

Manual Cluster: SLiM Project Timothé

*Disentangling how **selection**, **demography**, and **sex-biased behavior** influence **genetic diversity** using field study-informed **genetic simulations***

The folder of the project is: `/mnt/primevo/work/timothe_dandoy/SLiM_Project/`

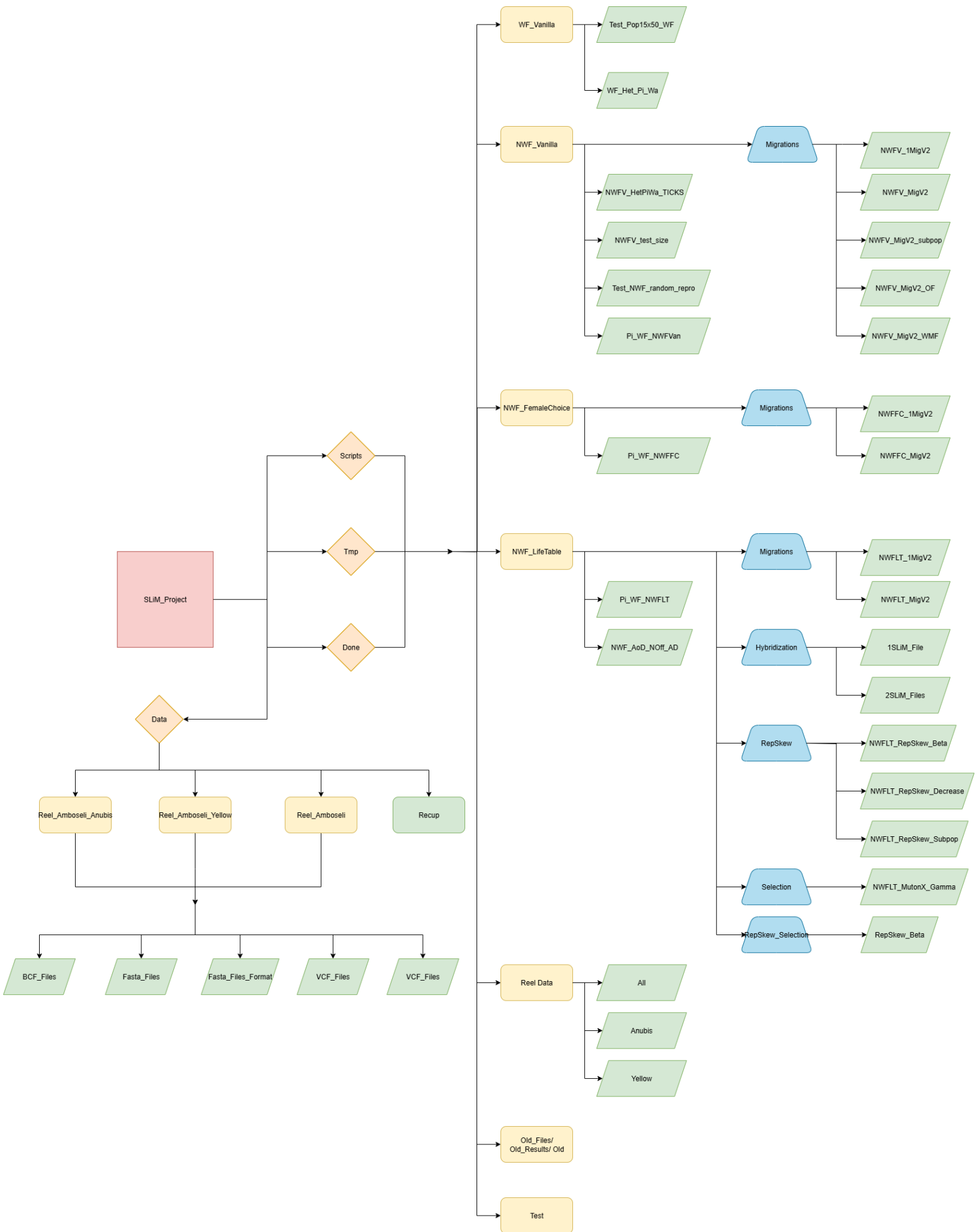
This manual aims to explain how this folder is organized. The folder is divided into 7 subfolders, but only 4 are used:

