About Me

Zoe Erkenbeck

Contents

Who I am and where I came from

I grew up in a small town called Cricklade in the South West of England and moved to Colorado in 2016. I've always been interested in Biology as I grew up on nature documentaries and Cricklade has a Site of Special Scientific Interest (SSSI) that I used to spend a lot of time in called North Meadow. It is a traditionally managed lowland hay meadow and has the largest population (80%) of *Fritillaria meleagris* (snake's-head fritillary) in Britain.

I did my undergrautate degree in Biological Sciences at the University of the West of England in Bristol, UK and became increasingly interested in genetics of species of conservation concern, culminating in a final project studying genetic markers for use in conservation in *Malacochersus tornieri* (Pancake tortoise).

After a few years of work including as a Temporary nutritionist at Denver Zoo and an Animal Care tech at Anschutz I decided to persue a Master's degree to continue with my interest in wildlife genetics and get back into research. I also became a bit of a bird nerd joining Denver and Colorado Field Ornithologists groups so decided to focus my master work on an avian species of conservation concern across its range.

I'm a second year Master's student in Dr. Mike Wunder's lab at CU Denver and am co-advised by Dr. Sara Oyler-McCance at the USGS in Fort Collins. My thesis is focused on comparative genomics of *Charadrius montanus* (Mountain plover), a migratory (not exactly) shorebird. I have my de novo genome assembly (the first whole genome sequence for the species) and plan to use it for comparative genomics with other plover species and a population study comparing migratory breeding populations from the US (such as SouthPark,CO), possibly with a resident breeding population in Nuevo Leon, MX if I can get the samples though from Mexico in time. Permitting has been an interesting process..

When I'm not in school I enjoy: 1. Fencing épée (make new friends and then stab them) 2. Travelling, hiking and exploring new areas with my husband 3. Taking care of my fish, my honey gourami Tiberius and grumpy betta Groucho 4. Planting and growing things - my favorite houseplants and my Pachypodiums and I'm planning out a predominantly native wildlife garden

Here's a picture of me with my study species at Chico Basin Ranch and me and my husband exploring in Grand Teton NP.


LEUPOLD
LEUPOLD


Canon

JPEGs that include EXIF information might be incorrectly rotated - this happened. Resaved after rotating in a seperate program to fix but sure there is a better way.

Research Interests

I'm interested in conservation genomics, particularly of species of concern and with my latest project have been getting down a rabbit hole of thinking about genetics in relation to migratory behavior and how genomics could influence conservation decisions. Working with non-model species presents an interesting research challenge.

Influential papers

When I first started thinking about migratory behavioral difference as a possible avenue for comparative work the review paper Merlin and Liedvogel (2019) was a big jumping off point to getting back to the foundational papers and considering differences between taxa. I have written my own review that I hope to get to publication soon with a bird exclusive slant: Considering genetic associations and other possible basis of migration – From the perspective of Aves.

Barnhill-Dilling and Delborne (2021) and the whole special issue the article belongs to in Conservation Science and Practice really got me thinking about the place of genomics in conservation and the future of conservation projects. I have a second review in progress based around the ideas from the issue: The current role of genetics and genomics in conservation: benefits, challenges to implementation and integration into standard practice and what actions could help.

The mathematics behind my research

Basic starting point for effective population size

$$N_e = \frac{4N}{2 + S_k^2}$$

FST- proportion of the total genetic variance contained in a subpopulation relative to the total genetic variance

$$F_{st} = \frac{H_t - H_s}{H_t}$$

$$H = 1 - \Sigma {p_i}^2$$

My computing experience

I had no programming experience until graduate school.

