The nerodevelopmental genes *alan shepard* and *Neuroglian* contribute to female mate preference in African *Drosophila melanogaster*

Paula R. Roy *[1] and Dean M. Castillo *[2]

***** School of Biological Sciences, University of Utah, Salt Lake City UT

1 Current affiliation: Department of Neurobiology, University of Utah, Salt Lake City UT

2 Corresponding Author, Current affiliation: Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln, Lincoln NE, dmc79@cornell.edu

Abstract:

Mate choice is a key trait that determines fitness for most sexually reproducing organisms, with females often being the choosy sex. Female preference often results in strong selection on male traits that can drive rapid divergence of traits and preferences between lineages, leading to reproductive isolation. Despite this fundamental property of female mate choice, very few loci have been identified that contribute to mate choice and reproductive isolation. We used a combination of population genetics, quantitative complementation tests, and behavioral assays to demonstrate that *alan shepard* and *Neuroglian* contribute to female mate choice, and could contribute to partial reproductive isolation between populations of *Drosophila melanogaster*. Our study is among the first to identify genes that contribute to female mate preference in this historically important system, where female preference is an active premating barrier to reproduction. The identification of loci that are primarily known for their roles in neurodevelopment provides intriguing questions of how female mate preference evolves in populations via changes in sensory system and higher learning brain centers.

Introduction

For sexually reproducing organisms mate choice is critical to balance maximizing fitness with avoiding costly heterospecific mating events (Mendelson and Shaw, 2012; Gray, 2022). Strong selection on male courtship traits and female preferences can drive rapid divergence between closely related lineages (Fisher, 1930; Andersson, 2019), providing a link between sexual selection and the evolution of reproductive isolation. A large body of work has focused on documenting and describing female preference and identifying genes controlling male courtship traits (Chenoweth and Blows, 2006; Yamamoto and Koganezawa, 2013; Neelon et al., 2019; Munson et al., 2020). However, there have been fewer studies that have identified specific genes that contribute to female mate preference, with most instead focusing on quantitative trait loci (Bay et al., 2017; Svensson et al., 2017; Blankers et al., 2019; Xu and Shaw, 2019; Chowdhury et al., 2020) and some identifying specific genes (Chowdhury et al., 2020). Identifying and characterizing loci contributing to female preference is a critical first step to understand how courtship signal detection, signal processing, and higher brain functions (i.e., learning and memory) contribute to sexual selection and female mate preference evolution.

One challenge in identifying the genetic basis of female choice lies in the complexity of preference-based behaviors (Andersson, 2019). During mate choice females are receiving many, potentially multimodal, signals that then require processing before a decision to accept or reject a mate is made (Gray, 2022; Munson et al., 2020). The complexity of female preference implies that this trait is polygenic, but it is unclear how many loci we would expect to contribute to this trait. Specifically, it remains unclear whether multiple loci interact in an emergent fashion, contributing to female preference as a whole, or if individual loci contribute to a specific facet (perception of a specific signal/cue) of female preference (Boake et al., 2002; Chenoweth and

Blows, 2006). Additionally, while premating isolation evolves rapidly and is often thought to evolve before other barriers to reproduction (Coyne and Orr, 1989; Turissini et al., 2018), it can be difficult to determine if the behaviors that currently act as a mating barrier contributed to the speciation event or evolved after species divergence (Coyne and Orr, 2004; Safran et al., 2013; Kopp et al., 2018). Focusing on populations that vary in the degree of mating isolation can overcome some of these challenges and can be used as a model to understand the early stages of divergence and identify the loci that contribute to speciation and reproductive isolation.

Drosophila melanogaster originated in Southern Africa but is now a cosmopolitan species with a worldwide distribution. This broad range has resulted in structured and cryptic populations, even within Africa (Coughlan et al 2021), and strong premating isolation between some specific populations (Wu et al., 1995). For example, females from strains collected in Southern Africa strongly reject males from strains that are collected from non-African localities in both choice and no-choice mating experiments (Wu et al., 1995; Hollocher et al., 1997; Coughlan et al., 2021). These strong premating isolation behaviors are also observed in other parts of the *D. melanogaster* range including the Southeast United States and the Caribbean (Yukilevich and True, 2008). These mating preferences and behaviors that occur outside of Africa are potentially a product of African ancestry in these populations and admixture with non-African populations (Kao et al 2015; Bergland et al 2016). Combined, these observations of strong and variable female preference and population structure provide an ideal system to explore the evolution of female mate preference.

In this study, we leverage these *D. melanogaster* populations by combining population genomics and behavioral assays to identify genes that contribute to female mate preference and partial reproductive isolation. We focus on two loci, *alan Shephard* (*shep*) and *Neuroglian* (*Nrg*)

and demonstrate a role for both genes in female mate preference. First, we use previously published genomic data from Kao et al. (2015) to demonstrate that these loci show a clinal pattern of allele frequency in an admixed secondary contact zone, suggesting a role in reproductive isolation (Harrison and Larson, 2014). We then provide genetic evidence that these two loci contribute to female mate preference using quantitative complementation tests. Our study is one of a small number to identify genes important for female mate preference, and in this case it involves populations where premating isolation is an active barrier to gene flow.

Methods

Genomic cline analysis

To understand how alleles potentially involved in reproductive isolation are structured we compared allele frequencies across the Southeastern United States and Caribbean using previously published data (Kao et al 2015). We focused on these populations because this clinal structure allowed us to make explicit expectations for a gene that contributes to reproductive isolation compared to the rest of the genome (Harrison and Larson 2014). Specifically, because previous studies have demonstrated a correlation between latitude and African ancestry, with more African ancestry in southern populations, we predict that genes involved in behavioral isolation will have a steeper slope (i.e., larger regression coefficient) compared to the rest of the genome. Using the genotype matrix (genotype calls for all variable positions they identified) from Kao et al. (2015), we fit linear regression models between SNP allele frequency and latitude for the populations. After calculating regression coefficients for each SNP, we found that two genes previously identified as important for female behavior (*alan shepard* (*shep*) Chen et al., 2014 and *Neuroglian* (*Nrg)* Carhan et al., 2005) had steep clines. Previous data for these

genes demonstrated that null mutant females rejected wild-type male courtship and were mating deficient. We determined how the cline regression coefficients for *Nrg* and *shep* compared to the genomic distribution and the number of SNPs that were outliers and the SNP locations within these loci.

