Finalize the methods for MM (MLH1) (which function ect

Summarize – the results (Predictions not met,

# MLH1 and Mixed Model

Given that – gwRR varies – what are the patterns of variation? (Do they fit any specific evolutionary models?)

(What evolutionary models fit/explain the patterns of variation for gwRR?

(major forces

**Sex –**(if sex / heterochiasmy is the driving force for gwRR variation --- this effect will be significant)

**Divergence** – constant (different evolution on different branches)

(mice / inbred strains from more divergent strains will be more dissimilar than mice from the same sub species

**Polymorphism** – (effect which allows variation within subspecies )

THE FINAL MODEL:

Examine, mouse averages for MLH1 mean and variance

Subsp: 3 (unordered factors)

Sex:

Strain: (nested within subsp, still random, can just one be random if it’s nested?)

**Re-run the Mixed model with lmer!!!**

1. lme(mean\_co ~ subsp\*sex, data=DF.HetC.MixedModel.HQ, random=list(strain=pdDiag(~sex) ) )
2. lme(mean\_co ~ subsp \* sex, data=DF.HetC.MixedModel.HQ, random = ~1|strain)
3. lmer(mean\_co ~ subsp \* sex + (1|strain), data=DF.HetC.MixedModel.HQ, random = ~1|strain)
4. lm(mean\_co ~ subsp \* sex \* strain, data=DF.HetC.MixedModel.HQ)
5. lm(mean\_co ~ sex \* strain, data=DF.HetC.MixedModel.HQ)

Are there ways to distinguish the models:

1. and 2. They have random effects …

3. and 4. Simple glm model for comparing effects of strain and subsp

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model old (remove)** | **intercept** | **subsp** | **sex** | **subsp\*sex** | **Strain (random)** | **Interpert** |
| 1.MLH1  Mouse.av | 26.36 | 1.053918 (Mol)  -0.611786 (Musc) | -1.5988 | 1.879296 (mol)  2.886234 (Musc) | Need to test random effects | Fixed effects not sig |
| p values |  | 0.422654  0.577404 | 0.298647 | 0.455353  0.191630 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model 2**  **Lme – fixed effects** | **intercept** | **subsp** | **sex** | **subsp\*sex** | **Strain (random)** | **Interpert** |
| MLH1  mouse.av | 26.348630 | -0.520512  -0.724686 | -1.62829 | 3.274996  2.909844 | ExactRLRT  RLRT = 26.933 | Sex is a significant fixed effect, |
| p values | 0.000000 | 0.774041  0.649822 | **0.000943** | **0.002097**  **0.000149** | < 2.2e-16 | Musc and Mol have sig interaction effects  Compared to model 1, random sex variance is shifted to this interaction term. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model 2**  **Lmer – random p value** | **intercept** | **subsp** | **sex** | **subsp\*sex** | **Strain (random)** | **Interpert** |
| MLH1  mouse.av | 26.3486 | -0.5205  -0.7247 | -1.6283 | 3.2750  2.9098 | ExactRLRT  RLRT = 26.933 |  |
| p values |  | --  (if the 2 are additive, 0.5244 ) | **--**  if the 2 are additive, 0.5876 ) | **0.0002045 \*\*\***  **(drop1)** | < 2.2e-16 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model 3, all fixed effects** | **subsp** | **Sex** | **strain** | **Sex\*subsp** | **Sex\*strain** | **Interpert** |
| 3. lm() ALL 3 fixed effects and interactions | 2.907357  0.855690 | -0.156443 | **3.297524**  **1.696468**  0.302533  0.069857  0.370667 | -3.222557  -0.807890 | **-3.171906**  -1.190954  **4.400842**  **6.946393**  2.262333 | Strain effects by themselves are more significant (LEW, G) |
| p values | 0.096333  0.234412 | 0.822008 | **0.000002** G  **0.019408** LEW  0.669318 PWD  0.967945 MSM  0.834162 SKIVE | 0.431760 | **0.001651**  0.278086  **0.000054**  **0.000920**  0.256521 | The sex\*strain is sig in 3 strains, G-female higher, MSM-PWD male. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **4. JUST sex and strain** | **sex** | **strain** | **Sex\*strain** | **Interpret** |
| Coefficients |  | 3.297524 -G  1.696468 -LEW  1.158224 -PWD  2.977214 -MSM  2.907357 -MOLF  1.226357 -SKIVE  0.855690 -KAZ | **-3.171906** (male-G)  -1.190954  **3.592951**  **3.723836**  -3.222557  1.454443  -0.807890 (male-KAZ) |  |
| p= | 0.822008 | **0.000002** -G  **0.019408** -LEW  **0.065356** -PWD  **0.000007** -MSM  **0.096333** -MOLF  0.480969 -SKIVE  0.234412 -KAZ | **0.001651**  0.278086  **0.000535**  **0.001962**  0.099050  0.460095  0.431760 | Sex and strain interactions – PWD and MSM male higher  G female higher |

Summary of the mixed models for mean CO

\*

M1: variance due to strain effect, is greater than 0. Sex and the interaction effects are significant. (but just subp isn’t)

M2:

M3: G female, PWD and MSM males are significant strain \* sex effect. (Is there a change)

# Female Specific gwRR Pattern

Female Table

WSB G LEW PWD MSM MOLF SKIVE KAZ CAST

14 12 9 15 14 1 1 9 1

Run these mouse av, observations through these 2 models

1. lm( mouse.av.CO ~ subsp \* strain )
2. lm( mouse.av.CO ~ strain )

|  |  |  |  |
| --- | --- | --- | --- |
| Female lm 1 | Intercept (WSB) | G | LEW |
| Co-efficient | 24.71164 | 3.29752 | 1.69647 |
| p value | < 2e-16 \*\*\* | 1.06e-05 \*\*\* | 0.0273 \* |

None of the subsp factors are significant. Two of the strain effects are significant, none of the subp-strain interaction effects are significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Female lm 2 | Intercept (WSB) | G | LEW | PWD | MSM |
| Co-efficient | 24.7116 | 3.2975 | 1.6965 | 1.1582 | 2.9772 |
| p value | < 2e-16 \*\*\* | 1.06e-05 \*\*\* | 0.0273 \* | 0.0810 . | 3.02e-05 \*\*\* |

G female, is 1.07 greater than the average female mean CO count per cell.

