Minimum Fitness

Shashank Pritam

Table of contents

0.1 Introduction

What is the impact of insertion bias on the minimum fitness of a population during the invasion of transposable elements (TEs), within the parameter space of bias and cluster size?

0.1.1 Initial conditions

0.2 Materials & Methods

version: invadego0.1.3

0.2.1 Commands for the simulation

The simulations were generated using the code from:

- sim_storm.py

0.3 Visualization in R

0.3.0.1 Setting the environment

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.2     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(RColorBrewer)
library(ggplot2)
library(patchwork)
library(dplyr)
theme_set(theme_bw())
```

0.3.0.2 Data loading and parsing

```
# Define and load DataFrame with column names
column_names <- c("rep", "gen", "popstat", "spacer_1", "fwte", "avw", "min_w", "avtes", "avpo
df <- read_delim('/Users/shashankpritam/github/Insertion-Bias-TE/Simulation-Results_Files/sir
```

```
Rows: 296 Columns: 22
-- Column specification --------------------------------------------------------
Delimiter: "\t"
chr  (8): popstat, spacer_1, spacer_2, phase, spacer_3, 3tot, 3cluster, spac...
dbl (13): rep, gen, fwte, avw, min_w, avtes, avpopfreq, fixed, fwcli, avcli,...
lgl  (1): X22

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Convert specific columns to numeric
numeric_columns <- c("rep", "gen", "fwte", "avw", "min_w", "avtes", "avpopfreq", "fixed", "fw
df[numeric_columns] <- lapply(df[numeric_columns], as.numeric)
```

0.3.1 Data Loading

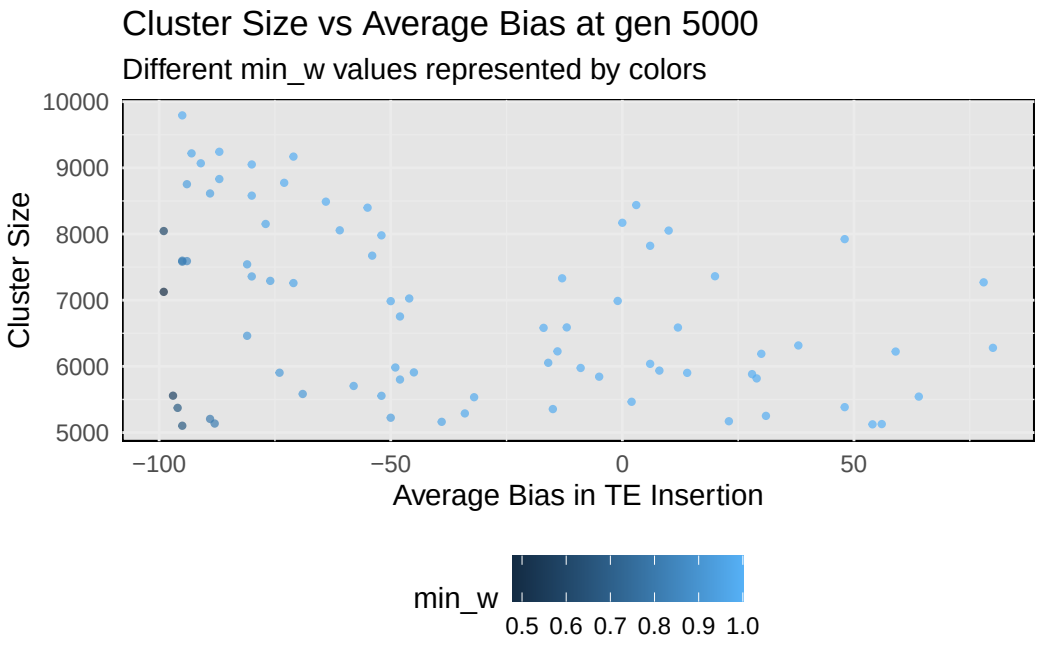
```
# Define color gradient functions
color.gradient <- function(x, colors=c("#D7191C","#FDAE61","#A6D96A","#1A9641"), colsteps=100
# Assign colors based on the 'min_w' column
df$col <- color.gradient(df$min_w)
df[df$popstat == "fail-0",]$col <- "grey"
df$col <- as.factor(df$col)

