Supplemental Material for An Exploration of Exploration: Measuring the ability of lexicase selection to find obscure pathways

Jose Guadalupe Hernandez, Alexander Lalejini, and Charles Ofria

2021-06-15

Contents

Chapter 1

Introduction

TODO

1.1 About our supplemental material

TODO

Chapter 2

Diagnostic cardinality

2.1 Overview

```
# Relative location of data.
working_directory <-
  "experiments/2021-05-27-cardinality/analysis/"
# working_directory <- "./"

# Settings for visualization
cb_palette <- "Set2"
# Create directory to dump plots
dir.create(paste0(working_directory, "imgs"), showWarnings=FALSE)
```

2.2 Analysis dependencies

```
library(ggplot2)
library(tidyverse)
library(cowplot)
library(viridis)
library(RColorBrewer)
source("https://gist.githubusercontent.com/benmarwick/2a1bb0133ff568cbe28d/raw/fb53bd97121f7f9ce9
```

These analyses were conducted in the following computing environment:

```
print(version)
```

```
##                _
## platform       x86_64-pc-linux-gnu
## arch           x86_64
```

```
## os             linux-gnu
## system         x86_64, linux-gnu
## status
## major          4
## minor          1.0
## year           2021
## month          05
## day            18
## svn rev        80317
## language       R
## version.string R version 4.1.0 (2021-05-18)
## nickname       Camp Pontanezen
```

```
data_loc <- paste0(
  working_directory,
  "data/timeseries-res-1000g.csv"
)
data <- read.csv(
  data_loc,
  na.strings="NONE"
)

data$cardinality <- as.factor(
  data$OBJECTIVE_CNT
)
data$selection_name <- as.factor(
  data$selection_name
)

data$elite_trait_avg <-
  data$ele_agg_per / data$OBJECTIVE_CNT

data$unique_start_positions_coverage <-
  data$uni_str_pos / data$OBJECTIVE_CNT

####### misc #######
# Configure our default graphing theme
theme_set(theme_cowplot())
```

2.4 Performance (max)

Raw aggregate performances. Note that different cardinalities have different score potentials.

```
ggplot(data, aes(x=gen, y=ele_agg_per, color=cardinality)) +
  stat_summary(geom="line", fun=mean) +
  stat_summary(
    geom="ribbon",
    fun.data="mean_cl_boot",
    fun.args=list(conf.int=0.95),
    alpha=0.2,
    linetype=0
  ) +
  scale_y_continuous(
    name="Elite aggregate performance",
    limits=c(0, 10000)
  ) +
  scale_x_continuous(
    name="Generation"
  ) +
  scale_fill_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  scale_color_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  ggsave(
    paste(
      working_directory,
      "imgs/elite_agg_performance_ot.pdf",
      sep=""
    )
  )
```

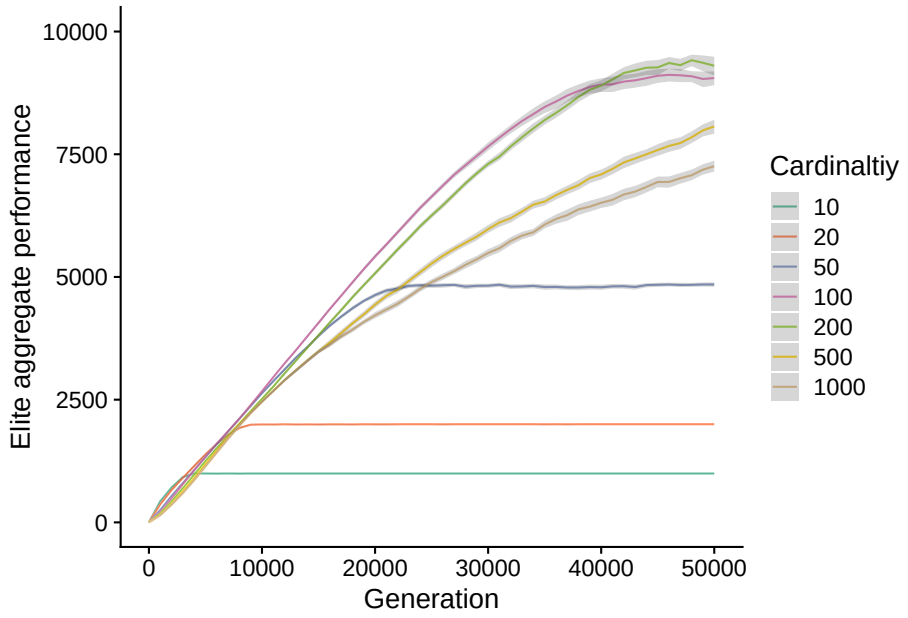
```
## Saving 6.5 x 4.5 in image

## Warning: Removed 115 rows containing non-finite values (stat_summary).

## Warning: Removed 115 rows containing non-finite values (stat_summary).

## Warning: Removed 115 rows containing non-finite values (stat_summary).

## Warning: Removed 115 rows containing non-finite values (stat_summary).
```

