



ORANGUTAN GENETIC LOAD AND THE EFFECTS OF POPULATION SIZE

Kendal Chatard (Timothy H. Webster, Ph.D.)

Department of Anthropology

ABSTRACT

Genetic load is a relative measure of the quantity of harmful mutations within a genome, and a population's demographic history can affect this measure. Small populations that have undergone a severe bottleneck have increased inbreeding. These factors can decrease genetic diversity and fix deleterious mutations in the population, increasing the realized genetic load. In this study, we analyzed whole genome sequencing data for 17 individuals from Sumatran, Bornean, and Tapanuli orangutans. The Tapanuli orangutan is the most endangered great ape species and has the smallest population size of the three species of orangutans. We hypothesized that the Tapanuli orangutan would have both the largest relative genetic load and the greatest realized load. To compare the genetic load among these three species, we focused on three types of variants: loss of function indels, loss of function SNPs, and missense SNPs. We found that the Sumatran orangutans have the largest genetic load followed by the Tapanuli, then the Bornean. However, when examining realized load, the Bornean orangutan had the largest. These findings did not support our hypothesis but add to the growing pool of research contradicting the long-standing theory that larger effective population sizes should be associated with smaller genetic load. In the specific case of the Tapanuli orangutan, very recent population declines might not have had the chance to shape genetic load yet. Future research should include more individual genomes to further explore the connections between genetic diversity, realized genetic load, and inbreeding.

INTRODUCTION

Genetic diversity is central to evolutionary and conservation genetics, as it gives researchers insight into both the evolutionary past and the future viability of species. Of particular interest are severe population declines, during which inbreeding increases greatly, thereby decreasing genetic diversity in the population. In addition to decreasing genetic variation, this has the potential to affect a population's genetic load. Genetic load refers to the quantity of deleterious (harmful) alleles carried within the genomes of a population. As genome-scale data have become available for non-model species, researchers have begun studying this phenomenon in wild populations, including those of endangered primates like gorillas and snub-nose monkeys (van der Valk et al., 2019; Xue et al., 2015; Zhou et al., 2016).

While there are many ways to define genetic load, they all describe processes which result in a decrease of species viability. Two components of genetic load are commonly considered. The first—realized load—is the load of deleterious alleles that have viability decreasing phenotypic effects on an individual. The second, called the masked load, is the load that often exists as recessive deleterious mutations masked in heterozygous states. It is important to keep both components of genetic load in mind to understand an organism's total viability loss (Bertorelle et al., 2022).

In a 2015 paper, Xue *et al.* introduced a specific procedure for calculating the genetic load of the mountain gorilla populations they studied (Xue *et al.*, 2015). For each population (eastern lowland, western lowland, and mountain gorillas), researchers first identified single nucleotide polymorphisms (SNPs) in individual genomes when compared to a reference western gorilla genome. These mutations were then classified as synonymous, missense, or loss of function (LoF) based on their predicted phenotypic effects. Synonymous mutations are mutations in coding regions that do not affect amino acids, while missense and LoF mutations do alter proteins and therefore have phenotypic effects that are thought to most commonly be deleterious. A mutation is classified as missense when the change in DNA sequence causes a change in the amino acid sequence of the protein coded for. To be classified as a LoF mutation, the altered protein must no longer function in the capacity it was meant to.

This method for determining relative genetic load can be directly applied to data from other organisms. Of particular interest are endangered species such as orangutans. There are three species of orangutans: Sumatran (*Pongo abelii*), Bornean (*Pongo pygmaeus*), and Tapanuli (*Pongo tapanuliensis*). The last was recently discovered in 2017 (Nater *et al.*, 2017) and has an extremely small population size. While all three orangutans are critically endangered, the Tapanuli is the most endangered great ape, with only about 800 individuals left living in North Sumatra (see Figure 1) (Wich *et al.*, 2019). The variation in effective population size and demographic history (see Figure 2) among the orangutan species presents an opportunity to study how those factors influence genetic load and better understand the genetic variation present in these endangered species.