- *Scripts*: With all the SLiM scripts, workflow, and Jupyter scripts
- *Tmp*: With all the resulting text files of the SLiM simulations
- *Done*: Graphs images of the results
- *Data*: Reel data (BCF Files) from reference panel baboon genomes

Scripts, tmp and done follow basically the same structure. The manual will present in detail the structure of scripts, and only the differences with tmp and done.

A mind map of all the folders found in the SLiM_Project is shown on the next page. The green lozenges are the final folders containing the files.

If a file or folder is missing in this notice, maybe it is not relevant.



I. scripts

script folder contains all the slim script and workflow gwf to launch the simulation. It also contains the Jupyter files to analyze the output of simulations and construct the graphs.

a. WF_Vanilla

WF_Vanilla folder contains the scripts files for the WF model.

i. Test_Pop_15x50_WF

Test_pop_15x50_WF folder contains the script to run and analyze the diversity in **WF_Vanilla simulations** with a **subdivided population** and **no migration**. Three files are in the folder: the slim script, the workflow and the Jupyter script.

ii. WF_Het_Pi_Wa

WF_Het_Pi_Wa folder contains the script to run and analyze the diversity in **WF_Vanilla simulations** with **different numbers of Ticks** and **3 different estimators: Pi, Wa and heterozygosity**. Three files are in the folder: the slim script, the workflow and the Jupyter script.

b. NWF_Vanilla

NWF_Vanilla folder contains the scripts files for the NWF_V model.

i. Migrations

Migrations folder contains scripts for the NWF_V model with a subdivided population and migrations.

1. NWFV_1MigV2

NWFV_1MigV2 folder contains scripts to run and analyze the diversity in **NWF_V simulations** with **only male migrations** and **one migration in a life of a male**. This model is looking at the **whole population and the one subpopulation diversities**. Simulations are done with **2 different migration modes: Star and Circle**. The folder contains slim scripts for both migration modes, the workflow and the jupyter script (first cell for the whole population diversity, and second cell for the one subpopulation one).

2. NWFV_MigV2

NWFV_MigV2 folder contains scripts to run and analyze the diversity in **NWF_V simulations** with **only male migrations**. This

model is only looking at the **whole population diversity**. Simulations are done with **3 different migration modes: Star, Circle and Steppingstones**. The folder contains slim scripts for the three migration modes, the workflow and two jupyter scripts:

-*NWFFV_MigV2_analyze*: Analyze with probability to migrate from 0.05 to 1.0.

-*NWFFV_MigV2_lowproba_analyze*: Analyze with low probability of migrations. This Files may not be updated and usable.

3. *NWFFV_MigV2_subpop*

NWFFV_MigV2_subpop folder contains scripts to run and analyze the diversity in **NWF_V simulations** with **only male migrations**. This model is only looking at the **one subpopulation diversity**. Simulations are done with **3 different migration modes: Star, Circle and Steppingstones**. The folder contains slim scripts for the three migration modes, the workflow and two jupyter scripts:

- *NWFFV_MigV2_subpop_analyze*: Analyze with probability to migrate from 0.05 to 1.0. The first cell of the script shows the mean diversity ratio of all the subpopulations, the second one the diversity ratio for only one subpopulation.

- *NWFFV_MigV2_lowproba_subpop_analyze*: Analyze with low probability of migrations. This Files may not be updated and usable.