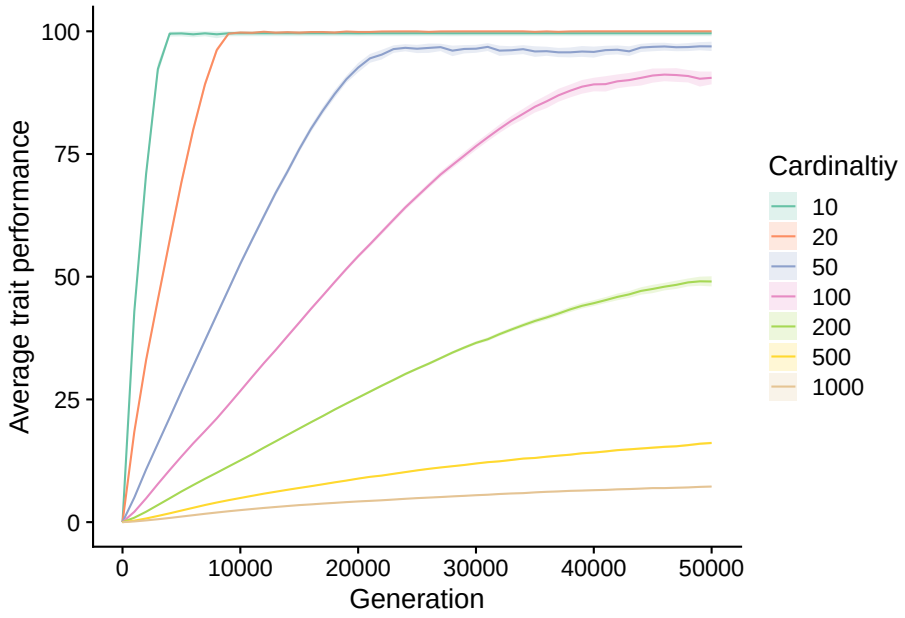



```
elite_trait_ave_fit <- ggplot(
    data,
    aes(
      x=gen,
      y=elite_trait_avg,
      color=cardinality,
      fill=cardinality
    )
  ) +
  stat_summary(geom="line", fun=mean) +
  stat_summary(
    geom="ribbon",
    fun.data="mean_cl_boot",
    fun.args=list(conf.int=0.95),
    alpha=0.2,
    linetype=0
  ) +
  scale_y_continuous(
    name="Average trait performance",
    limits=c(0, 100)
  ) +
  scale_x_continuous(
    name="Generation"
  ) +
  scale_fill_brewer(
```

```
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  scale_color_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  ggsave(
    paste(working_directory, "imgs/elite_trait_average_ot.pdf", sep="")
  )
```