My first experience was a class in Biological Data Analysis which used R. This was a lab for thinking about model selection using the University of California Berkeley admissions dataset

```
require(AICcmodavg)
ucb <- as.data.frame(UCBAdmissions)
ucb <- tidyr::spread(ucb, Admit, Freq) # use Hadley's excellent tidyr to reshape
ucb[order(ucb$Dept), ]
f1 <- glm(cbind(Admitted, Rejected) ~ Gender, ucb, family='binomial')
f2 <- glm(cbind(Admitted, Rejected) ~ Dept, ucb, family = 'binomial')
f3 <- glm(cbind(Admitted, Rejected) ~ Dept + Gender, ucb, family = 'binomial')
```

```
f4 <- glm(cbind(Admitted, Rejected) ~ Dept * Gender, ucb, family = 'binomial')
glm.list <- list(f1,f2,f3,f4)
glm.names <- as.character(unlist(lapply(glm.list,formula)))
(glm.results <- aictab(glm.list,modnames=glm.names))

exp(-0.5*glm.results$Delta_AICc[4])/sum(exp(-0.5*glm.results$Delta_AICc))

1/exp(-0.5*glm.results$Delta_AICc[4])/sum(exp(-0.5*glm.results$Delta_AICc))
summary(f1)
summary(f2)
summary(f3)
summary(f4)
```

I've been learning to work on a Linux machine (my computer is caller Pepper, it runs mint) and bash script to run programs on summit. This is for running cactus although I'm still having some issues with the program timing out with the walltime so I don't have alignment results yet

```
#!/bin/bash

#SBATCH --nodes=1
#SBATCH --ntasks=22
#SBATCH --time=168:00:00
#SBATCH --partition=shas
#SBATCH --qos=long
#SBATCH --mail-type=end
#SBATCH --mail-user=zoe.erkenbeck@cuanschutz.edu
#SBATCH --output=cactus_plovers.out

module purge

module load singularity/3.6.4

#Scripting here
SCRATCH_DIR=/scratch/summit/zazzy\@xsede.org
PROJECTS_DIR=/projects/zazzy\@xsede.org

JOB_STORE=/scratch/summit/zazzy\@xsede.org/jobstore$$
WORK_DIR=/scratch/summit/zazzy\@xsede.org/cactusWorkDir

export XDG_RUNTIME_DIR=$SCRATCH_DIR/
export SINGULARITY_CACHEDIR=$SCRATCH_DIR/.singularity
export SINGULARITY_TMPDIR=$SCRATCH_DIR/.singularity

rm -rf $JOB_STORE
rm -rf $WORK_DIR

#mkdir -p $JOB_STORE
mkdir -p $WORK_DIR

singularity pull --dir $SINGULARITY_CACHEDIR docker://quay.io/comparative-genomics-toolkit/cactus:v2.0.4
singularity run -B $PROJECTS_DIR:$PROJECTS_DIR,$WORK_DIR:$WORK_DIR,$SCRATCH_DIR:$SCRATCH_DIR --env TMPDI

module purge
```

What I hope to get out of this class

I hope to:

- Have a better idea of the workflow for the bioinformatics programs I will use for my thesis
- Meet like-minded people working on similar problems and learn about their research
- Make more progress than trying to figure out how to do analysis on my own