Quantitative complementation test

To determine if *Nrg* and *shep* alleles contribute to the female mate preference of African strains, we used a quantitative complementation test combined with a binary mate choice assay. Quantitative complementation tests have previously been used for no-choice tests (Chowdhury et al., 2020) and binary choice tests (Comeault et al., 2017). The basic premise of the quantitative complementation test is to cross wild type strains with a strain containing a mutant allele of the gene of interest to isolate the effects of different wild type alleles (Mackay, 2001). This is achieved by comparing progeny that are hemizygous for the locus of interest to individuals, ideally siblings, that are heterozygous at that locus. The complication of this test is that in *Drosophila* these mutants and deletions/deficiencies are often maintained over a balancer chromosome that has a marker with a dominant visible phenotype. This causes the heterozygote individuals, but not hemizygote individuals, to have a phenotype that could affect the trait of interest and must be taken into account during analysis (Turner et al., 2011).

Using a binary choice test, we introduced a single female of a given genotype to two males from two representative strains: Z53 representing an African male and DGRP882 representing a non-African. Trios were observed for 60 minutes or until copulation occurred. Upon copulation of a pair of individuals, the non-copulating male was removed, and the strain identity was determined (see behavioral assay conditions, below). After the pair was finished

copulating, we also examined this male to verify which strain the female chose. Our *a priori* expectation was that the females hemizygous for the African allele will prefer African males, whereas females hemizygous for the non-African allele will mate indiscriminately.

Drosophila strains and crosses

To determine whether *Nrg* and *shep* alleles contribute to female preference behavior between populations of *D. melanogaster* we needed to identify wild-type strains that show strong female mate preference and complementary strains that show no preference. Strains from Southern Africa were once considered to be a single lineage from the ancestral range and were referred to as Z-type. All other strains that migrated from the ancestral range (non-African and some Northern African strains) were considered a single cosmopolitan lineage and referred to as M-type. The M-type behavior contrasted with Z-type behavior in that M-type females often mate indiscriminately, showing no preference, whereas Z-type females strongly reject M-type males (Wu et al., 1995). Since this initial description, the demographic and evolutionary relationships between *D. melanogaster* populations have been shown to be more complex (Coughlan et al., 2021), and it is unlikely that all Z-type strains are homogenous lineage.

We were able to choose strains based on geographic location and evaluated the female preference behavior for wild type females prior to using these strains in our crosses. For our representative African strain, we chose Z53 because its behavior profile is well documented and has been consistent across studies (Wu et al., 1995; Hollocher et al., 1997; Moran, 2006; Jin et al., 2022). For our representative non-African strain, we chose DGRP882 which we had previously used and had expected baseline behavior (Jin et al., 2022). The African admixture has been estimated for individuals of the DGRP panel (Pool, 2015) and DGRP882 has among the

least amount of African ancestry making it most likely to be differentiated African strains. In addition to these representative lines, we also chose two additional African strains: Z30 and a more recently collected strain CH11. Canton-S was used as an additional non-African strain because it is commonly used in behavioral studies (Tompkins et al., 1982; Lasbleiz et al., 2006; Ng and Kopp, 2008; Kohlmeier et al., 2021).

To create heterozygote and hemizygote genotypes we used loss of function mutant lines for both *shep* and *Nrg* that had been previously characterized as null alleles (*shep* Chen et al., 2014; *Nrg* Bieber et al., 1989; Enniking et al., 2013). We used the $shep^{BG00836}$ null mutation (BDSC #12513), hereafter $shep^{-}$. Homozygous null mutant females for this allele are viable but have mating deficiencies where they do not accept matings as virgins (Chen et al., 2014). We balanced this mutation over the TM6B chromosome. This balancer chromosome carries the dominant visible *Hu* marker. We inbred progeny from this original cross for three generations before using this strain in our behavioral assays. The mutant line and the line that contributed the TM6B chromosome were not the same genetic background so residual heterozygosity could be present in this strain. Any effects of this heterozygosity on the X chromosome and 2nd chromosome would occur in all of the crosses/treatments that we examined. A *shep* deficiency line (BDSC #7583), already balanced over the TM6B, was tested in preliminary experiments, but female mating behavior in control crosses with DGRP882 produced highly variable results. Additional large deficiencies are available but would have precluded us from testing the role of *shep* directly. For this reason, we chose to use the balanced strain we had created. We crossed the *shep*/TM6B line to both DGRP882 and Z53 males to generate F1 females that were either heterozygous (had humoral phenotype) or hemizygous (had wild-type phenotype) for *shep*.

We used the *Nrg*^{17} allele, which is synonymous with *Nrg*^{2} and hereafter *Nrg*$^{-}$, balanced over the FM7 chromosome, which carries the *Bar* dominant visible marker (BDSC #5595). The *Nrg*^{17}requires balancing because this mutation is homozygous lethal. We crossed the *Nrg*$^{-}$/FM7 line to DGRP882 and Z53 similar to the design for *shep.* For experiments with *Nrg*, we tested two additional African and one non-African strains (described above) for two reasons. First, we were not sure how or if the *Bar* locus, which affects eye morphology, would affect female preference. This is particularly important because the impact of vision for female mating behavior could vary between strains. We tested the effect of *Bar* explicitly by including Canton S as an additional strain that should not exhibit strong mating preference. Second, even though our cline analysis identified *Nrg* as a strong outlier, this locus was only a marginal outlier in and independent study that identified putative female preference loci (Coughlan *et al.*, 2021). This could indicate that the effect of *Nrg* differs between strains and populations, or multiple alleles of *Nrg* segregate in African populations with only a subset contributing to female preference phenotypes. Either scenario could result in a situation where there is not a single high frequency allele contributing to female preference in African strains motivating us to test multiple African strains for this locus.

Behavioral assay conditions

All stocks and virgins were kept on standard cornmeal molasses media in an incubator on a 12:12 light cycle held at a constant 20 degrees Celsius and 50% humidity. Virgin males and females were collected within the first 0-4 hours post eclosion to ensure they had not mated. Females were phenotyped based on the presence of the dominant visible marker and housed in groups of 5-10. Males were housed individually because group housing can potentially change

courtship vigor (Dixon et al., 2003). All individuals were allowed to age 7-10 days, before experiments because this is the timeframe that produces the maximum number of matings for the African strains (Jin et al 2022). All behavioral assays were performed in a room held at a constant 20C degrees Celsius and 50% humidity within two hours of the incubator lights turning on to maximize the number of copulations we could observe.

