

# 20211117\_LocationPatterns1

December 1, 2021

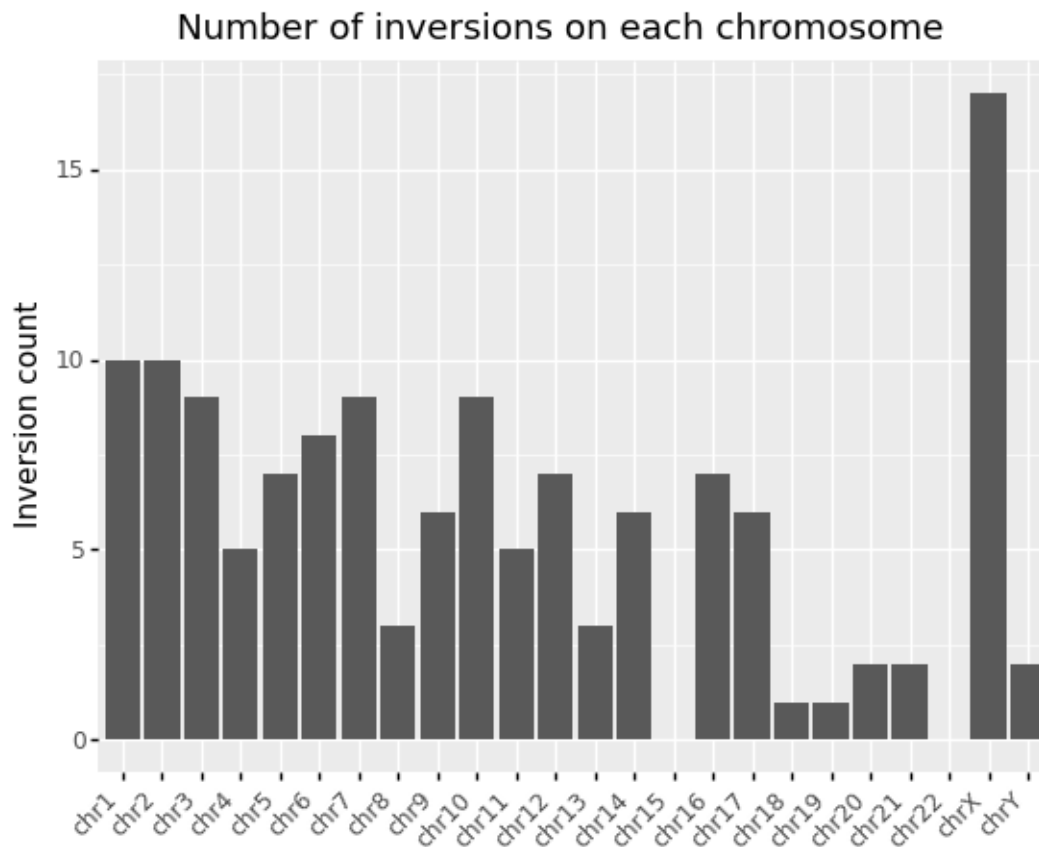
## 1 Location patterns 1 - global chromosome characteristics

This notebook explores the distribution of inversions along chromosomes and between chromosome types. The origin mechanism of inversions strongly determines their characteristics in terms of size, recurrence, and, as observed in this document, location.

We will be analyzing 135 inversions, 55 NAHR and 80 NH.

### 1.1 Differences in distribution between chromosomes

Smaller chromosomes have higher recombination rates than big ones to make sure that there are enough chiasmata to correctly segregate. Probably because of this difference, small autosomes have more NAHR inversions per Mb than big autosomes, while NH inversions are randomly distributed (Figures 3 & 4). As it is already known, the excess of repeats in chromosome X causes a significant enrichment of NAHR inversions. There are no differences in inversion sizes between chromosome types (Figure 5).



```
<ggplot: (8755126464841)>
```

Figure 1: Number of inversions for each chromosome in the dataset

```
R[write to console]: `geom_smooth()` using formula 'y ~ x'
```