4. *NWFFV_MigV2_OF*

NWFFV_MigV2_OF folder contains scripts to run and analyze the diversity in **NWF_V simulations** with **only female migrations**. This model is looking at the **whole population and the one subpopulation diversities**. Simulations are done with **3 different migration modes: Star, Circle and Steppingstones**. The folder contains slim scripts for the three migration modes, the workflow and the jupyter script (first cell for the whole population diversity, and second cell for the one subpopulation one).

5. *NWF_MigV2_WMF*

NWFFV_MigV2_WMF folder contains scripts to run and analyze the diversity in **NWF_V simulations** with **male and female**

migrations. This model is looking at the **whole population and the one subpopulation diversities**. Simulations are done with **3 different migration modes: Star, Circle and Steppingstones**. The folder contains slim scripts for the three migration modes, the workflow and the jupyter script (first cell for the whole population diversity, and second cell for the one subpopulation one).

ii. NWFV_HetPiWa_TICKS

NWFV_HetPiWa_TICKS folder contains the script to run and analyze the diversity in **NWF_V simulations** with **different numbers of Ticks** and **3 different estimators: Pi, Wa and heterozygosity**. Three files are in the folder: the slim script, the workflow and two Jupyter scripts:

- *NWFV_HetPiWa_analyze*: boxplots of the genetic diversity and X-to-A ratio for each Number of Ticks and estimator
- *NWFV_OnlyPi_Analyze*: boxplots of the genetic diversity and X-to-A ratio at each Number of Ticks for only the Pi estimator.

iii. NWFV_test_size

NWFV_test_size folder contains the script to run and analyze the diversity and the time of simulation in **NWF_V simulations** with **different population sizes** and **mutations rate coefficients**. The goal is to find the good couple of parameters to have the diversity we want. Three files are in the folder: the slim script, the workflow and the Jupyter script.

iv. Test_NWF_random_repro

Test_NWF_random_repro folder contains the script to run and analyze the diversity in **NWF_RandomRepro** (totally random for males and females) **and WF simulations** with **different sex-ratio**. Four files are in the folder: the slim scripts for both models, the workflow and the Jupyter script.

v. Pi_WF_NWFFan

Pi_WF_NWFFan folder contains the script to run and analyze the diversity in **NWF_V and WF simulations** with **different sex-ratio**. Four files are in the folder: the slim scripts for both models, the workflow and the Jupyter script.

c. NWF_FemaleChoice

NWF_FemaleChoice folder contains the scripts files for the NWF_FC model.

i. Migrations

Migrations folder contains scripts for the NWF_FC model with a subdivided population and migrations.

1. NWFFC_1MigV2

NWFFC_1MigV2 folder contains scripts to run and analyze the diversity in **NWF_FC simulations** with **only male migrations** and **one migration in a life of a male**. This model is looking at the **whole population and the one subpopulation diversities**. Simulations are done with **2 different migration modes: Star and Circle**. The folder contains slim scripts for both migration modes, the workflow and the jupyter script (first cell for the whole population diversity, and second cell for the one subpopulation one).

2. NWFFC_MigV2

NWFFC_MigV2 folder contains scripts to run and analyze the diversity in **NWF_FC simulations** with **only male migrations**. This model is looking at the **whole population and the one subpopulation diversities**. Simulations are done with **3 different migration modes: Star, Circle and Steppingstones**. The folder contains slim scripts for the three migration modes, the workflow and the jupyter script (first cell for the whole population diversity, and second cell for the one subpopulation one).

ii. Pi_WF_NWFFC

Pi_WF_NWFFC folder contains the script to run and analyze the diversity in **NWF_FC and WF simulations** with **different sex-ratio**. Four files are in the folder: the slim scripts for both models, the workflow and the Jupyter script.

d. NWF_LifeTable

NWF_LifeTable folder contains the scripts files for the NWF_LT model

i. Migrations

Migrations folder contains scripts for the NWF_FC model with a subdivided population and migrations.