The strains used in the experiment are phenotypically indistinguishable, so to be able to tell males apart we placed them on food containing blue food coloring 48 hours prior to the behavioral assay. This is a non-invasive, robust identification method (Wu et al., 1995; Hollocher et al., 1997; Serrato-Capuchina et al., 2021). In these studies, all males received food dye treatment, but in our experiment, we only marked one male per mating with blue food dye. We found that we could more easily distinguish between a dyed and non-dyed male rather than distinguishing between males fed two different colors. We tested the effect of this treatment on female preference and found no effect (Supplementary Information).

We used males from the representative strains Z53 and DGRP882 for all of our binary choice tests. We collected all genotypes for a given locus in each block of our trials in equal numbers, but due to differences in viability and the number of females that made choices, our final number of females that were included differed for each genotype (Table 1), but all genotypes were represented in each block. While the rearing conditions and behavior rooms conditions were very well controlled, our behavioral protocol also included completing a set of DGRP882 females in parallel to our experimental crosses. For each block of trials, 20 DGRP882 females were used as a behavioral control and we consistently had 8-12 (representing 40-60%) choosing Z males consistent with the random mating for this genotype, allowing us to conclude that the environment was not affecting our matings and pool replicates over blocks.

Statistical analysis of behavioral data

The analysis typically conducted for quantitative complementation tests compares the difference in phenotype for heterozygote genotypes to the difference in hemizygous genotypes, using two strains that maximize phenotype divergence (Pasyukova and Mackay 2000; Mackay 2001). Statistically, this has been accomplished by using a model that includes a strain effect, an effect for carrying the balancer chromosome (which also captures hemizygosity), and the interaction of these terms. This model works nicely for quantitative traits, but the interpretation of interaction effects is not straightforward for non-linear models or models based on logistic regression (Ai and Norton 2003; Chen 2003). For behavioral studies that have used complementation tests the preference is to use logistic regression as this best captures the binary nature of choice data, but not all studies have tested the interaction term explicitly. One study compared hemizygotes to heterozygotes for an interspecies cross using the wildtype strain of interest (Comeault et al., 2017), but this framework does not control for either hemizygosity effects or the effects of dominant mutations carried on balancer chromosomes. This would have required a within species control cross, whereas the study instead focused on interspecies hybrids. Another study included multiple strains in their statistical models to evaluate hemizygosity and balancer effects but still focused on the significance between hemizygote and heterozygote genotypes that were generated from the interspecific crosses (Chowdhury et al., 2020). This study modeled the interaction effect, but only reported the significance of this effect, which is not sufficient to understand the sign of the interaction effect without recalculating the odds ratios for the comparisons of interest (Ai and Norton 2003; Chen 2003). The interpretation of interaction effects in logistic regression cannot be made based on this single coefficient

because the interaction is conditional on other effects (Ai and Norton 2003). Since this is not an issue in statistical models using quantitative traits (i.e., linear regression) authors must be cautious when applying interaction models to logistic data.

For our analysis, we used a statistical framework that can be readily interpreted and parallels the intent of the original analysis outlined for quantitative complementation tests (Pasyukova and Mackay 2000; Mackay 2001), but avoids modelling the interaction effect, and the difficulty in the interpretation of this coefficient. By comparing each genotype in a set of crosses to a single baseline genotype, we can test for a strain effect, a balancer effect, and the effect of the hemizygote of interest. If a strain effect is detected a post-hoc test can be used to determine if there is an interaction between strain and hemizygosity in a similar fashion of what would be required to properly interpret a statistical interaction effect for a logistic regression.

We used logistic regression where the response variable was 0 if the female mated with the DGRP882 male and 1 if they mated with the Z53 male. Female genotype was included as a categorical variable in the model. We first determined whether any female genotypes (for a given locus) were significantly different that the DGRP882 heterozygote genotype that we used as baseline in our model. Comparing the heterozygote genotypes to the baseline allowed us to test for any strain effects. If a given African strain heterozygote was not significantly different from baseline, we could infer there was no strain difference. These heterozygotes would have one common haploid genome, and the other complement coming from the strain used in the cross. Testing the balancer effect, which also tests for hemizygosity effects, was completed by comparing the DGRP882 heterozygote that carries a balancer chromosome, to the DGRP882 hemizygote. In the absence of balancer and heterozygote effects we would expect these genotypes to show the same, random mating preference. For all crosses the DGRP882

hemizygote was never significantly different from the DGRP882 heterozygote, allowing us to focus on whether any of the African hemizyogtes were significantly different from baseline. If the African strain hemizygote was significantly different from the baseline, and DGRP882 hemizygote but the African strain heterozygote was not different from baseline we could infer that this allele contributes to female preference behavior. would fulfill the criteria of the difference in hemizygotes being greater than the difference in heterozygotes (Mackay, 2001). If both the specific African strain heterozygote and African strain hemizygote were significantly different from baseline we proceeded to the second analysis step. This involved computing a linear contrast for the difference in these genotypes from their representative DGRP882 counterparts. This second step was only necessary for the African strain CH11 using in the *Nrg* experiment.

Results

Female behavior loci are genomic outliers

To generate candidate genes for functional validation, we leveraged genomic data from populations of *D. melanogaster* that exhibit a cline in African ancestry (Kao et al., 2015; Bergland et al., 2016) and African-like female preference behavior (Yukilevich and True, 2008). Our goal was to identify outlier genes that showed a steeper cline in African ancestry compared to the genome wide average. We also compared our list of candidates to independent analyses that used populations from Africa to identify female preference alleles (Coughlan et al., 2021).

We retained the top 1% of outliers for loci that had the steepest clines with African alleles being more frequent at lower latitudes. Many of these SNPs did not occur in annotated genes, but two interesting candidates emerged, *alan shepard* (*shep*) and *Neuroglian* (*Nrg*). There was a

single SNP in *shep* that was above the top 1% threshold and this SNP was in an intron of *shep*. There were three additional SNPs close to the 1% threshold that were also in introns. For *Nrg* there were two intronic SNPs that were in the top 1% and these SNPs were within 2 kb of each other. Interestingly these loci were also outliers in the study by Coughlan et al (2021) with shep in the top 1% of outliers and *Nrg* in the top 5% of outliers. Previous behavioral analysis of homozygous viable mutant strains for these loci found evidence that these genes were involved female behavior, where virgin females rejected male courtship to remain unmated (Carhan et al., 2005; Chen et al., 2014). Taken together, our observation of a strong cline in allele frequency and previous studies suggesting a role in behavior led us to functionally test these loci for a role specifically in African female mate preference behavior.