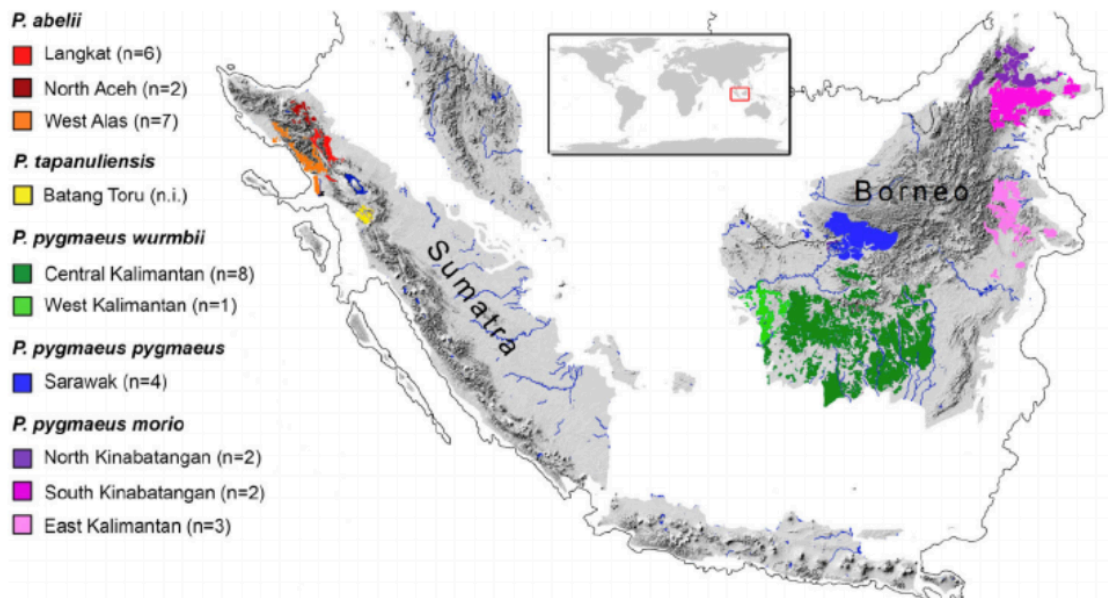


Figure 1: Distribution of genus *Pongo* in Indonesia, adapted from Nater *et al.*, 2017

In this research project, we used the framework developed by Xue *et al.* to measure genetic load in all three species of orangutans using different types of variants. Specifically, in addition to SNPs, which are typically the focus of studies like this, we also investigated indels. Indels are insertions or deletions of nucleotides in DNA and can have highly disruptive effects. While inserting or deleting chunks of DNA can obviously alter protein sequences, they can also cause frameshift mutations which alter the reading frame of the protein to be translated and often cause the protein to lose its function completely (Hämälä *et al.*, 2021; Montgomery *et al.*, 2013). This is because the entire sequence of amino acids downstream from the mutation will be changed due to the frameshift.

