**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

we quantified the **mouse average** variance and coefficient 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 localize the above male specific crossover 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.

-only spermatocytes, choice of strains

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.

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**

- 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

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

(SKIVE-KAZ)

**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 undersampling>

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 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

brief ref to the Q1 findings

(while testing the sex differences, -- there were results which suggest that the musculus and molossinus strains had longer SC compared to the domesticus strains – these two strains also contain the high rec strains – motivating a model/hypothesis that sc length / chromatin compaction evolution may proceed – facilitate the rapid male specific evolution in gwRR)

framework

we apply t.tests to test if with subspecies the high rec strains have significantly more/longer sc metrics.

We run these comparisons on the total sc area pre cell and two reduced bivalent datasets, -long and short.

Plots (total sc)

KAZ has longer total sc means, the other (new) strains I added in have higher total sc means. The lack of significant difference between the high and low strains, is driven by higher total sc levels in tom, ast, and Czech.

t.tests – High vs Low

1. total SC – All- significant difference, musc – NS, mol sig longer
2. long biv --- ALL NS, MUSC -- NS, MOL -- NA not enough observations
3. short biv – ALL NS, MUSC -- NS, MOL -- NA not enough observations

BIV-level, High vs Low

1. ALL p-value = 2e-13

2. long p-value = 0.4

3. short p-value = 0.7

t.tests Dom vs Mol and Dom vs Musc

1. total SC -- dom-musc p=.5, dom-mol p=.1

2. short SC – Dom-musc NS, dom-mol p=0.09

3. long SC -- Dom-musc p=0.08, dom-mol NS

BIV-level

1. ALL DOM-MUSC p-value = 4e-13, DOM-MOL p-value = 6e-10

2. long Dom-MUSC p-value = 0.2, Dom – Mol p-value = 0.9

3. short Dom – Musc p-value =0.8 , Dom – Mol p = 0.06

**f-Weak support of dom sc being shorter than the other subspecies – the significant differences are when bivalent-level (instead of mouse av) where there are significant differences between Dom-and (musc& mol) and also High and low groups**

- **the high and low strains CANT be distinguished by SC metrics – only with the total SC for all pooled measures (PWD+MOL+SKIVE**

t-test results not totally clear ---with the reduced bivalent datasets, it could be lack of power (these tests don’t distinguish between models where all strains within subspecies have evolved longer SC

so follow up with glms

GLMs mouse av.

1. total SC – M1 Musc, Mol, G LEW MSM, SKIVE KAZ M2. All of the strain effects are significant
2. long biv -- M1 – musc and msm are significant, M2. MSM and SKIVE and CZECH have significant effects
3. short biv – M1. Musc and Mol are significant, G strain effect is sig M2. ALL strains have significant

GLMs single bivalent // cell

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**

Motivation: how foci are placed on the SC

brief ref to the Q1 findings

t.test – High Low

glmers

(mean\_IFD.2CO\_PER ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM)) M2- all of the high rec strains have significantly longer mouse mean normalized IFD.

The glms which subset by subspecies support that model, (Dom: no significant strain effects, Musc: KAZ and CZECH have significant negative effects (compared to SKIVE and PWD) Mol: MOLF has slightly significant negative effect.)

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

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

Bivalent level measures – (with chromosome length incorporated)

-PWD, SKIVE, MSM, MOLF have sig effects/ p values

The single fixed effects can be interpreted as, with each unit increase in SC length, the IFD^PER^ also increases (for the high rec strains + MOLF). All of the high rec strains and MOLF have significantly longer IFD^PER^.

There are also significant interaction effect of strain and chromosome length, in the high rec strains and MOLF, with each increase in bivalent length, the IFD^PER^ decreases slightly.

-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**

Brief ref to the Q1 findings, (motivation is driven by the lack of positive correlation for CO~SC relationship. (terminal CO position – is similar to stronger interference / while medial positions are analogous to weaker interference).

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

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

-Musc and Mol aren’t clearly different from Dom strains

**None of the t-test were significant, normalized 1CO position is not a good metric for distinguishing strains with rapid gwRR evolution**

(chromosome size is a significant effect within the 1CO positions – what’s the direction? What does this mean? How would the different proportions affect this?

USING models for the single bivalent level – (can incorporate SC length)

Strain \* chromosome size effect – (LEW)