Supplementary Information File 1

Long-term monitoring of the Iberian ibex population in the Sierra Nevada of the southeast Iberian Peninsula

R code used in the Usage Notes section

We provided a reproducible example using the data stored at GBIF.

Load packages

We used the packages `finch` [@finch], `tidyverse` [@tidyverse], `knitr` [@knitr] and `here` [@here]

```r
library("tidyverse")
library("here")
library("finch")
library("knitr")
```

Read the data from DwCA

The first step was to download the Darwin Core Archive (`.zip` file) of the dataset from the IPT doi: 10.15470/3ucqfm. Then, using the `finch` package [@finch] we processed the Darwin Core Archive (DwC-A) and load the datasets.

```r
# https://ipt.gbif.es/resource?r=iberianibex

f <- finch::dwca_read("https://ipt.gbif.es/archive.do?r=iberianibex&v=1.4")

# To see the files included in the DwC, type:
# f$files

# Read the data files
eventRaw <- read_delim(f$data[1], delim = "\t") # event.txt
occRaw <- read_delim(f$data[2], delim = "\t") # occurrence.txt
mofRaw <- read_delim(f$data[3], delim = "\t") # extendedmeasurementorfact.txt

# Or download the DwCore Archive and unzip. Specify the folder in the next lines

# occRaw <- read_delim(here::here("UNZIPFOLDER/occurrence.txt"), delim = "\t")
# eventRaw <- read_delim(here::here("UNZIPFOLDER/event.txt"),  delim = "\t")
# mofRaw <- read_delim(here::here("UNZIPFOLDER/extendedmeasurementorfact.txt"), delim = "\t")
```

In the following steps we computed the population structure along the study period and explored several population parameters, such as sex ratio and birth rate.

Population structure

Our interest is to explore the time course of population structure. For this, each year, we computed the percentage of individuals belonging to certain age class. Also, we computed the average of each age class

for the study period. First, from the `Occurrence` table (from the DarwinCoreArchive), we selected the field `lifeStage` which indicates the age of the individual. For individuals belonging to "kid" lifeStage, we considered the half are males. Then, we computed the number of individuals by year and age class, and the percentage.

```r
# Compute the counts and percentage by age-class
age_aux <- occRaw %>%
  filter(sex == "male") %>%
  mutate(year = lubridate::year(eventDate)) %>%
  dplyr::select(year, lifeStage, organismQuantity) %>%
  group_by(year, lifeStage) %>%
  summarise(count = sum(organismQuantity)) %>%
  mutate(age_class = case_when(
    lifeStage %in% c("1", "2") ~ "1_youngs",
    lifeStage %in% c("3", "4") ~ "2_subadults",
    lifeStage %in% c("5", "6", "7", "8") ~ "3_adults",
    lifeStage %in% c("9", "10", "11", "12") ~ "4_olds"
  ))

# Compute the kids (male) by year (0.5 all kids)
aux_kid_male <- occRaw %>%
  filter(lifeStage == "kid") %>%
  mutate(year = lubridate::year(eventDate)) %>%
  dplyr::select(year, lifeStage, organismQuantity) %>%
  group_by(year, lifeStage) %>%
  summarise(count = round(sum(organismQuantity) / 2)) %>%
  mutate(age_class = "0_kid")

# Join and compute percentge of counts
age_classes <- age_aux %>%
  bind_rows(aux_kid_male) %>%
  filter(!is.na(age_class)) %>%
  group_by(year, age_class) %>%
  summarise(counts = sum(count)) %>%
  mutate(per = counts / sum(counts))
```

For visualization purposes, we averaged the age-classes frequency along the period of our data set.

```r
# Average values for age-class during the study period
global_g_classes <- age_classes %>%
  group_by(age_class) %>%
  summarise(per_avg = mean(per))
global_g_classes <- global_g_classes %>% expand(global_g_classes, year = 1995:2018)
```

Then we plot the structure of population for every year.

```r
# Structure of population plot
cp_classes <- age_classes %>%
  dplyr::select(-counts) %>%
  full_join(global_g_classes, by = c("year", "age_class")) %>%
  gather(key = "var", value = "per", -year, -age_class)

# plot_structure_classes <-
```

```r
plot_structure_classes <- cp_classes %>%
  filter(year != 1995 & year != 1999 & year != 2005 & year != 2006) %>%
  ggplot(aes(x = age_class, y = per, fill = var)) +
  facet_wrap(~year, strip.position = "top") +
  geom_bar(stat = "identity", position = position_dodge(width = 0.6)) +
  theme_classic() +
  scale_fill_manual(values = c("black", "gray")) +
  scale_x_discrete(
    labels =
      c(
        "0_kid" = "kids",
        "1_youngs" = "youngs",
        "2_subadults" = "subadults",
        "3_adults" = "adults",
        "4_olds" = "olds"
      )
  ) +
  theme(
    legend.position = "none",
    strip.background = element_blank(),
    strip.placement = "inside",
    axis.text.y = element_text(size = 10),
    axis.text.x = element_text(
      size = 10,
      angle = 90, face = "bold"
    )
  ) +
  xlab("Age classes") + ylab("Percentage")

plot_structure_classes
```

Sex Ratio and Birth rate

First we computed the male and female numbers grouped by year. We also computed the number of *kids* per year, using the field `lifeStage` included in the `Occurence` table.

