

## Research Notes

**Absence of strong signal of background selection affecting nucleotide diversity in *Drosophila pseudoobscura*.**

**Hish, Alexander J., and Mohamed A.F. Noor.** Biology Department, Duke University, Durham, NC 27708 USA; noor@duke.edu.

**Summary**

Nucleotide variation correlates positively with regional rates of recombination in the genomes of many species, and intensive effort has been devoted to deciphering the evolutionary forces driving this association. The effects of sweeps associated with the spread of advantageous mutations and background selection associated with the removal of detrimental mutations are exceedingly hard to decouple. Here, we examine patterns of nucleotide variation in a subset of *Drosophila pseudoobscura* genes that are unlikely to have been affected by one extreme class of selective sweeps, and test for residual signal that may be attributable to background selection. While our statistical power is low due to the number of genes fitting our criteria and small numbers of sequences, the proportion of variation explained by our tests is extremely low and sometimes opposite in direction of predictions of background selection. We tentatively conclude that most of the association of nucleotide variation with recombination rate in this species is probably not attributable to background selection.

**Introduction**

One of the most consistent patterns observed in molecular evolutionary studies of diverse species is that regions of the genome experiencing low recombination also exhibit little sequence variation among individuals (see review in Smukowski and Noor, 2011). Researchers have been seeking the cause of this pattern since it was first observed in *Drosophila melanogaster* over 20 years ago (Aguade *et al.*, 1989; Stephan and Langley, 1989; Begun and Aquadro, 1992). A positive association between recombination rate and nucleotide diversity could result from recombination being mutagenic, for which there is some evidence in yeast with respect to mitotic recombination (Strathern *et al.*, 1995). However, results of detailed studies in *Drosophila* species in particular suggest that much of this pattern is explained by natural selection eliminating variation in regions of low recombination (*e.g.*, Begun and Aquadro, 1992; Begun *et al.*, 2007; McGaugh *et al.*, 2012).

At least two forms of natural selection can explain the positive association between recombination rate and nucleotide diversity. First, selective sweeps reduce variation at closely-linked neutral loci. Alleles linked to the advantageous mutation become fixed along with the spreading mutation, and a single haplotype replaces multiple ancestral haplotypes (Maynard Smith and Haigh, 1974; Kaplan *et al.*, 1989). This effect varies with the degree to which the mutation and neutral loci are linked, since the size of the spreading haplotype will be larger in regions of reduced recombination. An alternative explanation is background selection (BgS), wherein deleterious mutations reduce variation at closely-linked neutral loci because alleles linked to the mutation are eliminated from the population when the mutation is eliminated (Charlesworth *et al.*, 1993). Similar to the effect of a selective sweep, the BgS effect will be amplified in regions of reduced recombination, as the haplotypes being eliminated are larger where recombination is low, therefore reducing the effective population size at more loci.

The similar predictions of the two models have led to disagreement over the relative impact of BgS on reducing genetic diversity (see reviews in Stephan, 2010; Charlesworth, 2013). Unambiguous effects of selective sweeps are documented repeatedly-- researchers observe new haplotypes that spread in local

populations and the associated reductions in variation at nearby sites (*e.g.*, see review in Sabeti *et al.*, 2006). In contrast, although results **consistent with** particular formulations of the BgS model have been observed, unambiguously ascribing a local reduction in nucleotide variation to BgS as distinct from selective sweeps has remained an elusive goal (Innan and Stephan, 2003; Loewe and Charlesworth, 2007; Kaiser and Charlesworth, 2009). For example, a recent analysis of diversity in the human genome found a large reduction in average diversity near conserved sites, but was unable to definitively ascribe these effects to background selection as distinct from possible confounding effects of positive selection (McVicker *et al.*, 2009). Because the selective sweep model predicts a shift of the site frequency spectrum towards low-frequency or rare nucleotide variants, whereas the spectrum produced by BgS models is essentially neutral, much attention has been paid to these types of analyses in distinguishing the models. Nonetheless, efforts using these methods have failed to definitively determine that a particular region of the genome has lost variation as a direct result of BgS.

