MLH1 Results Outline

3/1/2020

# Genome wide recombination rate estimates for both sexes

We used counts of MLH1 foci per cell to estimate the genome- wide recombination rates across our panel of in 14 wild- derived inbred strains sampled from three subspecies within the house mouse species complex; *M. m. domesticus,* *M. m. musculus* and *M. m. molossinus*. Mean MLH1 foci counts for each XX mice mouse were quantified from an average of XX spermatocytes per male (for a total of 1,867 spermatocytes) and XX oocytes per female (for a total of 1867 spermatocytes and 1,409 oocytes).

Graphical comparisons between the two sexes reveal several patterns (Figure 1A). First, recombination rate differs between females and males in most strains. Second, the direction and magnitude of heterochiasmy varies among strains. Although the majority of strains show higher recombination rates in females (following the pattern in laboratory mice), two *musculus* strains and one *molossinus* strain exhibit male-biased heterochiasmy. Hence, relative recombination rates in the two sexes are evolutionarily labile.

The general patterns of heterochiasmy in house mouse are displayed in Figure 1A. Taking note of the direction and the magnitude of the sex differences, our results confirm two general patterns: i) genome wide recombination rates averages are greater in females compared to males (female biased heterochiasmy) and ii) the degree of heterochiasmy (Female:Male ratio) is generally low, ranging from 1.17 in *domesticusG* to 1.02 in *domesticusLEW*. Three notable exceptions of male biased heterochiasmy are the strains *musculusPWD*, *molossinusMSM* and *musculusSKIVE*, with heterochiasmy values of 0.88, 0.9 and 0.96 respectively.

Separately examining the mouse means of MLH1 foci per cell separately for each sex reveal points to distinct patterns of variation (Figure 1 B- and C). Female recombination rates are evenly distributed around the sex-wide mean of approximately 25 MLH1 foci per cell (Figure 1 B). In stark contrast, males specific means separate more clearly into two groups of strains with high genome wide recombination rates (near 30 MLH1 foci per cell) and low genome wide recombination rates (near 23 MLH1 foci per cell) (Figure 1 C).

# Partitioning variation in recombination rate

To situate variation in recombination rate within an evolutionary framework, we fit a series of models including subspecies, strain, and sex, to mean MLH1 foci counts from 187 mice. We began with a full mixed model (M1, see Methods), which showed that strain (random effect p < someSmallNumber), sex (p = 3.9710^{-5}), subspecies (p= 1.7210^{-4}), and subspecies \* sex (p = 3.110^{-5}) each significantly affect recombination rate.

After a general linear model including all factors as fixed effects (M2) revealed only weak contributions of subspecies, we focused on additional models designed to illuminate the role of strain and sex. A general linear model with these two variables (M3) identified two strains with particularly strong effects on recombination rate: *musculusMSM* (p= 3.9910^{-6}) and *domesticusG* (p= 1.0410^{-6}). **In addition, two strains exhibit strain-by-sex interactions:  *molossinusMSM* (p = 1.2610^{-4}) and *musculusPWD* (p = 3.8610^{-4}).**

We next fit general linear models separately for 192 males and 144 females (M4). In the male dataset, three strains significantly affect recombination rate: *molossinusMSM* (p = 2.2310^{-14}; effect = X MLH1 foci), *musculusPWD* (p = 7.3710^{-10}; effect = X foci), and *musculusSKIVE* (p = 0.01; effect = X foci). These three strains point to rapid evolution in recombination rate in spermatocytes; we subsequently refer to them collectively as “high-recombination” strains. Analysis of the female dataset points to four strains with significant effects on recombination rate: *domesticusG* (p = 2.510^{-6}), *molossinusMSM* (p = 6.2410^{-6}), *domesticusLEW* (p = 0.01), and *musculusPWD* (p=0.02). Strain effect sizes in females are modest in magnitude (ranging from 1 to 4 foci) compared to those in males.

Together, these results demonstrate heritable differences in the genome-wide recombination rate evolving in a highly sex-specific manner over short evolutionary timescales.

