Predicting Bat Recombination Rate with ReLERNN

Emma Wade

ReLERNN Installation

Requirements: Install Tensorflow with CUDA and CUDNN support

I pulled from these instructions https://gretel.ai/blog/install-tensorflow-with-cuda-cdnn-and-gpu-support-in-4-easy-steps (https://gretel.ai/blog/install-tensorflow-with-cuda-cdnn-and-gpu-support-in-4-easy-steps)

```
qrsh -l gpu,RTX2080Ti,h_rt=2:00:00,cuda=1
conda create --name=tf python=3.9
conda activate tf
conda install -c conda-forge cudatoolkit=11.2.2 cudnn=8.1.0
#tensorflow/2.2.0, cudatoolkit/10.1.243, and cudnn/7.6.5

mkdir -p $CONDA_PREFIX/etc/conda/activate.d
echo 'export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$CONDA_PREFIX/lib/' > $CONDA_PREFIX/etc/conda/activate.d/env_vars.sh

# Sign out and sign back in via SSH or close and re-open your terminal window. Reactivate your conda session.
qrsh -l gpu,RTX2080Ti,h_rt=2:00:00,cuda=1
conda activate tf
python3 -m pip install tensorflow==2.10 # pybind11>=2.12
#pip install "tensorflow==2:10[and-cuda] pybind11>=2.12" # on another day I had to use this to get it to work
pip install "numpy<2.0" # I also had to downgrade numpy

# Verify install and GPU link:
python3 -c "import os; os.environ['TF_CPP_MIN_LOG_LEVEL'] = '3'; import tensorflow as tf; print('Num GPUs Available: ', len(tf.config.list_physical_devices('GPU')))"

#Hopefully you get "Num GPUs Available: 1!!!
```

Install relernn *inside tf conda environment

```
git clone https://github.com/kr-colab/ReLERNN.git
cd ReLERNN
pip install .
```

Running ReLERNN Example

```
cd examples
#Added /u/home/e/eewade/software/ReLERNN/ReLERNN to beginning of ReLERN simulate, etc in
# pipeline file
./example_pipeline.sh
# Didn't end up finishing it but got like half way through training and seemed fine
```

Preparing data

Choosing four male and four female bats

```
library("readxl")
library("knitr")
bat_meta <- read_excel("/u/home/e/eewade/project-klohmuel/bats/20241007_MYYU_SampleMetadata.xlsx", sheet = 2)
# Choose 4 male and 4 female bats that were in 33ish latitude and -117ish longitude
bats_iwant_males = c("071007_hiddentr_myyu", "MYYU_CA2018_015", "MYYU_CA2018_017","MYYU_CA2018_036")
bats_iwant_females = c("MYYU_CA2008_004", "MYYU_CA2012_018", "MYYU_CA2017_003", "MYYU_CA2018_004")
kable(bat_meta[bat_meta$`Sample Name` %in% bats_iwant_females | bat_meta$`Sample Name` %in% bats_iwant_males,])
```

Sample Name	SPP	FA	Sex	Age	Latitude	Longitude	Coordinate Resolution	State/Country	Locality	Collection Year	Contributor	VCFTools Depth	Depth Bin	VCFTools Missingness	Extraction Protocol	Elution Buffer	S
071007_hiddentr_myyu	MYYU	NA	M	J	33.14682	-117.0088	Site	CA	Escondido, San Diego County, CA	2007	SDMNH	11.78640	NA	0.0403343	Omega MagBind Blood and Tissue	10mM TrisCl	G
MYYU_CA2008_004	MYYU	NA	F	A	37.42830	-121.9790	Site	CA	Alviso Cannery, Santa Clara County, CA	2008	D. Johnson	17.98880	NA	0.0103850	Omega MagBind Blood and Tissue	10mM TrisCl	N
MYYU_CA2012_018	MYYU	NA	F	A	39.99049	-122.1735	Site	CA	Dye Creek Preserve, Tehama County, CA	2012	D. Johnson	15.35340	NA	0.0126671	Qiagen DNeasy Blood and Tissue	10mM TrisCl	N
MYYU_CA2017_003	MYYU	NA	F	A	38.50308	-121.9875	Site	CA	Putah Creek Oxbow, Solano Count, CA	2017	Harris	6.98649	NA	0.2930920	Omega MagBind Blood and Tissue	10mM TrisCl	A
MYYU_CA2018_004	MYYU	NA	F	A	33.99884	-119.7151	Site	CA	Cañada Del Puerto, Santa Cruz Island, Santa Barbara County, CA	2018	P. Brown	9.17378	NA	0.1715920	Omega MagBind Blood and Tissue	10mM TrisCl	N
MYYU_CA2018_015	MYYU	NA	M	J	33.63797	-117.6002	Site	CA	Rancho Santa Margarita, Orange County, CA	2018	CDPH	11.92260	NA	0.0432267	Omega MagBind Blood and Tissue	10mM TrisCl	T
MYYU_CA2018_017	MYYU	NA	M	J	33.92152	-117.8838	City	CA	Brea, Orange County, CA	2018	CDPH	8.20283	NA	0.1913700	Omega MagBind Blood and Tissue	10mM TrisCl	N
MYYU_CA2018_036	MYYU	NA	M	A	33.78748	-117.8525	City	CA	Orange, Orange County, CA	2018	CDPH	14.00400	NA	0.0187172	Omega MagBind Blood and Tissue	10mM TrisCl	A