**Within Mouse Variance**

Lew – is significant for both models (significantly more within mouse variance) (try this is Q12)

# Male Specific gwRR Pattern

Male table (extra males which don’t have female observations)

WSB G LEW PERC PWD MSM MOLF SKIVE KAZ TOM AST CZECH CAST HMI

12 18 10 1 8 8 6 6 13 2 3 3 2 4

Run these mouse av, observations through these 2 models

1. lm( mouse.av.CO ~ subsp \* strain )
2. lm( mouse.av.CO ~ strain )

Significant strain factors for the male data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Male lm 1 | Intercept (WSB) | PWD | MSM | SKIVE | AST |
| Co-efficient | 24.45317 | 6.11271 | 7.21425 | 3.88183 | 2.10533 |
| p value | < 2e-16 \*\*\* | 3.00e-08 \*\*\* | 5.44e-14 \*\*\* | 0.000359 \*\*\* | 0.084032 . |

This is kinda different than the last time I ran the model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Male lm 2 | Intercept (WSB) | PWD | MSM | SKIVE |
| Co-efficient | 24.45317 | 4.85321 | 7.00108 | 2.62233 |
| p value | < 2e-16 \*\*\* | 2.48e-10 \*\*\* | < 2e-16 \*\*\* | 0.000626 \*\*\* |

I use these results as the primary evidence that PWD, MSM and SKIVE should be classified as ‘High Rec’ group.

PWD\_male gwRR is 1.2, MSM\_male gwRR is 1.292392 and SKIVE\_male gwRR is 1.112478 compared to the other male gwRR means.

Variance for males neither of the two models gave significant results (or very slight sig).

# MLH1 countVariance Models

# (use variance instead of CO) – Testing what effects within Mouse variance in CO counts (focus of chapter

HetC.MixedModel.HQ <- lme(VAR ~ subsp\*sex, data=DF.HetC.MixedModel.HQ, random=list(strain=pdDiag(~sex) ) )

1. Reduced.strain.HQ <- lmer (VAR ~ subsp \* sex + ( 1|strain ), data=DF.HetC.MixedModel.HQ )
2. lm(mean\_co ~ subsp \* sex \* strain, data=DF.HetC.MixedModel.HQ)
3. strain.fixed <- lm(mean\_co ~ sex \* strain, data=DF.HetC.MixedModel.HQ)

**–** main signal to look for across quality bins is if the female.var > male.var (sex effect for variance is maintained)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| M1. Var  **Lmer()!!** | intercept | subsp | sex | Subsp\*sex | strain | Interpt |
| 2.**Var** | 15.148325 | -0.494389  -1.953908 | -7.923511 (male) | 0.589028  1.595256 |  |  |
| Pvalues | 0.000000 | 0.84388 0.376361 | 0.000000 | 0.836295  0.461209  0**.7551 (drop1)** | **0.0375** | P values on top line come from lme model  Lmer pvalues |
| **M1. cv**  **Lmer()!!** | 14.0930 | -0.2205  -0.4109 | -3.6022 | -0.4417  -0.5531 |  |  |
| P values |  |  |  | 0**.8843 (drop1)** | **0.0505 exact()** |  |
| **Q12 M1. Var** | 15.14832 | -0.494389  -1.953908 | -7.923511 | 0.589028  1.595256 |  |  |
| pvalues | 0.000000 | 0.843889  0.376361 | 0.000000 | 0.836295  0.461209  **0.03049 \* drop1()** | **p-value = 1** |  |
| **Q12**  **M1 cv** | 14.91782 | -2.137928  -1.642817 | -5.323350 | 1.456652  -0.571496 |  |  |
| pvalues |  |  |  | **0.359** | **0.3883** |  |

The fixed sex effect has the largest effect sizes. All are negative (for male). (I’m not sure if I should switch to testing the additive fixed effects by themselves) – I’m pretty sure the outcome would be that the sex effect gets very large in those models.

The strain effects are only slightly significant for within mouse variance measures in the full dataset (includes low quality cells)

(For the next two sets of models, which have many interaction effects, I only report the p values for significant ones)

M3 lm(mean\_co ~ subsp \* sex \* strain, data=DF.HetC.MixedModel.HQ)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **M3** | **Subsp** | **Sex** | **strain** | **Sex\*subsp**  **(too many to list)** | **Sex\*strain** |
| **var** | 6.333  3.197 | -4.646 | Too many? | Too many? |  |
| **P values summary** | 1.63e-12 \*\*\* (int) | 0.044181 \* (male) | 0.000169 \*\*\* (LEW) |  | 0.027610 \* (male LEW) |
| p values  (anova) | 0.36414 | 4.675e-12 | 0.05789 . | 0.74179 | 0.17383 |
| **cv** | 2.21049  1.63817 |  |  |  |  |
| **Pval. summary** |  | 0.02244 | 0.00407 (LEW) |  | --? |
| **p values**  (anova) | 0.35157 | 3.188e-13 \*\*\* | 0.08555 . | 0.80793 | 0.53015 |
|  |  |  |  |  |  |
| **Q12 var** |  |  |  |  |  |
| Summary() |  | 0.000228 \*\*\* (male) |  |  |  |
| p values  (anova) | 0.10536 | 2.362e-11 \*\*\* | 0.48325 | 0.05775 . | 0.22231 |
| **Q12 cv** |  |  |  |  |  |
| Summary() |  | 3.16e-05 \*\*\* (male) | 0.0919 . (G) |  | 0.0534 . (G male) |
| p values  (anova) | 0.06403 . | 2.429e-13 \*\*\* | 0.52698 | 0.37513 | 0.12912 |

For LEW – females must have high variance, for CV Lew strain alone has increased cv. Consider plotting the variance – and diving deeper into this pattern

Most of the strain specific differences in within mouse variance go away in the Q12 dataset.