Shep contributes to female mate preference in an African strain

We used quantitative complementation tests to determine if *alan shepard* (*shep*) contributes to female mate preference. We were able to validate that *shep* contributes to female mate preference by specifically demonstrating that females that are hemizygous for the Z53 African allele have significantly different behavior than the other control genotypes. These African hemizygote females show a significantly higher preference for African males over non-African males (Fig. 2). The rest of the genotypes were not significantly different from our baseline genotype, the DGRP882 heterozygote which carried the DGRP882 *shep* allele and the *shep* allele on the TM6B chromosome (Table 1). The non-African hemizygote, that only carried the DGRP882 allele of *shep* was not significantly different from the baseline (β=-0.374, P=0.1761), neither was the Z53 heterozygote genotype (β=0.522, P=0.0552). The only genotype that was significantly different was the Z53 hemizygote (β=1.068, P=0.0022). This genotype preferred

Z53 males over DGRP882 males in a similar direction, but to a lesser degree, as the wild-type Z53 parental strain indicating that this allele of *shep* contributes to female mate preference.

Nrg contributes female mate preference in African strains

We also used quantitative complementation tests to validate the role of *Nrg* alleles on female mate preference in African strains. Our design and logic were identical for what we used to analyze *shep* and for this gene we included more genotypes to determine if the visible marker, *Bar*, on the balancer chromosome was affecting behavior, and whether different *Nrg* alleles might show different effects on female mate preference. We used the non-African genotype DGRP882 heterozygote which carries the *Nrg* allele from DGRP882 and the *Nrg* allele from the FM7 chromosome as our baseline.

To test for the effects of the *Bar* allele specifically on female behavior we tested the behavior of another non-African strain, Canton S, and compared it to the DGRP882 strain. We used Canton S because we had a clear expectation: that both female genotypes from a cross between Canton S and the Nrg^{-}/FM7 strain should have the same behavior as the DGRP882 genotypes, and they should mate randomly and not show preference for either male genotype. We in fact observed that these genotypes mated indiscriminately, and we did not detect any significant difference in behavior comparing DGRP882 genotypes and Canton S genotypes (Fig 2; Table 1).

We then tested *Nrg* alleles from three African strains for effects on female mate preference. The first strain CH11 provided inconclusive results because females from both the heterozygote and hemizygote genotypes showed strong preference for Z53 males compared to the DGRP882 baseline (Fig 2; Table 1). The strength of this preference resulted in both

genotypes being significantly different from the baseline, and indistinguishable from each other. The dominant female preference from this strain precluded our ability to test the effect of the CH11 *Nrg* allele specifically.

The remaining two African strains had a pattern consistent with *Nrg* contributing to female mate preference (Fig 3; Supplemental Table 1). Both the heterozygous female genotypes for the Z30 and Z53 strain were not significantly different from the baseline (Z30 β=0.936, P=0.0862; Z53 β=0.288; P=0.5752). The hemizygous female genotypes for both strains were significantly different from the baseline (Z30 β=1.986, P=0.0004; Z53 β=1.413; P=0.0072) indicating that when females only have these African strain alleles they exhibit significant preference for African males similar to the wild-type African parental strains.

Discussion

Identifying genes that contribute to female mate preference is an important first step in understanding how this important behavior evolves and contributes to the evolution of reproductive isolation and speciation. In this study, we demonstrate that *shep* and *Nrg* play a role in female mate preference in some African strains. Our current study is one of very few studies that has identified loci with direct genetic evidence for female mate preference (Chowdhury et al., 2020) and importantly in a system where premating isolation is the strongest and most relevant barrier to reproduction. These African populations of *D. melanogaster* have been a historically important system in the study of speciation (Wu et al., 1995; Coyne et al., 1999; Greenberg et al., 2003), and we have now identified, in part, genes that could contribute to premating isolation in this system. Since the genetic basis of behavior is complex, and we observed variation in behavior among strains (see below), it will be important to test the roles of

shep and *Nrg* in more strains from different locations within Africa. Regardless, these remain promising candidates that we can leverage to understand the evolution of behavior. For example, when this *D. melanogaster* system was initially described, mapping studies were only able to identify large chromosomal regions on chromosome 3, the largest contributor to premating isolation, and the X chromosome (Hollocher et al., 1997; Takahashi and Ting, 2004). Combining population genetics with complementation tests, we identified a gene on chromosome 3, *shep* is on 3L, and the X chromosome, *Nrg*, that contribute to female preference behavior. Getting gene scale resolution will allow us to further understand how female mate preference has evolved. Since both *Nrg* and *shep* are essential for neurodevelopment and female mating behavior (Carhan et al., 2005; Chen et al., 2014), the identification of their role in female preference in African strains can potentially hint at sensory pathways that are important for mating decisions and partial reproductive isolation.

Female preference can evolve when structures/genes that are involved in sensory perception evolve, when brain centers important for learning/decision making evolve, or when both evolve in tandem (Stevens, 2013; Schaefer and Ruxton, 2015; Munson et al., 2020). Both *shep* and *Nrg* regulate neurological development, specifically neural remodeling during metamorphosis, contributing heavily to the formation of higher learning centers, specifically those responsible for processing olfactory signals (Carhan et al., 2005; Chen et al., 2014). Our results complement a gene ontology analysis of outliers for female mate preference by Coughlan et al. (2021) that found significant outliers involved in sensory perception, including three olfactory receptors, as well as genes important for learning and memory, cognition, and neurological development. Both analyses are consistent with previous work in these populations of *D. melanogaster* has suggested a large role for olfaction in female mate preference (Grillet et

al., 2012; Moran, 2006; Jin et al., 2022). Given these complementary observations it will be important to understand how variation *Nrg* and *shep* alleles are structured in strains and populations that show high or low mate discrimination preference. Identifying different alleles could provide insight into the structural and/or functional changes in the neural networks of females and whether this is correlated with the olfactory cues used in mate preference. Regardless, while we observed large effects of these single loci, the genetic basis of female preference is likely complex, with additional loci contributing to isolation in these strains.

One interesting observation that could suggest that more than these two loci contribute to female preference is the variation in dominance that we observed for female mating preference in hybrids between non-African and African strains. In our experiment, we used three African strains that show strong preference for African males and never choose non-African males in choice tests. We observed that female mate preference was recessive in F1 hybrids for two of the strains and dominant in F1 hybrids the third strain when crossed to the same non-African strain (see Results). The dominance in the CH11 strain precluded us from testing the effects of *Nrg* on female preference for that strain. When assessing how important *shep* and *Nrg* are in reproductive isolation across strains different experimental designs can be used to circumvent this issue. Nevertheless, testing the dominance of female behavior in more strains will be informative in and of itself and could highlight how general this phenomenon is which will be important because the dominance of reproductive barriers can have strong impacts on the outcome of speciation (Thompson et al., 2021) and might indicate how particular mating barriers evolve.