Pulling the bats from the bigger VCF file

```
qrsh -l h_data=10G,h_rt=6:00:00
module load bcftools
#this took a while
bcftools view -s Myoyum_071007_hiddentr_myyu,Myoyum_MYYU_CA2018_015,Myoyum_MYYU_CA2018_017,Myoyum_MYYU_CA2018_036 \
   /u/home/e/eewade/project-klohmuel/bats/GCA_028538775.1_mMyoYum1.0.hap1_PassSNPs_115samples.vcf.gz \
   -Oz -o /u/home/e/eewade/project-klohmuel/bats/male_bats.vcf.gz

bcftools view -s Myoyum_MYYU_CA2008_004,Myoyum_MYYU_CA2012_018,Myoyum_MYYU_CA2017_003,Myoyum_MYYU_CA2018_004 \
   /u/home/e/eewade/project-klohmuel/bats/GCA_028538775.1_mMyoYum1.0.hap1_PassSNPs_115samples.vcf.gz \
   -Oz -o /u/home/e/eewade/project-klohmuel/bats/female_bats.vcf.gz
```

Formatted the bed file of chromosome lengths

```
cat "/u/home/e/eewade/project-klohmuel/bats/chrom_lengths-copy.bed"
```

```
JAPQVT010000001.1  0  240344003
JAPQVT010000002.1  0  217333615
JAPQVT010000003.1  0  215164148
JAPQVT010000005.1  0  108188207
JAPQVT010000006.1  0  99144700
JAPQVT010000007.1  0  93950265
JAPQVT010000008.1  0  93940565
JAPQVT010000009.1  0  85204546
JAPQVT010000010.1  0  84518507
JAPQVT010000011.1  0  73253886
JAPQVT010000012.1  0  62293450
JAPQVT010000013.1  0  60843581
JAPQVT010000014.1  0  58814807
JAPQVT010000015.1  0  54424144
JAPQVT010000016.1  0  53638781
JAPQVT010000017.1  0  44179438
JAPQVT010000018.1  0  41023509
JAPQVT010000019.1  0  28724713
JAPQVT010000020.1  0  20234284
JAPQVT010000021.1  0  19002367
JAPQVT010000022.1  0  17375685
JAPQVT010000023.1  0  15528544
JAPQVT010000024.1  0  12354025
JAPQVT010000025.1  0  2204611
JAPQVT010000026.1  0  1950662
JAPQVT010000027.1  0  1305348
```

Running ReLERNN with Bat Data

Module 1 : Simulate

```
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_simulate_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_simulate_$TASK_ID.eo
#$ -j y
## Edit the line below as needed:
#$ -l h_data=75G,h_rt=23:59:00
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 4
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-2:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/male_bats.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput/"
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/female_bats.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput/"
fi

echo $VCF

SIMULATE="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_SIMULATE"
SEED="42"
MU="2.366e-9" # mutation rate
GENTIME=2 # generation time in years
NCPU=4
URTR="1" # default
GENOME="/u/home/e/eewade/project-klohmuel/bats/chrom_lengths-copy.bed"

# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf

# Simulate data
${SIMULATE} \
--vcf ${VCF} \
--genome ${GENOME} \
-d ${DIR} \
-u ${MU} \
-l ${GENTIME} \
--upperRhoThetaRatio ${URTR} \
--seed ${SEED}
```

Job-array task 5959754.1 (bats_simulate.sh) Complete User = eewade Queue = pod_smp.q@n1080 (mailto:pod_smp.q@n1080) Host = n1080.hoffman2.idre.ucla.edu Start Time = 11/01/2024 11:05:00.172 End Time = 11/01/2024 11:18:25.742 User Time = 00:53:08 System Time = 00:05:10 Wallclock Time = 00:13:25 CPU = 00:58:19 Max vmem = 260.851G Max rss = NA Exit Status = 0

Module 2 : Train

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_train_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_train_$TASK_ID.eo
## Edit the line below as needed:
#$ -l h_data=10G,h_rt=23:59:00
#$ -l A100,gpu,gpu_mem=80G,cuda=2
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 64
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-2:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/male_bats.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput/"
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/female_bats.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput/"
fi

echo $VCF

SIMULATE="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_SIMULATE"
TRAIN="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_TRAIN"
PREDICT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_PREDICT"
BSCORRECT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_BSCORRECT"
SEED="42"
MU="2.366e-9"
GENTIME=2
NCPU=64
URTR="1"
GENOME="/u/home/e/eewade/project-klohmuel/bats/chrom_lengths-copy.bed"

# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf

# Train network
${TRAIN} \
--projectDir ${DIR} \
--seed ${SEED} \
-t ${NCPU}
```

Job-array task 5961870.2 (bats_train.sh) Complete User = eewade Queue = gpu_a100.q@g13 (mailto:gpu_a100.q@g13) Host = g13.hoffman2.idre.ucla.edu Start Time = 11/01/2024 17:55:14.685 End Time = 11/01/2024 23:42:56.685 User Time = 03:17:41 System Time = 10:48:45 Wallclock Time = 05:47:42 CPU = 14:06:27 Max vmem = 26943.599G Max rss = NA Exit Status = 0

Training results