M4. strain.fixed <- lm(mean\_co ~ sex \* strain, data=DF.HetC.MixedModel.HQ)

The summary for lm’s will give p values --- the are different than the anova p values

|  |  |  |  |
| --- | --- | --- | --- |
| **M4** | **sex** | **strain** | **Sex\*strain** |
| Coefficients | -4.6455 | 9.1510 | -8.0239 |
| P values  (summary) | 0.044181 \* (male) | 0.000169 \*\*\* (LEW) | 0.027610 \* (LEW male) |
| P (anova) | 3.195e-12 \*\*\* | 0.1133 | 0.3044 |
| **CV** | -2.60550 | 3.40510 (LEW) |  |
|  | 0.02244 (male) | 0.00407 \*\* |  |
| P (anova) | 3.195e-12 \*\*\* | 0.1133 | 0.3044 |
|  |  |  |  |
| **Q12Var**  – coeff | -13.4472 | -8.7385 (PWD) | 10.1708 |
| Summary | 0.000228 \*\*\* | 0.010686 \* | 0.034838 \* (male PWD) |
| anova | 2.487e-11 \*\*\* | 0.24657 | 0.08073 . |
| **Q12**  **cv** | -7.05868 (male) | -2.5841 (G)  -4.36531 (PWD)  -2.64563 (MSM) | 4.0256 (male G)  4.14951 (male PWD) |
| Summary | 3.16e-05 \*\*\* | 0.0919 . (G)  0.00573 \*\* (PWD  0.08002 . (MSM | 0.05336 . (male G)  0.06073 . (male PWD) |
| anova | 4.009e-13 \*\*\* | 0.1299 | 0.1700 |

Do I have a good justification for using the summary p value and anova?

Even with the inconsistent strain effects, females still have more within mouse variance. (How much does strains differ for females in full and Q12?)

**Qualitative Descriptions (Summary for within mouse variance on CO count per cell)**

*-all strains / data sets show that females have more within mouse variance for MLH1 counts per cell.*

-Female observations have some variation (but are largely similar)

-male specific rapid evolution of two inbred strains from two subspecies, (note Musc and Mol share ancestry)

-Musc has gradient of polymorphism for gwRR across inbred strains (PWD > SKIVE > KAZ)

(WANT to say something like – multiple types of models are applied but most all of them back up the qualitative discriptions )

# Q1 Single Bivalent Heterochiasmy Predictions

*Which bivalent level traits will be sexually dimorphic?*

1. **SC length, *SC length will be sexually dimorphic***

**(result**) A. single SC are longer in females than male

B. (this differences in SC size is beyond the effect of XX), (the expected difference if XX was the only source of difference), ALL bivalents (by class) are longer in females than males. The distributions for female SCs within cells are all longer. (The bottom 25% of bivalents are longer in females than males)

**2) Normalized CO positions**

**A. *1CO normalized positions will be sexually dimorphic***

Telomere bias and uniform 1CO landscapes are (pretty conserved in all strains) (look closer at a few male Dom strains)

**B. Sis-co-ten (sister cohesin tension) will also be sexually dimorphic:** this metric reflects the overall uniform vs telomere pattern documented in males and females.

**C. Centromere and telomere distances will be sexually dimorphic**; males having more telomere positioned COs

**3) IFD,  *No predicted difference in physical distance between foci on the same bivalent across sexes.***

A. raw IFDs are longer in females, and normalized IFDs are longer in males. This is driven by more observations of lower normalized IFDs in females. (Try running the MM without PWD, MSM and SKIVE – Is sex a sig factor for normalized IFD) (I am not sure if I could classify the close IFD positions are different classes)

B. the high Rec strains (and SKIVE) have a male specific 30% normalized IFD lower threshold. (Males in these strains lack bivalents with very close foci on 2 CO bivalents)

Y = subsp \* sex + random(strain)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Traits** | Fixed Effects | | | Random Effect | Predictions met |
| **subsp** | **sex** | **subsp\*sex** | **strain** |
| SC length  All pooled | 0.364038  NS | -28.769236  Highly Sig | 6.037493  NS | Slightly significant | Yes, female bivalents are longer (still need to test XX adjustment) |
| Long.biv | -0.126264  NS | -23.719327  High sig | 6.096744  NS |  | (this model started throwing error) |
| 1CO | 1.616610  NS | -24.979951  Highly sig | 1.646027  NS |  |  |
| 2CO | -- | - | - |  | Can’t get these models to work |
| 3CO | -- | - | - |  | Can’t get these models to work, Na.omit() |
| 2CO\_IFD.raw | 0.448646  P= 0.8551 | -6.358345  P= 0.04751 | 3.661211  P= 0.38178 | X | **No**, the male effect coeff is  -6.5. Female IFDs are slightly longer. |
| 2CO\_IFD.PER  (1-0) | 0.004682 p= 0.737 | 0.072634  **p= 0.00553** | 0.026687  p= 0.41905 | p=0.0548 |
| 2CO\_IFD.PER  (-PWD,MSM,  SKIVE) | 0.008505  p=0.707425 | 0.071628  **p=0.000003** | -0.027552  p=0.320879 |  | Males slightly higher, sex is still significant. |
| Nrm.1CO.pos | X | p=0 | X | p=0 | Yes, the CO position (‘landscape’) traits are sexually dimorphic |
| Sis.co.ten | -0.981277  p=0.712167 | -20.767638  p=0 | 9.928147  p=0.005 | p=0.0018 |
| Telomere.dist | X | p=0 | X | p=0.0021 |
| Cent.dist | X | X | p=0.0009 | X |

-These results confirm that i) SC length/chromatin condensation and ii) differing recombination ‘landscapes’ are conserved chromosome level aspects of heterochiasmy.