Differences in dominance for female mate preference could reflect differences in the selective environment caused by local mating dynamics. This scenario would be analogous to

parallel vs non-parallel evolution of ecological traits in different environments (Oke et al., 2017; Bolnick et al., 2018). If different populations experience different selective pressures, then different alleles of *shep* and *Nrg* could segregate in these populations. African lineages are quite diverse and while strong premating isolation is found in many populations it is likely that female preferences and associated alleles could be segregating within Africa (Yukilevich and True, 2008; Coughlan et al 2021) Another possibility, that is not mutually exclusive, is that the genetic architecture of female preference in hybrids is genotype-dependent because different genes contribute to female mate preference in these different populations. Given the limited data on the genetic basis of female preference it is difficult to infer the likelihood of this process, but polymorphic incompatibilities are common in other systems (Cutter, 2012; Castillo and Barbash, 2017).

One example in *Drosophila* suggests that the genetic architecture of premating isolation can be variable. When different strains of *D. simulans* were crossed and tested for mate rejection of *D. melanogaster* different numbers, locations, and dominance of quantitative trait loci (QTL) were identified that were strain specific (Uenoyama and Inoue, 1995; Carracedo et al., 1998a; Carracedo et al., 1998b; Carracedo et al., 2000). The variation in dominance and the number of loci contributing to mate preference in this *D. simulans* example might be consistent with this barrier evolving after the speciation event that separated these lineages (Laturney and Moehring, 2012). Moving forward it will be important to determine how common it is for variable loci to contribute to premating isolation or uniform genetic architecture for reproductive isolation between populations. Either way, the difference in dominance suggests there may be standing variation for female mate preference that selection can act on, which is known to facilitate the

evolution of premating isolation through sexual selection (Mendelson et al., 2014; Castillo and Delph, 2016).

Overall, our results provide evidence for two neurodevelopmental genes, *Nrg* and *shep*, contributing to female mate preference in African strains, and could contribute to partial reproductive isolation between populations of *D. melanogaster*. Our results connect population genetics from a system that exhibits a cline in behavior and allele frequency, to functional validation of the effect of these genes on behavior. The differences in alleles for both loci in non-African vs African populations should be further explored and the testing of more strains are needed to determine whether these could be considered speciation genes. Understanding allelic diversity will help develop experiments that can test the functional consequences on neural circuits and sensory perception for specific alleles. Neurobiological experiments could provide insight into how female preference evolved in this system and provide a framework for exploring these types of data in other systems. Ultimately, this could lead to understand common patterns for the evolution of female mate preference and the genetic and neural level.

Literature Cited

Ai, C.R., Norton, E.C., 2003. Interaction terms in logit and probit models. Economic Letters 80, 123-129.

Andersson, M., 2019. Sexual selection. Princeton University Press.

Bay, R.A., Arnegard, M.E., Conte, G.L., Best, J., Bedford, N.L., McCann, S.R., Dubin, M.E., Chan, Y.F., Jones, F.C., Kingsley, D.M., Schluter, D., Peichel, C.L., 2017. Genetic Coupling of Female Mate Choice with Polygenic Ecological Divergence Facilitates Stickleback Speciation. Current Biology 27, 3344-3349.e4.

Bergland, A.O., Tobler, R., González, J., Schmidt, P., Petrov, D., 2016. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. Mol Ecol 25, 1157–1174.

Bieber, A.J., Snow, P.M., Hortsch, M., Patel, N.H., Jacobse, J.R., Traquina, Z.R., Schilling, J., Goodman C.S., 1989. *Drosophila* neuroglian: a member of the immunoglobulin superfamily with extensive homology to the vertebrate neural adhesion molecule L1. Cell 59, 447-460.

Blankers, T., Berdan, E.L., Hennig, R.M., Mayer, F., 2019. Physical linkage and mate preference generate linkage disequilibrium for behavioral isolation in two parapatric crickets. Evolution 73, 777–791.

Boake, C.R., Arnold, S.J., Breden, F., Meffert, L.M., Ritchie, M.G., Taylor, B.J., Wolf, J.B., Moore, A.J., 2002. Genetic tools for studying adaptation and the evolution of behavior. The American Naturalist 160, S143–S159.

Bolnick, D.I., Barrett, R.D.H., Oke, K.B., Rennison, D.J., Stuart, Y.E., 2018. (Non)Parallel Evolution. Annu. Rev. Ecol. Evol. Syst. 49, 303–330.

Carhan, A., Allen, F., Armstrong, J.D., Goodwin, S.F., O'Dell, K.M.C., 2005. Female receptivity phenotype of icebox mutants caused by a mutation in the L1-type cell adhesion molecule neuroglian. Genes Brain Behav 4, 449–465.

Carracedo, M., Asenjo, A., Casares, P., 1998a. Brief communication. Inheritance mode of *Drosophila simulans* female mating propensity with *D. melanogaster* males. Journal of Heredity 89, 101–104.

Carracedo, M.C., Suarez, A., Asenjo, A., Casares, P., 1998b. Genetics of hybridization between *Drosophila simulans females* and *D. melanogaster males*. Heredity 80, 17–24.

Carracedo, M.C., Suarez, C., Casares, P., 2000. Sexual isolation between *Drosphila melanogaster*, *D. simulans* and *D. mauritiana*: sex and species-specific discrimination. Genetica 108, 155–162.

Castillo, D.M., Barbash, D.A., 2017. Moving Speciation Genetics Forward: Modern Techniques Build on Foundational Studies in *Drosophila*. Genetics 207, 825–842.

Castillo, D.M., Delph, L.F., 2016. Male–female genotype interactions maintain variation in traits important for sexual interactions and reproductive isolation. Evolution 70, 1667–1673.

Chen, D., Qu, C., Bjorum, S.M., Beckingham, K.M., Hewes, R.S., 2014b. Neuronal Remodeling During Metamorphosis Is Regulated by the *alan shepard* (shep) Gene in *Drosophila melanogaster*. Genetics 197, 1267–1283.

Chen, J.J., 2003. Communication complex information: the interpretation of statistical interaction in multiple logistic regression analysis. American Journal of Public Health 93, 1376-1377.