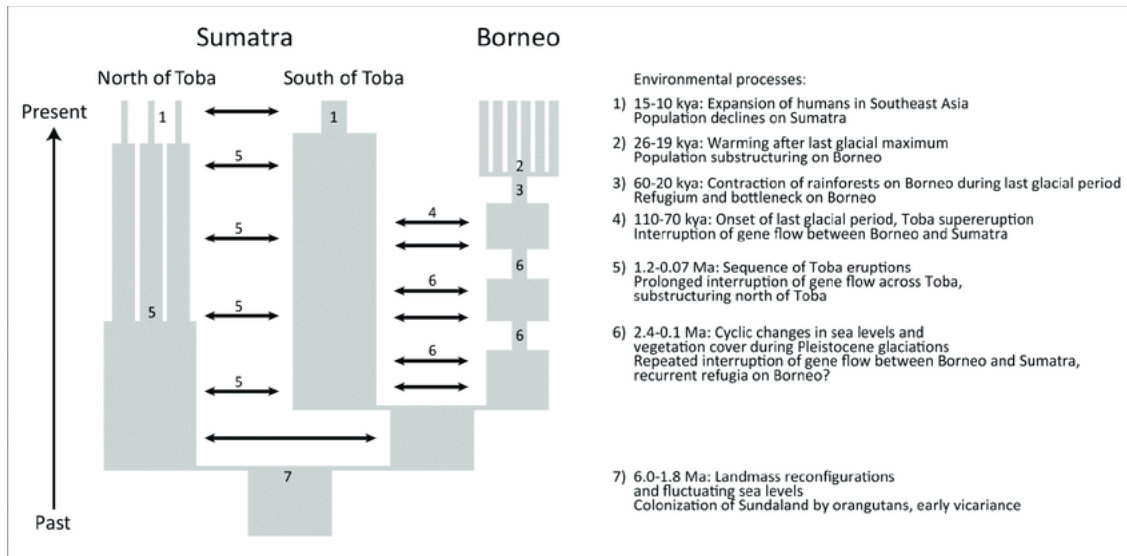


Figure 2: Demographic History of Orangutans. Figure from Nater, 2012

As mentioned above, most genetic load studies measure the quantity of missense and loss of function SNP mutations (Hu et al., 2020; Wang et al., 2021). Since SNPs are generally less deleterious than indels, they are more common in the genome. A larger sample increases statistical power, but by focusing on one type of mutation, the effects of other harmful variants are overlooked.

Recently, Bertorelle *et al.* argued that differentiating between realized load (expressed load) and masked load (inbreeding load) is required for a better understanding the effects of deleterious mutations in population genetics. The realized load is expressed phenotypically and it reduces the viability of the current generation. The masked load encompasses the deleterious alleles that could become expressed in future generations if inbreeding were to increase (Bertorelle *et al.*, 2022). Delineating the realized from the masked load in the data for these populations of orangutans could directly predict the phenotypic outcomes of the deleterious mutations present in the gene pool. The realized load is predicted to increase after a population bottleneck, while the masked load is expected to be high in a larger population before a bottleneck, where inbreeding has not yet become harmful.

The Tapanuli orangutan is thought to have undergone a massive population decline in just the past 100 to 150 years. Their current range is estimated to be only 2.5% of their 1890s range and 5.0% of their 1940s range (Meijaard et al., 2020). Moreover, another study found extensive runs of homozygosity in the Tapanuli orangutan (Nater et al., 2017). The combination of a severe population decline and increased homozygosity would predict fixed deleterious mutations and therefore a high realized genetic load.

Since the effective population size of the Tapanuli orangutan is significantly smaller than the Sumatran and Bornean orangutans, we hypothesize that both the total genetic load and the realized load of loss of function indels, loss of function SNPs, and missense SNPs will be greater in the Tapanuli orangutan than the other two species.

METHODS

We used previously published high-coverage whole genome sequencing data for 17 orangutan individuals from all three species (Nater et al., 2017). This dataset included six Sumatran orangutans (*Pongo abelii*), ten Bornean orangutans (*Pongo pygmaeus*), and one Tapanuli orangutan (*Pongo tapanuliensis*).

After first assessing the overall quality of reads using FastQC (Andrews, 2010) and MultiQC (Ewels, Magnusson, Lundin, & Källér, 2016), we discovered many reads with low mean quality scores and high N base content that needed to be trimmed and filtered. Using BBduk (BBMap, 2022), we removed sequencing adapters and filtered and trimmed low-quality reads using the parameters: “ktrim=r k=21 minq=11 hdist=2 qtrim=rl trimq=15 minlen=50 maq=20 tbo tbe.” A second assessment of quality with FastQC and MultiQC revealed that after trimming, read quality was satisfactory.

Next, we mapped reads and processed alignments using BWA-MEM (Li, 2013) and SAMtools (Li, et al., 2009; Marshall, et al., 2009), the latter of which was used to fix mate pairs, sort, merge, and index the BAM files. We then used Picard (Picard Tools, n.d.) to mark duplicate reads, after which we used SAMtools to calculate statistics from the resulting BAM files. The reference genome we used is for *Pongo abelii* (version Susie_PABv2), and we downloaded it from NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCF_002880775.1/).