-The predication for the IFD differences didn’t meet the null expectation. The IFD tests indicate that female IFD’s are generally longer compared to males.

-There might be more nuanced patterns and effects due to strains for the rec landscape and IFD. I plan to follow up these tests with post-hoc comparisons within each strain.

Do I have predictions for within strain comparisons?

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SC Length**  **Means** | **Fmean**  **1CO** | **F**  **N**  **1CO** | **Mmean**  **1CO** | **M**  **N**  **1CO** | **Fmean**  **2CO** | **F**  **N**  **2CO** | **Mmean**  **2CO** | **M**  **N**  **2CO** | **Fmean**  **3CO** | **F**  **N**  **3CO** | **M.mean**  **3CO** | **M**  **N**  **3CO** |
| WSB | 92.58 | 540 | 68.86 | 472 | 112.58 | 187 | 85.62 | 107 | 121.70 | 31 | -- | -- |
| LEW | 94.965 | 433 | 73.92 | 409 | 112.57 | 245 | 91.868 | 106 | 125.43 | 28 | 89.50 | 10 |
| G | 101.77 | 457 | **78.455** | 797 | 123.55 | 242 | 94.49 | 183 | 135.58 | 19 | 95.20 | 10 |
| PWD | 100.25 | 698 | 77.49 | 365 | 119.44 | 286 | **97.24** | 301 | 118.41 | 36 | **99.33** | 9 |
| MSM | 104.69 | 362 | **80.48** | 77 | 127.51 | 159 | **95.406** | 91 | **150.62** | 21 | **97.000** | 5 |
| KAZ | 96.20 | 351 | **79.006** | 502 | 122.51 | 128 | 91.631 | 84 | **166.00** | 5 | 92.57 | 7 |
| SKIVE | 92.304 | 352 | 73.05 | 510 | 110.88 | 113 | 91.956 | 253 | 116.5 | 20 | **98.0** | 6 |
| MOLF |  |  | **78.58** | 326 |  |  | 92.794 | 97 |  |  | 94.6 | 5 |

(make a standardized name for these tables in R).

The 1CO pattern in males is a bit nuanced. The male high rec strains have longer 2CO bivalents. G also have higher average SC.

Run all these thru mixed models

--I should try to subsample these

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **IFD** | **F.n** | **M.n** | **F.**  **mean**  **raw** | **M.IFD.**  **mean**  **raw** | **F.**  **norm** | **M.**  **norm** |
| WSB | 187 | 107 | 52.50538 | 47.14019 | 0.4602812 | 0.5362821 |
| LEW | 245 | 106 | 53.96327 | 51.80952 | 0.4791878 | 0.5533493 |
| G | 242 | 183 | 60.44583 | 52.55738 | 0.4886198 | 0.5477403 |
| PWD | 286 | 301 | 57.15088 | **58.81333** | 0.4770388 | **0.6018426** |
| MSM | 159 | **91** | 60.67089 | 56.25275 | 0.4711622 | 0.**5874693** |
| KAZ | 128 | **84** | 63.27344 | 48.94048 | 0.5087810 | 0.5239928 |
| SKIVE | 113 | 253 | 53.93805 | **57.61508** | 0.4846905 | 0.**6266287** |
| MOLF |  | 97 |  | 47.36458 |  | 0.5133379 |

The SKIVE male is normalized IFD is the highest! SKIVE male also has high Norm IFD, like PWD and MSM. – Is it strange that SKIVE has intermediate gwRR (but maybe some aspects are like MSM and PWD). \*\*SKIVE has (IFD) patterns that is closer to the high MSM-PWD\*\*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CO.Positions** | **F.n** | **M.n** | **F.mean** | **M.mean** | **Test (follow up on the sig fixed effect)** |
| WSB |  |  |  |  |  |
| LEW |  |  |  |  |  |
| G |  |  |  |  |  |
| PWD |  |  |  |  |  |
| MSM |  |  |  |  |  |
| KAZ |  |  |  |  |  |
| SKIVE |  |  |  |  |  |

(which metrics are used for this … normalized 1CO position) – In prev MS – I ran KS test to compare the density plots.

(centromere and telomere – these might not be the best metrics)

(sis-co-ten)? Sis-co-ten by chromosomeLength --- There are distinct patterns by across sexes – In males the chrm classes (1CO, 2CO) separate/cluster very cleanly – where as they don’t separate / cluster in females.

**Sis-co-ten** – Even with in the normalized siscoten, I think the male data shows more clustering. I’m not sure

Sex and interaction are significant. – what does normalized sis.co.ten mean? I’m not sure that it’s most biologically relevant since I don’t think cohesion adjusts to size of chrms.

