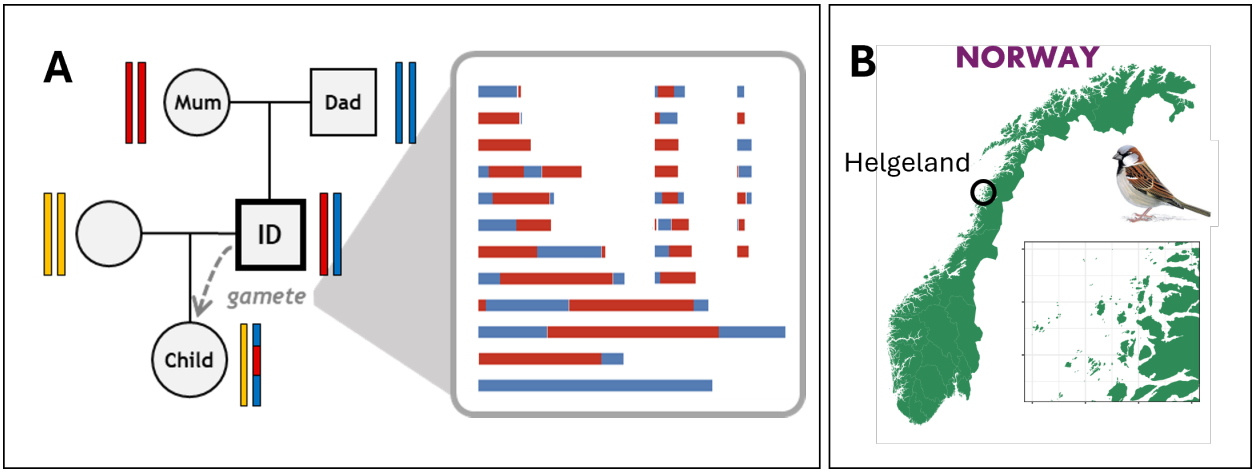
Genomic Prediction of Sparrow Recombination