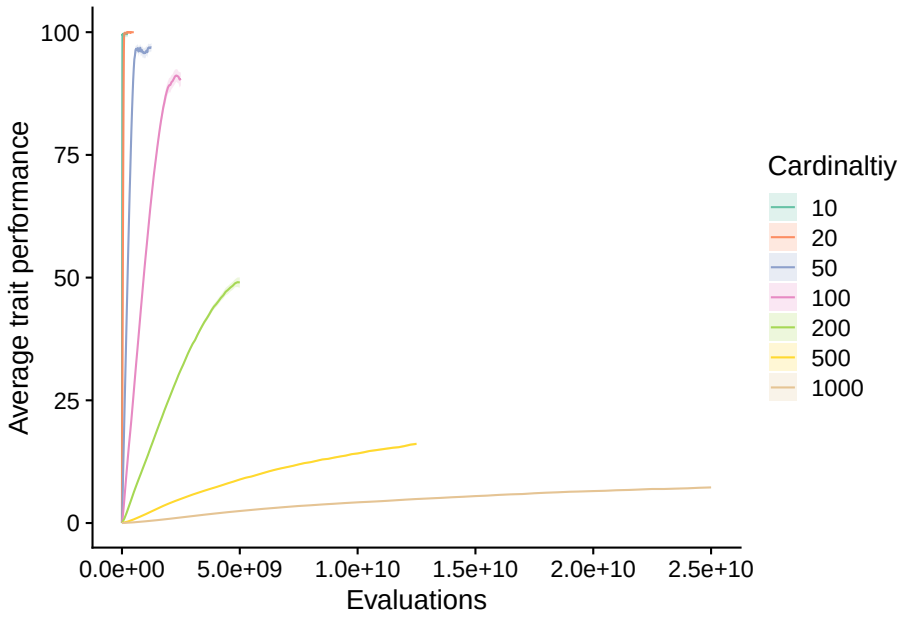
```
## Saving 6.5 x 4.5 in image
```

```
elite_trait_ave_fit
```



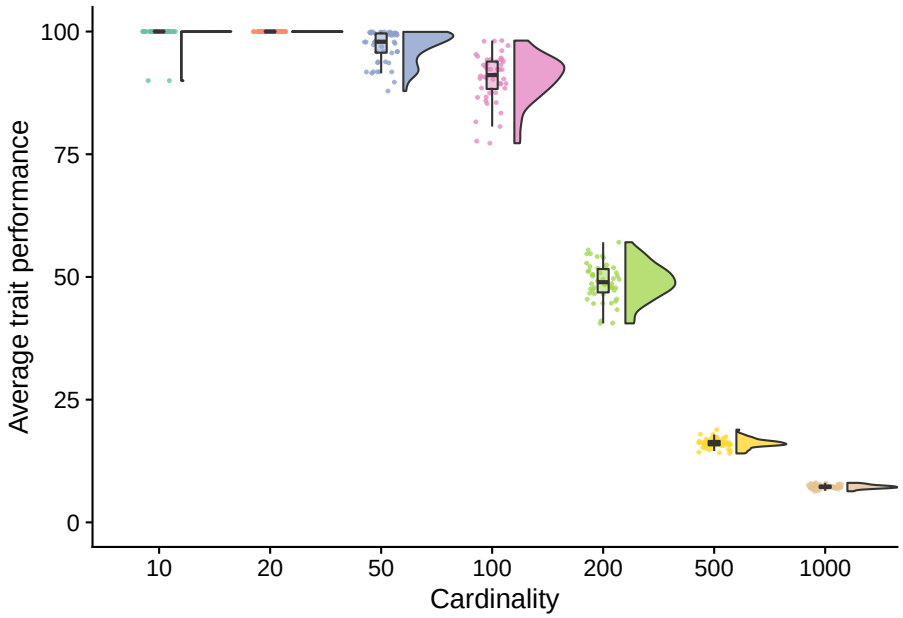

```
ggplot(
    data,
    aes(
      x=evaluations,
      y=elite_trait_avg,
      color=cardinality,
      fill=cardinality
    )
  ) +
  stat_summary(geom="line", fun=mean) +
  stat_summary(
```

```
    geom="ribbon",
    fun.data="mean_cl_boot",
    fun.args=list(conf.int=0.95),
    alpha=0.2,
    linetype=0
  ) +
  scale_y_continuous(
    name="Average trait performance",
    limits=c(0, 100)
  ) +
  scale_x_continuous(
    name="Evaluations"
  ) +
  scale_fill_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  scale_color_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  )
```