We called variants across all samples using GATK4 (Poplin, et al., 2018). We used default settings for HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs, with the exception of turning off physical phasing to ensure the resulting VCF was compatible with tools used in downstream analyses. Finally, we filtered all variants to ensure at least a per-sample read depth of 10, a genotype quality score of 30, and a site-wide mapping quality score of 30.

To identify the variant classes and consequences, we used Ensembl’s Variant Effect Predictor (VEP) (Ensembl Variant Effect Predictor (VEP), 2021). We used a cache directory to access genome annotations for the same *Pongo abelii* reference genome as above (http://ftp.ensembl.org/pub/current_variation/indexed_vep_cache/pongo_abelii_merged_vep_105_Susie_PABv2.tar.gz). We ran VEP for each species (*P. abelii*, *P. pygmaeus*, *P. tapanuliensis*) and for each individual sample. We used many nonstandard flags to provide as much useful information as possible in the results. These included check_existing, numbers, symbol, regulatory, canonical, protein, biotype, variant_class, flag_pick, and tab (Variant Effect Predictor- Running VEP, 2021). The flag_pick was particularly important because it forces VEP to pick one consequence per variant.

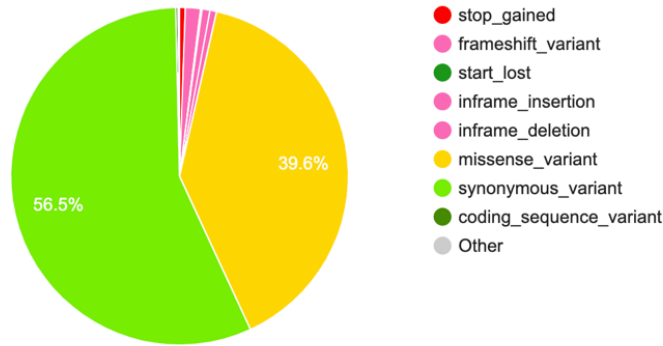
Variant classes



Variant class	Count
substitution	10,732
indel	75,723
insertion	327,378
deletion	349,837
SNV	6,240,815
sequence_alteration	11,912,733

Figure 3: VEP Variant class results for *Pongo tapanuliensis*

Coding consequences



Consequence type	Count
stop_gained	888
frameshift_variant	2,368
stop_lost	101
start_lost	229
inframe_insertion	1,230
inframe_deletion	1,011
protein_altering_variant	87
missense_variant	64,365
synonymous variant	91,912

Figure 4: VEP Coding consequence results for SNPs in *Pongo tapanuliensis*

The VEP software not only identified the type of variant but also its consequence on the gene (Figures 3 and 4). For our purposes, variants classified as transcript_ablation, splice_donor_variant, splice_acceptor_variant, stop_gained, frameshift_variant, inframe_insertion, inframe_deletion, splice_region_variant, stop_lost, and start_lost were labeled as non-synonymous mutations. Synonymous variants are those labeled synonymous_variant and the intergenic variants are those labeled upstream_gene_variant, downstream_gene_variant, and intergenic_variant.

$$1. \quad f_i^A = \frac{d_i^A}{n_i^A} \quad 2. \quad L_{A,B}(C) = \frac{\sum_{i \in C} f_i^A (1 - f_i^B)}{\sum_{j \in I} f_j^A (1 - f_j^B)} \quad 3. \quad R_{\frac{A}{B}}(C) = \frac{L_{A,B}(C)}{L_{B,A}(C)}$$

Equations 1-3: Where d_i^A is the number of derived alleles called at site i in population A, and n_i^A is the total number of alleles called, f_i^A is the observed derived allele frequency. Given C is a set of protein coding sites, and intergenic sites I , we define $L_{A,B}(C)$. The ratio $R_{A/B}(C)$ is therefore a measure of the relative number of derived alleles in population A compared to population B.