Susan Johnston

2024-09-10

1. Introduction - recombination in house sparrows.

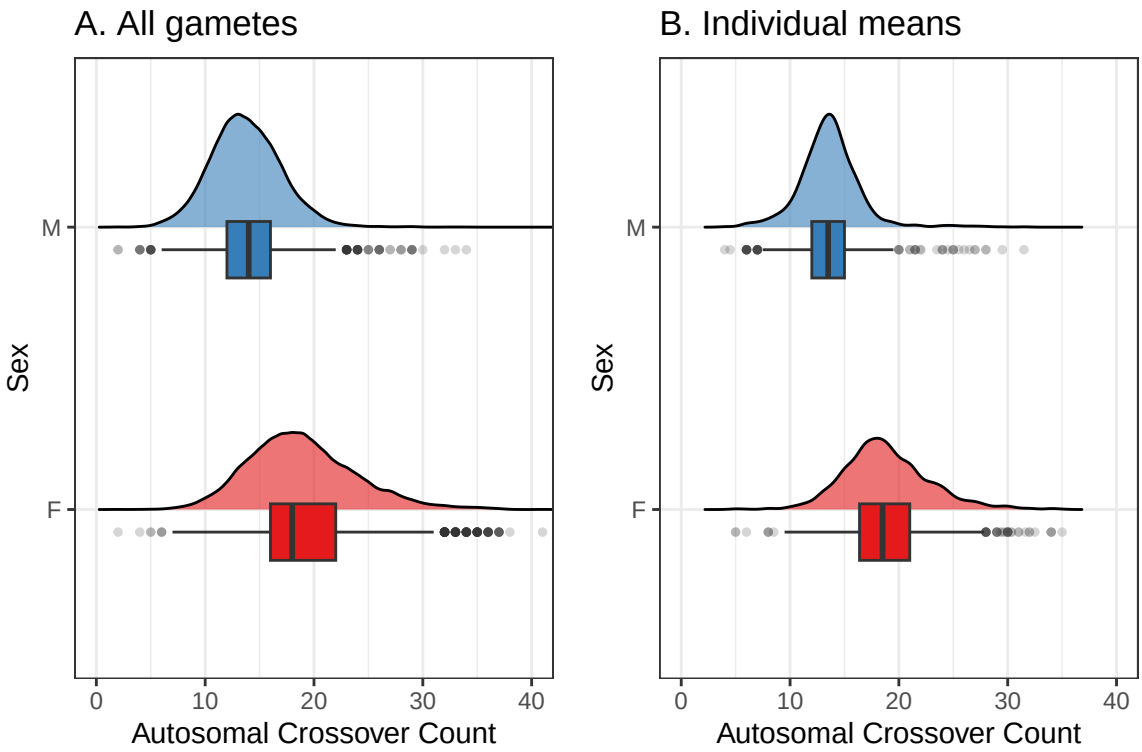


The positions of meiotic crossover events are measured by tracking the inheritance patterns of alleles between parents and offspring and determining the positions of changes in phase. Our dataset `recsumm` has estimates of crossover count (`co_count`) and intra-chromosomal allelic shuffling (`intra-shuff`) for 13042 `offspring` resulting from gametes transmitted from 2652 unique `id`. Our data also includes the `total_coverage` which is the length of the genome informative for identifying crossover events. **In this document, I will solely focus on crossover count.**

```
head(recsumm)
```

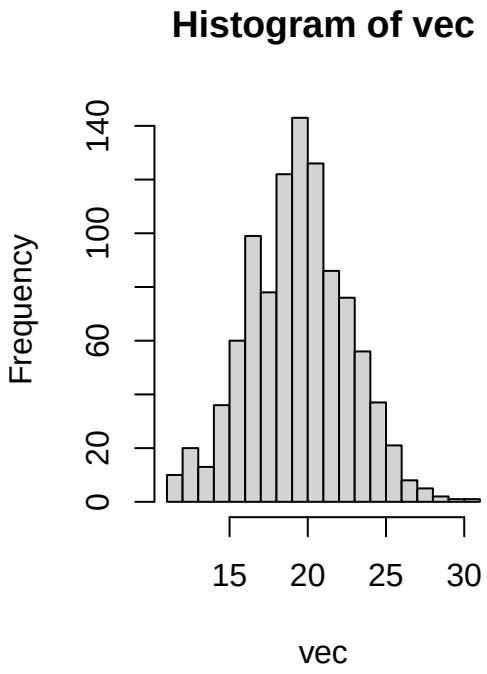
```
##              meiosis      id  offspring sex total_coverage co_count intra_shuff
## 1 8118424_8L64240_A3 8118424 8L64240_A3   M      901765568       17  0.03330310
## 2    8118424_8L64241 8118424    8L64241   M      903819159       11  0.02020535
## 3 8118424_8L64242_C3 8118424 8L64242_C3   M      902947627       16  0.02541074
## 4    8118424_8L89644 8118424    8L89644   M      903580120       17  0.02776473
## 5    8118424_8L89675 8118424    8L89675   M      903886149       16  0.02607208
## 6    8118424_8L89683 8118424    8L89683   M      903886149       16  0.02943656
```

This is the distribution of crossover counts across all gametes (A) vs an individual's mean crossover count (B). As you can see, there is a clear sex difference in the trait distribution:

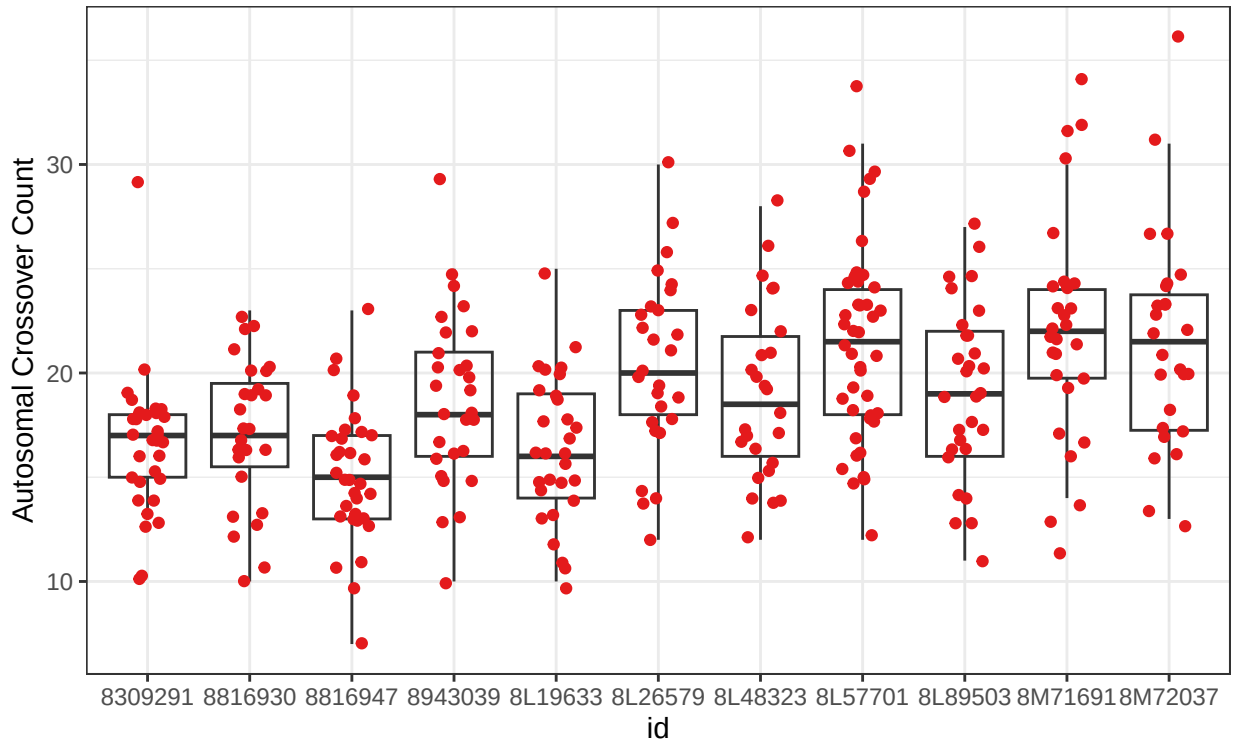


2. Recombination has a high sampling variance, which leads to weird quirks.

Crossover count is an unusual phenotype as there is more intra-individual variation than inter-individual variation. This is because for each crossover event that happens during meiosis, we can assume that is ~50:50 chance that it segregates into the egg or sperm. If you have 40 crossovers, an average of 20 will end up in the gamete. But that still means there is decent variation around that mean. For example, for 1000 gametes coming from an individual with 40 crossovers in every meiosis, the distribution of crossovers in the gametes would be:



As I stated above, this translates into more intra-individual variation than inter-individual variation. Let's look at the female sparrows with the most unique measures:



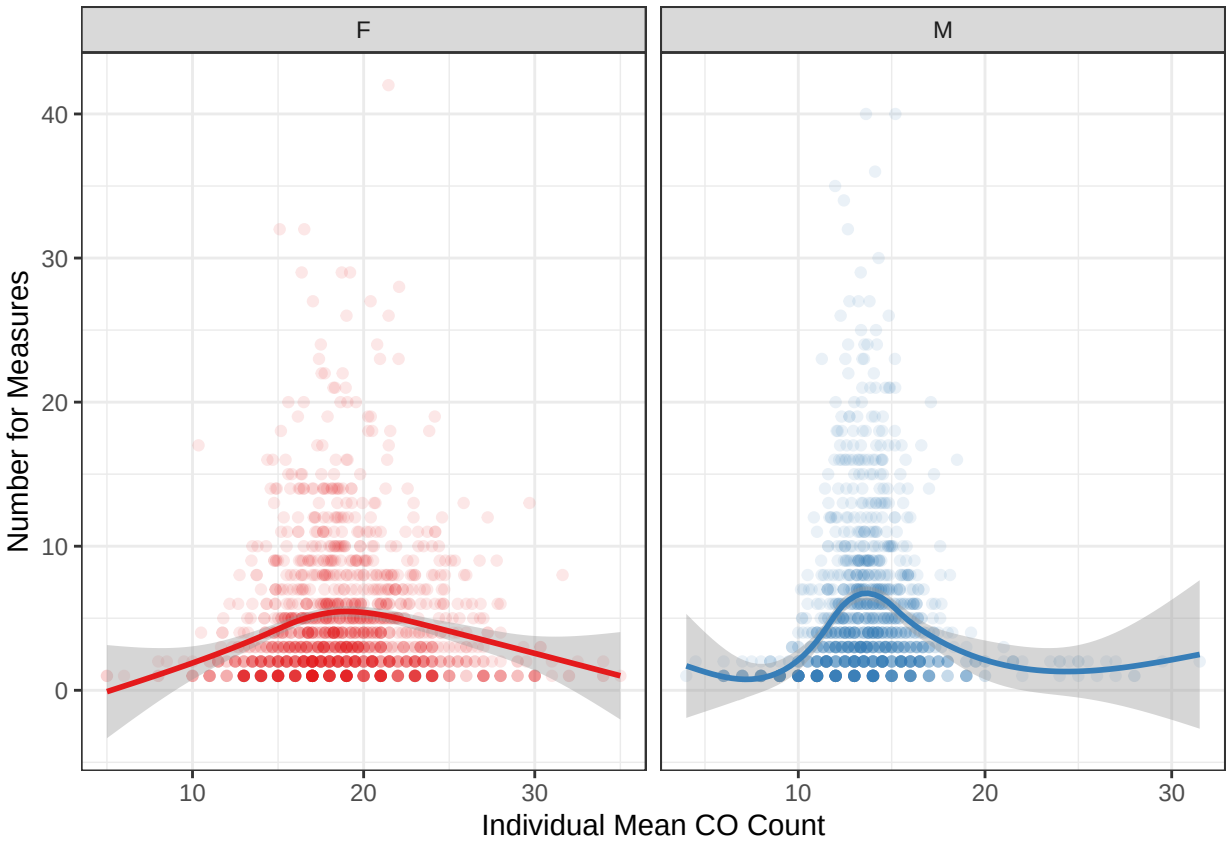
The variance across individual means is 5.5438493, whereas the variance within individuals is:

```
## # A tibble: 11 x 2
##    id      `var(co_count)`
##    <fct>             <dbl>
##  1 8309291            11.1
##  2 8816930            12.0
##  3 8816947            10.4
##  4 8943039            16.2
##  5 8L19633            12.0
##  6 8L26579            18.6
##  7 8L48323            17.7
##  8 8L57701            23.1
##  9 8L89503            17.5
## 10 8M71691            30.5
## 11 8M72037            27.3
```