Chenoweth, S.F., Blows, M.W., 2006. Dissecting the complex genetic basis of mate choice. Nature Reviews Genetics 7, 681–692.

Chowdhury, T., Calhoun, R.M., Bruch, K., Moehring, A.J., 2020. The *fruitless* gene affects female receptivity and species isolation. Proc. R. Soc. B. 287, 20192765.

Comeault, A.A., Serrato-Capuchina, A., Turissini, D.A., McLaughlin, P.J., David, J.R., Matute, D.R., 2017. A nonrandom subset of olfactory genes is associated with host preference in the fruit fly *Drosophila orena*: Behavioral Specialization In *D. Orena*. Evolution Letters 1, 73–85.

Coughlan, J.M., Dagilis, A.J., Serrato-Capuchina, A., Elias, H., Peede, D., Isbell, K., Castillo, D.M., Cooper, B.S., Matute, D.R., 2021. Patterns of and processes shaping population structure and introgression among recently differentiated *Drosophila melanogaster* populations (preprint). Evolutionary Biology.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer associates Sunderland, MA.

Coyne, J.A., Orr, H.A., 1989. Patterns of speciation in *Drosophila*. Evolution 43, 362–381.

Coyne, J.A., Wicker-Thomas, C., Jallon, J.-M., 1999. A gene responsible for a cuticular hydrocarbon polymorphism in *Drosophila melanogaster*. Genet. Res. 73, 189–203.

Cutter, A.D., 2012. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. Trends in Ecology & Evolution 27, 209–218.

Dixon, S.M., Coyne, J.A., Noor, M.A.F., 2003. The evolution of conspecific sperm precedence in Drosophila. Molecular Ecology 12:1179-1184.

Enneking, E.M., Kudumala, S.R., Moreno, E., Stephan, R., Boerner, J., Godenshwege, T.A., Pielage, J., 2013. Transsynaptic coordination of synaptic growth, function, and stability by the L1-type CAM *Neuroglian*. Plos Biology 11, e1001537.

Fisher, R.A., 1930. The genetical theory of natural selection. Clarendon Press, Oxford.

Gray, D.A., 2022. Sexual selection and 'species recognition' revisited: serial processing and order-of-operations in mate choice. Proc. R. Soc. B. 289, 20212687.

Greenberg, A.J., Moran, J.R., Coyne, J.A., Wu, C.-I., 2003. Ecological Adaptation During Incipient Speciation Revealed by Precise Gene Replacement. Science 302, 1754–1757.

Grillet, M., Everaerts, C., Houot, B., Ritchie, M.G., Cobb, M., Ferveur, J.-F., 2012. Incipient speciation in *Drosophila melanogaster* involves chemical signals. Sci Rep 2, 224.

Harrison, R.G., Larson, E.L., 2014. Hybridization, Introgression, and the Nature of Species Boundaries. Journal of Heredity 105, 795–809.

Hollocher, H., Ting, C.-T., Wu, M.-L., Wu, C.-I., 1997c. Incipient Speciation by Sexual Isolation in *Drosophila melanogaster:* Extensive Genetic Divergence Without Reinforcement. Genetics 147, 1191–1201.

Jin, B., Barbash, D.A., Castillo, D.M., 2022. Divergent selection on behavioural and chemical traits between reproductively isolated populations of *Drosophila melanogaster*. J of Evolutionary Biology jeb.14007.

Kao, J.Y., Zubair, A. Salomon, M.P., Nuzhdin, S.V., Campo, D., 2015. Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from the south-eastern United States and Caribbean Islands. Molecular Ecology 24, 1499-1509.

Kohlmeier, P., Zhang, Y., Gorter, J.A., Su, C.Y., Billeter, J.C., 2021. Mating increases *Drosophila melanogaster* females' choosiness by reducing olfactory sensitivity to a male pheromone. Nature Ecology and Evolution 5:1165-1173.

Kopp, M., Servedio, M.R., Mendelson, T.C., Safran, R.J., Rodríguez, R.L., Hauber, M.E., Scordato, E.C., Symes, L.B., Balakrishnan, C.N., Zonana, D.M., van Doorn, G.S., 2018. Mechanisms of Assortative Mating in Speciation with Gene Flow: Connecting Theory and Empirical Research. The American Naturalist 191, 1–20.

Lasbleiz, C., Ferveur, J.F., Everaerts, C., 2006. Courtship behavior of *Drosophila melanogaster* revisited. Animal Behavior 72:1001-1012.

Laturney, M., Moehring, A.J., 2012. The Genetic Basis of Female Mate Preference and Species Isolation in *Drosophila*. International Journal of Evolutionary Biology 2012, 1–13.

Mackay, T.F.C., 2001. Quantitative trait loci in *Drosophila*. Nat Rev Genet 2, 11–20.

Mendelson, T.C., Martin, M.D., Flaxman, S.M., 2014. Mutation-order divergence by sexual selection: diversification of sexual signals in similar environments as a first step in speciation. Ecol Lett 17, 1053–1066.

Mendelson, T.C., Shaw, K.L., 2012. The (mis) concept of species recognition. Trends in ecology & evolution 27, 421–427.

Moran, J.R., 2006. The Genetics of Pre-mating Isolation at the Incipient Stage of Speciation in Drosophila Melanogaster. University of Chicago, Committee on Genetics.

Munson, A.A., Jones, C., Schraft, H., Sih, A., 2020. You're Just My Type: Mate Choice and Behavioral Types. Trends in Ecology & Evolution 35, 823–833.

Neelon, D.P., Rodríguez, R.L., Höbel, G., 2019. On the architecture of mate choice decisions: preference functions and choosiness are distinct traits. Proc. R. Soc. B. 286, 20182830.

Ng, C.S., Kopp, A., 2008. Se combs are important for male mating success in *Drosophila melanogaster*. Behavior Genetics 38:195-201.

Oke, K.B., Rolshausen, G., LeBlond, C., Hendry, A.P., 2017. How Parallel Is Parallel Evolution? A Comparative Analysis in Fishes. The American Naturalist 190, 1–16.

Pasyukova, E.G., Vieira, C., Mackay, T.F., 2000. Deficiency mapping of quantitative train loci affecting longevity in *Drosophila melanogaster*. Genetics 156:1129-1146.

Pool, J.E., 2015. The mosaic ancestry of the *Drosophila* genetic reference panel and the D. melanogaster reference genome reveals a network of epistatic fitness interactions. Molecular biology and evolution 32, 3236–3251.