Equations from Xue et al., 2015

Using Equations 1-3, we calculated the relative excess of deleterious variants in pairs of populations. This comparison uses the relative frequency of variants in the intergenic regions as a proxy for the baseline mutation rate for the species. At each protein coding site and each intergenic site, we counted the number of nonreference alleles (one for each heterozygous site, and two for each homozygous site), and the total number of reference and derived alleles to use

in Equation 1 for all populations. Then, for two populations at a time, we used those frequencies in Equations 2 and 3. We ran this calculation for three measures: the relative excess of loss of function indels, loss of function SNPs, and missense SNPs. We used the frequency of the indels or SNPs in the intergenic regions as the baseline rate by inputting those into the denominator of Equation 2. In addition, we used the counts of homozygous alleles for these three mutations to estimate the realized load. By using the relative rates of excess deleterious mutations (i.e., comparing the rate of the variants in the coding region to the intergenic region), we control for the mapping bias due to the fact that the reference genome used for all three species is from *Pongo abelii*.

RESULTS

We found that the Sumatran orangutan (*P. abelii*) had the highest genetic load of the three species when considering the average genetic load of indels and biallelic SNPs together (Table 1).

SPECIES COMPARISON	BIALLELIC SNPS	INDELS
PA/PT	1.114	1.117
PP/PT	0.949	0.782
PP/PA	0.906	1.094

Table 1: Average Relative Genetic Load for SNPs and Indels [*Pongo pygmaeus* (PP), *Pongo abelii* (PA), and *Pongo tapanuliensis* (PT)]. A value greater than 1 indicates a higher load in the population in the numerator, while a value less than 1 indicates a higher load in the denominator population.

In the graph below (Figure 5), the genetic load of SNPs is split into missense and loss of function consequences. *P. abelii* has the greatest excess of missense SNPs and loss of function SNPs. The genetic load of indels was similarly highest in *P. abelii* with an exception of *P. pygmaeus* having 1.094 times more indels than *P. abelii*. While the Bornean orangutans (*P. pygmaeus*) had the smallest overall genetic load, they had the largest realized load followed by *P. abelii* and then *P. tapanuliensis* (Table 2).

SPECIES COMPARISON	LOF INDELS	LOF SNPS	MISSENSE SNPS
PA/PT	1.169	1.402	1.231
PP/PT	1.424	1.748	1.465
PP/PA	2.396	1.501	1.851

Table 2: Relative Genetic Load of Homozygous Alleles [*Pongo pygmaeus* (PP), *Pongo abelii* (PA), and *Pongo tapanuliensis* (PT)]. A value greater than 1 indicates a higher load in the population in the numerator, while a value less than 1 indicates a higher load in the denominator population.