Table of proportions below 30% IFD PER

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2CO | 2CO | Below  30% | Below  30% | Female ratio 2CO IFD  PER below 30% | Male ratio 2CO IFD  PER below 30% | F - M |
| **WSB** | 187 | 107 | 16 | 35 | 0.18716578 | 0.14953271 | 0.03763307 |
| **G** | 242 | 183 | 15 | 28 | 0.11570248 | 0.08196721 | 0.03373527 |
| **LEW** | 245 | 106 | 10 | 30 | 0.12244898 | 0.09433962 | 0.02810936 |
| **PWD** | 286 | 301 | 4 | 46 | 0.16083916 | 0.01328904 | 0.14755012 |
| **KAZ** | 128 | 84 | 14 | 12 | 0.09375000 | 0.16666667 | **-0.07291667** |
| **SKIVE** | 113 | 253 | 0 | 11 | 0.09734513 | 0.00000000 | 0.09734513 |
| **MSM** | 159 | 91 | 3 | 20 | 0.12578616 | 0.03296703 | 0.09281913 |
| **MOLF** |  |  |  |  | (need female data) | 0.08247423 |  |

i) All high rec strains have a large difference between the proportions of bivalents which have short normalized IFDs. The ‘high rec strains’

ii) KAZ is the only strain with more short normalized IFDs in males than females.

MOLF – sex difference



# Q2 Male Musc Polymorphism

Which traits distinguish high and low recombining males in Musc strains? (High: MSM and PWD, Low: KAZ, SKIVE, MOLF). Consider more Musc strains: PWD, SKIVE, CZECH, KAZ (TOM, AST)

Male number of mouse (consider removing the cast strains)

WSB G LEW PERC PWD MSM MOLF SKIVE KAZ TOM AST CZECH CAST HMI

12 18 10 1 8 8 6 6 13 2 3 3 2 4

Number of Bivalents

Long Biv Dataset (show the number of Bivalents)

WSB G LEW PWD MSM MOLF SKIVE KAZ CZECH

19 47 10 21 4 16 33 45 13

For this section results from 3 models will be applied

M1. Lm(mouse.av.meteric ~ subsp \* strain)

M2. Lm(mouse.av.meteric ~ strain) - divided by subsp

M3. Logistic regression ( rec group ~ mouse.av.metric )

The data will be subdivided in the following ways:

-full data set, chrm classes (1CO, 2CO, 3CO) and long biv data set

3 general predictions for how the strains will very within each subsp (based on the gwRR patterns)

1. Dom strain no change (WSB = LEW = G)

2. Musc strains (PWD > SKIVE > KAZ, CZECHII, the rest )

3. Mol strains (MSM > Mol)

1.SC length positive correlation with high Rec

2. norm CO position – no predicted difference (no good prediction)

3. COI / IFD negative correlation with high Rec

## Q2 Variation in the SC Lengths

**SC Length**

1. All 1 way anova’s have significant values for the Dom strains (doesn’t meet prediction)

Male Mean SC lengths across classes

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SC Length**  **Means** | **Total**  **obs** | **Mean total**  **sc** | **Mean**  **0CO** | **nCO** | **mean**  **1CO** | **n**  **1CO** | **mean**  **2CO** | **n**  **2CO** | **M.mean**  **3CO** | **n**  **3CO** | **Interpretation** |
| WSB | 625 | 71.7328 | 59.33333 | 12 | 68.86653 | 472 | 85.61682 | 107 |  |  |  |
| G | 1031 | 81.6033 |  |  | 78.45546 | 797 | 94.4918 | 183 | 95.2 | 10 |
| LEW | 558 | 77.89068 | 64.5 | 6 | 73.91932 | 409 | 91.86792 | 106 | 89.5 | 10 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| PWD | 718 | 86.68384 |  |  | 77.49041 | 365 | 97.24585 | 301 | 99.33333 | 9 | Expected order not followed |
| SKIVE | 813 | 79.35547 | 76.83333 | 6 | 73.04902 | 510 | 91.95652 | 253 | 98 | 6 |
| KAZ | 740 | 79.99054 |  |  | 79.00598 | 502 | 91.63095 | 84 | 92.57143 | 7 |
| CZECHII | 256 | 82.61328 | 52.5 | 4 | 79.40223 | 179 | 96.5 | 50 | 100.6667 | 6 |
| AST | -- |  |  |  |  |  |  |  |  |  |  |
| TOM | -- |  |  |  |  |  |  |  |  |  |  |
| MSM | 178 | 88.20787 |  |  | 80.48052 | 77 | 95.40659 | 91 | 97 | 5 | Expectation followed |
| MOLF | 466 | 81.29185 | 68.66667 | 9 | 78.57669 | 326 | 92.79381 | 97 | 94.6 | 5 |

**Q2.SC all**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **all** | **Dom** | **Musc** | **Mol** |  |
| M2. Glm()  SC ~ strain | All strains sig  Int (WSB), G, PWD, MSM are most sig | Strain G, LEW are sig  Intercept (WSB) - < 2e-16 \*\*\*  G - 9.05e-05 \*\*\*  LEW - 0.0055 \*\* | 4 (strains)  -intercept(PWD) -  < 2e-16 \*\*\*  SKIVE - 0.00254 \*\*  KAZ - 0.01256 \*  CZECH - 0.05200 . | molf is sig  0.0284 \* | WSB and G were not expected to be sig.  Dom doesn’t meet prediction. I think Musc and Mol met predictions |
| SC ~ strain \* subsp | **??** | --NA-- | --NA-- | --NA-- |  |
| Log regression high vs low  P value | 0.00544 \*\*  0.00665 \*\* | --NA-- | 0.425  0.396 | 0.283  0.283 |  |

My prediction for variation across Dom strains isn’t met, there is more variation across strains for mouse average SC length. (I need to address the fact that there might be variation due to the number of mice or cells)

**SC Length Chrm Class**

Log regression( High vs Low ~ SC Length )

|  |  |  |  |
| --- | --- | --- | --- |
| Log regression  SC predict high vs Low  p values | SC length 1CO | SC length 2CO | SC length 3CO |
| All | 0.910  0.999 | 0.0363 \*  0.0441 \* | 0.179  0.221 |
| Musc | 0.0685 . (wrong direction)  0.0700 . | 0.364 (right direction, not sig)  0.333 | 0.367  0.351 |
| Mol | 0.4  0.4 | 0.294  0.295 | 0.808  0.807 |

This result for

*Shaper clustering of SC lengths across chromosome classes in the high Rec males* (1CO bivalents are shorter – because longer chromosomes move into the 2CO bin, whereas in the low rec group both physical long and shorter chroms have 1CO). A good way to display this pattern is the SC length y axis and CO number on the x axis



Each point is a mouse average SC. When all bivalents are pooled, the logistic regression SC lengths predict the groups. When the data is divided by CO number, 1CO’s have a higher SC length in the low group – in the low group more chrms, including the longer ones have 1CO – whereas in the high group, only shorter chrms (SC) are in the 1CO class.