```
### Female bats
knitr::include_graphics("/u/project/klohmuel/eewade/bats/femalebats_trainingresults.pdf")
```

```
### Male bats
knitr::include_graphics("/u/project/klohmuel/eewade/bats/malebats_trainingresults.pdf")
```

Module 3 and 4: Predict and Correct

```
if [[ "$SGE_TASK_ID" -eq 2 ]]; then
```

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_predict_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_predict_$TASK_ID.eo
## Edit the line below as needed:
#$ -l h_data=10G,h_rt=23:59:00
#$ -l A100,gpu,gpu_mem=80G,cuda=1
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 8
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-2:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/male_bats.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput/"
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/female_bats.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput/"
fi

echo $VCF

SIMULATE="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_SIMULATE"
TRAIN="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_TRAIN"
PREDICT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_PREDICT"
BSCORRECT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_BSCORRECT"
SEED="42"
MU="2.366e-9" # mutation rate
GENTIME=2 # generation time in years
NCPU=4
URTR="1" # default
GENOME="/u/home/e/eewade/project-klohmuel/bats/chrom_lengths-copy.bed"

# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf

# Predict
${PREDICT} \
  --vcf ${VCF} \
  --projectDir ${DIR} \
  --seed ${SEED}

# Parametric Bootstrapping
${BSCORRECT} \
  --projectDir ${DIR} \
  --nSlice 2 \
  --nReps 2 \
  --seed ${SEED}
```

Job-array task 5983120.2 (bats_predict_correct.sh) Complete User = eewade Queue = gpu_a100.q@g13 (mailto:gpu_a100.q@g13) Host = g13.hoffman2.idre.ucla.edu Start Time = 11/04/2024 11:34:05.997 End Time = 11/04/2024 11:37:51.597 User Time = 00:03:18 System Time = 00:01:27 Wallclock Time = 00:03:45 CPU = 00:04:46 Max vmem = 30791.806G Max rss = NA Exit Status = 0 ### Plot Results

```r
library(data.table)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.3     ✓ dplyr   1.0.7
## ✓ tidyr   1.1.3     ✓ stringr 1.4.0
## ✓ readr   2.0.0     ✓ forcats 0.5.1
```

```
## ── Conflicts ──────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

```r
male = fread("/u/project/klohmuel/eewade/bats/male_bats.PREDICT.BSCORRECTED.20241104.txt",h=T)
male$sex = "Male"
mean(male$recombRate)
```

```
## [1] 5.534073e-11
```

```r
female = fread("/u/project/klohmuel/eewade/bats/female_bats.PREDICT.BSCORRECTED.20241104.txt",h=T)
female$sex = "Female"
mean(female$recombRate)
```

```
## [1] 1.354759e-13
```

```r
allbats = rbind(male, female)

ggplot(allbats[allbats$chrom %in% c("b'JAPQVT010000001.1'"),], aes(x = start, y = recombRate, color=sex)) +
  geom_point(alpha=0.3) +
  geom_errorbar(aes(ymin = CI95LO, ymax = CI95HI), alpha = 0.3) +
  labs(title = "Recombination Rate Across JAPQVT01000001.1",
       x = "Position",
       y = "Recombination Rate") +
  theme_minimal()
```

Recombination Rate Across JAPQVT01000001.1
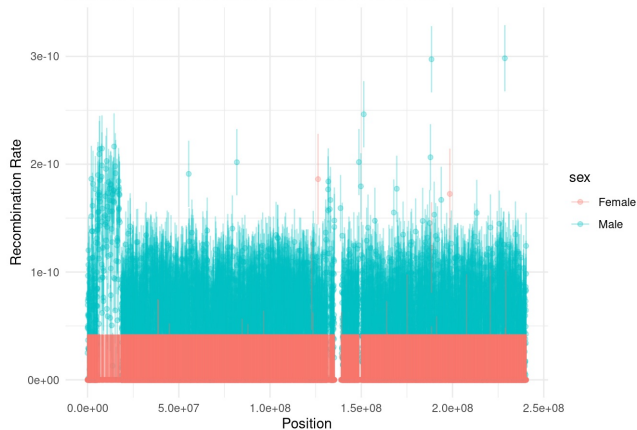
Second Attempt with More Samples, Less Chromosomes

```r
library("readxl")
library("knitr")
bat_meta <- read_excel("/u/home/e/eewade/project-klohmuel/bats/20241007_MYYU_SampleMetadata.xlsx", sheet = 2)
bat_meta$`Sample Name`[bat_meta$Sex == "M"]
```

```
##  [1] "043017_harbison_myyu"        "071007_hiddentr_myyu"
##  [3] "092007_chulavista_myyu"      "092020_elcajon_myyu"
##  [5] "092221_catalinabarn_myyu1"   "102121_whitewater_myyu1"
##  [7] "15A"                         "20230722_SteelheadPreserve_myyu1"
##  [9] "hartpark_093021_myyu1"       "MVZ146766a_MYVE"
## [11] "MYYU_CA2018_015"             "MYYU_CA2018_017"
## [13] "MYYU_CA2018_036"             "MYYU_CA2020_CCGP"
## [15] "MYYU_CA2021_003"             "MYYU_CA2021_007"
## [17] "MYYU_CA2021_034"             "NE90_A_MYYU"
## [19] "TK163190"                    "TK199418"
## [21] "TK48571"
```

