**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.

-number of cells (spermatocyte / oocytes)

- description of MLH1 dataset and measures of repeatability

- male rates are adjusted (+1) for the PAR – since the XX is a confounding factor for female cells.

- table of the means and summary statistics, ranges

- Our results are similar to previously reported measures, and report novel measures from wild derived inbred strains in house mouse.

-how close the means are to the minimum.

**- models / results for mouse mean**

**Models / results for mouse var**

- table’s --

(In order to – make more comparable summary comparisons – assessed the sex specific patterns initially. /(the variation across strains) with glms for each sex)

**-female:**

-**male**: ranges, of strain means, (ranges of mouse means)

PWD, MSM, and SKIVE have significant strain effects. They are grouped into the ‘High Rec’ group. PWD, MSM, and SKIVE have evolved 20, 30% and 10% higher than other means respectively.

- for females, G, LEW, and MSM has significant strain effects

are 7% higher than other female means, MSM is 6% greater than the other female means (G and MSM could be the high rec females)

(interpretation)

*- Sex specific evolution is the major pattern.*

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

(below might need to be moved to the methods section)

**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)

(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 )

( in order to follow up with the mixed models – we used post hoc glms to more specifically test the strain effects )

-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. These results suggest there is a lack of support for uniform divergence in the trait which would be the expected pattern under a simple neutral evolution

After mixed models, we ran post hoc fixed effect models

-glm: The G strain effect is the largest effect for gwRR. The sex\*strain interaction effects for PWD and MSM are also significant. Additional glm models show that PWD\*male and MSM\*male increase the gwRR and G\*male decrease the gwRR. Qualitatively the difference between the G female and male means are greater than other strains.

1. *Predictions for ‘uniform’ patterns across all Mus musculus strains are not met; (ie. strains within subspecies didn’t diverge uniformly)*
2. *Sex is a significant effect, but not in a uniform manner, the significant fixed effects are interactions (strain \* sex).*

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

(in order to compare the within organism variance – we quantified the **mouse average** variance and coeffieient of variance (cV) of MLH1 foci per cell)

- 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 locatize the above male specific crossoever number evolution within the meiotic pathway we quantified a marker for DSBs, DMC1, in early prophase meiocytes.

-basic stats, strains quantified from juvenile mice, cells selected based on stage of prophase. Means and number of cells reported in table.

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

Boring results; 1. Early staged cells have significant more foci (DSBs) than the later stage.

Unexpected / new results

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).

2. TABLE OF CO : NCO RATIOS ( DSB / MLH1 = estimated proportion of NCO. ) –**correlation with MLH1**

L based ratios: WSB, G and KAZ: 7.3 and 6.5, 6.6\* PWD and MSM: 6.1 and 7.4

Z based ratios: WSB, G and KAZ: 5.4 and 5.9, 6.9 PWD and MSM: 4.8 and 5.3

+ t.tests of ratios

\*\* the above ratios don’t track the Crossover number variation – (this is lack of support of the evolution being associated with the repair decision process – and rather might indicate that – the important metric is upstream before DSBs are established)\*\*

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

( in order to decompose the cell wide rate, we decided to look at the proportion or chromosomes with different numbers of COs)

- Review limits and comparisons of the number of COs per chromosome. Most species have a range of 1 to 3 COs per chromosome (Stapley et al 2017).

- The male high rec strains have significantly more 2CO bivalents than low rec strains. This skew in chromosome class proportions isn’t seen in female strains.

*- This is motivation to investigate more traits at the chromosome level*

**Single Bivalent Level Dataset**

- Review lit and previous findings for single bivalent measures (FISH, tetra and polar body sequencing).

- Description of the data set and brief description of the biological relevance for the following metrics:

1. SC Length,
2. Normalized 1CO position (rec landscape)
3. CO interference via interfocal distance (IFD) of 2CO bivalents

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?

Validity of comparing bivalent observations

While the automated software doesn’t isolate all bivalents/chromosomes from each cell (on average 17), we assume that the isolation process is not biased. Because there are hundreds of observations per category, we assume that each of the 19 autosomes (chromosomes) is equally represented in the dataset of single bivalents.