- Extract year from `eventDate` field.
- Group data by year and determine the male and female counts

Then we computed the **sexratio** (sr) as female count / male count and the **birth rate** (**br**) as the *kid* count / female count. We used the variables `eventDate`, `sex`, `organismQuantity` and `lifeStage` from `Ocurrence` table (from the DarwinCoreArchive).

```r
aux <- occRaw %>%
  filter(lubridate::year(eventDate) != 1993) %>%
  mutate(year = lubridate::year(eventDate)) %>%
  dplyr::select(year, sex, organismQuantity) %>%
  group_by(year, sex) %>%
  summarise(count = sum(organismQuantity)) %>%
  spread(key = "sex", value = "count")
```
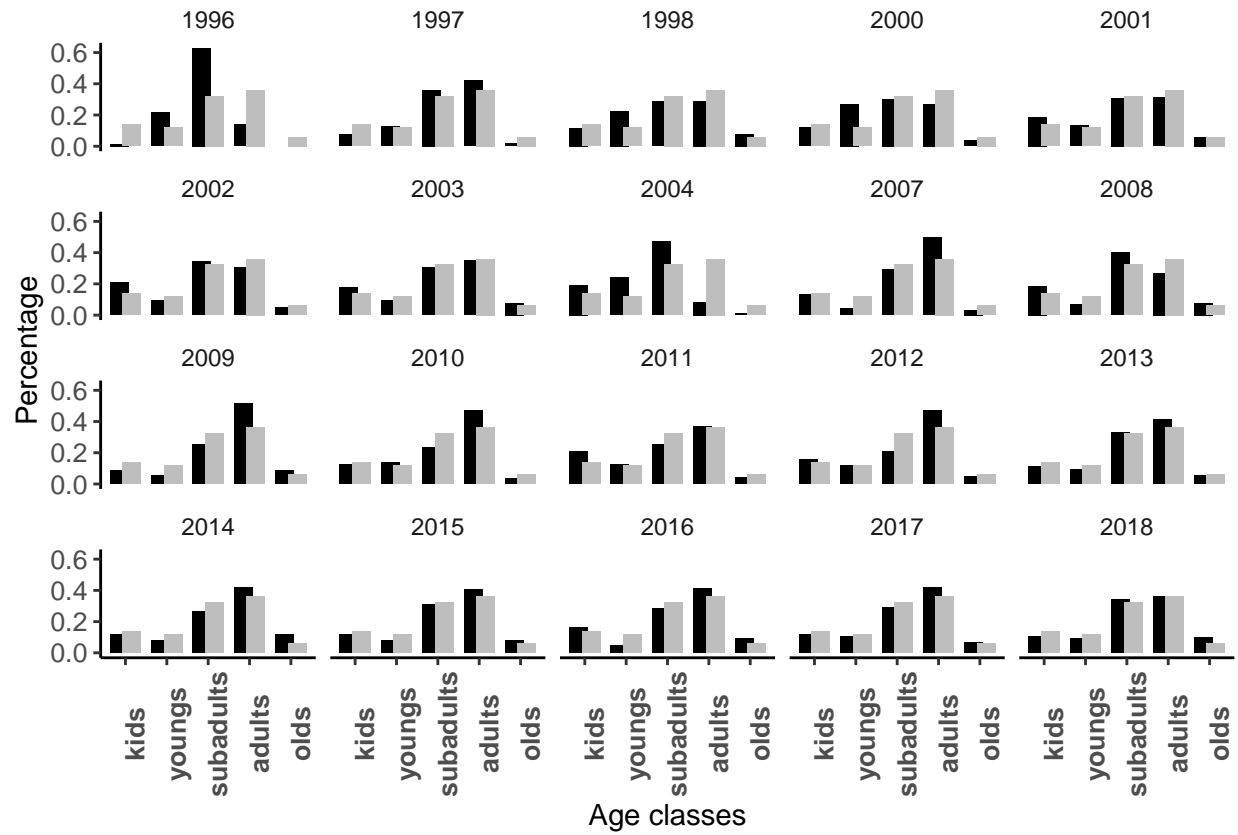
Figure 1: Annual population structure (age classes) of Capra pyrenaica at Sierra Nevada. Black bars indicate individual frequency for each age class. Gray bars indicate the average frequency for each age-class during the period 1995-2018

```r
aux_kid <- occRaw %>%
  filter(lubridate::year(eventDate) != 1993) %>%
  filter(lifeStage == "kid") %>%
  mutate(year = lubridate::year(eventDate)) %>%
  dplyr::select(year, lifeStage, organismQuantity) %>%
  group_by(year, lifeStage) %>%
  summarise(count = sum(organismQuantity)) %>%
  spread(key = "lifeStage", value = "count")

df <- aux %>%
  inner_join(aux_kid, by = "year") %>%
  mutate(
    sr = female / male,
    br = kid / female
  ) %>%
  dplyr::select(year, sr, br)
```