Safran, R.J., Scordato, E.S.C., Symes, L.B., Rodríguez, R.L., Mendelson, T.C., 2013. Contributions of natural and sexual selection to the evolution of premating reproductive isolation: a research agenda. Trends in Ecology & Evolution 28, 643–650.

Schaefer, H.M., Ruxton, G.D., 2015. Signal Diversity, Sexual Selection, and Speciation. Annu. Rev. Ecol. Evol. Syst. 46, 573–592.

Serrato-Capuchina, A., Schwochert, T.D., Zhang, S., Roy, B., Peede, D., Koppelman, C., Matute, D.R., 2021. Pure species discriminate against hybrids in the Drosophila melanogaster species subgroup. Evolution 75, 1753–1774.

Stevens, M., 2013. Sensory ecology, behaviour, and evolution. Oxford University Press.

Svensson, O., Woodhouse, K., van Oosterhout, C., Smith, A., Turner, G.F., Seehausen, O., 2017. The genetics of mate preferences in hybrids between two young and sympatric Lake Victoria cichlid species. Proc. R. Soc. B. 284, 20162332.

Takahashi, A., Ting, C.-T., 2004. Genetic Basis of Sexual Isolation in *Drosophila melanogaster*. Genetica 120, 273–284.

Thompson, K.A., Urquhart-Cronish, M., Whitney, K.D., Rieseberg, L.H., Schluter, D., 2021. Patterns, Predictors, and Consequences of Dominance in Hybrids. The American Naturalist 197, E72–E88.

Turissini, D.A., McGirr, J.A., Patel, S.S., David, J.R., Matute, D.R., 2018. The rate of evolution of postmating-prezygotic reproductive isolation in Drosophila. Molecular Biology and Evolution 35, 312–334.

Turner, T.L., Stewart, A.D., Fields, A.T., Rice, W.R., Tarone, A.M., 2011. Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in Drosophila melanogaster. PLoS Genet 7, e1001336.

Uenoyama, T., Inoue, Y., 1995. Genetic studies on premating isolation in *Drosophila simulans*. I. A *D. simulans* line highly crossable with *D. melanogaster*. The Japanese Journal of Genetics 70, 365–371.

Wu, C.-I., Hollocher, H., Begun, D.J., Aquadro, C.F., Xu, Y., Wu, M.-L., 1995. Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. Proceedings of the National academy of sciences 92, 2519–2523.

Xu, M., Shaw, K.L., 2019. Genetic coupling of signal and preference facilitates sexual isolation during rapid speciation. Proc. R. Soc. B. 286, 20191607.

Yamamoto, D., Koganezawa, M., 2013. Genes and circuits of courtship behaviour in *Drosophila* males. Nat Rev Neurosci 14, 681–692.

Yukilevich, R., True, J.R., 2008. African Morphology, Behavior and Phermones Underlie Incipient Sexual Isolation between US and Caribbean *Drosophila melanogaster*. Evolution 62, 2807–2828.

Table 1. The behavioral differences in female mate preference from the baseline genotype for both *shep* and *Nrg*. The baseline genotype in each case was the DGRP882 heterozygote which carried the DGRP882 allele for the specific locus and the allele from the corresponding balancer chromosome. The NA represents that this genotype was used as the baseline. The remaining coefficients represent log odds ratios, estimated from logistic regression, describing the increase or decrease in a genotypes propensity to mate with Z53 males. The significance of each coefficient was tested independently against the baseline and ** represents $P<0.001$.

shep				
Strain-Genotype	Females tested	Proportion choose	β coefficient	*P-value*
DGRP882 homozygote	40	0.48	NA	NA
Z53 heterozygote	53	0.64	0.552	0.0552
DGRP882 hemizygote	54	0.41	-0.3747	0.1761
Z53 hemizygote	43	0.74	1.068	0.0022**
Nrg				
Strain-Genotype	Females tested	Proportion choose Z	β coefficient	*P-value*
DGRP882 homozygote	29	0.58	NA	NA
Canton S homozygote	24	0.54	0.515	0.35464
Z30 heterozygote	28	0.64	0.936	0.0862
Z53 heterozygote	33	0.48	0.288	0.5752
CH11 heterozygote	22	0.82	1.852	0.0056**
DGRP882 hemizygote	37	0.46	0.186	0.7107
Canton S hemizygote	29	0.55	0.556	0.2947
Z30 hemizygote	43	0.84	1.986	0.0004**
Z53 hemizygote	39	0.74	1.413	0.0072**
CH 11 hemizygote	30	0.80	1.735	0.0034**

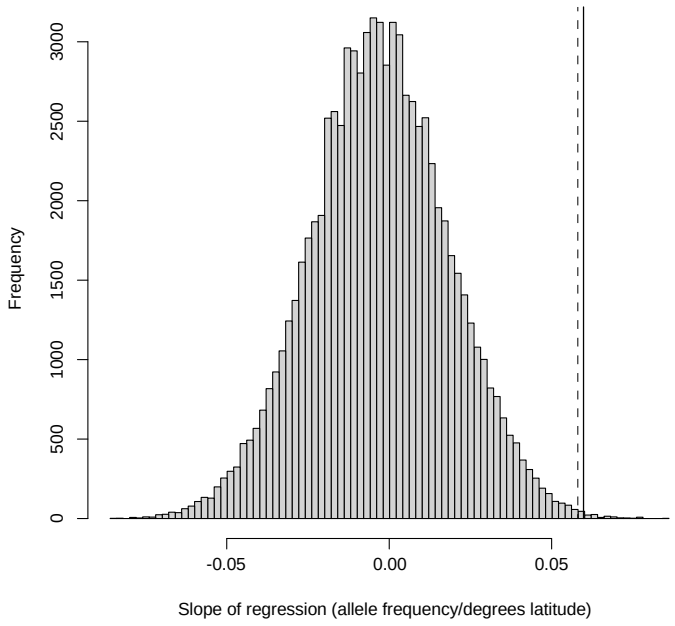

Figure 1. The genes *alan shepard* (*shep*) and *Neuroglian* (*Nrg*) are genomic outliers with a stronger relationship in African allele frequency and latitude in an analysis using genomes from populations collected in the Southeastern United States and the Bahamas (data previously collected in Kao et al 2015). The dashed line corresponds to the value for *shep* and the solid line corresponds to the value for *Nrg*. These regressions are for individual SNPs within these loci that are in the top 1% for slope in this analysis.