We also observe that individuals with more measures will have means close to the population mean, which has implications for further analysis of fitness!

```
ggplot(recmeans, aes(mean_co_count, n, colour = sex)) +
  geom_point(alpha = 0.1) +
  stat_smooth() +
  scale_color_brewer(palette = "Set1") +
  facet_wrap(~sex, scales = "free_x") +
  theme_bw() +
  theme(legend.position = "none") +
  labs(x = "Individual Mean CO Count", y = "Number for Measures")
```

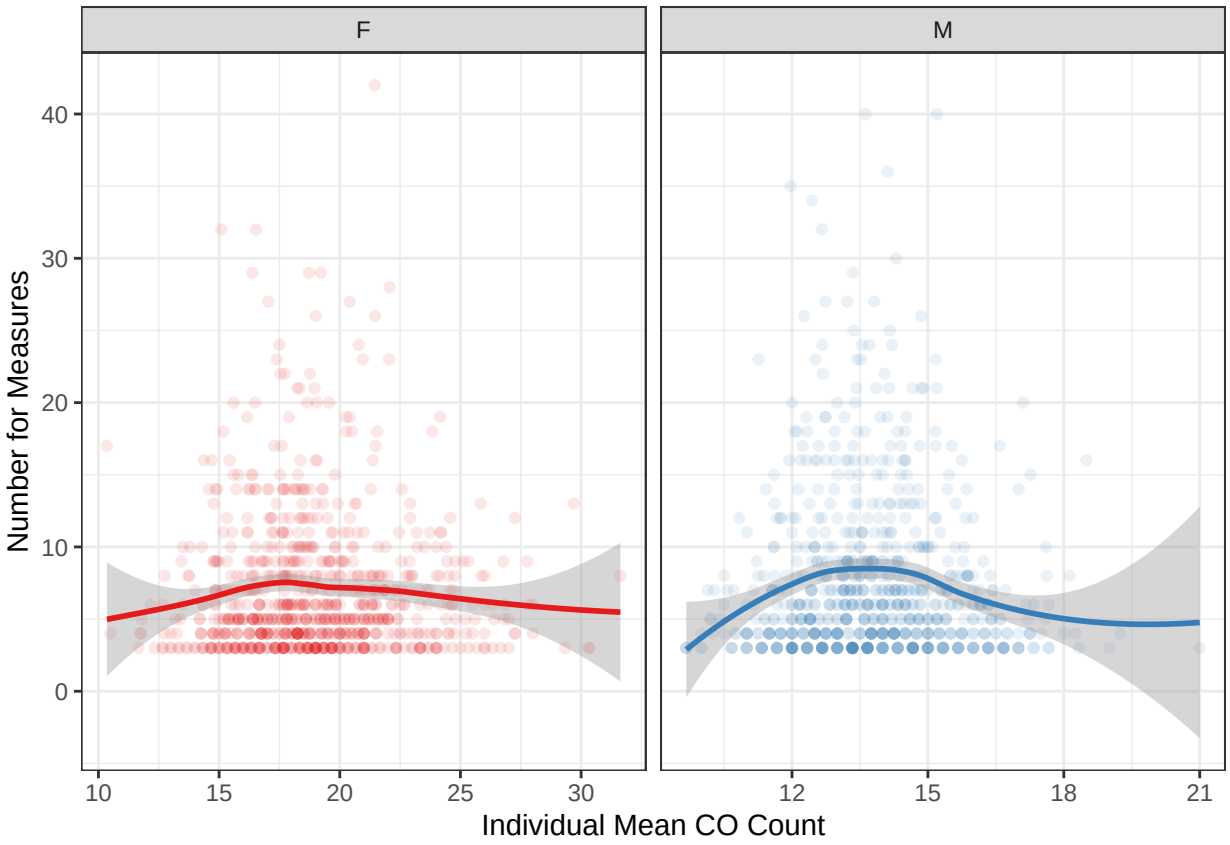
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