2.4.1 Final performance

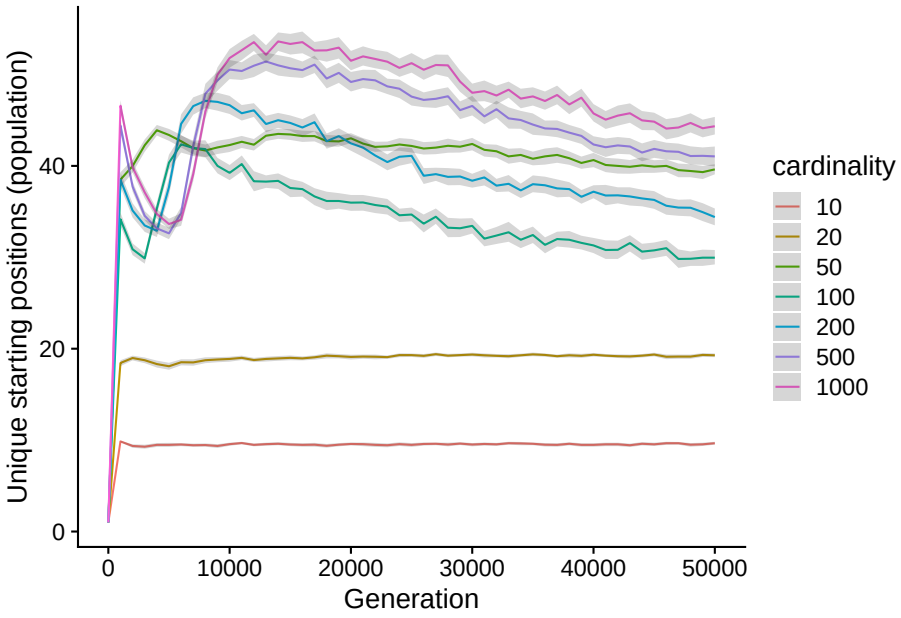
```
final_data <- filter(data, gen==max(data$gen))
elite_trait_ave_fit_final <- ggplot(
    final_data,
    aes(x=cardinality, y=elite_trait_avg, fill=cardinality)
  ) +
  geom_flat_violin(
    position = position_nudge(x = .2, y = 0),
    alpha = .8,
    scale="width"
  ) +
  geom_point(
    mapping=aes(color=cardinality),
    position = position_jitter(width = .15),
    size = .5,
    alpha = 0.8
  ) +
  geom_boxplot(
    width = .1,
    outlier.shape = NA,
    alpha = 0.5
  ) +
  scale_y_continuous(
    name="Average trait performance",
    limits=c(0, 100)
  ) +
  scale_x_discrete(
    name="Cardinality"
  ) +
  scale_fill_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  scale_color_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  theme(
    legend.position="none"
  )
elite_trait_ave_fit_final
```



2.5 Unique starting positions (population)

```
ggplot(data, aes(x=gen, y=uni_str_pos, color=cardinality)) +
  stat_summary(geom="line", fun=mean) +
  stat_summary(
    geom="ribbon",
    fun.data="mean_cl_boot",
    fun.args=list(conf.int=0.95),
    alpha=0.2,
    linetype=0
  ) +
  scale_y_continuous(
    name="Unique starting positions (population)",
  ) +
  scale_x_continuous(
    name="Generation"
  ) +
  ggsave(
    paste(working_directory, "imgs/pop_unique_starting_positions_ot.pdf", sep="")
  )
```

```
## Saving 6.5 x 4.5 in image
```