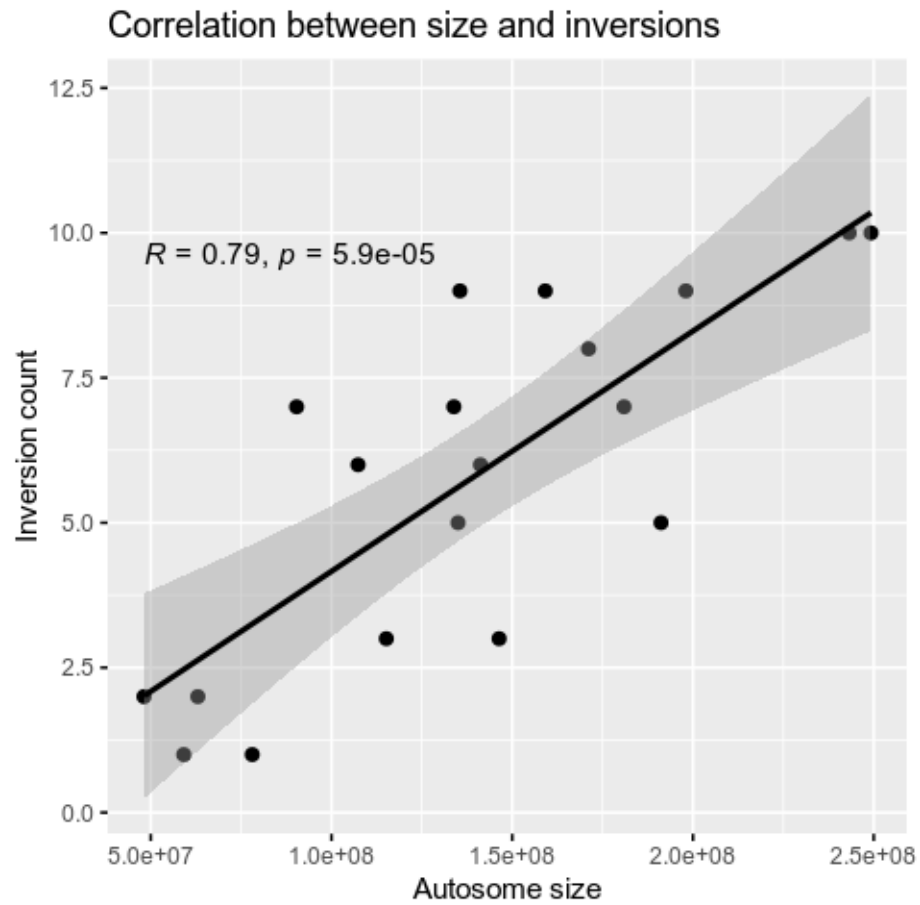
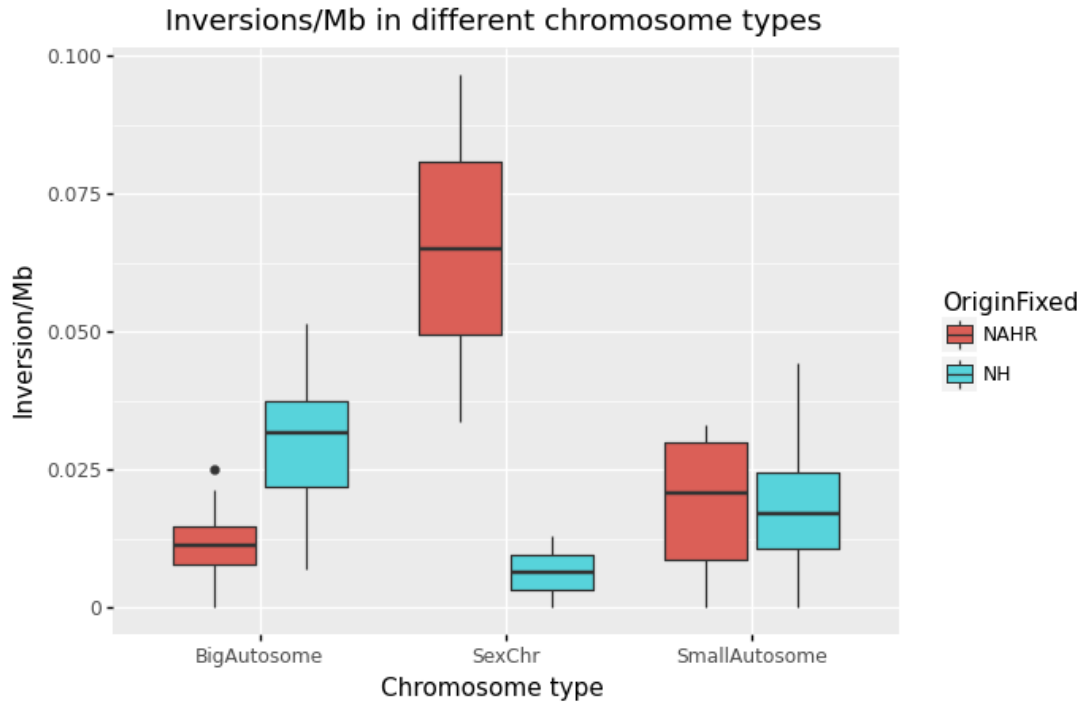