-we primarily use the mouse average of the choosen bivalent metric … because

**Heterochiasmy starts at the recombination landscape of single bivalents (SC, rec landscape, and interference)**

In order to assess of SC lengths vary between sexes – we com

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

Our data set provides an opportunity to test if the canonical pattern of females having higher gwRR and more SC area) – have been uncoupled in the instance of rapid male specific evolution in the gwRR of PWD and MSM.

-boring results out of the way

-positive correlation with length and CO per bivalent (sc length across bivalent classes)

-strong sex specific signal using the full single bivalent data set, – but watch out the interpertations is complicated by sex chromosomes, and difference chrm class proportions

1. Total sc, (motivation )Total sc compares the SC area of the whole cell

Basic stats; X number from all the original,

Female average total sc area

t.tests

lmer

glm

1. short bivalent

-use the short bivalent dataset to control for XX in females

basic stats

t.test :

in all but SKIVE, the female mean length of short bivalents were significantly longer. For SKIVE (p= NS) qualitatively the female means are lower compared to the other musc female means (KAZ and PWD).

Lmer-- Sex is the most significant effect

Glm --- Sex is the most significant effect

1. <Simple model / prediction not met, suggesting a DECOUPLING of broad summaries of SC length and gwRR – female have longer SC metrics even in strains with males have more COs per cell. 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)

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)

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

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

- logic and biological significance, normalized single foci positions from 1CO bivalents

-Basic stats, X number of single bivalents for each category, mouse averages calculated..

**-**The sex effect is highly significant for glms of the normalized foci 1 position. In models for the full dataset, LEW, MOLF, and KAZ had significant strain effects, which MOLF being the largest. The LEW and the MOLF strain effects are replicated in models sub setting the subspecies.

-Males have more telomeric positions of single foci while female single foci are closer positioned to 50% of the total length.

-Some strains MOLF and LEW, have significant strain effects, indicating both male and female have positions closer to the middle of the bivalent than the total average.

- *SC length and the normalized 1CO position are conserved chromosome level traits of heterochiasmy in house mouse (and other species).*

**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 differences compared to Molossinus (– with the dom strains being more telomeric in males compared to mol )**

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

- Review of CO interference measures (approaches and species). Example tiger vs elephant shrew show a clear negative correlation between interference strength and genome wide recombination rate (Segura et al).

– Note that sex specific comparisons of IFD / interference in the physical scale (SC-AE) are very rare, De Boer et al 2006 (maybe Hassold).

- Review genetic map measures sex differences in COI and how this is also connected to more uniform REC in females and localized REC in males.

- Female normalized IFDs are an average of 45 to 50% of the length of the SC-AE across strains. Male normalized IFDs are on average 51 to 60% the length of the SC-AE across strains.

- Generally all females have enrichment of short normalized IFD (except KAZ). This enrichment is most pronounced in strains with high rec males, (PWD, MSM, SKIVE) there is a cut off of low normalized IFDs ~30%. In the remaining strains the normalized IFD ranges overlap between males and females.

- *Females have weaker interference as indicated by normalized IFD.*

*- 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**

-General pattern of positive correlation with SC lengths and chrm class – any deviations can be attributed to low sample sizes

- CHROMOSOME SIZE EFFECT // chromosome proportion differences confound comparisons of chromosome classes across strains

-mouse effects!!

- compare sc mean by chromosome classes for strains with similar proportions

-compared reduced bivalent data sets (long and short biv data)

Lmer / GLM

- All strains effects are significant (for mouse averages)

Full

Short

long

**Weak correlation with lower gwRR and terminal CO landscape**

-Review why only 1CO bivalents and mouse average normalized measure is used.

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

-High rec strains have more central normalized Foci1 pos, but so do other strains: G, **PWD**, **MOLF**, **SKIVE** and KAZ, are significant strain effects in the full model.

-High and low strains are not clearly predicted by logistic regression modes (for mouse average normalized Foci1).

**Higher rec strains have stronger interference**

- PWD, MSM, and SKIVE are significant strain effects in glm for raw IFD (mouse averages).

-PWD and MSM are significant strain effects for normalized IFD.

-Mouse average for normalized IFD is significant in predicting high and low rec strains in the model where all strains are pooled.

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