This study seeks to identify effects of BgS as distinct from effects of selective sweeps using data from *Drosophila pseudoobscura*. Previous work showed that this species exhibits a positive association between recombination rate and nucleotide diversity (Kulathinal *et al.*, 2008; McGaugh *et al.*, 2012). Here, we focused our analyses on genes without fixed differences between *D. pseudoobscura* and its close relative *D. miranda*, and stratified these genes *a priori* based on population genetic parameters that correlate with degree of purifying selection (and thus, presumably, background selection) on their coding regions. We then tested for associations between these correlates of purifying selection on coding sequence and nucleotide variation in introns. Our test is conservative in some respects, because background selection can occur in genes that bear fixed sequence differences between related species, but our focus on such genes excludes at least one class of selective sweeps ("hard, complete sweeps", strong selection driving new mutations to rapid fixation; see Pritchard *et al.*, 2010).

## Methods

### Candidate Gene Identification

To seek possible effects of BgS, we identified genes in the *D. pseudoobscura* genome that were unlikely to have been affected by "hard, complete" selective sweeps. We focused our analyses on genes bearing one intron less than 1000 bp in length, because linkage disequilibrium (and thus likely effects of BgS) decays rapidly in *Drosophila*, even in regions of low recombination (*e.g.*, Langley *et al.*, 2000).

We used aligned coding sequence data from 11 *D. pseudoobscura* and 3 *D. miranda* individuals (sequence data downloaded from Pseudobase: <http://pseudobase.biology.duke.edu/>; McGaugh *et al.*, 2012; McGaugh and Noor, 2012) to identify all genes with no fixed differences between the two species in coding regions or introns. This limitation on fixed differences precludes the possibility that a novel advantageous mutation arose and spread to fixation in the lineage leading to *D. pseudoobscura* after divergence from *D. miranda*. To increase sample size, 15 genes were included in the study that contained small numbers of fixed differences between *D. pseudoobscura* and *D. miranda*. In this subset, all fixed differences were inferred to have arisen in *D. miranda* via parsimony: the *D. pseudoobscura* variant was shared with the outgroup species *D. lowei*. Therefore, similar to the case discussed above, a recent, hard, complete selective sweep in *D. pseudoobscura* was unlikely to have occurred in these genes. The final set of 43 genes fitting these criteria, along with their FlyBase identifiers and associated statistics, is presented in a Supplementary Table in Dryad (<http://datadryad.org/>).

### Statistical Analyses

For the final suite of candidate genes, we used DnaSP (Librado and Rozas, 2009) to obtain various population genetic parameters including the average pairwise difference in sequence within introns among individuals within species (intron  $\pi$ ), intron Tajima's D (Tajima, 1989),  $d_{xy}$  (Nei, 1987) between *D. pseudoobscura* and distant outgroup species *D. lowei* in the coding region, and Ka/Ks (Nei and Gojobori, 1986) between *D. pseudoobscura* and *D. lowei* in the coding region. We then performed a series of regressions to test if the data supported patterns predicted the BgS model.

## Results

We predicted that genes that experienced BgS would have reduced intronic nucleotide diversity (low intron  $\pi$ ) and disproportionately negative Tajima's D (indicating population size expansion and/or purifying selection). We found that the average values of intron  $\pi$  from our candidate genes were close to those estimated for the genome as a whole (autosomes: 0.016 for candidate genes vs. 0.015 for whole genome). Most genes in *D. pseudoobscura* have a negative Tajima's D (Moriyama and Powell, 1996; Hamblin and Aquadro, 1999; Machado *et al.*, 2002), so we predicted a **disproportionately** negative Tajima's D, relative to other *D. pseudoobscura* genes, in those genes experiencing BgS (Charlesworth *et al.*, 1993; Charlesworth *et al.*, 1995). We tested the association of each of these metrics to three parameters correlated with the expected degree of BgS: Ka/Ks,  $d_{xy}$ , and recombination.

