**Results Outline.v2**

**Estimates of gwRR from the house mouse species complex**

To estimate the gwRR for inbred house mouse strains we calculated the mean of the average number of MLH1 foci per cell.

- Description of MLH1 dataset, number of cells (spermatocyte / oocytes), and measures of repeatability, how close the means are to the minimum.

1. Strain means for females, have low variance, while male strain means have more variation in strain means. (The largest female difference is G, 1.07X the largest male difference is PWD and MSM which are 1.3 and 1.2 respectively.)
2. (At the mouse mean level within strain – males have lower variance? Compared to females var(mouse means) within strains)

-Comparisons to previous measurements,

-house mouse close to minimum of 1 obligate CO per chrm

**Evolutionary framework, Analysis using a mixed model framework for examining patterns of heterochiasmy**

( in order to test – describe the patterns of evolution we used a mixed model, subspecies = divergence, strain = random effect – as wild derived inbred strains simulated as random samples of standing genetic variation)

-models basics, (strains, logic of effects)

1. mixed model, all effects significant, Sex, interaction effect with subspecies were significant and the random strain effect were significant indicating the variance due to strain effect (genotype) is not 0.

2. –glm’s (slight differences between the 2 models), but main pattern – MSM and PWD have significant strain \* male interaction effects

3. Sex specific models – show significant strain effect models, but the range of effects are larger in males compared to females. (for males, MSM, PWD and SKIVE have significant effects (M4. Just strain effects). For females, G, LEW, MSM, PWD, and MOLF).

***- Sex specific evolution is the major pattern.***

***-3 High rec males in Musc and mol subspecies – rapid sex specific evolution***

**Within mouse variance for MLH1 count per cell**

(in order to understand the variance within an evolutionary framework – we fit the mouse gwRRs to a mixed model, which specified subspecies, sex and the interaction as fixed effects. The subspecies effect – is a proxy for measuring the degree of divergence across subspecies. The strains were coded as a random effect to –simulate / approximate / the standing genetic variation across subspecies )

-logic and motivation for using within mouse variance

- applied same model framework

- The same mixed models and glms were ran with within mouse variance (and coefficient of variance) of MLH1 counts per cell as the dependent variable.

- Across models, sex had the largest effect and smallest p values. Some strains and interaction fixed effects had moderately significant values, but these varied across models.

- Models using datasets with higher quality cells had a similar pattern of the sex effect being the largest factor while some strain and interaction effects had moderately significant p values.

**VARIATION IN DSB NUMBER**

In an attempt to localize the above male specific crossover number evolution within the meiotic pathway we quantified a marker for DSBs, DMC1, in early prophase meiocytes. (we choose strains and sex observations to leverage the – understand the difference between the high and low rec strains (observed above)

()

1. **Lep (early Z) mean number of DMC1 foci per cell has evolved, in the direction predicted by number of COs.**

1. The high rec group has more foci for the early L stage, the no significant difference is observed for the later Z stage There is a stronger correlation between the number of foci in the early L stage (r= 0.8736143) than the later Z stage (r = 0.284302).

**Genome wide recombination rate variation translates to the chromosome level**

gwRR evo -> chrm proportion

1. general pattern – vast majority of bivalents in house mouse have one or two foci

2. The proportions of 1CO and 2CO distinguish the high rec male strains from low re strains

Report the chi square tests, for PWD-MSM, PWD-KAZ and MSM-MOLF.

3.explain why this complicates further comparisons (chromosome size effect \* chromosome class)

**Single Bivalent Level Dataset**

1.-motivation – de-construct the cell wide pattern – apply an algorithm to the MLH1 stained meiocyte images

2 - Algorithm stats and performance (n bivalents (with hand foci), (Table 3 --- Ncells, Nbivs, X, proportion chrm class, )

3. average number of bivalents isolated-measures and estimates of error -- (high rate of replication) (estimates of error compare to manual measures – ref Peterson 2019, table and figure--(not all bivalents isolated average number of bivalents isolated per image, --but such large dataset – assume across all the bivalent datasets across categories ARE EQUALLY REPRESENTATIVE OF GENERAL PATTERNS.

4. <transition to driving questions>

Main motivating questions / We will use this data to address 2 questions:

Q1. Which bivalent level traits will be sexually dimorphic?

Q2. Which traits distinguish high and low recombining strains in males?

**Q1. 1. SC length (Chromatin Compaction Differences)**

(motivation, set up prediction for the (reversed heterochiasmy strains

-logic for the reduced dataset and basic stats (Get around the XX – reduced single bivalent data set – from shortest 5 bivalents from a single cell, the XX is thought to be 3rd longest in reference genome (mb). The XY in males and distinguishable and can be is filtered out from the single bivalent data set)

**Oocytes have longer SC**

**Shortest bivalents**

**Total SC area per cell**

**Sex effect of models are the most significant**

SKIVE was the only strain without significant differences for the short biv dataset, … the short biv samples for skive females are lower than other female means – this could be due to sampling / lower power of the reduced dataset. But this strain does have significant for total SC.

Models - total

Lmer: all are effects significant

Glm m2;

Glm m3

Models – short

Lmer: significant sex and sub\*sex

M2 glm: sex male most sig, skive strain and skiv\*male are slightly sig

M3 glm: sex male is the only significant effect

<the skive effects are likely due to under sampling>

indicate strain effects have evolved – molossinus and musculus male sc is longer than domesticus male (significant sex\*subspecies from mixed model) (also lew and g have v significant strain effects

1. Total SC ---Ttests, lmer, and glm – FEMALE HAVE SIG MORE SC\_AE
2. Female have longer SC metrics even in strains with males have more COs per cell. Suggesting a DECOUPLING of broad summaries of SC length and gwRR. This is not a complete decoupling since in all strains the positive correlation of SC lengths across bivalent classes is held.