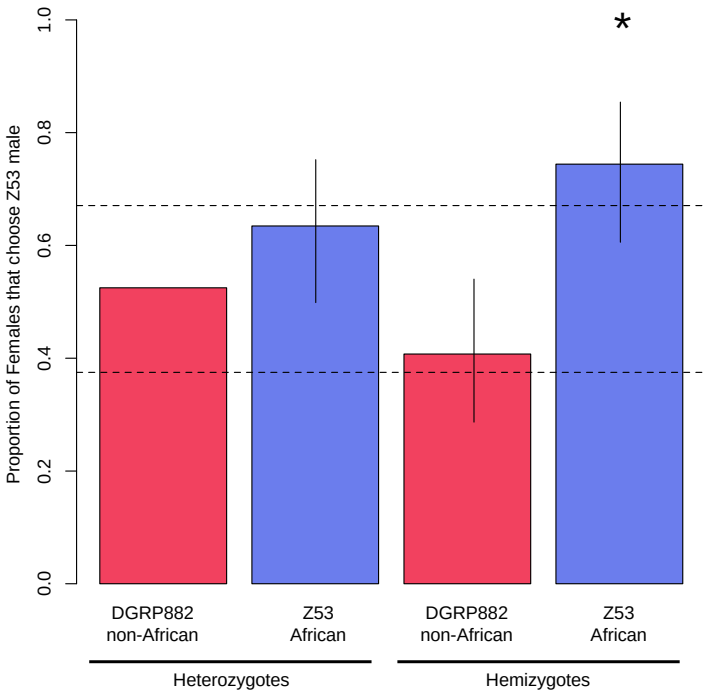

Fig 2. *alan shepard* (*shep*) contributes to African female mate preference. Female flies that are hemizygous and carry only the African allele of *shep* show increased preference for African males, similar to wild-type African females. The horizontal dashed lines are the 95% confidence interval for the baseline genotype, DGRP882 heterozygote. Each remaining genotype was compared to the baseline in a logistic regression where * represents a significant difference from the baseline at $P<0.05$. Confidence intervals are Wilson rank intervals from the binomial distribution.

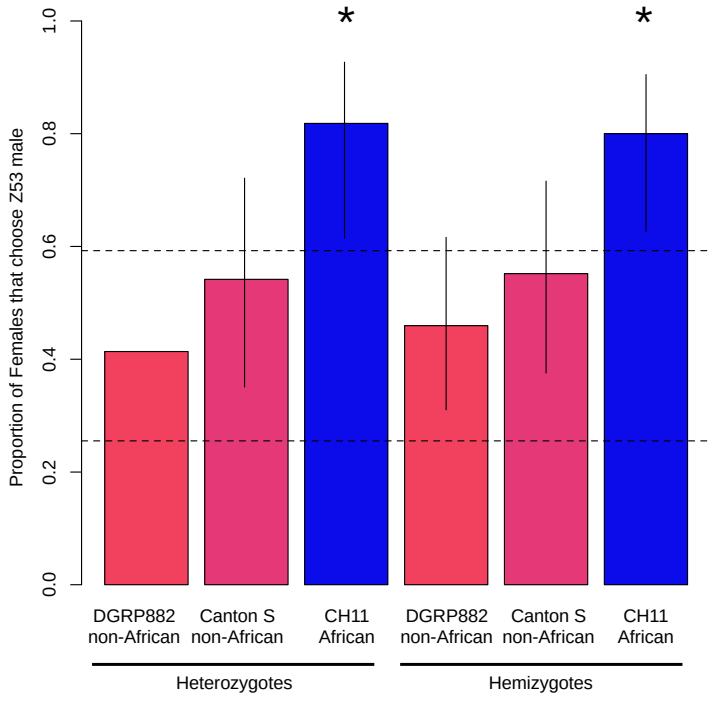

Fig 3. The dominant visible marker, *Bar*, on FM7 does not affect female mate preference for either non-African or African strains. Canton S was included as a control because we expected it to behave similarly to the DGRP882 strain. The CH11 strain was included to test multiple African *Nrg* alleles. Because this strain exhibited a dominant female preference phenotype for both heterozygous and hemizygous females these genotypes were indistinguishable. The horizontal dashed lines are the 95% confidence interval for the baseline DGRP882 heterozygotes. Each remaining genotype was compared to this baseline in a logistic regression where * represents a significant difference from the baseline at $P<0.05$. Confidence intervals are Wilson rank intervals from the binomial distribution.

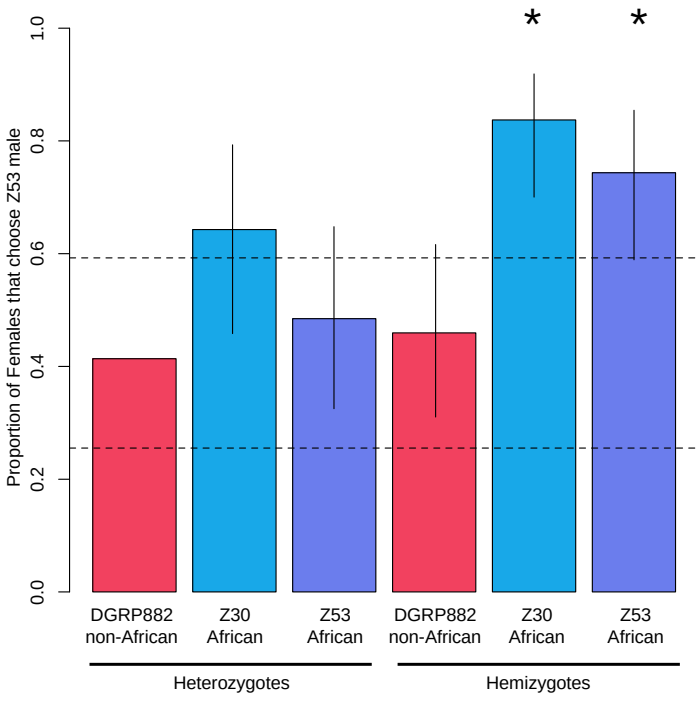

Fig 4. *Neruoglian* (*Nrg*) contributes to African female mate preference. Female flies that are hemizygous and carry only the African allele of *Nrg* show increased preference for African males, similar to wild-type African females. The horizontal dashed lines are the 95% confidence interval for the baseline DGRP882 heterozygote. Each remaining genotype was compared to this baseline with a logistic regression where * represents a significant difference from the baseline at $P<0.05$. Confidence intervals are Wilson rank intervals from the binomial distribution.