# Create and plot the ggplot object
# Subset the data for gen 5000
df_gen_5000 <- df[df$gen == 5000,]
```

0.3.2 Plot 1

```
# Plot avbias vs sampleid (Cluster Size) with min_w as color
g_avbias_cluster_size <- ggplot(df_gen_5000, aes(x = avbias, y = sampleid, color = min_w)) +
  geom_point(alpha = 0.7, size = 0.8) +
  ylab("Cluster Size") +
  xlab("Average Bias in TE Insertion") +
  labs(title = "Cluster Size vs Average Bias at gen 5000",
       subtitle = "Different min_w values represented by colors") +
  scale_color_gradient(name = "min_w") +
  theme_minimal() +
  theme(legend.position = "bottom", panel.background = element_rect(fill="grey90"))

# Display the plot
plot(g_avbias_cluster_size)
```

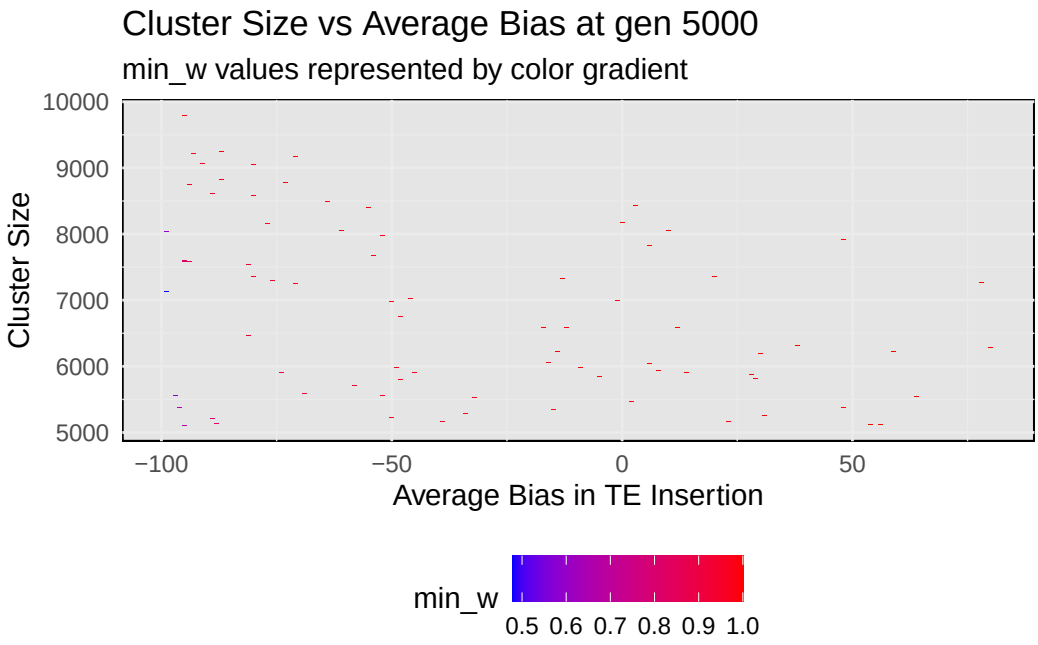


0.3.3 Plot 2

```
library(ggplot2)

# Plot avbias vs sampleid with min_w as color
g_avbias_cluster_size <- ggplot(df_gen_5000, aes(x = avbias, y = sampleid, fill = min_w)) +
  geom_tile() +
  scale_fill_gradient(low = "blue", high = "red", name = "min_w") +
  ylab("Cluster Size") +
  xlab("Average Bias in TE Insertion") +
  labs(title = "Cluster Size vs Average Bias at gen 5000",
       subtitle = "min_w values represented by color gradient") +
  theme_minimal() +
  theme(legend.position = "bottom", panel.background = element_rect(fill="grey90"))

# Display the plot
plot(g_avbias_cluster_size)
```


Cluster Size vs Average Bias at gen 5000
min_w values represented by color gradient
Cluster Size
10000
9000
8000
7000
6000
5000
−100
−50
0
50
Average Bias in TE Insertion
min_w
0.5 0.6 0.7 0.8 0.9 1.0