For the 2CO class of bivalents – all the SC are longer, the high group averages were longer than the low.

and the mean 2COs SC are higher in the high strains

(this must have something to do with the physical size effect)

**(insert means of long chrm data set above?)**

**Long Biv Data set**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **all** | **Dom** | **Musc** | **Mol** |  |
| M2. Glm()  SC ~ strain | Int 4.35e-10 \*\*\*  MSM 0.0089 \*\*  SKIVE 0.0667 .  CZECH 0.0361 \* | G 0.32  LEW 0.41 | SKIVE 0.602  KAZ 0.842  CZECH 0.267 | NA  (not enough observations?) | subsetted by subsp – I’m not surprised there are no average difference for the binned long chrms |
| SC ~ strain \* subsp |  | --NA-- | --NA-- | --NA-- |  |
| Log regression high vs low  P value | Int 0.155  Mean\_SC 0.181 |  |  |  | Not enough obs to split by subsp, |

For the full data set --- MSM long biv SC’s are significantly longer? --- this is in agreement with the main finding of longer SC. There is limited variation in the long biv group to detect effects due to strain or subspecies.

Number of long biv observations by chrm class

WSB G LEW PERC PWD MSM MOLF SKIVE KAZ TOM AST CZECH CAST

0 0 0 0 0 0 0 0 0 1 0 0 0 0

1 12 29 5 0 3 2 9 18 21 0 0 4 0

2 7 16 4 0 17 1 4 11 11 0 0 8 0

3 0 1 1 0 0 1 0 2 0 0 0 0 0

How well does SC length predict chromosome class?

Log regression( 1CO vs 2CO ~ SC Length) (within categories / male strains)

|  |  |  |  |
| --- | --- | --- | --- |
| Log regression  SC predict 1CO | 2CO  p values |  |  |  |
|  |  |  |  |
| Musc |  |  |  |
| Mol |  |  |  |

Prediction, males will be better than females.

High strain SC will be better at predicting the number of COs (in agreement with the clustering of SC lengths across chromosome classes). **Come back to this later….**

## Q2 CO placement (1CO)

-Has the male 1CO landscape evolved across house mouse?

-Is there a consistent pattern for the high Rec strains?

Mouse av 1CO F1\_PER (are any of the strains significant?)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nrm.1CO** | **Intercept** | **Subsp** | **strain** |  |
| M1. SC ~ strain \* subsp | 0.73155 | -0.04064  -0.14008 | -0.06198 (G)  -0.03912 (LEW)  0.09707 (MSM) |  |
| pvals | < 2e-16 \*\*\* | 0.119031  7.8e-07 \*\*\* | 0.002990 \*\*  0.084331 .  0.000368 \*\*\* | MOLF strain effect is NA, |
|  |  |  |  |  |
| M2. SC ~ strain | 0.73155 |  | -0.06198  -0.03912  -0.06309  -0.04300  -0.14008  -0.05197  -0.05225  -0.04064 | WSB has the most terminal F1 |
| pvals | < 2e-16 \*\*\* |  | 0.00299 \*\* (G)  0.08433 . (LEW)  0.00701 \*\* (PWD)  0.07412 . (MSM)  7.8e-07 \*\*\* (MOLF)  0.01877 \* (SKIVE)  0.04744 \* (KAZ)  0.11903 (CZECH) | WSB is most telomeric.  MOLF is very centrally positioned.  G – is also pretty central. |

-maybe the molf strain F1 has evolved?

-WSB has most telomeric F1 position

**Lm**

**For mouse average, Norm1CO**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **all** | **Dom** | **Musc** | **Mol** |  |
| Log regression high vs low |  | 0.73155  -0.06198  -0.03912 | 0.66846 (int PWD)  0.01112 (SKIVE)  0.01084 (KAZ)  0.02245 (CZECH) | 0.68855  -0.09707 |  |
| P value | 0.00299 \*\* (G)  0.08433 . (LEW)  0.00701 \*\* (PWD)  0.07412 . (MSM)  7.8e-07 \*\*\* (MOLF)  0.01877 \* (SKIVE)  0.04744 \* (KAZ)  0.11903 (CZECH) | < 2e-16 \*\*\*  0.00256 \*\*  0.06427 . | 3.89e-16 \*\*\*  0.557  0.635  0.332 | 1.47e-07 \*\*\*  0.0328 \* |  |