## Within-mouse variance in CO count per cell

Counting MLH1 foci in multiple oocytes for each female and multiple spermatocytes for each male allowed us to examine determinants of the within-mouse variance in recombination rate. To do this, we considered the same models as above, but replaced mean MLH1 foci count with within-mouse variance in MLH1 foci count as the dependent variable. Sex is the only variable that significantly affects recombination rate in both the mixed model (M1) (p < smallNumber) and general linear model (M2) (p = 2.310^{-4}). In general, females have almost twice as much variance in MLH1 foci per cell compared to males (Figure 1). Since estimates of within-mouse variance may be more susceptible to technical error from the staining protocol, we repeated the analyses using a subset of cells with higher quality scores (criteria?). The results are similar: sex is the strongest effect (p < smallNumber, M1; p = 2.310^{-4}, M2). When both quality-curated and full datasets are considered, strain does not significantly and consistently affect variance in MLH1 foci count per cell in either sex. These results suggest that within-mouse variance in recombination rate evolves independently of mean recombination rate.

# Evolution of genome-wide recombination rate is associated with evolution of double strand breaks

In an attempt to localize the male-specific evolution of crossover number to steps of the meiotic pathway, we counted foci from a marker for double strand breaks (DSBs), DMC1, in prophase spermatocytes. DMC1 foci were scored from a total of 76 leptotene-stage and 75 zygotene-stage spermatocytes from juvenile mice (12 to 18 days) from three low-recombination strains (*musculusKAZ* , *domesticusWSB* , and *domesticusG*) and two high-recombination strains (*musculusPWD* and *molossinusMSM*).

The high-recombination strains have significantly more DMC1 foci than the low-recombination strains in leptotene cells (t-test, p<someSmallNumber; one-way ANOVA, p = 0.00027). In contrast, the two strain groups do not differ in DMC1 foci counted in zygotene cells (later prophase)(t-test, p = 0.66; one-way-anova, p = 0.15).

After DSB formation, DSBs are repaired as either non-crossovers (NCO) or as crossovers (COs), with the vast majority being repaired as NCOs. Thus the ratio of CO:DSB is a partial indicator of the proportion of DSBs which are designated as COs. The ratios, calculated for DMC1 means from both stages, are not significantly different between the high and low strain groups (t-test, p = 0.94 and p = 0.11 for leptotene and zygotene ratios, respectively). This comparison raises the possibility that the evolution of crossover number is primarily due to processes that precede the crossover/non-crossover decision.

# Evolution of genome-wide recombination rate is reflected at the single chromosome level

To examine the connection between evolution of the genome-wide recombination rate and changes to the recombination landscape, we used an image analysis pipeline to measure properties of single bivalents (Peterson et al. 2019). This algorithm substantially speeds the accurate measurement of bivalents, but has the limitation that not all bivalents per cell can be isolated due to overlapping bivalents. In this dataset, isolation rates per cell range from 0.51 (*molossinusMSM* male) to 0.72 (*musculusKAZ* female). From the total set of cell images, 10,458 bivalent objects were isolated by the image analysis software. After a human curation step (Peterson et al. 2019), 9,829 single-bivalent observations remained. We assume that the isolation of bivalents within cells is unbiased. Given the large number of single-bivalent observations, **we assume that each of the datasets are equally representative of general patterns.**

**An additional challenge of the MLH1 framework is that the identities of individual autosomes and the XX in females cannot be easily obtained (the male XY is distinct).**

Ninety-six percent of single bivalents in our pooled dataset (n = 34,982) have either one or two crossovers (Figure X). The proportion of one-crossover (1CO) to two-crossover (2CO) bivalents distinguishes the high vs. low recombining strains (Figure X). High-recombination strains are enriched for 2CO bivalents at the expense of 1CO bivalents: proportions of 2CO bivalents are 0.33 in *musculusSKIVE* , 0.44 in *musculusPWD* , and 0.53 in *molossinusMSM* . Following patterns in the genome-wide recombination rate, male *musculusPWD* and male *molossinusMSM* have 2CO proportions that are more similar to each other than to strains from their own subspecies (chi-square tests; *musculusPWD* vs. *musculusKAZ* p = 3.1510^{-33}; *molossinusMSM* vs. *molossinusMOLF* p = 4.7210^{-13}).

Next, we use this single-bivalent dataset to focus on aspects of the recombination landscape along chromosomes. We address two main questions. First, which traits are sexually dimorphic? Second, which traits differ between males from high-recombination vs. low-recombination strains?

# Q1 Sex Differences in the Length of the Synaptonemal Complex

In many mammalian species, the synaptonemal complex (SC) is longer in females. Yet, the majority of these observations come from species with female-biased heterochiasmy. Our wider survey of recombination in house mice provides an opportunity to determine whether sex differences in chromatin compaction (SC length) are reversed when heterochiasmy is male-biased. In addition, if SC length is a strong determinant of the genome-wide recombination rate, male *musculusPWD* and male *molossinusMSM* should have longer SC lengths than the other strains.