Evaluating some R code

This is a range map for Mountain Plover showing the breeding range in yellow, wintering range in blue and migration period range in green. I'm still working on playing with mapping in R and my sites may change but this base can be updated for use later

```
library(ggplot2)
library(ggspatial)
library(maptools)
```

```
## Loading required package: sp
```

```
## Checking rgeos availability: FALSE
## Please note that 'maptools' will be retired by the end of 2023,
## plan transition at your earliest convenience;
## some functionality will be moved to 'sp'.
##      Note: when rgeos is not available, polygon geometry     computations in maptools depend on gpcli
##      which has a restricted licence. It is disabled by default;
##      to enable gpclib, type gpclibPermit()
```

```
library(mapdata)
```

```
## Loading required package: maps
```

```
library(rgdal)
```

```
## Please note that rgdal will be retired by the end of 2023,
## plan transition to sf/stars/terra functions using GDAL and PROJ
## at your earliest convenience.
##
## rgdal: version: 1.5-28, (SVN revision 1158)
## Geospatial Data Abstraction Library extensions to R successfully loaded
## Loaded GDAL runtime: GDAL 3.2.1, released 2020/12/29
## Path to GDAL shared files: C:/Users/Zazzy/Documents/R/win-library/4.1/rgdal/gdal
## GDAL binary built with GEOS: TRUE
## Loaded PROJ runtime: Rel. 7.2.1, January 1st, 2021, [PJ_VERSION: 721]
## Path to PROJ shared files: C:/Users/Zazzy/Documents/R/win-library/4.1/rgdal/proj
## PROJ CDN enabled: FALSE
## Linking to sp version:1.4-6
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,
## use options("rgdal_show_exportToProj4_warnings"="none") before loading sp or rgdal.
## Overwritten PROJ_LIB was C:/Users/Zazzy/Documents/R/win-library/4.1/rgdal/proj
```

```
library(maps)
library(sf)
```

```
## Linking to GEOS 3.9.1, GDAL 3.2.1, PROJ 7.2.1; sf_use_s2() is TRUE
```

```
library(rgdal)

MOPLpath <- "./shapefiles/Charadrius_montanus.shp"

MOPLplot <- readOGR(MOPLpath)
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\Zazzy\Documents\Graduate School\Spring 2021 Data Analysis\bioinf-rmarkdown-introdu
## with 3 features
## It has 21 fields
## Integer64 fields read as strings:  vertex Field
```

```
MOPLdf <- fortify (MOPLplot)
```

```
## Regions defined for each Polygons
```

```
MOPLdf$group <- factor(MOPLdf$group,levels = c("2.1","2.2","2.3","2.4","1.1","1.2","1.3","1.4","1.5","1
MOPLdf$fill <- 2
MOPLdf[MOPLdf$group == '0.1'|MOPLdf$group == '0.2'|MOPLdf$group == '0.3'|MOPLdf$group == '0.4'|MOPLdf$gr
MOPLdf[MOPLdf$group == '1.1'|MOPLdf$group == '1.2'|MOPLdf$group == '1.3'|MOPLdf$group == '1.4'|MOPLdf$gr
MOPLdf[MOPLdf$group == '2.1'|MOPLdf$group == '2.2'|MOPLdf$group == '2.3'|MOPLdf$group == '2.4',]$fill <-

(sites <- data.frame(longitude = c(-100.9536431, -103.5646914, -104.3386891, -105.8656448, -107.760596)
                     latitude = c(26.3780515, 37.6552239, 38.3679908, 39.2060909, 47.971252)))
```

```
##   longitude latitude
## 1 -100.9536 26.37805
## 2 -103.5647 37.65522
## 3 -104.3387 38.36799
## 4 -105.8656 39.20609
## 5 -107.7606 47.97125
```