In females, once you have more than 3 measures, this effect starts to flatten out. In males it doesn't, perhaps they are more sensitive to this problem because they have fewer crossovers..?

```
ggplot(subset(recmeans, n > 2), aes(mean_co_count, n, colour = sex)) +
  geom_point(alpha = 0.1) +
  stat_smooth() +
  scale_color_brewer(palette = "Set1") +
  facet_wrap(~sex, scales = "free_x") +
  theme_bw() +
  theme(legend.position = "none") +
  labs(x = "Individual Mean CO Count", y = "Number for Measures")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



2. How heritable is crossover rate?

Because there are differences in distribution between the sexes, let's just focus on **females** for the moment. Heritabilities are estimated using a pedigree-based relationship matrix (this gives very similar results to the GRM used in the McAuley et al 2024 MBE paper). I use `library(asreml)` for the following models.

Below I consider two phenotypes:

1. What is the heritability of the number of crossovers in a sampled gamete?
2. What is the heritability of the mean number of crossovers across all gametes?

Let's begin. Here, the heritability is estimated using the crossovers in all sampled gametes (value for `vm(id, ainv)`):

```
recsumm_f <- subset(recsumm, sex == "F")

female.acc.ped <- asreml(fixed = co_count ~ total_coverage,
                         random = ~ vm(id, ainv) + ide(id),
                         residual = ~idv(units),
                         data = recsumm_f,
                         trace = F)

asreml4pin(female.acc.ped)
```

```
##      Estimate         SE       Effect
## h1 0.25948288 0.03196861 vm(id, ainv)
## h2 0.03335797 0.02544646      ide(id)
## h3 0.70715915 0.01633110  units!units
```

And here is the mean rate per individual (value for `vm(id, ainv)`):

```
recmeans_f <- subset(recmeans, sex == "F")

female.acc.ped.mean <- asreml(fixed = mean_co_count ~ mean_total_coverage + n,
                      random = ~ vm(id, ainv),
                      residual = ~idv(units),
                      data = recmeans_f,
                      trace = F)

asreml4pin(female.acc.ped.mean)
```

```
##     Estimate         SE       Effect
## h1 0.4993435 0.06423255 vm(id, ainv)
## h2 0.5006565 0.06423255  units!units
```

And here is the mean rate per individual with >2 measures (value for `vm(id, ainv)`):

```
female.acc.ped.mean <- asreml(fixed = mean_co_count ~ mean_total_coverage + n,
                      random = ~ vm(id, ainv),
                      residual = ~idv(units),
                      data = subset(recmeans_f, n > 2),
                      trace = F)

asreml4pin(female.acc.ped.mean)
```

```
##    Estimate         SE       Effect
## h1 0.607963 0.08171855 vm(id, ainv)
## h2 0.392037 0.08171855  units!units
```

As you can see, the heritability basically doubles when using the mean measures!