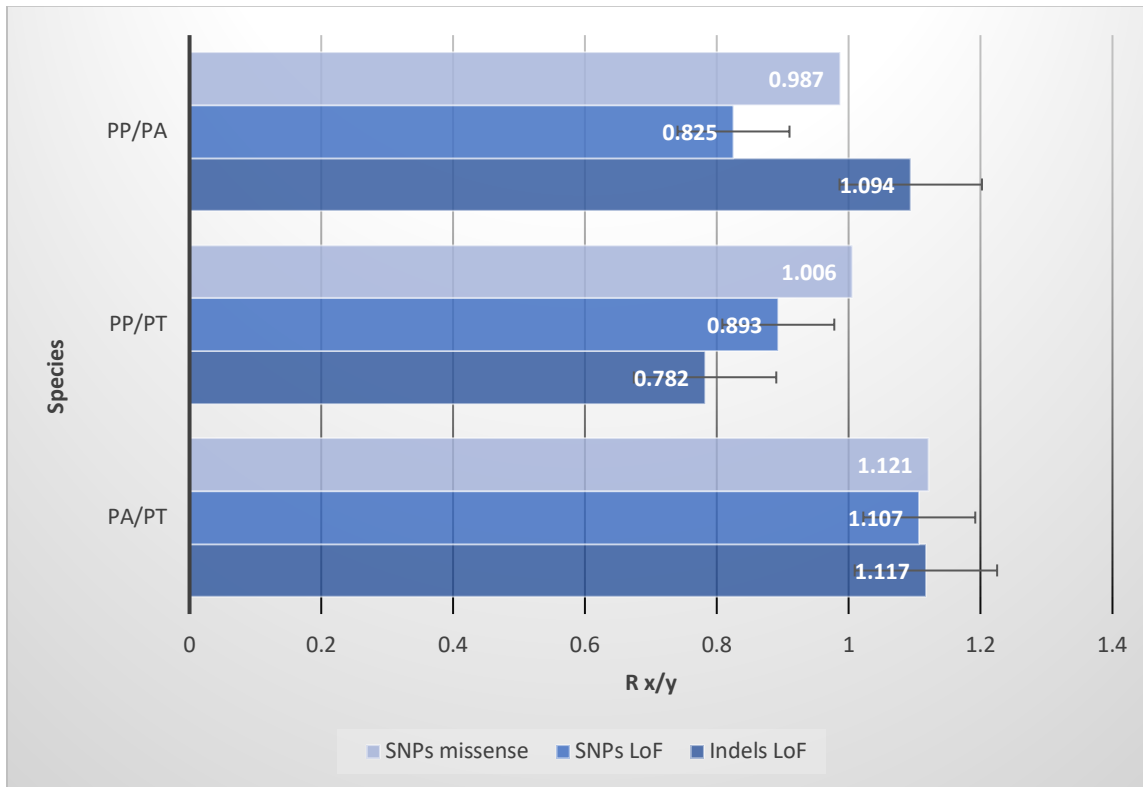


Figure 5: Relative Genetic Load- number of derived alleles at missense and loss of function (LoF) sites for indels and biallelic SNPs that are frequent in one species over another. A value greater than 1 indicates a higher load in the population in the numerator, while a value less than 1 indicates a higher load in the denominator population. [Pongo pygmaeus (PP), Pongo abelii (PA), and Pongo tapanuliensis (PT)]

DISCUSSION

When considering the conservation of orangutans and biodiversity generally, genetic diversity is a central component. The literature supports the idea that maintaining a high level of variation across the genome is just as important for healthy biodiversity as maintaining high species and ecosystem diversity (DeWoody et al., 2021). The correlation between genetic diversity and overall fitness has been widely accepted as a ‘fact’ since the 1980s and is regularly used as justification for many conservation genetics research projects and papers. However, this theory in conservation genetics that greater genome-wide genetic diversity in a population directly increases the fitness of that population has been challenged. Recently, better genomic analysis tools used to estimate the number of deleterious mutations in a genome have been made. Additionally, the complexity surrounding how those harmful mutations arise, persist, and are purged in populations has become clearer.

Genetic diversity depends, in part, on effective population size. When a population is successfully surviving and reproducing the effective population size (N_e) is also expected to be high, which means there are more opportunities for novel mutations in the genome to appear. The mutations can be neutral, deleterious, or beneficial, but their presence is what drives adaptations to the environment. In theory, when a population undergoes a bottleneck and N_e is reduced, variation is lost. During bottlenecks, harmful recessive mutations which were previously masked within the population are expressed more often due to increased homozygosity from inbreeding, transforming the masked load into realized load. The realized load could then reduce the viability of the current generation.

Our findings differ from these expectations and contradict our hypothesis that the Tapanuli orangutans, with the smallest population size, would have the largest genetic load. Instead, the Sumatran orangutan, which has significantly larger effective population size than the Tapanuli, had the largest genetic load. When focusing on the realized load—the portion of genetic load that is found in a homozygous state and therefore phenotypically expressed—the Bornean orangutans carried the greatest load.

There are multiple possible explanations for our results. The more recent extreme population declines in the Tapanuli orangutan might have purged some of the realized load from the genome, which would decrease the total genetic load in addition to the realized load. Another possibility is that orangutans' long lifespans and therefore long generation times have not yet let the effects of population decline leave a signature in the genome, as the decline happened within the last 100 years, or approximately four generations. The Bornean orangutan's population bottlenecks were more ancient, which would allow time for realized load from inbreeding to appear in the data.

Multiple studies have been published recently that also argue against the generalization that genetic diversity and fitness are intimately related (Kyriazis, 2021; Teixeira, 2021). Instead, it might not be genetic diversity but the types of mutation consequences present in the gene pool and their frequency in the population that better predict viability (Kyriazis, 2021).