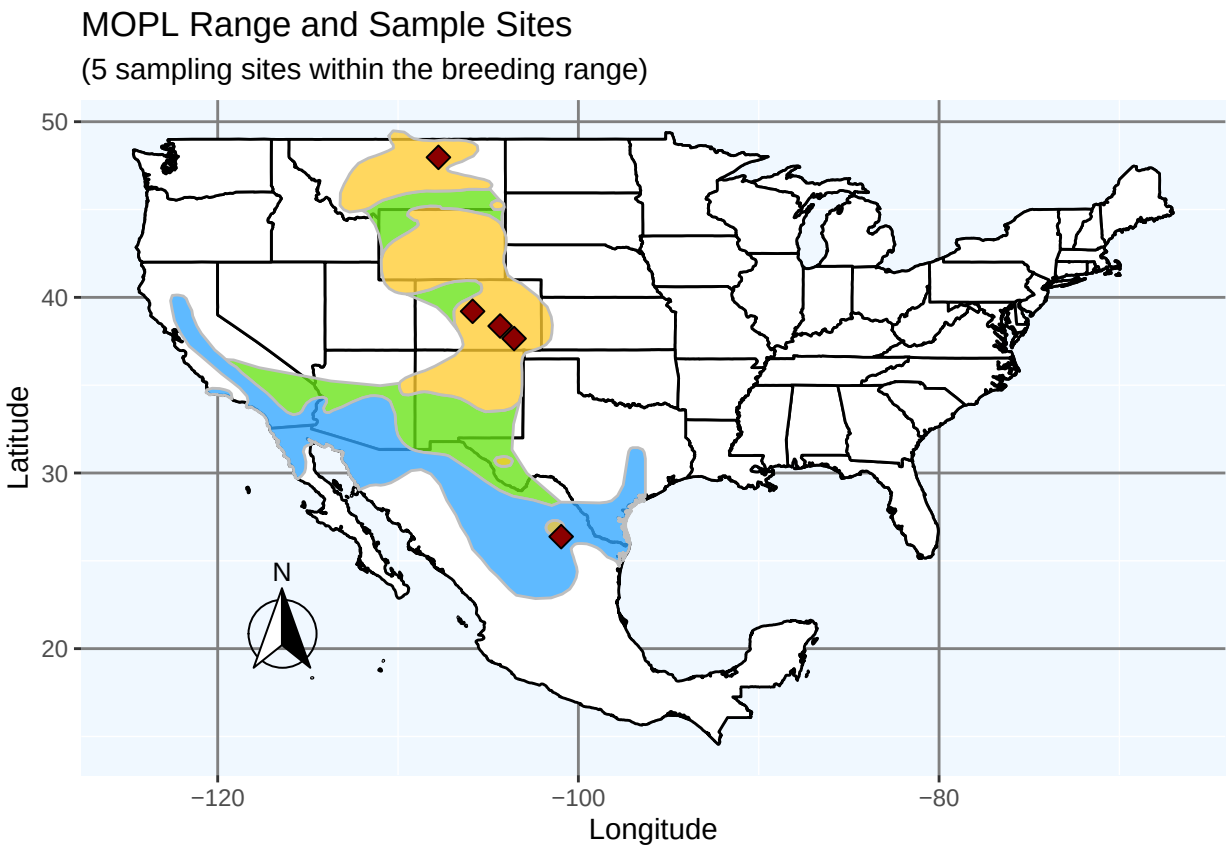
```
usa <- map_data("usa")
states <- map_data("state")
mexico <- map_data("worldHires" , "Mexico")

NAmap <- ggplot() + geom_polygon(data = usa,
                                 aes(x=long, y = lat, group = group),
                                fill = "white",
                                color="black") +
  geom_polygon(data = states,
               aes(x=long, y = lat, group = group),
               fill = "white",
               color="black") +
  geom_polygon(data = mexico,
```

```
                aes(x=long, y = lat, group = group),
                fill = "white",
                color="black") +

  geom_polygon(data = MOPLdf,
                aes(x=long,y=lat,group= group),
              fill= MOPLdf$fill,
              color="gray") +
  geom_point(data = sites, aes(x = longitude, y = latitude), size = 3,
              shape = 23, fill = "darkred")+
 # annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
                         pad_x = unit(0.75, "in"), pad_y = unit(0.5, "in"),
                         style = north_arrow_fancy_orienteering) +
xlab("Longitude") + ylab("Latitude") +
  ggtitle("MOPL Range and Sample Sites", subtitle = "(5 sampling sites within the breeding range)") +
            theme(panel.grid.major = element_line(color = gray(0.5),
                                                  size = 0.5), panel.background = element_rect(fill = "
NAmap
```

```
## Warning in f(...): True north is not meaningful without coord_sf()
```



References

Barnhill-Dilling, Kathleen S, and Jason A Delborne. 2021. "Whose Intentions? What Consequences? Interrogating 'Intended Consequences' for Conservation with Environmental Biotechnology." *Conservation Science and Practice* 3 (4).

Merlin, Christine, and Miriam Liedvogel. 2019. "The Genetics and Epigenetics of Animal Migration and Orientation: Birds, Butterflies and Beyond." *The Journal of Experimental Biology* 222 (February): jeb191890. https://doi.org/10.1242/jeb.191890.