WSB is the most telomereic of the strains.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nrm.1CO** | **Intercept** | **Subsp** | **strain** |  |
| M1. SC ~ strain \* subsp | 0.73155 | -0.04064  -0.14008 | -0.06198 (G)  -0.03912 (LEW)  0.09707 (MSM) |  |
| pvals | < 2e-16 \*\*\* | 0.119031  7.8e-07 \*\*\* | 0.002990 \*\*  0.084331 .  0.000368 \*\*\* | MOLF strain effect is NA, |
|  |  |  |  |  |
| M2. SC ~ strain | 0.73155 |  | -0.06198  -0.03912  -0.06309  -0.04300  -0.14008  -0.05197  -0.05225  -0.04064 | WSB has the most terminal F1 |
| pvals | < 2e-16 \*\*\* |  | 0.00299 \*\* (G)  0.08433 . (LEW)  0.00701 \*\* (PWD)  0.07412 . (MSM)  7.8e-07 \*\*\* (MOLF)  0.01877 \* (SKIVE)  0.04744 \* (KAZ)  0.11903 (CZECH) | WSB is most telomeric.  MOLF is very centrally positioned.  G – is also pretty central. |

-can’t detect any clear connection between the 1CO normalized position and the Rec.groups

**Logistic regression – can the 1CO normalized position predict if the Rec Group**

|  |  |  |
| --- | --- | --- |
| Log regression  1CO position | All (mice) | Interpt |
| All (mice)  p values | 0.802  0.910 | When all mice are analyzed, |
| Musc All (mice)  p values | 0.437  0.466 |  |
| Mol All (mice)  p values | 1  1 |  |
|  |  |  |
| Bivalent level measures | 0.802  0.910 |  |
| Bivalent level measures  Musc | 0.00447 \*\*  0.15912 | The intercept is significant. |
| Bivalent level measures  Mol | 1  1 |  |

The rec group can’t be accurately predicted based on the mouse averages normalized 1CO positions

What do these logistic regression plots look like? – Is it important – Maybe just conclude that Norm.1CO isn’t relevant for (predicting gwRR variation)



This plot is from bivalent level observations. From the plot you can see that PWD, SKIVE and MSM have more centralized / centromeric Foci

1. WSB is the most telomeric strain (the position has evolve? What about mouse specific patterns)
2. The high rec strains have more central foci.

What if this is due to chromosome size effect? (smaller chrms are more likely to be 1CO – smaller chrms have more uniform landscape)

This second plot is made from the mouse averages. They generally show the same pattern. MOLF is very central.

The Musc high Rec are a bit lower than the other Musc strains.

## Q2 IFD and Interference

**M1. (**mouse av IFD metric ~ subsp \* strain **)**

**M2. (**mouse av IFD metric ~ strain) \*split up by subsp?

**M3.** ( Rec. group **~** mouse av IFD metric **)**

The dataset used for this section will be only the 2CO bivalent measures of IFD. The raw and normalized measures will be compared.

Mean 2CO IFD will be shorter in high rec strains

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Glm() |  |  |
| Anova ( glm ) |  |  |
| Log regression high vs low |  |  |

Metric ~ subsp\*strain ?

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Dom (no dif)** | **Musc (high shorter than low)** | **Mol**  **(MSM < molf)** |
| Glm() |  |  |  |
| Anova ( glm ) |  |  |  |
| Log regression high vs low |  |  |  |

consider – all Musc strains – even if they don’t have female) – run a new mixed model --- all male, all musc – There are no other effects

mouse.av\_MLH1 ~ random(strain)

mouse.var\_MLH1 ~ strain

the new models indicate that PWD, MSM and SKIVE were significant fixed effects – so I can group them into a ‘high\_rec’ group (then apply to logistic regression for bivalent level analysis).

Compared to // given all the mouse strains,

-the 4 strains in Dom are not different

- PWD, MSM and SKIVE are different from the others

-there might be a KAZ male outlier ‘10jul19\_KAZ\_m3’ (look at these image files)

# Single Bivalent

1. **SC length**
   1. **SC length will be longer in high strains.**
2. **Normalized CO positions**
   1. **1CO normalized positions, will be the same (Null expectation)**
   2. **Pooled Sis-co-ten: higher in High rec strains because there are more 2COs,**

**Sis-co-ten separated by chromosome class, the strains will not be different (null expectation).**

* 1. **Pooled Centromere and telomere distances, – High rec will have shorter telomere and cent distance on average due to more 2COs.**
  2. **Separated by Chrm Class, the strains will not be different**

1. **IFD**
   1. **Shorter IFD in high rec strains to allow denser packing of COs.**

Two tests

1. Logistic regression, Rec Group ~ trait
2. T.tests (high vs low)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Traits** | **Logistic regression** | | **t.tests** | |  |
| **^all bivs** | **Mouse average** | **Single bivalent** | **Mouse average** | **Single**  **bivalent** | **Predictions** **met** |
| ^SC length | 0.0237 \* | <2e-16 \*\*\* | 0.001567 | < 2.2e-16 | Yes |
| 1CO | 0.0973 . | 0.179 | 0.1905 | < 2.2e-16 |  |
| 2CO | 0.0176 \* | 3.01e-05 \*\*\* | 0.03376 | < 2.2e-16 | No (2CO are distinct?) |
| 3CO | 0.223 | 0.567 | 0.3661 | 9.889e-08 |  |
| 2CO\_IFD.raw | 0.0366 \* | 9.98e-05 \*\*\* | 0.009133 | 3.595701e-05 | No, High Rec groups have longer, not shorter IFDs. |
| 2CO\_IFD.PER | 0.1143 | 0.0443 \* | 0.06068 | 1.304e-07 |
| Nrm.1CO.pos | 0.942 | 0.355 | 0.2558 | 0.3529 | Yes, (NS)  Yes, pooled pattern is sig, Separated chrm classes NS  No, (redo with chrm separated tests) |
| ^Sis.co.ten | 0.00725 \*\* | <2e-16 \*\*\* | 8.823e-12 | 4.74e-16 |
| Sis.co.ten\_1CO |  |  | 0.6619 |  |
| Sis.co.ten\_2CO |  |  | 0.1787 |  |
| Sis.co.ten\_3CO |  |  | 0.4327 |  |
| ^Telomere.dist | 0.0555 . | 0.00902 \*\* | 0.0007462 | 0.007452 |
| ^Telo.Dist\_PER |  | 1.32e-05 \*\*\* |  | 5.425e-06 |
| ^Cent.dist | 0.00368 \*\* | 5.21e-09 \*\*\* | 0.006659 | 3.658e-09 |
| ^Cent.Dist\_PER |  | 6.74e-16 \*\*\* |  | 4.52e-16 |  |

1. The general increase of SC length along with more DSB meets the model of increase gwRR stemming from changes early in the pathway (more SC area -> more DSB -> more COs).