Figure 2: There is a moderate, significant positive correlation between autosome size and inversion count when all inversions and autosomes are considered.



```
<ggplot: (8755125692010)>
```

Figure 3: Different chromosome types have different distribution of inversion origins. Big Autosomes are chromosomes 1-12, Small Autosomes are chromosomes 13-22, inversion counts are corrected by chromosome size. The general NAHR:NH ratio is 1:1.53, while in chromosome X is 1:0.134, in Big chromosomes the mean ratio is 1:2.841 and in Small chromosomes the mean ratio is 1:0.767.

```
R[write to console]: `geom_smooth()` using formula 'y ~ x'
```

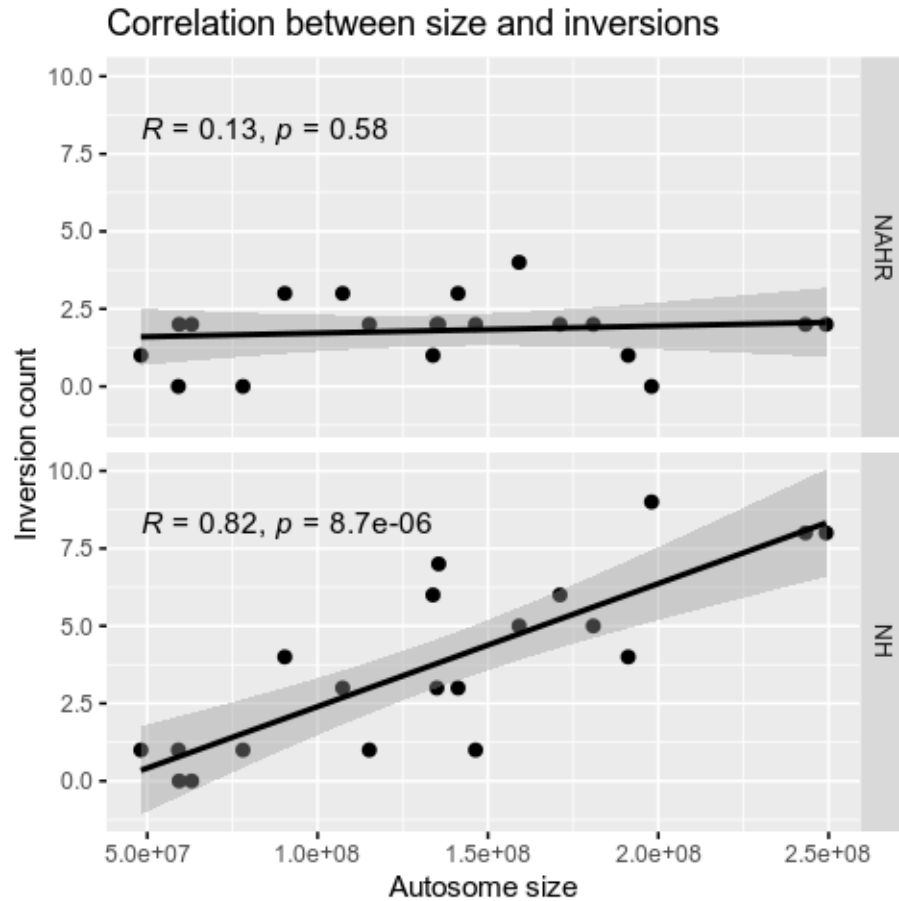
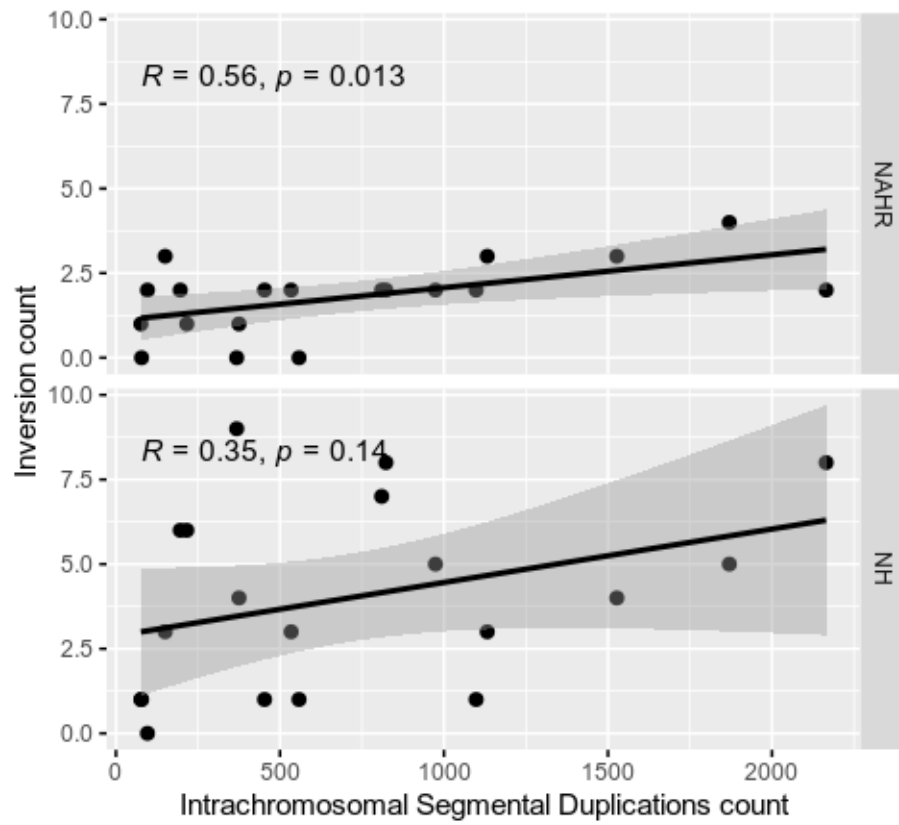
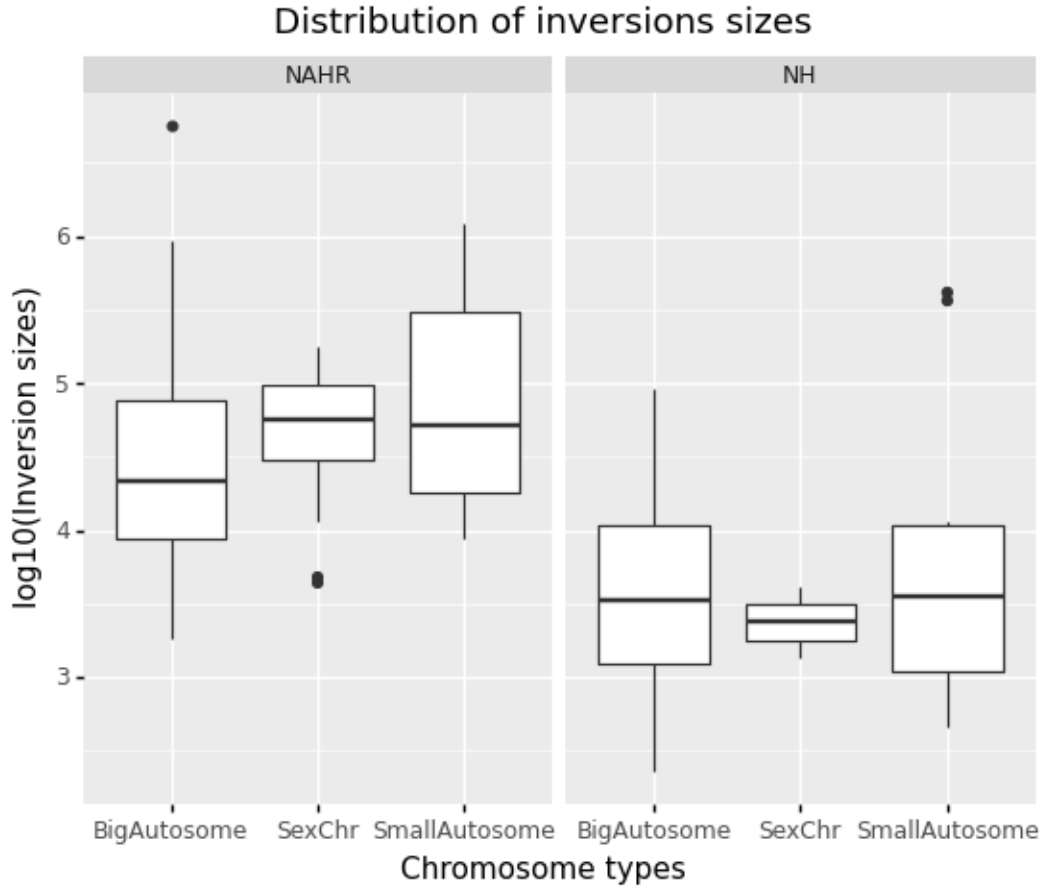