1. NWFLT_1MigV2

NWFLT_1MigV2 folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with **only male migrations** and **one migration in a life of a male**. This model is looking at the **whole population and the one subpopulation diversities**. Simulations are done with **2 different migration modes: Star and Circle**. The folder contains slim scripts for both migration modes, the workflow and the jupyter script (first cell for the whole population diversity, and second cell for the one subpopulation one).

2. NWFLT_MigV2

NWFLT_MigV2 folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with **only male migrations**. This model is looking at the **whole population and the one subpopulation diversities**. Simulations are done with **3 different migration modes: Star, Circle and Steppingstones**. The folder contains slim scripts for the three migration modes, the workflow and the jupyter script (first cell for the whole population diversity, and second cell for the one subpopulation one).

ii. Hybridization

Hybridization folder contains scripts for the NWF_FC model with hybridization between two diverged populations.

1. 1SLiM_File

1SLiM_File folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with **hybridizations**. Three simulations are possible: **only male hybridization** (OM), **only female hybridization** (OF) or **with male and female hybridization** (WMF). This model is looking at the **genetic diversity** of one population during all the simulations, and the **Fst value** when the two populations diverge. These graphs are done with a burn-in period of 60000 Ticks. The folder contains slim scripts for the three hybridization modes, the workflow and five jupyter scripts:

- *All_in_1_Hyb_OM_Analyze*: Analyze of the X-to-A ratio for the only male hybridization. Each cell corresponds to a different number of migrants.

- *All_in_1_Hyb_OF_Analyze*: Analyze of the X-to-A ratio for the only female hybridizations. Each cell corresponds to a different number of migrants.

- *All_in_1_Hyb_WMF_Analyze*: Analyze of the X-to-A ratio for male and female hybridization. Each cell corresponds to a different number of migrants.

- *NWFLT_Pi_BP_Hyb*: Analyze of the X-to-A ratio after the first burn-in period. The value corresponds to the NWF_LT X-to-A ratio.

- *New_Test*: Just a test file

2. 2SLiM_Files

2SLiM_Files folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with **hybridizations**. Three simulations are possible: **only male hybridization** (OM), **only female hybridization** (OF) or **with male and female hybridization** (WMF). Another model is available (NextBP) to simulate the drift of one population with no hybridizations event. The simulations in two parts: 1) a burning period which is shared with all the models, 2) separation, hybridization and drift periods. This model is looking at the **genetic diversity** of one population during all the simulations, and **the Fst value** when the two populations diverge. These graphs are done with a burn-in period of 20000 Ticks. The folder contains slim scripts for the burn-in period, for the three hybridization modes and for the next burn-in period, the workflow and five jupyter scripts:

- *Hyb_OM_Analyze*: Analyze of the X-to-A ratio and diversity for the only male hybridization. Each cell corresponds to a different number of migrants.

- *Hyb_OF_Analyze*: Analyze of the X-to-A ratio for the only female hybridization. Each cell corresponds to a different number of migrants.

- *Hyb_WMF_Analyze*: Analyze of the X-to-A ratio with male and female hybridization. Each cell corresponds to a different number of migrants.

- *No_Hyb_BP_Analyze*: Analyze of the X-to-A and diversity ratio with no hybridization.

- *Hyb_BurningPeriod_Analyze*: Analyze of the X-to-A and diversity ratio at the end of the burn-in period.
- *FST_Hyb_OM_Analyze*: Analyze of the Fst value between P0 and P1 with only male hybridization.

iii. RepSkew

RepSkew folder contains scripts for the NWF_FC model with different modes of reproductive skew in males.

1. NWFLT_RepSkew_Beta

NWFLT_RepSkew_Beta folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with **male reproductive skew** and a **constant tag value**. This model is looking at the **genetic diversity**, and the **demography outputs** (age of death and number of offsprings). The folder contains the slim script, the workflow and two jupyter scripts:

- *Analyze_Variance_RepSkew*: Analyze of the variance in reproductive skew in males and females when beta =1.
- *NWFLT_RepSkew_analyze*: Analyze of the X-to-A ratio, genetic diversity, reproductive skew and number of offsprings for every beta.