```r
bat_meta$`Sample Name`[bat_meta$Sex == "F"]
```

```
##  [1] "08082021_patandbill1_myyu"  "08082021_patandbill6_myyu"
##  [3] "090607_chulavista_myyu"     "092807_poway_myyu"
##  [5] "20210616_MODOC_myyu_ca1"    "20230823_Temecula"
##  [7] "26A"                        "DP87_myyu"
##  [9] "DP96_myyu"                  "DSJ_15"
## [11] "DSJ_9"                      "JW06_A"
## [13] "MK33_A"                     "MYYU_CA2008_004"
## [15] "MYYU_CA2012_018"            "MYYU_CA2017_003"
## [17] "MYYU_CA2018_004"            "MYYU_CA2018_014"
## [19] "MYYU_CA2018_034"            "MYYU_CA2018_040"
## [21] "MYYU_CA2018_043"            "MYYU_CA2018_046"
## [23] "MYYU_CA2018_048"            "MYYU_CA2018_056"
## [25] "MYYU_CA2018_062"            "MYYU_CA2018_063"
## [27] "MYYU_CA2018_065"            "MYYU_CA2021_035"
## [29] "MYYU_CA2021_042"            "MYYU_CA2021_046"
## [31] "TD9845"                     "TK186214"
## [33] "TK48589"                    "UAM_Mamm_120402"
## [35] "UH20_A"                     "WA351"
## [37] "WA358"
```

Pulling second attempt bats – 21 each

```
qrsh -l h_data=10G,h_rt=6:00:00
module load bcftools
bcftools index /u/home/e/eewade/project-klohmuel/bats/GCA_028538775.1_mMyoYum1.0.hap1_PassSNPs_115samples.vcf.gz
# I guess needed to use --regions?
#this took a while
# Males
bcftools view -s Myoyum_043017_harbison_myyu,Myoyum_071007_hiddentr_myyu,Myoyum_092007_chulavista_myyu,Myoyum_092
020_elcajon_myyu,092221_catalinabarn_myyu1,Myoyum_102121_whitewater_myyu1,Myoyum_15A,Myoyum_20230722_SteelheadPre
serve_myyu1,Myoyum_hartpark_093021_myyu1,Myovel_MVZ146766a_MYVE,Myoyum_MYYU_CA2018_015,Myoyum_MYYU_CA2018_017,Myo
yum_MYYU_CA2018_036,Myoyum_MYYU_CA2020_CCGP,Myoyum_MYYU_CA2021_003,Myoyum_MYYU_CA2021_007,Myoyum_MYYU_CA2021_034,
Myoyum_NE90_A_MYYU,Myoyum_TK163190,Myoyum_TK199418,Myoyum_TK48571  \
  --regions JAPQVT010000001.1,JAPQVT010000002.1,JAPQVT010000003.1,JAPQVT010000005.1 \
  /u/home/e/eewade/project-klohmuel/bats/GCA_028538775.1_mMyoYum1.0.hap1_PassSNPs_115samples.vcf.gz \
  -Oz -o /u/home/e/eewade/project-klohmuel/bats/all_male_bats_4chrs.vcf.gz

#Females
bcftools view -s 08082021_patandbill1_myyu,Myoyum_08082021_patandbill6_myyu,Myoyum_090607_chulavista_myyu,Myoyum_
092807_poway_myyu,20210616_MODOC_myyu_ca1,Myoyum_20230823_Temecula,Myoyum_26A,Myoyum_DP87_myyu,Myoyum_DP96_myyu,M
yoyum_DSJ_15,Myoyum_DSJ_9,Myoyum_JW06_A,Myoyum_MK33_A,Myoyum_MYYU_CA2008_004,Myoyum_MYYU_CA2012_018,Myoyum_MYYU_C
A2017_003,Myoyum_MYYU_CA2018_004,Myoyum_MYYU_CA2018_014,Myoyum_MYYU_CA2018_034,Myoyum_MYYU_CA2018_040,Myoyum_MYYU
_CA2018_043 \
  --regions JAPQVT010000001.1,JAPQVT010000002.1,JAPQVT010000003.1,JAPQVT010000005.1 \
  /u/home/e/eewade/project-klohmuel/bats/GCA_028538775.1_mMyoYum1.0.hap1_PassSNPs_115samples.vcf.gz \
  -Oz -o /u/home/e/eewade/project-klohmuel/bats/all_female_bats_4chrs.vcf.gz
```

Simulate

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_simulate_2_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_simulate_2_$TASK_ID.eo
#$ -j y
## Edit the line below as needed:
#$ -l h_data=75G,h_rt=23:59:00
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 4
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-2:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_4chrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_2/"
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_4chrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_2/"
fi

echo $VCF

SIMULATE="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_SIMULATE"
SEED="42"
MU="2.366e-9" # mutation rate
GENTIME=2 # generation time in years
NCPU=4
URTR="1" # default
GENOME="/u/home/e/eewade/project-klohmuel/bats/selected_chr_lengths.bed"

# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf

# Simulate data
${SIMULATE} \
--vcf ${VCF} \
--genome ${GENOME} \
-d ${DIR} \
-u ${MU} \
-l ${GENTIME} \
--upperRhoThetaRatio ${URTR} \
--seed ${SEED}
```