Figure 4: The correlation between Autosome size and inversion count is affected by the differential distributions observed in Figure 3. NH inversion counts show a strong, significant positive correlation with chromosome size, while no such correlation is observed for NAHR inversions because of their excess in Small chromosomes and depleat Big chromosomes.

R[write to console]: ``geom_smooth()`` using formula 'y ~ x'

Correlation between segdups and inversion count (autosomes)



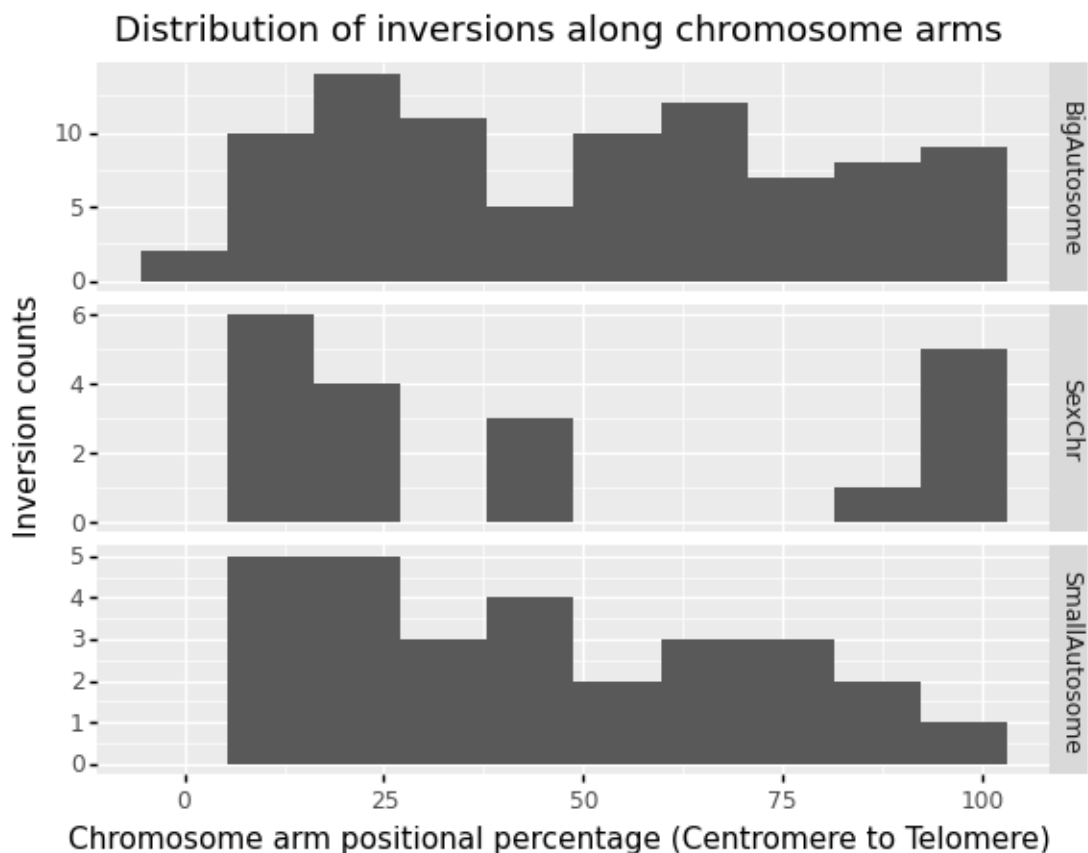


```
<ggplot: (8755117181085)>
```

Figure 5: Inversion sizes seem to be evenly distributed across chromosome types.

## 1.2 Differences in distribution within chromosomes

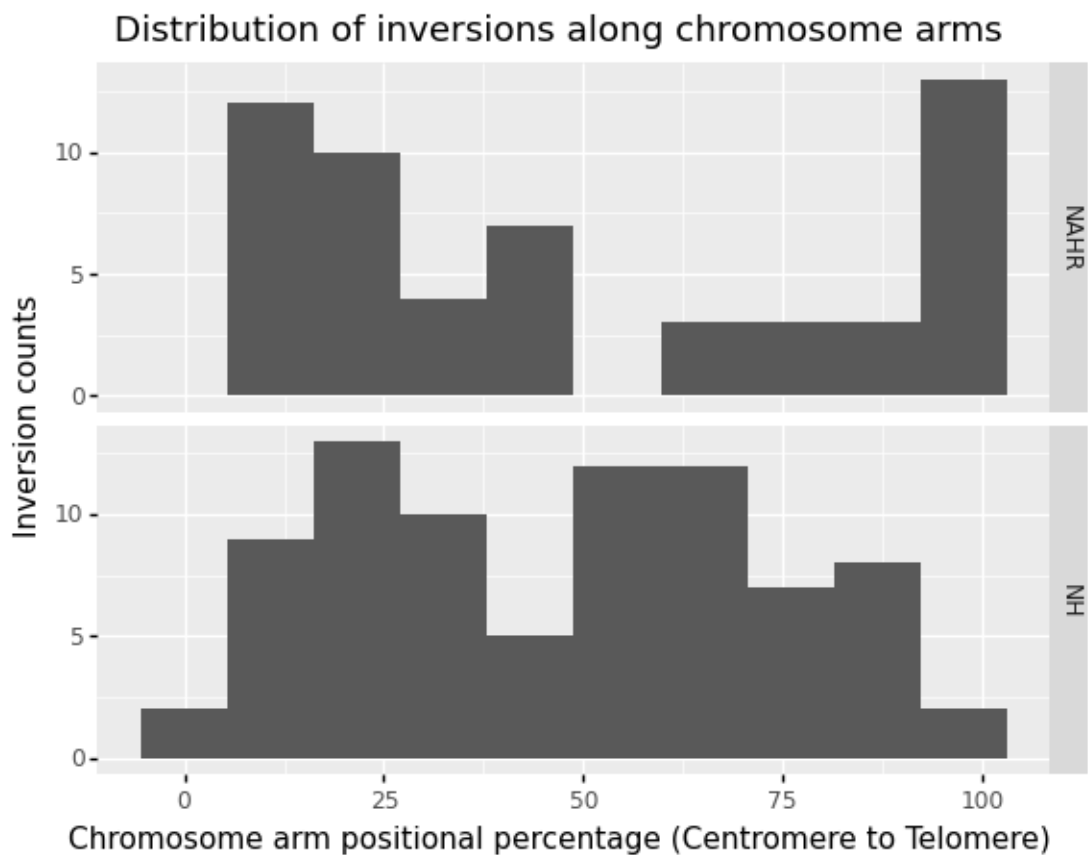
Similarly, the known, large-scale recombination patterns within chromosome arms affect inversion incidence. We don't have much power in small autosomes to identify location patterns, but in big autosomes there is an important proportion of NAHR inversions occurring near telomeres, where recombination rates are higher. On the other hand, NH inversions tend to concentrate around the center of the chromosome arm (Figures 6,7,8). NH inversions distribution is similar to a simulated distribution that exemplifies how the probability of having two breaks at a suitable distance to generate an inversion is higher around the center of the arm than near centromeres and telomeres (Figure 9).



```
<ggplot: (8755117137830)>
```