2. Longer SC-AE in females is – a consistent feature across all strains (t-tests, model’s (large sex effect)) for a cell wide summary and (a reduced single bivalent data set) (consistent across other species – a conserved aspect of oogenesis)

\*\*3. males from Musc and mol strains have significantly longer SC metrics than males from Dom. Suggests at chromatin compaction – between these subspecies and may (be a requirement for the rapid evolution seen in the 2 musc and mol strains) \*\*\* later I say this isn’t true!!

<transition> -- following up on SC area – with how COs are place --- placement and number of COs on single bivalents

1. Transition, --- the SC-AE area can be thought of as the ‘area / available real-estate for COs

-- these results suggest that this area is ‘utilized’ in different ways across sexes and strains – so we next investigate the recombination landscape – (ie the relationship between the placement and number of COs along chromosomes).

2. We focus on two metrics/ aspects i) the placement of single foci along a bivalent (just 1CO) and ii) the placement of two foci on the same bivalent / the interfocal distance of 2CO bivalents (reflects crossover interference)

**2. Rec landscape differences (Q1. 1CO pos)**

In order to test if there has been evolution of the typical rec landscape differences, we examined the normalized forci position of single foci bivalents (1COs).

Framework notes

-Normalized positions to control for differences in SC length (across sexes)

-(stick to full biv data set) – if results not sig, try smaller data sets

-basic stats – table X – number of 1CO bivalents per category

Results

1. Sex is the most significant factor influencing the normalized placement of single foci along a bivalent (by t-tests, mixed and linear models effects).

2. (Dom and musc strains more significant sex dimorphism compared to Molossinus (– with the dom strains being more telomeric in males compared to mol males)

**3. Sex differences in cytological measures of CO interference**

In order to further test differences in the rec landscape between males and females – due to // attributed to crossover interference, we compared the IFD of 2CO bivalents

1. t.test for sex differences

2. mixed models and glms for evolution and strain effects

Results

-t.tests !!

- Female normalized IFDs are an average of 45 to 50% of the length of the SC-AE across strains (weaker interference)

- Male normalized IFDs are on average 51 to 60% the length of the SC-AE across strains (stronger)

*- High rec males have stronger interference, in terms of raw and normalized IFD measures. They also have a lower threshold for IFDs, 30%, this could indicate stricter control over the REC landscape to enrich for more 2CO bivalents.*

**Variation in gwRR across house mouse males, Framework for Q2**

Which single bivalents traits distinguish high and low recombining males in Musc strains?

- Review dataset of male, additional strains which didn’t have female observations

- General predictions based on the gwRR results:

I. Dom strain will not have significant effect (WSB = LEW = G)

II. Musc strains (PWD > SKIVE > KAZ, CZECHII )

III. Mol strains (MSM > Mol)

-glms and logistic regression models

**Q2. SC length, High rec strains have longer SC lengths, but depends on how you measure it**

motivation

Main results:

1. general positive correlation of SC length and more COs per chromosome is confirmed … (test? Figure?)
2. but evidence that evolution of SC length metrics are decoupled to an extent, In most comparisons of high and low rec strains (using mouse means) the SC metrics show no significant difference.
3. Qualitative patterns from Looking at the figure for **Total.SC mouse.av** :
   * domesticus has lower mouse averages compared to high rec AND (KAZ mouse means are close to the Dom means), MSM strain mean is significantly higher total SC area.
   * BUT, the other low rec strains, have ‘high’ total sc means despite having low gwRR
   * PWD (has quite a bit of variance across the mouse means.

Plot(the strain averages for x=gwRR and y=total.Sc

1. Only at the level of single bivalent SC metrics are the expected contrasts significant.

--‘clean’ predictions based off of gwRR aren’t met –(but the glms for the SC metrics …. Do they indicate / support evolution

-subspecies effects are significant for total sc and reduced dataset

(What evolutionary patterns are supported?

Divergence…. Dom < (Musc and MOL) the M1 model for total sc – has significant subspecies effects (longer SC) (same result for the short and bivalent dataset, in the long biv M1 model just musc is significant

-M2 models : most strains have significant effects – (wide spread evolution)

* Evolution in totalSC area per cell which is decoupled from the gwRR evolution. … but there is some (weak) support for musc and mol effects on sc metrics (total sc)

- **the high and low strains CANT be distinguished by mouse averages for SC metrics, only with the total SC for all pooled measures**

**The evolution of interference correlates/is paired with evolution in crossover interference,**

**Higher rec strains have stronger interference**

(Motivation: how foci are placed on the SC, most predictions for this rec landscape metric are that bivalents with weaker interference will be more room / space to fit more crossovers.)

Main results:

t.test – High Low: significant, mouse averages and single bivalent level significant

glmers – only the 3 strains have significant effects

rates of short IFDs (<30%) drives the differences in IFD distributions less than 5% rate in high rec,

~10 to 15% in high rec strains

(caveats and explanation: )

***-High Rec strains have stronger interference measured via longer normalized IFDs.***

**1CO rec landscape evolution is decoupled from gwRR evolution**

motivation:

Main results:

1. WSB has the most terminal 1CO rec landscape and MOLF has the most central.

2. Evidence for evolution of the 1CO rec landscape, but it is decoupled from the evolution in gwRR we observe in figure 1

2.5 The normalized 1CO position is not a good metric for distinguishing strains with rapid gwRR evolution