Job-array task 6083278.2 (bats_simulate.sh) Complete User = eewade Queue = msa_smp.q@n7062 (mailto:msa_smp.q@n7062) Host = n7062.hoffman2.idre.ucla.edu Start Time = 11/11/2024 17:21:24.256 End Time = 11/11/2024 17:41:15.535 User Time = 00:45:47 System Time = 00:03:14 Wallclock Time = 00:19:51 CPU = 00:49:01 Max vmem = 187.572G Max rss = NA Exit Status = 0 ### Train

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_train_2_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_train_2_$TASK_ID.eo
## Edit the line below as needed:
#$ -l h_data=10G,h_rt=23:59:00
#$ -l A100,gpu,gpu_mem=80G,cuda=2
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 64
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-2:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_4chrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_2/"
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_4chrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_2/"
fi

echo $VCF

SIMULATE="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_SIMULATE"
TRAIN="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_TRAIN"
PREDICT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_PREDICT"
BSCORRECT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_BSCORRECT"
SEED="42"
MU="2.366e-9"
GENTIME=2
NCPU=64
URTR="1"
GENOME="/u/home/e/eewade/project-klohmuel/bats/selected_chr_lengths.bed"

# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf

# Train network
${TRAIN} \
--projectDir ${DIR} \
--seed ${SEED} \
-t ${NCPU}
```

Job-array task 6083611.1 (bats_simulate.sh) Complete User = eewade Queue = gpu_a100.q@g13 (mailto:gpu_a100.q@g13) Host = g13.hoffman2.idre.ucla.edu Start Time = 11/11/2024 21:21:48.964 End Time = 11/12/2024 00:25:26.283 User Time = 02:23:59 System Time = 06:40:20 Wallclock Time = 03:03:37 CPU = 09:04:19 Max vmem = 31465.695G Max rss = NA Exit Status = 0 #### Training results

```r
### Female bats
knitr::include_graphics("/u/project/klohmuel/eewade/bats/femalebats_trainingresults2.pdf")
```

```bash
if [[ "$SGE_TASK_ID" -eq 2 ]]; then
```

```
### Male bats
knitr::include_graphics("/u/project/klohmuel/eewade/bats/malebats_trainingresults2.pdf")
```

Predict and Correct

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_predict_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_predict_$TASK_ID.eo
## Edit the line below as needed:
#$ -l h_data=10G,h_rt=23:59:00
#$ -l -l gpu,A6000,cuda=1
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 8
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-2:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_4chrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_2/"
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_4chrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_2/"
fi

echo $VCF

SIMULATE="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_SIMULATE"
TRAIN="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_TRAIN"
PREDICT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_PREDICT"
BSCORRECT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_BSCORRECT"
SEED="42"
MU="2.366e-9"
GENTIME=2
NCPU=64
URTR="1"
GENOME="/u/home/e/eewade/project-klohmuel/bats/selected_chr_lengths.bed"
# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf

# Predict
${PREDICT} \
  --vcf ${VCF} \
  --projectDir ${DIR} \
  --seed ${SEED}

# Parametric Bootstrapping
${BSCORRECT} \
  --projectDir ${DIR} \
  --nSlice 2 \
  --nReps 2 \
  --seed ${SEED}
```

ob-array task 6117114.2 (bats_simulate.sh) Complete User = eewade Queue = gpu_a6000.q@g14 (mailto:gpu_a6000.q@g14) Host = g14.hoffman2.idre.ucla.edu Start Time = 11/12/2024 07:47:10.753 End Time = 11/12/2024 07:49:27.147 User Time = 00:01:04 System Time = 00:00:30 Wallclock Time = 00:02:16 CPU = 00:01:35 Max vmem = 1089.085G Max rss = NA Exit Status = 0 ### Visualize Results

```r
library(data.table)
library(tidyverse)
male = fread("/u/project/klohmuel/eewade/bats/all_male_bats_4chrs.PREDICT.BSCORRECTED.20241112.txt",h=T)
male$sex = "Male"
mean(male$recombRate)
```

```
## [1] 4.269323e-11
```

```r
female = fread("/u/project/klohmuel/eewade/bats/all_female_bats_4chrs.PREDICT.BSCORRECTED.20241112.txt",h=T)
female$sex = "Female"
mean(female$recombRate)
```

```
## [1] 6.425288e-11
```

```r
allbats = rbind(male, female)

ggplot(allbats[allbats$chrom == "b'JAPQVT010000001.1'" & allbats$sex == "Male",], aes(x = start, y = recombRate, color=sex)) +
  geom_point(alpha=0.5) +
  facet_wrap(vars(chrom),scales = "free") +
  geom_errorbar(aes(ymin = CI95LO, ymax = CI95HI), alpha = 0.3) +
  labs(x = "Position",
       y = "Recombination Rate") +
  theme_minimal()
```