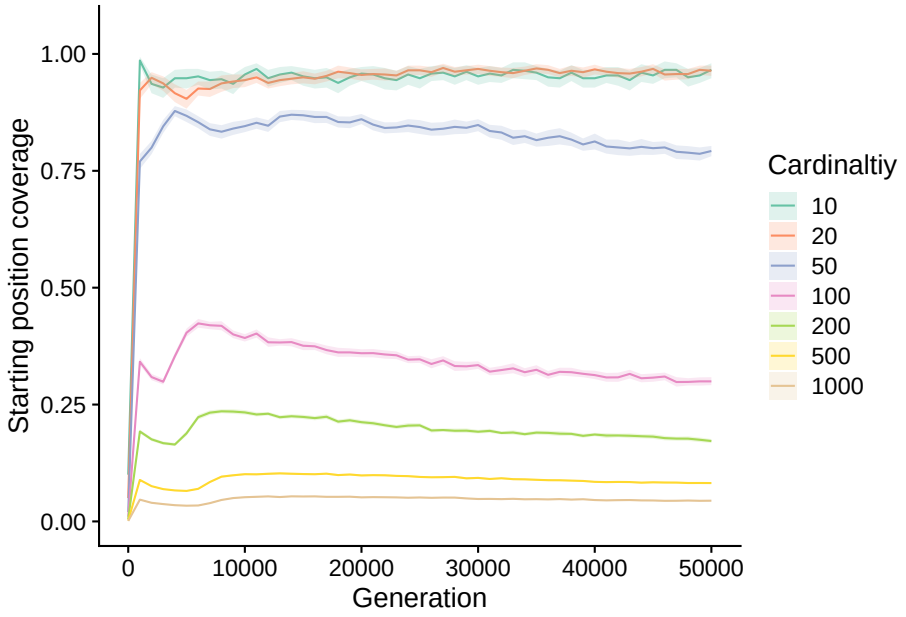



```
unique_start_positions_coverage_fig <- ggplot(data, aes(x=gen, y=unique_start_positions_coverage,
  stat_summary(geom="line", fun=mean) +
  stat_summary(
    geom="ribbon",
    fun.data="mean_cl_boot",
    fun.args=list(conf.int=0.95),
    alpha=0.2,
    linetype=0
  ) +
  scale_y_continuous(
    name="Starting position coverage",
    limits=c(0.0, 1.05)
  ) +
  scale_x_continuous(
    name="Generation"
  ) +
  scale_fill_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  scale_color_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  ggsave(
```

```
    paste(working_directory, "imgs/pop_unique_starting_position_coverage_ot.pdf", sep="
  )
```

```
## Saving 6.5 x 4.5 in image
```

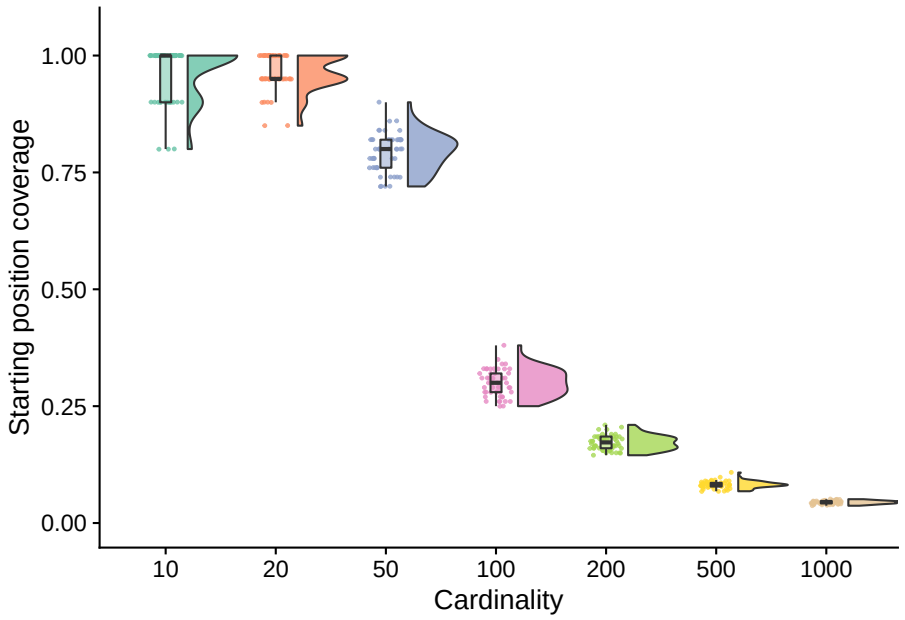
```
unique_start_positions_coverage_fig
```



