**Neo-sex chromosome shapes introgression in a hybrid swarm**

Silu Wang1, Matt J. Nalley1, Kamalakar Chatla1, Russ Corbett2, Doris Bachtrog1

1Integrative Biology

University of California, Berkeley

3040 VLSB, Berkeley, CA, USA 94720

2Biomolecular Engineering and Genomics Institute

University of California, Santa Crutz

Corresponding author email: dbachtrog@berkeley.edu

**Introduction**

Speciation is the fundamental process that generates the diversity of life (Darwin 1859). The genomic mechanisms shaping introgression at species boundary of diverging and/or recently diverged lineages discloses the intrinsic mechanisms of speciation. Species-specific inheritance can substantially regulates gene flow at species boundaries (polyploidy, inversion, cite the toad paper, ploidy change, etc.). It is the primary layer of genetic mechanism of reproductive isolation, yet we are only at the beginning understanding how they influence speciation at species boundary.

Here we investigated the role of inheritance on introgression between recently diverged sister species with the multi-generational hybrid population generated by interbreeding the parental lineages. Hybrid swarms experience recombination over multiple generations and thus harbor a variety of genomic combinations. Sampling and sequencing these hybrid populations at various generations allow time-series tracking the extent and direction of introgression in different parts of the genome, and reflect gene flow at different regions in the genome.

**Sex chrom**

The speciation between *Drosophila albomicans* and *D. nasuta* might be related to neosex chromosomes evolution. [background of muller CD and A fusion, geography, closely related species, divergence time]

Here we generated hybrid swarms of D. albomicans (abbreviated as alb hereon) and its sister taxon, D. nasuta (abbreviated as nas hereon) which harbors extensive genomic differences and sequenced the random samples from the hybrid swarm for 28 generations to investigate the effect of neosex chromosome on gene flow.

We investigate how are ancestry blocks get or not get broken down by hybridization. In particular, we look for candidate barrier genomic modules.

**Methods**

*Hybrid swarm*

N= of each sex of each species were crossed

Plexiglass (dimension), the arena is maintained at room temperature, humidity, light-dark cycle.

Reference haplotypes

*Sequencing*

We sequenced at early generation (0-5, N=21), generation 10 (N=1), generation 18 (2), 21 (N=22), 27 (N=44), and 28 (N=37).

*Sequence processing*

We aligned the reads to a *sulfurigaster* reference with bwa (Li 2010). Trim the reads, read quality cutoff =20. Individuals with less than 50,000 reads were excluded from downstream analysis.

*Ancestry calling*

Ancestry HMM was used to infer local genomic ancestry among hybrids in the hybrid swarm experiment. Following setting was employed: *-a 2 0.5 0.5 -p 0 -3 0.5 -p 1 -3 0.5 -r 0.000005*. In particular, we assumed equal parental ancestry contributions, and recombination rate being 5X10-6, and estimated the generations before present in which the ancestry pulse occurred. Ancestry genotype was called if the posterior probability > 0.9.

We first find the fixed difference (SNP-based FST =1) between nasuta and albomicans. Then we prepared the input for ancestry HMM with a custom script. 0= nasuta, 0.5 = heterozygotes, 1 = albomicans.

*Neosex chromosome haplotyping*

Neo-X and Neo-Y specific alleles in albomicans were identified following Wei et al. (2019). The neo-sex chromosome SNPs should be hemizygous in males and homozygous in females, thus the ratio of coverage between female and male should be 2. For each individual, we estimated the haplotype proportion (proportion of sites of each haplotype across all the haplotype-informative sites). We refer the any single haplotype with > 0.5 haplotype proportion being the neosex chromosome ‘haplotype background’ of each individual. This way, we can association between neosex chromosome haplotypes with introgression and ancestry turnover rates.

*Introgression*

We considered introgression with respect to the extent of admixture inferred with heterozygosity. Heterozygosity (*h*) was calculated as the fraction of heterozygous ancestry-informative sites within each individual. Admixture proportion (*p*) was the fraction of alb alleles across ancestry-informative sites. We calculated admixture proportion and heterozygosity both genome-wide and within each Muller element. To investigate the introgression across various muller element, we calculated Muller element-specific alb-biased introgression as

.