2. NWFLT_RepSkew_Decrease

NWFLT_RepSkew_Decrease folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with **male reproductive skew** and a **decreasing tag value**. This model is looking at the **genetic diversity**, and the **demography outputs** (age of death and number of offsprings). The folder contains the slim script, the workflow and the jupyter script (X-to-A ratio, genetic diversity, reproductive skew and number of offsprings for every beta).

3. NWFLT_RepSkew_Subpop

NWFLT_RepSkew_Subpop folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with **male reproductive skew**, a **constant tag value** and a **subdivided populations with circle migrations**. This model is looking at the **genetic diversity**, and the **demography outputs** (age of death and number of offsprings). The folder contains the slim script, the workflow and the jupyter script (X-to-A ratio, genetic

diversity, reproductive skew and number of offsprings for every beta).

iv. Selection

Selection folder contains scripts for the NWF_FC model with selection.

1. NWFLT_MutonX_Gamma

NWFLT_MutonX_Gamma folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with a **selection only happening on the X**, with deleterious coefficient following a **gamma distribution**. This model is looking at genetic **diversity** and population **size**. The folder contains the slim script, the workflow and two jupyter scripts:

- *Low_Selection_Analyze*: Analyze of the X-to-A ratio, genetic diversity, and population size in heat maps with low selection. (This File is the main one).

- *NWFLT_Selection_BSX_analyze*: Analyze of the X-to-A ratio, genetic diversity, and population size in heat maps with high selection.

v. RepSkew_Selection

RepSkew_Selection folder contains scripts for the NWF_FC model with selection and reproductive skew at the same time.

1. RepSkew_Beta

RepSkew_Beta folder contains scripts to run and analyze the diversity in **NWF_LT simulations** with a **selection only happening on the X** and a **reproductive skew in males**. This model is looking at genetic **diversity** and population **size**. The folder contains the slim script and the workflow. The simulations have not yet run.

vi. Pi_WF_NWFLT

Pi_WF_NWFLT folder contains the script to run and analyze the diversity in **NWF_LT**, **NWF_Allfemale** (LT with all individuals with female mortality), and **WF simulations** with **different sex-ratio**. The folder contains slim scripts for the three models, the workflow and five Jupyter scripts:

- *Corrected_Pi_WF_NWF_analyze*: Analyze of the diversity of every referent model (using the burn-in period of hybridization for NWF_LT).
- *Pi_All_NWF_analyze*: Analyze of the diversity of every referent model (using NWF_LT simulations for NWF_LT), for different sex-ratio. This file needs the Pi_WF_NWFLT and Pi_WF_NWFVan to be run.
- *Diversity_NWFLT_Analyze*: Value of diversity and effective population size of NWF_LT model (using NWF_LT simulations).
- *Pi_analysewithage*: Analyze of diversity of NWF_LT and WF simulations for different sexratio.
- *WithAge_allfemale*: Analyze of diversity of NWF_AllFemale and WF simulations for different sexratio.

vii. *NWF_AoD_NOff_AD*

NWF_AoD_NOff_AD folder contains the script to run and analyze the diversity and demographic data in **NWF_LT**, **NWF_V** and **NWF_FC simulations**. The folder contains slim scripts for the three models, the workflow and four Jupyter scripts:

- *NWF_CalcRepSkew*: Estimations of every model reproductive skew, and of border models reproductive skew.
- *NWFFC_demog_analyse*: Analyze of the age of death, number of offsprings and age distribution of the NWF_FC model.
- *NWFLT_demog_analyse*: Analyze of the age of death, number of offsprings and age distribution of the NWF_LT model.
- *NWFFV_demog_analyse*: Analyze of the age of death, number of offsprings and age distribution of the NWF_V model.