b'JAPQVT010000001.1'
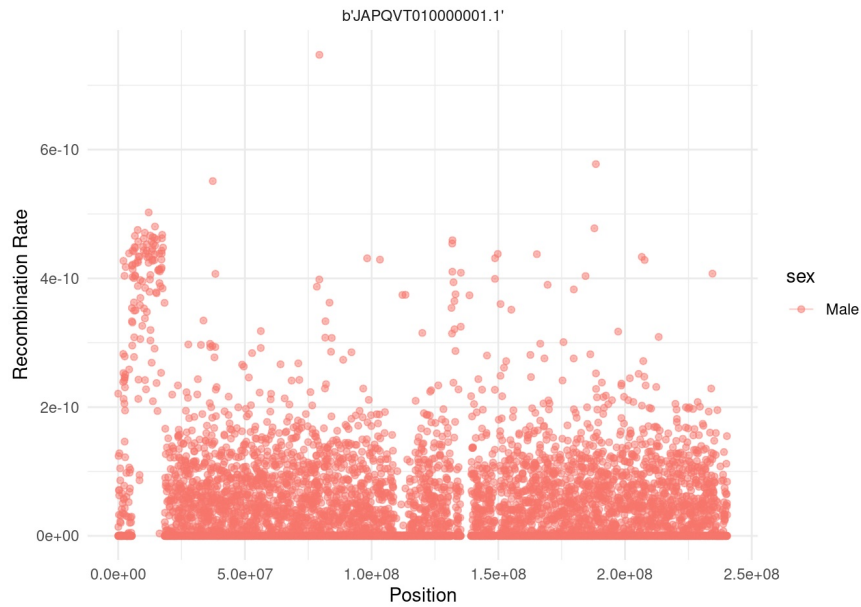
Try #3 and #4 and #5 w/ higher, regular, and lower mutation rate, all bats, all chromosomes

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/makevcfs_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/makevcfs_$TASK_ID.eo
## Edit the line below as needed:
#$ -l h_data=10G,h_rt=23:59:00
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-2:1

# load the job environment:
. /u/local/Modules/default/init/modules.sh
module load bcftools

if [ "$SGE_TASK_ID" -eq 1 ]; then
#this took a while
# Males
bcftools view -s Myoyum_043017_harbison_myyu,Myoyum_071007_hiddentr_myyu,Myoyum_092007_chulavista_myyu,Myoyum_092
020_elcajon_myyu,092221_catalinabarn_myyu1,Myoyum_102121_whitewater_myyu1,Myoyum_15A,Myoyum_20230722_SteelheadPre
serve_myyu1,Myoyum_hartpark_093021_myyu1,Myovel_MVZ146766a_MYVE,Myoyum_MYYU_CA2018_015,Myoyum_MYYU_CA2018_017,Myo
yum_MYYU_CA2018_036,Myoyum_MYYU_CA2020_CCGP,Myoyum_MYYU_CA2021_003,Myoyum_MYYU_CA2021_007,Myoyum_MYYU_CA2021_034,
Myoyum_NE90_A_MYYU,Myoyum_TK163190,Myoyum_TK199418,Myoyum_TK48571  \
  /u/home/e/eewade/project-klohmuel/bats/GCA_028538775.1_mMyoYum1.0.hap1_PassSNPs_115samples.vcf.gz \
  -Ov -o /u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
#Females
bcftools view -s 08082021_patandbill1_myyu,Myoyum_08082021_patandbill6_myyu,Myoyum_090607_chulavista_myyu,Myoyum_
092807_poway_myyu,20210616_MODOC_myyu_ca1,Myoyum_20230823_Temecula,Myoyum_26A,Myoyum_DP87_myyu,Myoyum_DP96_myyu,M
yoyum_DSJ_15,Myoyum_DSJ_9,Myoyum_JW06_A,Myoyum_MK33_A,Myoyum_MYYU_CA2008_004,Myoyum_MYYU_CA2012_018,Myoyum_MYYU_C
A2017_003,Myoyum_MYYU_CA2018_004,Myoyum_MYYU_CA2018_014,Myoyum_MYYU_CA2018_034,Myoyum_MYYU_CA2018_040,Myoyum_MYYU
_CA2018_043 \
  /u/home/e/eewade/project-klohmuel/bats/GCA_028538775.1_mMyoYum1.0.hap1_PassSNPs_115samples.vcf.gz \
  -Ov -o /u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf
fi
```

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_simulate_3_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_simulate_3_$TASK_ID.eo
#$ -j y
## Edit the line below as needed:
#$ -l h_data=75G,h_rt=23:59:00
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 4
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-6:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_lowmu/"
MU="2.366e-10" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_lowmu/"
MU="2.366e-10" # mutation rate
fi

if [ "$SGE_TASK_ID" -eq 3 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_regmu/"
MU="2.366e-9" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 4 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_regmu/"
MU="2.366e-9" # mutation rate
fi

if [ "$SGE_TASK_ID" -eq 5 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_highmu/"
MU="2.366e-8" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 6 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_highmu/"
MU="2.366e-8" # mutation rate
fi

echo $VCF

SIMULATE="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_SIMULATE"
SEED="42"
GENTIME=2 # generation time in years
NCPU=4
URTR="1" # default
GENOME="/u/home/e/eewade/project-klohmuel/bats/chrom_lengths-copy.bed"

# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf


# Simulate data
${SIMULATE} \
--vcf ${VCF} \
--genome ${GENOME} \
-d ${DIR} \
-u ${MU} \
-l ${GENTIME} \
--upperRhoThetaRatio ${URTR} \
--seed ${SEED}
```