To account for confounding effects of sex chromosomes from pooled samples of bivalents, we considered a reduced dataset including only bivalents with SC lengths below the 2nd quartile (for SC length) in each cell. This dataset included the four or five shortest bivalents and excluded the X bivalent in oocytes. A total of 678 ‘short’ bivalents were isolated from 103 oocytes and 37 spermatocytes. Although this smaller dataset has decreased power, it offers a more comparable set of single bivalents to compare between the sexes.

Short bivalents are significantly longer in females than males in all strains (t-test; p < 0.05) except *musculusSKIVE*. The female:male ratio ranges across strains from 1.15 (*musculusMSM*) to 1.49 (*domesticusWSB*).

As an additional metric of chromatin compaction, we computed the total (summed) SC length of all bivalents for single cells, using the image analysis algorithm of Wang et al. (2019). Mouse means were calculated from cell-wide total SC lengths in 2,984 out of 3,680 cells with MLH1 counts (Figure X). Females have significantly longer total SC than males in each strain tested separately (t-test; p < 0.05). That females have longer SCs is also supported by mixed models and general linear models with covariates, which identify sex as the most consistently significant effect (p < 0.05). Additionally, there are some significant subspecies and strain effects (p < 0.05), indicating that SC length has evolved among strains and subspecies.

In summary, two approaches for measuring and analyzing SC length indicate that females have longer SCs (chromosome axes), even in strains where males have more MLH1 foci per cell. Furthermore, joint consideration of MLH1 foci and total SC length suggests that males from high-recombination strains have less “space” in which to place their additional crossovers.

# Q1. Sex Differences in the Positions of Single Crossovers

In most of the strains we surveyed, the majority of bivalents are observed to contain one crossover (focus). In all strains, the landscape across 1CO bivalents is significantly different in females and males. Normalized foci positions tend to be more central in females (mean focus position = X) and closer to the telomere in males (mean focus position = Y) (t-test; p = 2.9210^{-22}). Sex is also the most significant effect on focus position in a mixed model (M1: p =1.2610^{-25}) and in general linear models (M2: p =1.3310^{-7}; M3: p = 1.3310^{-7}). These sex differences in the placement of foci on 1CO bivalents follow a pattern observed across a variety of mammalian species (Sardell Kirkpatrick).

# Q1. Sex Differences in CO Interference

There is no strong signal of sex differences in raw mean inter-focal distances (IFDraw)across the full set of strains. A marginally significant difference between the sexes (t-test; p = 0.07) is driven by one strain (t-test without *domesticusG*; p = 0.27). This result indicates that females and males exhibit a similar level of interference when it is measured in physical (SC) units. In contrast, males have significantly longer normalized mean inter-focal distances (IFDnorm) than females in seven out of eight strains (t-tests; p < 0.02), with only *musculusKAZ* showing no difference (t-test; p = 0.33). Examination of IFDnorm distributions indicates that female IFDnorm values are centered at approximately 50% and show a slight enrichment of low (<25%) values, whereas males are enriched for higher values.

Mixed models and general linear models of IFDnorm support the inference of stronger interference in males: sex is the most significant variable (LRT: p = 6.7410^{-14}). When interference is measured in physical SC units (IFDraw), the differences between sexes is low and only slightly significant (data not shown).

In summary, controlling for differences in chromatin compaction (SC length) using IFDnorm indicates that interference is stronger in males, whereas consideration of IFDraw shows that the sexes exhibit a similar level of interference on the physical (SC) scale.

While the comparisons of the IFDnorm metrics can reveal more general recombination landscape patterns while controlling for the underlying differences in chromatin compaction and SC length. This metric is closer to measures of interference from linkage maps which are also removed from physical scales in that they measure frequencies of crossovers. Crossover interference is stronger in male specific linkage maps compared to female maps in a variety of species (ref).

# Additional Determinants of Genome-wide Recombination Rate Evolution in Males

Next, we used the contrast between high-recombination strains (*musculusPWD* , *musculusSKIVE*, and *molossinusMSM*) and low-recombination strains to identify features of the recombination landscape associated with evolutionary transitions in the genome-wide recombination rate.

