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?)

1. HetC.MixedModel.HQ <- lme(mean\_co ~ subsp\*sex, data=DF.HetC.MixedModel.HQ, random=list(strain=pdDiag(~sex) ) )
2. Reduced.strain.HQ <- lme(mean\_co ~ subsp \* sex, data=DF.HetC.MixedModel.HQ, random = ~1|strain)
3. lm(mean\_co ~ subsp \* sex \* strain, data=DF.HetC.MixedModel.HQ)
4. strain.fixed <- 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 model for comparing effects of strain and subsp

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model 1** | **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** | **intercept** | **subsp** | **sex** | **subsp\*sex** | **Strain (random)** | **Interpert** |
| MLH1  mouse.av | 26.348630 | -0.520512  -0.724686 | -1.62829 | 3.274996  2.909844 | Need to test random effects | Sex is a significant fixed effect, |
| p values | 0.000000 | 0.774041  0.649822 | **0.000943** | **0.002097**  **0.000149** | Need to test random effects | Musc and Mol have sig interaction effects |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **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** | **Interpert** |
| 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

Even models without random effects --

**Variance Models** (use variance instead of CO)

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| 1.  MLH1 mouse **var** (within mouse var | 14.093000 | -0.220451  -0.410856 | -3.602197 | -0.441679  -0.553108 |  |  |
| p values |  | 0.853884  0.686413 | 0.000001 | 0.749139  0.598722 |  | Males have significantly less within mouse variance |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| 2.**Var** | 15.148325 | -0.494389  -1.953908 | -7.923511 | 0.589028  1.595256 |  |  |
| pvalues | 0.000000 | 0.84388 0.376361 | 0.000000 | 0.836295  0.461209 |  | This alternative model has the same conclusions. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model 3, all fixed effects** | **subsp** | **Sex** | **strain** | **Sex\*subsp** | **Sex\*strain** |  |
| 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** | **interpertation** |
| 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 |

**Qualitative Descriptions**

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

-Female have significant

-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)

# 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.

# 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)

Add in all the other excluded male strains?

# Musc Mixed Model

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

# Single Bivalent

1. **SC length**
   1. **SC length will be longer in high strains.**
2. **IFD**
   1. **Shorter IFD in high rec strains to allow denser packing of COs.**
3. **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**

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.

**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