This alb-biased introgression index ranges ranging from -1 (nas-biased introgression) and 1 (alb-biased introgression) with zero implying equal introgression from either ancestry. It controls for the fact that hybrids with more admixed genomes (lower heterozygosity, greater deviation of *y* from 1) and signature of backcrossing (positive deviation of *x* from 0.5) are more likely to demonstrate asymmetrical introgression.

*Ancestry turnover across the genome*

We estimated ancestry turnover rate (*r*) by the number of ancestry switches across each muller element and across the genome. However, hybrids of later generations of admixture will exhibit more recombination, while backcrosses exhibit less recombination. To ensure that each class of hybrids are comparable, we employed a transformation index based on the Euclidian distance between the position of each individual and the point (0.5, 0) in the triangle plot.

The greater the distance, the less recombination event is expected. Thus r \* is the admixture-corrected recombination estimate.

To compare ancestry turnover frequency among different Muller elements, the per informative site ancestry turnover rate was used, by dividing the ancestry switches by the number of informative sites in each Muller element. Muller F is excluded from this comparison because of the scarcity of informative sites.

*Genomic clines*

To detect genetic clusters that serve as barriers to gene flow, we fitted genomics clines (Gompert and Buerkle 2011). We first identified genetic clusters within which the ancestry blocks tend to co-segregate among all the hybrids detecting K-means clusters (Forgy 1965). Specifically, with R function *kmeans*, we iteratively incremented k (from k = 1) until 60% of the total variance was explained by between-clusters variance. Then we estimated cline parameters for each genetic cluster based on cluster-specific introgression relative to the genomic background. We focused on late generation (21, 27, 28) females, because the signature of selection tends to become more evident when the masking admixture effect becomes less pronounce after generations of recombination.

**Results**

*Asymetrical Introgression*

There was albomicans-biased genome-wide introgression, which was predominantly represented by Muller CD (Figure triangle plot; Figure BGC plot with muller cd haplotypes). Even though the genomes of many hybrids were admixed in the rest of the genome, muller CD is mostly dominated by the albomicans ancestry (Figure 1-3), which reveals albomicans-biased barrier effect of muller CD relative the rest of the genome (Figure 4A). Among males, the frequency of neo-X, neo-Y haplotype gets higher in the later generation (Figure 2 B). The *α* and *β* estimates ± SE of muller CD cline (Figure 3 A) is respectively 0.102 ± 0.056, 0.476 ± 0.169.