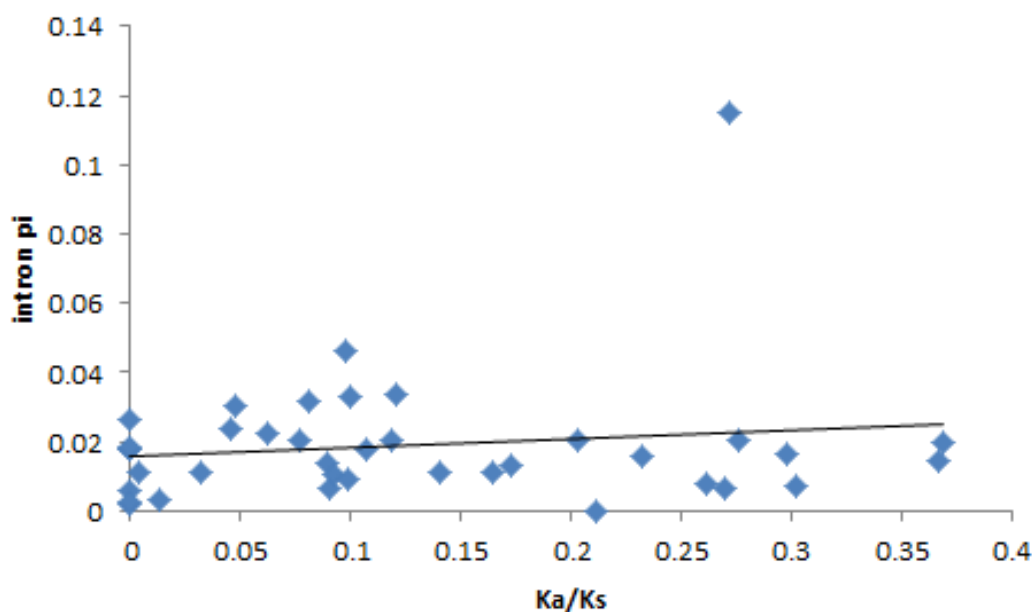


Figure 1. Correlation of intron  $\pi$  with coding region Ka/Ks (N = 38).

First, we tested for associations between intronic nucleotide diversity and Ka/Ks (Nei and Gojobori, 1986) between *D. pseudoobscura* and outgroup *D. lowei*. Ka/Ks provides a measure of evolutionary constraint on the coding region, with lower values indicating greater constraint and potentially greater background selection. Therefore, BgS predicts low values of intronic diversity to be associated with low values of Ka/Ks. We excluded genes with high ( $>1$ , 1 gene excluded) or moderate ( $>0.5$ , 4 genes excluded) values of Ka/Ks so as to limit possible effects of positive selection. Our data follows the trend predicted by BgS, but is not statistically significant ( $r^2 = 0.02$ ,  $p = 0.38$ , Figure 1). Removing the one extreme outlier with  $\pi = 0.11$  made the trend negative, opposite in prediction to BgS. Results were qualitatively similar when  $\log(\text{Ka/Ks})$  was used and when X-chromosomal and autosomal genes were split for separate analysis. We also tested for an association of Tajima's D of intron sequences to Ka/Ks, and again, all associations were weak and not statistically significant (see table in Dryad: <http://datadryad.org/>).

Second, we tested for associations between intronic diversity and the average number of nucleotide substitutions per coding region site between *D. pseudoobscura* and *D. lowei* ( $d_{xy}$ ; see Nei, 1987). Similar to Ka/Ks,  $d_{xy}$  potentially provides a measure of evolutionary constraint (low values indicating sequence similarity over a long evolutionary timespan), but  $d_{xy}$  estimates conservation more directly without *a priori* assumptions about synonymous or nonsynonymous differences. Again, BgS predicts low values of intronic diversity to be associated with low values of  $d_{xy}$  (indicating high constraint). As above, results followed the trend predicted

by BgS, but very weakly and with low statistical support ( $r^2 = 0.02$ ,  $p = 0.42$ , Figure 2). Removing the one extreme outlier with  $\pi = 0.11$  again made the trend negative, opposite in prediction to BgS. Results were qualitatively similar again for the association between Tajima's D and  $d_{xy}$ , and when the X-chromosomal and autosomal genes were split for analyses (see data in Dryad: <http://datadryad.org/>).