**Linear model M1**

**Linear model M2**

Previous empirical work suggests basic predictions for the relationship between the genome wide recombination rate and two aspects of the recombination landscape. SC length is expected to be positively associated with genome-wide recombination rate **because X. Crossover interference is expected to be negatively associated with genome-wide recombination rate because X**. Following this logic we predict (1) *musculusPWD* will have greater SC length and weaker interference than *musculusSKIVE*, which in turn will have longer SC and weaker interference compared to the other *musculus* strains, (2) *molossinusMSM* will have longer SC and weaker interference compared to *molossinusMOLF*, and 3) *domesticus* strains will have similar SC length and crossover interference.

# Q2 SC Length

We compared mouse means of three SC-based metrics: total SC length, mean short bivalent length, and mean long bivalent length. The long bivalent data set was isolated by choosing bivalents in the X quartile of the SC length distribution within cells. The short bivalent dataset was isolated as described above.

Confirming the basic predictions, **there is a positive correlation between mean MLH1 foci per cell and total SC** **([report spearman’s rho and p-value]).** Nevertheless, mean total SC only partially differentiates high-recombination and low-recombination strains (Figure X). While high-recombination strains have significantly more SC area in the total dataset (t-test; p = 0.01), separate tests by strain show that only within *molossinus* are high- and low-recombination strains significantly different (t-test; *molossinus*: p = 0.03; *musculus*: p = 0.87). Additionally, the means for the reduced (short and long) bivalent datasets are not significantly different between high-recombination and low-recombination strains (t-test; short: p = 0.88; long: p = 0.18).

In a general linear model with total SC as the dependent variable, two subspecies effects are significant (p = *musculus* 1.2410^{-6}, *molossinus* p = 10^{-6}). In general linear models with reduced bivalent means as dependent variables, several subspecies and strain effects reach significance (p < 0.05) but they are not consistent across models**, indicating to some extent the chromatin compaction evolution is decoupled from evolution in mean MLH1 foci per cell.**

# Q2.1CO rec landscape evolution is decoupled from gwRR evolution

The normalized 1CO position is not significantly different between high-recombination and low-recombination strains for the total pooled data (t-test; p = 0.24) and also when examined within subspecies (t-test; p = 0.41 and p = 0.07 for *musculus* and *molossinus*, respectively). **While there are significant strain effects for *domesticusWSB* and *molossinusMOLF* in a general linear model with normalized position as the dependent variable** (Figure X), this evolution of the 1CO positioning is decoupled from the total genome-wide recombination rate.

# Q2 Evolution of interference is associated with genome wide recombination rate evolution

Mouse averages for both IFDraw and IFDnorm are significantly longer in high-recombination strains (t-test; IFDnorm: p = 7.7410^{-7}; IFDraw: p = 8.7810^{-6}). This pattern is confirmed by separate comparisons within *musculus* (t-test; IFDnorm: p =2.0410^{-5}; IFDraw: p = 1.9410^{-4}) and within *molossinus* (IFDnorm: p= 0.17; IFDraw: p = 0.08). Similar results are seen with general linear models for both IFDraw and IFDnorm: only effects associated with high-recombination strains are significant (p < 0.05). That IFDraw and IFDnorm show similar patterns eliminates variation in SC lengths and bivalent sizes as primary explanations. We determined that the main difference in IFDnorm distributions between high-recombination and low-recombination strains is an enrichment of IFDnorm values under 30% in low-recombination strains. The frequency of IFDnorm values that fall below 30% ranges from 8.2% (*domesticusG*) to 16% (*musculusKAZ*) in low-recombination strains, whereas high-recombination strains all show such frequencies below 5% (0%, 1.3%, and 3.3% for *musculusSKIVE*, *molossinusMSM*, and *musculusPWD*, respectively).

In summary, the level of interference is a significant predictor of evolution in the genome-wide recombination rate, but SC length and crossover position on 1CO bivalents are not. However, the pattern is in the opposite direction to our prediction: high-recombination strains have stronger interference. At least on 2CO bivalents, crossovers are spaced further apart when the genome-wide recombination rate is higher.

# Q2 Summary

Our results show that the greater crossover interference is the strongest single bivalent-based predictor for the observed rapid evolution of mean MLH1 foci per cell. While these results do not conform initial predictions for how a recombination landscapes would accommodate more crossovers, the increased strength of interference aligns with our results on sex differences. The typical recombination landscapes for males and females results in divergence in the proportion of linked sites along chromosomes which segregate together. The stronger interference of the 2CO bivalents in the high recombining strains accentuates this effect. The measures of DSB and some comparisons of SC length between high and low recombining strains suggest that the SC length have evolved to be longer in high recombining strains, however this evolution of SC length is partially decoupled from the number of crossovers since similar amounts of SC length evolution are seen in low recombining strains.

# References