**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.

1. -female strain means are v close -male strain means are more varied!!

-Comparisons to references / literatures

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

2. The largest female difference is G, 1.07X the largest male difference is PWD and MSM which are 1.3 and 1.2 respectively.

**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)

(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 )

**- 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**: G and MSM

-**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*

-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

**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).

\*\* 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*

1. Basic stats – bivalent data numbers
2. Rang of co per chrm, mostky 1co 2co

gwRR evo -> chrm proportion

1. High rec strains have evolved a higher proportion of 2CO

For males Chi squared test, p = low group, p = MSM and PWD, SKIVE is intermediate)

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

(across strains with v. different proportions

**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?

**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

Ttests, lmer, and glm – FEMALE HAVE SIG MORE SC\_AE

1. **Short bivalent data set**

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

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

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 bivalens (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

Framework

-- raw and normalized measures used, (raw beceause as a predicted mechanical / physical force – this should be conserved,

And normalized values – to account for SC length differences // get a better idea for the ‘landscape’

-normalized ifd to account for inherent difference in the SC lengths

-basic stats of the number of observations (mouse averages used)

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.*

-1.conserved raw IFD values?

**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

framework

total SC area and reduced bivalent datasets

t.test – difference between the groups

or logistic regression?

GLMs

For each of these metrics

1. total SC
2. long biv
3. short biv

-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

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

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

**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.*

**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).

None of the t-test were significant, normalized 1CO position is not a good metric for distinguishing the two groups