e. *Reel_Data*

Reel_Data folder contains the scripts files for the diversity analysis of baboon genetic reel data. This folder contains the jupyter file (that analyzes the diversity and the X-to-Chr ratio of each chromosome in a graph), and two sh files (they must be run before running the workflow of subfolders):

- *Fasta_division*: Divides one fasta file with many chromosomes into many fasta files for every chromosome. It also puts every nucleotide in capital letter and removes all the problems.
- *Fasta_line*: Make the sequence in a fasta file be a line.

i. All

All folder contains the script to analyze the **bcf files** of **both the anubis and yellow baboons data**. The bcf files and fata files are used to create a population in SLiM in which we can calculate the diversity. We do this for every chromosome. The folder contains the slim script, a sh file that calculates the number of sampled baboons (it must be run before the workflow), the workflow and one python script that removes every non-SNP mutation on the VCF files.

ii. Anubis

Anubis folder contains the script to analyze the **bcf files** of **the anubis baboon data**. The folder has the same composition as the *All* folder.

iii. Yellow

Yellow folder contains the script to analyze the **bcf files** of **the anubis baboon data**. The folder has the same composition as the *All* folder, except two additional sh files that count the number of nucleotides in a fasta file sequence.

f. Old_Files

Old_Files folder contains some previous scripts files and folders that are not used anymore. They are not updated.

g. Test

Test folder contains some test files and folders that are not relevant for the project.

II. tmp

tmp folder contains all the outputs of the workflows. The structure of *tmp* is the same as the *scripts* one except for some differences. The differences are:

- *Old_Files* is named *Old_results*.
- An *Old* folder is in *Reel_Data*, which contains previous results.
- There is no *Test* folder.
- The outputs of *NWV_test_size* are in *Old_results*.

III. done

done folder contains all the figures from the jupyter analyzes in *scripts*. The structure of *done* is the same as the *scripts* one except for some differences. The differences are:

- *Migrations* of *NWF_FemaleChoice*, *RepSkew_Selection*, *NWFFV_Test_Size* and *Test_Pop15x50_WF* do not exist.
- All hybridization results are in *Hybridization*
- All migration results are in *Migrations*, for both *NWF_V* and *NWF_LT* models.
- All selection results are in *Selection*
- All reel data results are in *Reel_Data*

IV. data

data folder contains all the BCF files, VCF files and fasta files of the baboon reel data. The structure of the subfolders: *Reel_Amboseli_Anubis*, *Reel_Amboseli_Yellow*, *Reel_Amboseli* are the same.

a. *Reel_Amboseli_Anubis*

Reel_Amboseli_Anubis folder contains all the files of the anubis baboons (baboons are not from Amboseli).

i. *BCF_Files*

BCF_Files folder contains the bcf files of every chromosome.

ii. *Fasta_Files*

Fasta_Files folder contains the fasta files of every chromosome.

iii. *Fasta_Files_Format*

Fasta_Files_Format folder contains the fasta files of every chromosome after the division and linearization.

iv. *VCF_Files*

VCF_Files folder contains the vcf files of every chromosome, created with their bcf files.

v. *VCF_Files_SNP*

VCF_Files_SNP folder contains the vcf files of every chromosome, when keeping only the SNP mutations.

b. *Reel_Amboseli_Yellow*

Reel_Amboseli_Yellow folder contains all the files of the yellow baboons (baboons are not from Amboseli). The structure of the folder is exactly the same as the *Reel_Amboseli_Anubis* one.

c. *Reel_Amboseli*

Reel_Amboseli folder contains all the files of both anubis and yellow baboons (baboons are not from Amboseli). The structure of the folder is exactly the same as the *Reel_Amboseli_Anubis* one.

d. *Recup*

Recup folder is a transitory folder where Files coming from Iker's folder are stocked.