Note: Ran out of memory on the 48G GPU nodes :/

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_train_3_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_train_3_$TASK_ID.eo
#$ -j y
## Edit the line below as needed:
#$ -l h_data=10G,h_rt=23:59:00
#$ -l A100,gpu,gpu_mem=80G,cuda=2
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 64
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-6:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_lowmu/"
MU="2.366e-10" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_lowmu/"
MU="2.366e-10" # mutation rate
fi

if [ "$SGE_TASK_ID" -eq 3 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_regmu/"
MU="2.366e-9" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 4 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_regmu/"
MU="2.366e-9" # mutation rate
fi

if [ "$SGE_TASK_ID" -eq 5 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_highmu/"
MU="2.366e-8" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 6 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_highmu/"
MU="2.366e-8" # mutation rate
fi

echo $VCF

TRAIN="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_TRAIN"
SEED="42"
GENTIME=2 # generation time in years
NCPU=64
URTR="1" # default
GENOME="/u/home/e/eewade/project-klohmuel/bats/chrom_lengths-copy.bed"

# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf

# Train network
${TRAIN} \
  --projectDir ${DIR} \
  --seed ${SEED} \
  -t ${NCPU}
```

qsub -hold_jid 12345 job2.sh

```bash
#!/bin/bash
#$ -cwd
# error = Merged with joblog
#$ -e /u/scratch/e/eewade/logs/bats_predict_3_$TASK_ID.eo
#$ -o /u/scratch/e/eewade/logs/bats_predict_3_$TASK_ID.eo
#$ -j y
## Edit the line below as needed:
#$ -l h_data=10G,h_rt=23:59:00
#$ -l gpu,A6000,cuda=1
## Modify the parallel environment
## and the number of cores as needed:
#$ -pe shared 8
# Email address to notify
#$ -M eew226@g.ucla.edu
# Notify when
#$ -m bea
#$ -t 1-6:1

if [ "$SGE_TASK_ID" -eq 1 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_lowmu/"
MU="2.366e-10" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 2 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_lowmu/"
MU="2.366e-10" # mutation rate
fi

if [ "$SGE_TASK_ID" -eq 3 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_regmu/"
MU="2.366e-9" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 4 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_regmu/"
MU="2.366e-9" # mutation rate
fi

if [ "$SGE_TASK_ID" -eq 5 ]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_male_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/malebatsoutput_highmu/"
MU="2.366e-8" # mutation rate
fi

if [[ "$SGE_TASK_ID" -eq 6 ]]; then
VCF="/u/home/e/eewade/project-klohmuel/bats/all_female_bats_allchrs.vcf"
DIR="/u/scratch/e/eewade/femalebatsoutput_highmu/"
MU="2.366e-8" # mutation rate
fi

echo $VCF

BSCORRECT="/u/home/e/eewade/software/ReLERNN/ReLERNN/ReLERNN_BSCORRECT"
SEED="42"
GENTIME=2 # generation time in years
NCPU=8
URTR="1" # default
GENOME="/u/home/e/eewade/project-klohmuel/bats/chrom_lengths-copy.bed"

# load the job environment:
. /u/local/Modules/default/init/modules.sh
## Edit the line below as needed:
module load mamba

mamba activate tf

# Predict
${PREDICT} \
  --vcf ${VCF} \
  --projectDir ${DIR} \
  --seed ${SEED}

# Parametric Bootstrapping
${BSCORRECT} \
  --projectDir ${DIR} \
  --nSlice 2 \
  --nReps 2 \
  --seed ${SEED}
```

Training results

```
###### Low mu #######
### Female bats
knitr::include_graphics("/u/scratch/e/eewade/femalebatsoutput_lowmu/networks/testResults.pdf")
```

```
### Male bats
knitr::include_graphics("/u/scratch/e/eewade/malebatsoutput_lowmu/networks/testResults.pdf")
```

```
###### Reg mu ######
### Female bats
knitr::include_graphics("/u/scratch/e/eewade/femalebatsoutput_regmu/networks/testResults.pdf")
```

```
### Male bats
knitr::include_graphics("/u/scratch/e/eewade/malebatsoutput_regmu/networks/testResults.pdf")
```

```
###### High mu ######
### Female bats
knitr::include_graphics("/u/scratch/e/eewade/femalebatsoutput_highmu/networks/testResults.pdf")
```

```
### Male bats
knitr::include_graphics("/u/scratch/e/eewade/malebatsoutput_highmu/networks/testResults.pdf")
```

Move results