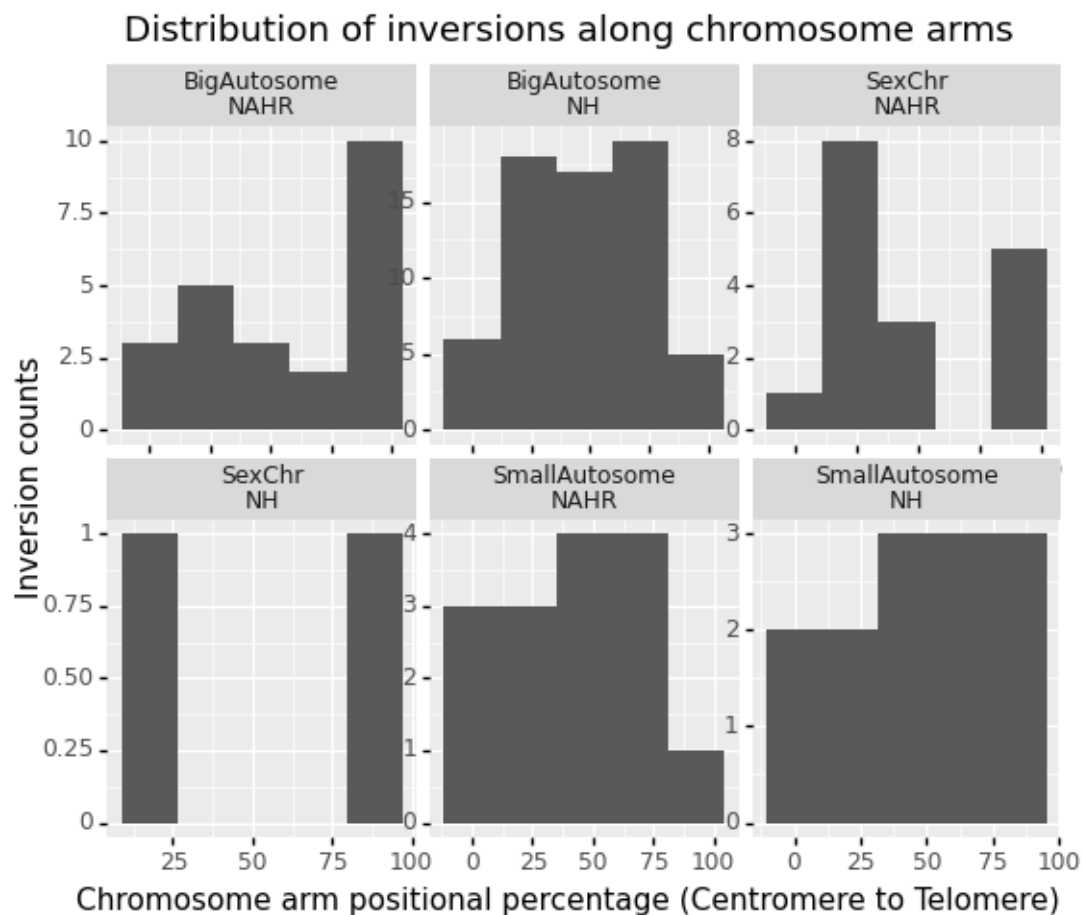
Figure 6: Distribution of inversion counts along chromosome arms, where 0 is the centromere and 100 is the telomere. Inversions seem to be located preferently around centromeres when not taking into account inversion origin, although in Sex chromosomes and Small Autosomes the first bin, most proximal to the centromere, is empty.