2. Higher rec strains have more space between foci on the same bivalent, which goes the logic and general pattern I just outlined above. This indicates CO interference strength has also evolved with higher gwRR. These results require some follow up.

**Male Strain predictions**

**1. Dom strains not different from each other**

**2. PWD and SKIVE are different from the other strains**

**3. MSM and MOLF are different from each other**

-Glm’s strains are usually sig

- pairwise t-tests

-lme4 (use RLTR exact() for testing random effects --- what are the random effects in these models?

-ANOVA? (one way anova – 1 factor in this group is different from the others

fit <- aov(y ~ A, data=mydataframe)

**ANOVA test hypotheses**:

* Null hypothesis: the means of the different groups are the same
* Alternative hypothesis: At least one sample mean is not equal to the others.

<http://www.sthda.com/english/wiki/one-way-anova-test-in-r>

1. **Non-sig anova**
2. **Sig anova**
3. **Sig anova**

**Pairwise t.test results for the 3 subsp**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1CO Pos** | **n.obs** | **Mean**  **F1** | **Mean**  **nrmF1** | **interp** |
| WSB | 472 | 49.59534 | 0.7278682 |  |
| G | 797 | 51.82058 | 0.6650317 |  |
| LEW | 409 | 51.05623 | 0.6918036 |  |
|  |  |  |  |  |
| PWD | 365 | 50.96986 | 0.6666695 | The higher strains are more centromeric  (is this difference stat sig) |
| SKIVE | 510 | 48.45383 | 0.6671023 |
| KAZ | 502 | 53.37649 | 0.6807107 |
| CZECHII | 179 | 54.15642 | 0.6870426 |
| AST |  |  |  |
| TOM |  |  |  |
|  |  |  |  |  |
| MSM | 77 | 52.1039 | 0.6627513 | MSM is more telomeric – subsampling might be an issue |
| MOLF | 326 | 46.4816 | 0.5963067 |

The Normalized positions do look that different – but still need formal test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **IFD** | **n.obs** | **Mean**  **F1** | **Mean**  **nrmF1** | **interp** |
| WSB | 107 | 47.14019 | 0.5362821 | (are these sig) – test mouse averages |
| G | 183 | 52.55738 | 0.5477403 |
| LEW | 106 | 51.80952 | 0.5533493 |
|  |  |  |  |  |
| PWD | 301 | 58.81333 | 0.6018426 | (looks like PWD and SKIVE are greater ) |
| SKIVE | 253 | 57.61508 | 0.6266287 |
| KAZ | 84 | 48.94048 | 0.5239928 |
| CZECHII | 50 | 52.54 | 0.5362667 |
| AST |  |  |  |
| TOM |  |  |  |
|  |  |  |  |  |
| MSM | 91 | 56.25275 | 0.5874693 | MSM loot sig greater than MOLF |
| MOLF | 97 | 47.36458 | 0.5133379 |

**Next Steps**

\*Follow up the SC Length differences with sub-sampling tests and **XX adjustment**.

\*t-tests show no difference between the siscoten scores when separated by chromosome class. Does this indicate that the t-test of pooled sis-co-ten is due to different proportions in the high and low strains? Follow up with sub-sampling permutations.

\*Run pairwise strain comparisons between these groups, currently I’m pooling MSM and PWD.

**Points for Paper**

- Heterochiasmy extends to chromosome level, (could stem from cell processes). SC / Chromatin compaction and placement of COs are conserved traits of heterochiasmy in house mouse.

-Interference (in the physical SC scale), may also be sexually dimorphic.

- Larger IFD/stronger interference might be act as a limit/suppressor of gwRR getting too high in the high Musc strains.

- Male High Rec strains in Musc have more 2CO bivalents per cell. It’s unclear if there are general patterns for the 2CO bivalents from a MSM and PWD cells that distinguish them from other Musc strains.

## Discussion

Table X, Current models and their predictions for the evolution of heterochiasmy

Table X, results from proposed predictions

(Figure X, cartoon of difference in bivalent on spindle for 1CO and 2COs)

1. Review main patterns
   1. Male specific polymorphism for gwRR in musculus and molossisnus, may not be a species wide optimum for gwRR
   2. More variance in females for meiotic features, resulting in greater variation in gwRR
   3. Rapid male specific evolution upstream of CO repair stage
2. SACE predictions and bivalent selection models are not mutually exclusive,
3. Importance of broad scale patterns for recombination

(centromere effects for mis-segregation rates) - (high rate of robertsonian translocation in Dom, and absent in Musc – maybe something about centromeres (encourages transloactions + suppresses 2CO (rec near centromere) in DOM

(that has changed in Musc, REC near centromere suppresses rates of robertsonian translocation)

**Discussion**

Table X, Current models and their predictions for the evolution of heterochiasmy

Table X, results from proposed predictions

(Figure X, cartoon of difference in bivalent on spindle for 1CO and 2COs)

1. Review main patterns
   1. Male specific polymorphism for gwRR in musculus and molossisnus, may not be a species wide optimum for gwRR
   2. More variance in females for meiotic features, resulting in greater variation in gwRR
2. SACE predictions and bivalent selection models are not mutually exclusive,
3. Importance of broad scale patterns for recombination