*Recombination associated with individual Muller CD haplotypes*

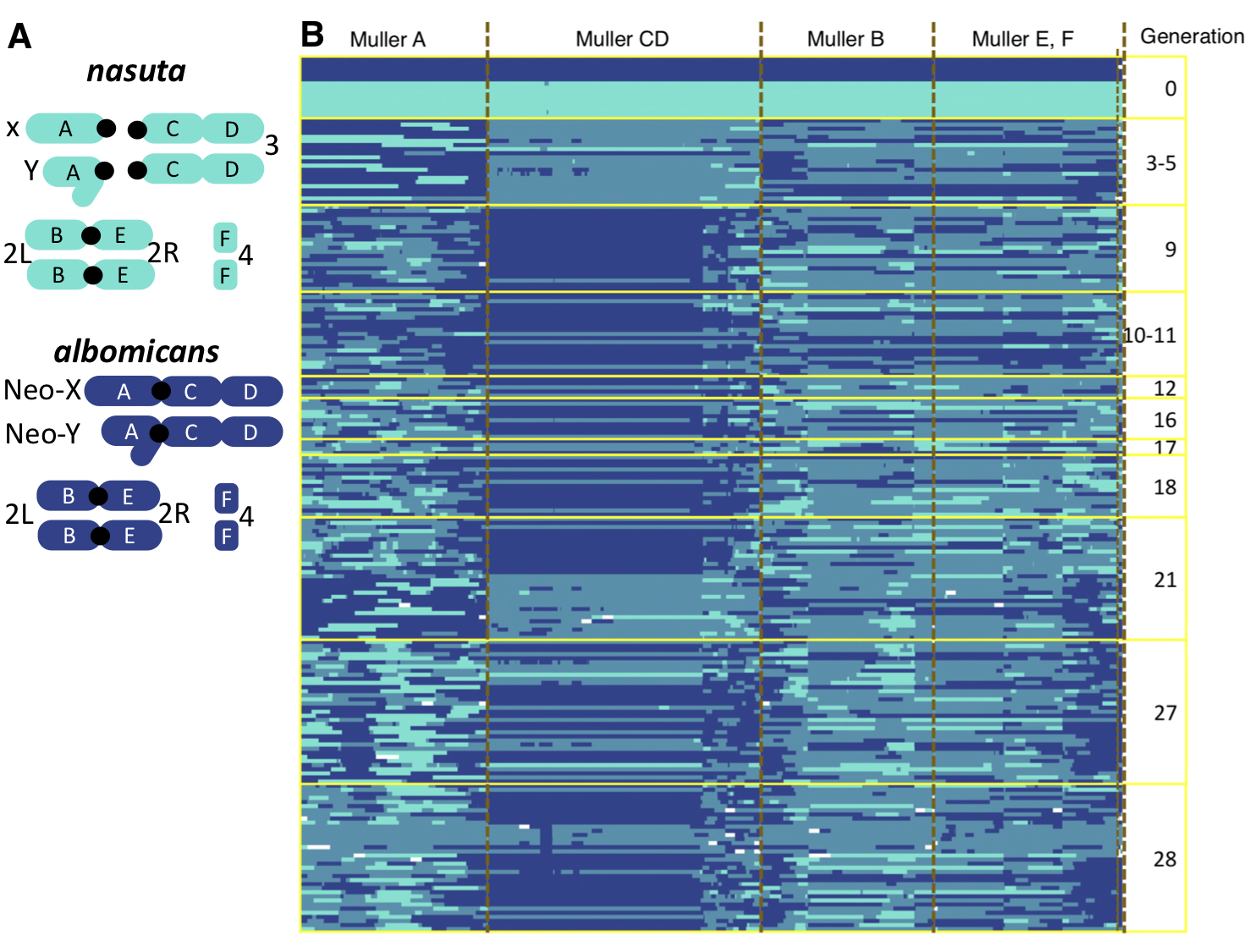
Genome-wide alb-biased introgression is significantly associated with Muller CD inheritance. There was significantly less recombination rate (Figure 4B) and greater alb-biased introgression (Figure 4C; p < 0.05) within Muller CD than other muller element. In particular, individuals of alb (neo-X, neo-X) and (neo-X, neo-Y) showed greater genome-wide alb introgression (*p* < 10-6 among females; Figure 4D), and they harbor more genome-wide ancestry turnovers than individuals with nasuta Muller CD (*p* < 0.05; figure 4E).

*Ancestry clusters*

The k-means cluster drastically reduced the number of ancestry-informative units in each muller element (Table S1). Due to the lack of recombination within Muller CD, this chromosome was collapsed into only 6 genetic clusters each respectively represents 10728, 10519, 148818, 1, 6716, and 11622 ancestry-informative sites.

*Genomic clines*

In early generations, signatures of admixture swamps the footprints of selection which settles down in later generations. In later generations (21, 27, and 28), the genetic cluster representing the inversion within Muller CD is repetitively exhibits signatures of selection against hybrids (positive beta values, and the 95% CIs don’t include zero) (Figure 5, right panel, yellow dots). The genetic cluster at the end of Muller CD and a suit of clusters towards the end of muller E exhibit alb-biased introgression (Figure 5, left panel, yellow dots), while the last genetic cluster of muller B (Figure 5, left panel, turquoise dots) demonstrate significant nas-biased introgression.



**Figure 1** **Karyotypes and** **admixture between albomicans and nasuta ancestry.** **A**, karyotypes of alb and nas, in which the Muller CD and Muller A that are separate in nas, are fused in alb forming neo-X or neo-Y. **B**, Ancestry-HMM haplotypes of haplotypes (in columns) in hybrids of various generations (rows). The turquoise and royal blue respectively represents homozygous nasuta and albomicans genotype, and the heterozygous genotypes are represented by pale blue.



**Figure 2 Muller CD haplotypes female (A) and male (B) hybrids sampled from different generation.**

Chart, diagram

Description automatically generated

**Figure 3** **Chromosome-specific admixture patterns.** The triangle plot with admixture proportion (pure nas =0, pure alb =1) on the x-axis and heterozygosity in different chromosome on the y-axis is variable in different muller element. In particular within Muller CD, alb-biased backcrossing is prevalent, while the same hybrids appear to be more admixed in other Muller elements.