2.5.1 Final coverage

```
final_unique_start_positions_coverage_fig <- ggplot(final_data,  aes(x=cardinality, y=u
  ) +
  geom_flat_violin(
    position = position_nudge(x = .2, y = 0),
    alpha = .8,
    scale="width"
  ) +
  geom_point(
    mapping=aes(color=cardinality),
    position = position_jitter(width = .15),
    size = .5,
    alpha = 0.8
  ) +
  geom_boxplot(
    width = .1,
    outlier.shape = NA,
```

```
    alpha = 0.5
  ) +
  scale_y_continuous(
    name="Starting position coverage",
    limits=c(0, 1.05)
  ) +
  scale_x_discrete(
    name="Cardinality"
  ) +
  scale_fill_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  scale_color_brewer(
    name="Cardinaltiy",
    palette=cb_palette
  ) +
  theme(
    legend.position="none"
  )
final_unique_start_positions_coverage_fig
```

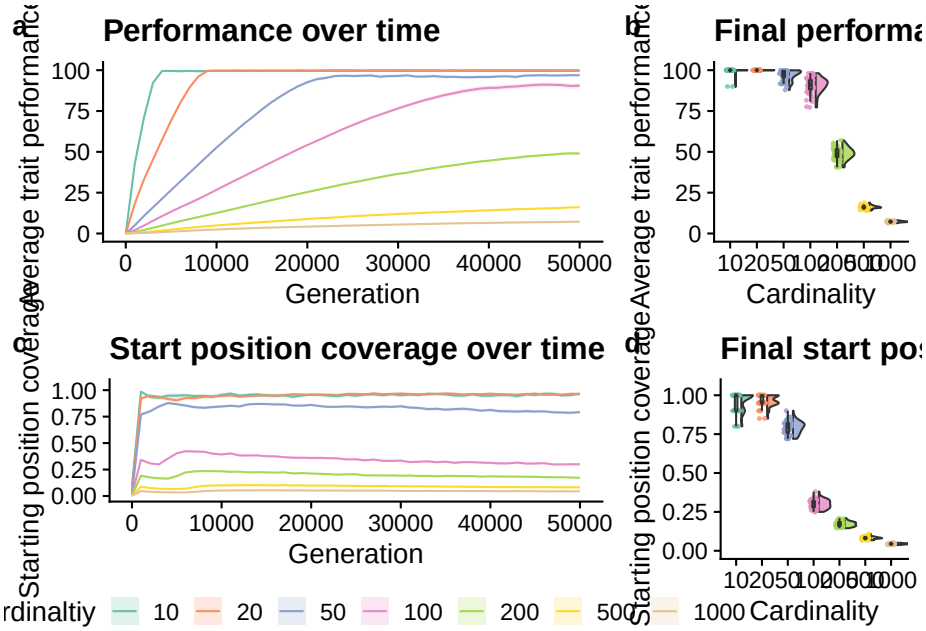


2.6 Manuscript figures

```
grid <- plot_grid(
  elite_trait_ave_fit +
    ggtitle("Performance over time") +
    theme(legend.position="none"),
  elite_trait_ave_fit_final +
    ggtitle("Final performance") +
    theme(),
  unique_start_positions_coverage_fig +
    ggtitle("Start position coverage over time") +
    guides(color = guide_legend(nrow = 1), fill=guide_legend(nrow = 1)) +
    theme(
      legend.position="bottom",
      legend.box="horizontal"
    ),
  final_unique_start_positions_coverage_fig +
    ggtitle("Final start position coverage") +
    theme(),
  nrow=2,
  ncol=2,
  rel_widths=c(2,1),
  labels="auto"
)

save_plot(
  paste(working_directory, "imgs/cardinality-panel.pdf", sep=""),
  grid,
  base_width=12,
  base_height=10
)

grid
```


Performance over time
Average trait performance
100
75
50
25
0
0
10000
20000
30000
40000
50000
Generation
Final performa
Average trait performance
Cardinality
Start position coverage over time
Starting position coverage
1.00
0.75
0.50
0.25
0.00
Generation
Final start po
Starting position coverage
Cardinality
rdinaltiy
10
20
50
100
200
500
1000