We plot the sexratio and the birth rate.

```r
# plot
label_variables <- c("br" = "Birth rate (kids/adult female)", "sr" = "Sex ratio")

cp_sexRatio <- df %>%
  filter(!(year %in% c(1996, 2002, 2003, 2004, 2007))) %>%
  gather(var, value, -year) %>%
  ggplot(aes(x = year, y = value)) +
  # geom_line(stat = "identity", colour = "gray") +
  geom_point(stat = "identity", size = 3) +
  facet_wrap(~var, nrow = 2, scales = "free_y",
    labeller = as_labeller(label_variables)
  ) +
  theme_classic() +
  xlab("Year") + ylab("") +
  theme(
    strip.background = element_rect(colour = "gray"),
    legend.position = "none",
    strip.placement = "inside",
    strip.text = element_text(face = "bold"),
    axis.text.y = element_text(size = 10),
    axis.text.x = element_text(size = 7)) +
  scale_x_continuous(breaks = 1995:2018)

cp_sexRatio_gam <- cp_sexRatio +
  geom_smooth(
    size = 1, method = "gam",
    formula = y ~ s(x, bs = "cs"),
    col = "darkgray", alpha = .2)

cp_sexRatio_gam
```
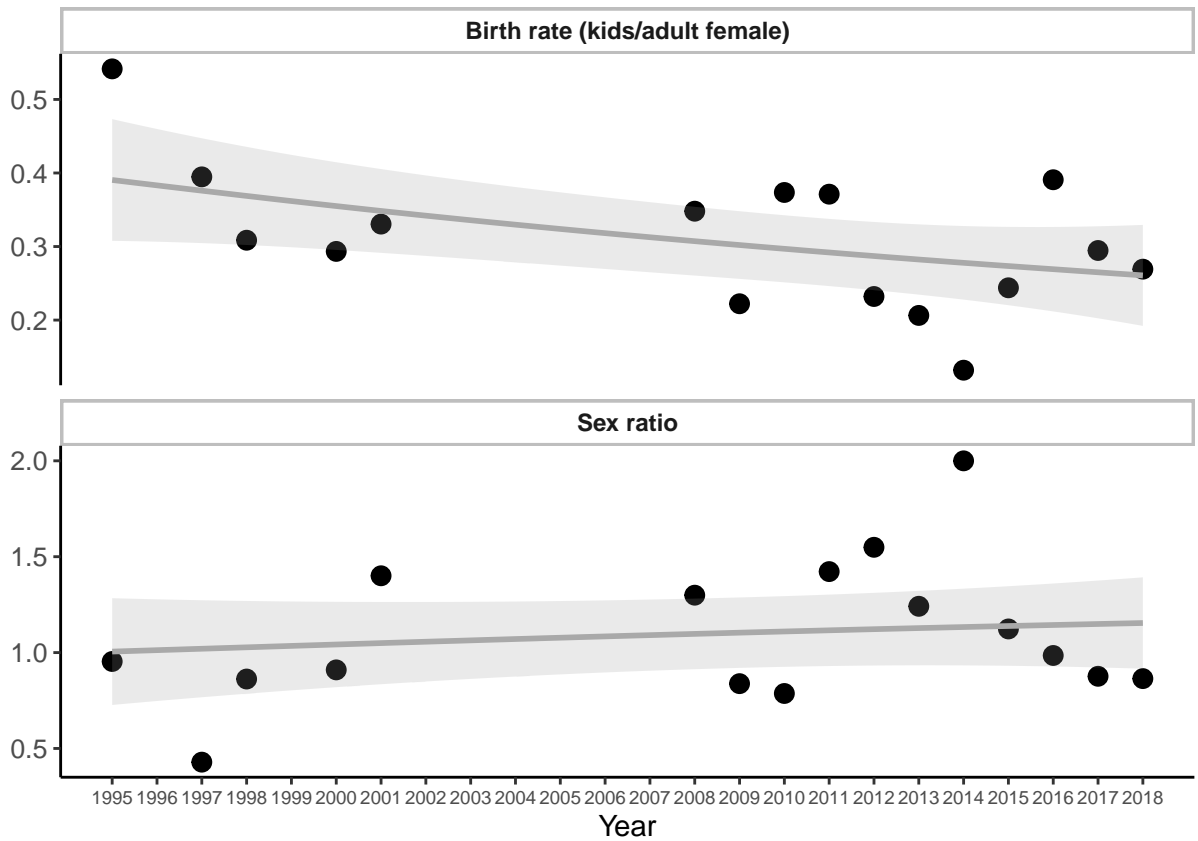
References

Figure 2: Temporal evolution of the birth rate (number of kids / adult female) (top) and sex ratio (bottom) of the Capra pyrenaica populations at Sierra Nevada. The code to generate the plot are included as supplementary material. A GAM smooth was added.