A case study of genetic rescue in the gray wolf population on Isle Royal illustrates this idea (Kyriazis, 2021). Genetic rescue is a tactic used by conservation scientists to “save” populations of endangered species by introducing new individuals from a larger population. The theory behind this is that adding more diverse genetic material will reverse some of the effects inbreeding has had on a population. However, in this study, one of the wolves introduced into that area brought with it lethal equivalents that spread into the population and through inbreeding, reduced the population size to two individuals (Kyriazis, 2021). The authors highlight that while population viability, large N_e , and greater genetic diversity are all positively related, the focus for conservation should be on minimizing deleterious genetic load, not maximizing genetic diversity (Kyriazis, 2021).

By taking a closer look at what kind of genetic diversity exists in a population, the paths for conservation will be different. It will be important to be able to identify what alleles in the gene pool are actually detrimental to the populations' survival. Knowing that neutral genetic diversity does not predict extinction risk (Teixeira & Huber, 2021) can also help inform conservation management strategies.

A 2009 study on the demography and conservation outlook for Bornean and Sumatran orangutans examined the effects of lethal equivalents in the genome (Marshall, et al., 2009). The authors found that in medium- and poor-quality habitats, inbreeding depression caused extinction events when frequencies of lethal equivalents were high. Future research could construct a clearer picture of the orangutans' fitness level by combining the more estimates of effective population size, species divergence, and inter-population migration (Nater et al., 2017) with the models presented by Marshall *et al.*

Additional research on genetic diversity, genetic load, and the Tapanuli orangutans should focus include more individuals. This study only included one Tapanuli genome and it is possible that the genome data is not representative of the whole species. While there were relative differences, genetic load was high in both the Sumatran and Tapanuli orangutans and with accelerating habitat loss (Nater et al., 2017), the extinction risk for these species is very real. It is critical that immediate steps are taken to preserve these endangered species.

REFERENCES

- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- BBMap*. (2022, 02 25). Retrieved from SourceForge: <https://sourceforge.net/projects/bbmap/>
- Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H. E., & van Oosterhout, C. (2022). Genetic load: genomic estimates and applications in non-model animals. *Nature Reviews Genetics* 2022, 1–12. <https://doi.org/10.1038/s41576-022-00448-x>
- Brown, A. (1978). Isozymes, plant population genetic structure and genetic conservation. *Theoretical Applied Genetics*, 145-157.
- DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance of genetic diversity in conservation. In *Molecular Ecology* (Vol. 30, Issue 17, pp. 4147–4154). John Wiley and Sons Inc. <https://doi.org/10.1111/mec.16051>
- Ensembl Variant Effect Predictor (VEP)*. (2021, 12). Retrieved from Ensembl: <https://uswest.ensembl.org/info/docs/tools/vep/index.html>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 3047-3048.
- García-Dorado, A., & Caballero, A. (2021). Neutral genetic diversity as a useful tool for conservation biology. *Conservation Genetics*, 541-545.
- Hämälä, T., Wafula, E. K., Guiltinan, M. J., Ralph, P. E., DePamphilis, C. W., & Tiffin, P. (2021). Genomic structural variants constrain and facilitate adaptation in natural populations of *Theobroma cacao*, the chocolate tree. *Proceedings of the National Academy of Sciences of the United States of America*, 118(35), 2102914118. <https://doi.org/10.1073/pnas.2102914118>
- Hu, Y., Thapa, A., Fan, H., Ma, T., Wu, Q., Ma, S., Zhang, D., Wang, B., Li, M., Yan, L., & Wei, F. (2020). Genomic evidence for two phylogenetic species and long-term population bottlenecks in red pandas. *Science Advances*, 6(9). https://doi.org/10.1126/SCIADV.AAX5751/SUPPL_FILE/AAX5751_TABLES_S11_TO_S16.XLSX
- Kardos, M., Armstrong, E., Fitzpatrick, S., Hauser, S., Hedrick, P., Miller, J., . . . Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *PNAS*, 33-47.
- Kyriazis, C. C. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters*, 33-47.
- Li, H. (2013, 05 26). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Retrieved from Cornell University: <https://arxiv.org/abs/1303.3997>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Subgroup, 1. G. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 2078-2079.