**Figure 4 Neo-sex chromosome inheritance underpins alb-biased introgression** **A**, Muller CD harbors barrier effect and alb-biased introgression than the rest of the genome. Muller CD demonstrates less ancestry turnovers per site (**B**) and greater alb-biased introgression (corrected for admixture, **C**) than Muller A, B, and E. **D**, There was greater alb-biased introgression (controlling for admixture) within Muller CD in individuals with (neo-X, neo-X), (neo-X, neo-Y) than individuals with other Muller CD haplotypes, but not for the rest of the genome (**E**). (**F**) There was greater haplotype switches (controlling for admixture) within Muller CD in individuals with (nas, neoY) haplotype than individuals with other haplotypes. (**G**) In the rest of the genome, individuals with (neo-X, neo-X) haplotype exhibited more ancestry switches than individuals with other haplotype. **D**-**G**, The letters “A-C” on the plots delineate significantly different (*p* < 0.05) groups, otherwise “N.S.” indicates no significant difference was observed among groups.

Diagram

Description automatically generated

**Figure 5 Cross-generational barrier effects and ancestry advantage at genetic clusters.** The plots demonstrate estimates (surrounded by 95% CI) of alpha and/or beta genomic cline parameters across K-means genetic clusters in the genome in later generations (21, 27, and 28) of hybrids. A genetic cluster in Muller CD demonstrate significantly positive beta values (right panel, yellow) consistently across generations (21 to 28), implying that is cluster is a barrier to introgression. A different set of genetic clusters (largely within Muller E) turn out to show significant alb-biased introgression (left panel yellow dots, positive alpha), suggesting potential advantage of alb ancestry at these regions. In contrast, a cluster in Muller B consistently showed significant nas-biased introgression (left panel turquoise dots, negative alpha), indicating nas advantage at this genetic cluster.

**Discussion**

With generations of hybrid swarms, we revealed the role of neo-sex chromosome shaping asymmetry of the species barrier.

*Species barrier*

Large-X effect genome-wide --large scale barrier effect

Low recombination regions are associated with barriers

Speciation scenario –allopatric

The second cluster on Muller CD encompassing 10519

*Asymmetrical introgression*

Asymmetrical introgression of parental background (Matute et al. 2020). We observed similar pattern, but not as extreme. Meiotive drive

*Step-wise prevalence of neo-X and neo-Y*

Neo-Y was almost absent in earlier generations of male hybrids, which could be due to the meiotic structural incompatibility, and can only become more abundant when neo-X is more common over nas. Neo-X and neo-Y are more ad

**Reference**

Campbell, C. R., J. W. Poelstra, and A. D. Yoder. 2018. What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biol. J. Linn. Soc. 124:561–583.

Darwin, C. 1859. On the Origin of the Species.

Forgy, E. 1965. Cluster analysis of multivariate data : efficiency versus interpretability of classifications. Biometrics 21:768–769.

Gompert, Z., and C. A. Buerkle. 2011. Bayesian estimation of genomic clines. Mol. Ecol. 20:2111–2127.

Li, H. 2010. Aligning new-sequencing reads by BWA BWA : Burrows-Wheeler Aligner. Slides, doi: 10.1002/pssa.200673542.

Matute, D. R., A. A. Comeault, E. Earley, A. Serrato-Capuchina, D. Peede, A. Monroy-Eklund, W. Huang, C. D. Jones, T. F. C. Mackay, and J. A. Coyne. 2020. Rapid and predictable evolution of admixed populations between two Drosophila species pairs. Genetics 214:211–230.

**Supplementary**

**Figure S1** Ancestry haplotypes, each column is a SNP, each row is an individual.

**Figure S2** ancestry haplotypes after imputation, each column is a SNP, each row is an individual. The generation number is on the right side.

Figure S4

A picture containing chart

Description automatically generated

**Figure S3** K-means clusters

**Table S1** Summery of numbers of ancestry informative sites and k-means clusters across each muller element.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Muller A** | **Muller B** | **Muller CD** | **Muller E** | **Muller F** |
| Ancestry-informative sites | 128,670 | 120,223 | 188,404 | 126,724 | 3,036 |
| K-means clusters | 18 | 14 | 5 | 42 | 5 |