```
mkdir /u/project/klohmuel/eewade/bats/predictions-varymu-20241122/

cp /u/scratch/e/eewade/femalebatsoutput_highmu/all_female_bats_allchrs.PREDICT.BSCORRECTED.txt /u/project/klohmue
l/eewade/bats/predictions-varymu-20241122/highmu_females.PREDICT.BSCORRECTED.txt

cp /u/scratch/e/eewade/malebatsoutput_highmu/all_male_bats_allchrs.PREDICT.BSCORRECTED.txt /u/project/klohmuel/ee
wade/bats/predictions-varymu-20241122/highmu_males.PREDICT.BSCORRECTED.txt

cp /u/scratch/e/eewade/femalebatsoutput_regmu/all_female_bats_allchrs.PREDICT.BSCORRECTED.txt /u/project/klohmuel
/eewade/bats/predictions-varymu-20241122/regmu_females.PREDICT.BSCORRECTED.txt

cp /u/scratch/e/eewade/malebatsoutput_regmu/all_male_bats_allchrs.PREDICT.BSCORRECTED.txt /u/project/klohmuel/eew
ade/bats/predictions-varymu-20241122/regmu_males.PREDICT.BSCORRECTED.txt

cp /u/scratch/e/eewade/femalebatsoutput_lowmu/all_female_bats_allchrs.PREDICT.BSCORRECTED.txt /u/project/klohmuel
/eewade/bats/predictions-varymu-20241122/lowmu_females.PREDICT.BSCORRECTED.txt

cp /u/scratch/e/eewade/malebatsoutput_lowmu/all_male_bats_allchrs.PREDICT.BSCORRECTED.txt /u/project/klohmuel/eew
ade/bats/predictions-varymu-20241122/lowmu_males.PREDICT.BSCORRECTED.txt
```

```python
import pandas as pd
import numpy as np
import os

rhorates = pd.DataFrame()

for sex in ["females", "males"]:
    for mu in ["low", "reg", "high"]:
        file_path = f"/u/project/klohmuel/eewade/bats/predictions-varymu-20241122/{mu}mu_{sex}.PREDICT.BSCORRECTE
D.txt"
        # Read the file
        thisone = pd.read_csv(file_path, sep="\t")  # Assuming tab-separated, change if needed
        print(f"Sex: {sex}, Mu: {mu}, Rho average: {thisone['recombRate'].mean()}")
        print(f"Chromosomes present: {', '.join(thisone['chrom'].unique())}")
        # Add new columns
        thisone['sex'] = sex
        thisone['mu'] = mu
        # Concatenate to the main dataframe
        rhorates = pd.concat([rhorates, thisone], ignore_index=True)

# Function to create overlapping windows
def sliding_window(data, window_size, step_size):
    for start in range(0, data['end'].max(), step_size):
        end = start + window_size
        yield data[(data['start'] >= start) & (data['end'] <= end)]

# Set parameters
window_size = 20_000_000  # 20 Mb
step_size = 50_000  # 50 kb

# Create a list to store results
results = []

# Group by chromosome and sex and mu
for (chrom, sex, mu), group in rhorates.groupby(['chrom', 'sex', 'mu']):
    for window in sliding_window(group, window_size, step_size):
        if not window.empty:
            avg_rate = np.average(window['recombRate'], weights=window['nSites'])
            avg_ci_lo = np.average(window['CI95LO'], weights=window['nSites'])
            avg_ci_hi = np.average(window['CI95HI'], weights=window['nSites'])
            results.append({
                'chrom': chrom,
                'start': window['start'].min(),
                'end': window['end'].max(),
                'avg_recombRate': avg_rate,
                'avg_CI95LO': avg_ci_lo,
                'avg_CI95HI': avg_ci_hi,
                'sex': sex,
                'mu': mu
            })

# Create a new DataFrame from results
new_df = pd.DataFrame(results)

# Sort the DataFrame
new_df = new_df.sort_values(['chrom', 'start', 'sex', 'mu'])

# Write to a new file
new_df.to_csv('/u/project/klohmuel/eewade/bats/predictions-varymu-20241122/myaveraged_20mbwindow_50kbstep_recombi
nation_rates_with_CI.csv', index=False)
```

Plot Results

```
library(data.table)
library(tidyverse)
library(MetBrewer)

rhorates = fread('/u/project/klohmuel/eewade/bats/predictions-varymu-20241122/myaveraged_20mbwindow_50kbstep_reco
mbination_rates_with_CI.csv', h=T)

rhorates$realmu = "mu = 2.366e-9"
rhorates$realmu[rhorates$mu == "low"] = "mu = 2.366e-10"
rhorates$realmu[rhorates$mu == "high"] = "mu = 2.366e-8"

ggplot(rhorates[rhorates$chrom %in% c("b'JAPQVT010000006.1'")], aes(x = start, y = avg_recombRate, color=sex)) +
  geom_line() +
  geom_ribbon(aes(ymin=avg_CI95LO, ymax=avg_CI95HI,fill=sex),alpha=0.5)+
  facet_wrap(vars(factor(realmu, levels=c("mu = 2.366e-8","mu = 2.366e-9","mu = 2.366e-10") ), chrom),scales = "f
ree")+
  labs(x = "Position",
       y = "Recombination Rate (crossovers/bp)") +
  theme_minimal()+
  scale_color_manual(values=met.brewer("Archambault"))+
  scale_fill_manual(values=met.brewer("Archambault"))
```