Li H, Sahu SK, Liu Z, Chen D, Liu Q, Wang A, Wang X, Ma Y, Li S, Zhu Y, Wang X, Ren X, Lu H, Huang Y, Yu J, Liu B, Wang Q, Zhang S, Xu X, Yang H, Liu D, Liu H, Xu Y. Chromosome-scale genomes reveal genomic consequences of inbreeding in the South China tiger: A comparative study with the Amur tiger. Mol Ecol Resour. 2023 Feb;23(2):330-347. doi: 10.1111/1755-0998.13669. Epub 2022 Jul 1. PMID: 35723950; PMCID: PMC10084155.