Marshall, A. J., Lacy, R., Ancrenaz, M., Byers, O., Husson, S. J., Leighton, M., & ... & Wich, S. A. (2009). Orangutans: Geographic variation in behavioral ecology and conservation. In *Orangutan population biology, life history, and conservation*. (pp. 311-326).

Meijaard, E., Ni'matullah, S., Dennis, R., Sherman, J., Onrizal, & Wich, S. A. (2020). How widespread was the Tapanuli orangutan and what led to its decline? *BioRxiv*, 2020.08.11.246058. <https://doi.org/10.1101/2020.08.11.246058>

Montgomery, S. B., Goode, D. L., Kvikstad, E., Albers, C. A., Zhang, Z. D., Mu, X. J., Ananda, G., Howie, B., Karczewski, K. J., Smith, K. S., Anaya, V., Richardson, R., Davis, J., MacArthur, D. G., Sidow, A., Duret, L., Gerstein, M., Makova, K. D., Marchini, J., ... Lunter, G. (2013). The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. *Genome Research*, 23(5), 749–761. <https://doi.org/10.1101/gr.148718.112>

Nater, Alexander. Processes underlying genetic differentiation and speciation in Orangutans (*Pongo* spp.). 2012, University of Zurich, Faculty of Science.

Nater, A., Mattle-Greminger, M. P., Nurcahyo, A., Nowak, M. G., de Manuel, M., Desai, T., Groves, C., Pybus, M., Sonay, T. B., Roos, C., Lameira, A. R., Wich, S. A., Askew, J., Davila-Ross, M., Fredriksson, G., de Valles, G., Casals, F., Prado-Martinez, J., Goossens, B., ... Krützen, M. (2017). Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. *Current Biology*, 27(22), 3487-3498.e10. <https://doi.org/10.1016/j.cub.2017.09.047>

Picard Tools. (n.d.). Retrieved from Broad Institute: <http://broadinstitute.github.io/picard/>

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., ... Roazen, D. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178.

Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *PNAS*.

van der Valk, T., Diez-del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. *Current Biology*, 29(1), 165–170. <https://doi.org/10.1016/j.cub.2018.11.055>

Variant Effect Predictor- Running VEP. (2021, 12). Retrieved from Ensembl: https://uswest.ensembl.org/info/docs/tools/vep/script/vep_options.html

Wang, P., Burley, J. T., Liu, Y., Chang, J., Chen, D., Lu, Q., Li, S. H., Zhou, X., Edwards, S., & Zhang, Z. (2021). Genomic Consequences of Long-Term Population Decline in Brown Eared Pheasant. *Molecular Biology and Evolution*, 38(1), 263–273. <https://doi.org/10.1093/MOLBEV/MSAA213>

Wich, S. A., Fredriksson, G., Usher, G., Kühl, H. S., & Nowak, M. G. (2019). The Tapanuli orangutan: Status, threats, and steps for improved conservation. *Conservation Science and Practice*, 1(6), e33. <https://doi.org/10.1111/csp2.33>

Xue, Y., Prado-Martinez, J., Sudmant, P. H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D. N., De Manuel, M., Hernandez-Rodriguez, J., Lobon, I., Siegismund, H. R., Pagani, L., Quail, M. A., Hvilsom, C., Mudakikwa, A., Eichler, E. E., ... Scally, A. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science*, *348*(6231), 242–245. <https://doi.org/10.1126/science.aaa3952>

Zhou, X., Meng, X., Liu, Z., Chang, J., Wang, B., Li, M., Wengel, P. O. Ter, Tian, S., Wen, C., Wang, Z., Garber, P. A., Pan, H., Ye, X., Xiang, Z., Bruford, M. W., Edwards, S. V., Cao, Y., Yu, S., Gao, L., ... Li, M. (2016). Population Genomics Reveals Low Genetic Diversity and Adaptation to Hypoxia in Snub-Nosed Monkeys. *Molecular Biology and Evolution*, *33*(10), 2670–2681. <https://doi.org/10.1093/molbev/msw150>