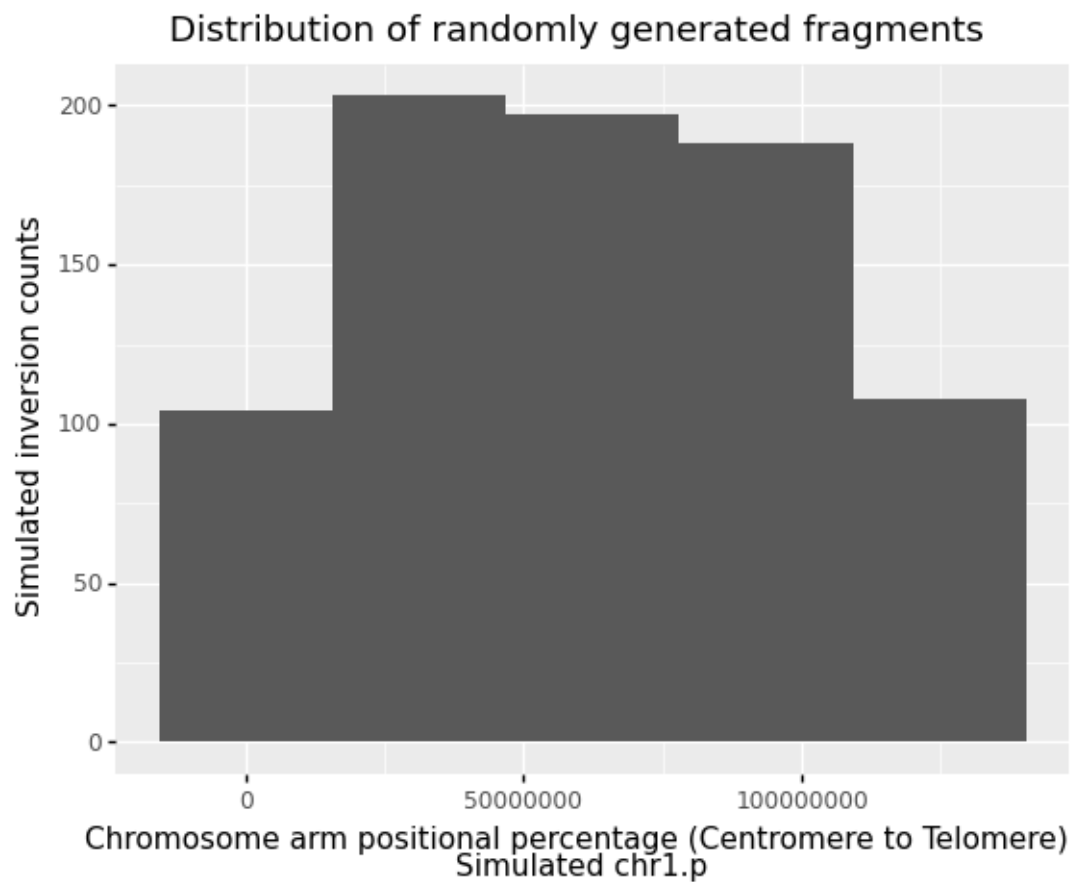
```
<ggplot: (8755117089178)>
```

Figure 7: Distribution of inversion counts along chromosome arms, where 0 is the centromere and 100 is the telomere. NAHR inversions are preferently located near centromeres and telomeres, while NH inversions seem to be more dsitributed along the chromosome arm.



```
<ggplot: (8755116962771)>
```

Figure 8: Distribution of inversion counts along chromosome arms, where 0 is the centromere and 100 is the telomere. When looking at all the categories, we loose definition, especially for small chromosomes. The most evident patterns can be observed in Big chromosomes: NH inversions concentrate in the middle of the chromosome arm while most NAHR generate near telomeres.



```
<ggplot: (8755136393072)>
```

Figure 9: Simulated distribution of inversions along a chromosome arm. One random point in the map was assigned to each of the 80 NH inversions, and then the corresponding breakpoints calculated according to inversion size. If inversion breakpoints fell outside the map limits, the inversion was discarded. This process was repeated 10 times.