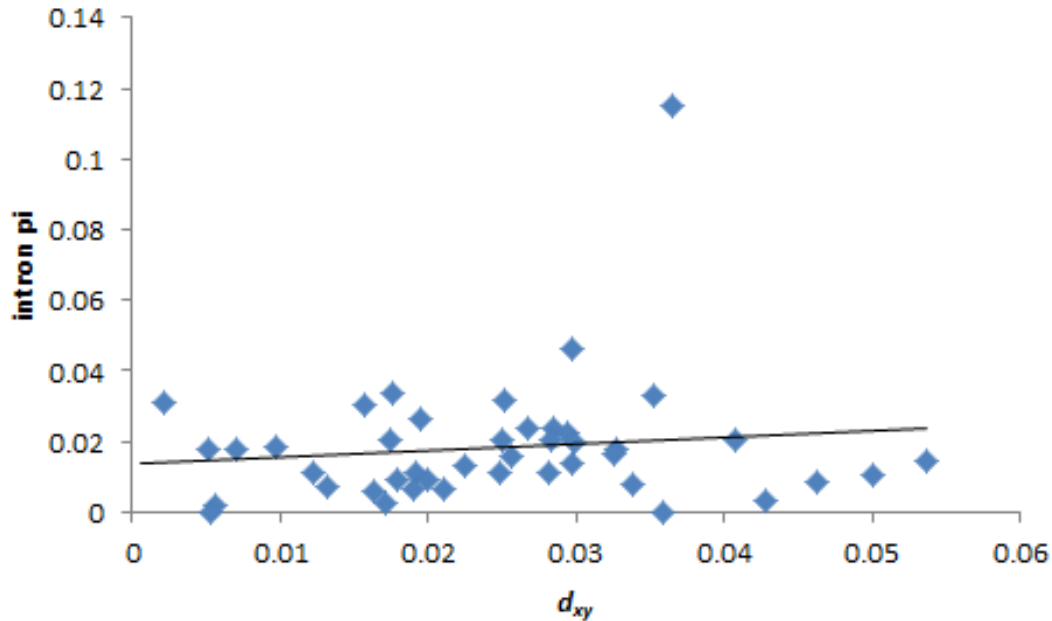


Figure 2. Correlation of intron  $\pi$  with coding region  $d_{xy}$  ( $N = 43$ ).

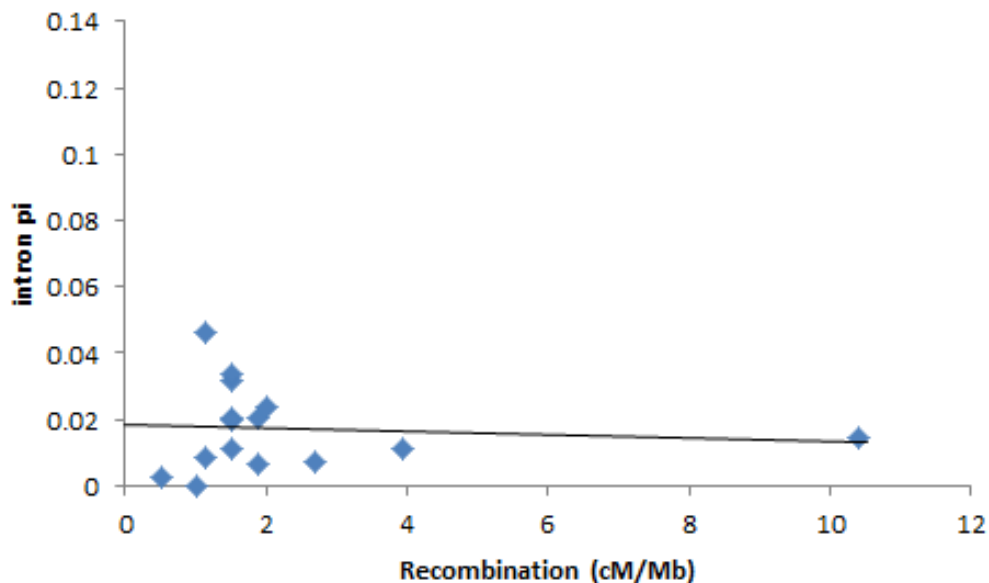


Figure 3. Correlation of intron  $\pi$  with recombination rate for X-chromosome genes ( $N = 15$ ).

Finally, we tested for associations between intronic diversity and recombination rate (measured in Kosambi cM/Mb). The recombination datasets for *D. pseudoobscura* consist of two very similar

recombination maps (McGaugh *et al.*, 2012), and we used the average of the values from the two maps. Assuming similar density of targets for purifying selection, background selection should be most intense in regions of low recombination (Charlesworth *et al.*, 1993), and so we predict a decrease in neutral variation (*i.e.*, intronic diversity) with a decrease in recombination rate. Instead, we observed a non-significant, negative association between recombination rate and intronic diversity, both for X-chromosomal ( $r^2 = 0.01$ ,  $p = 0.75$ , Figure 3) and autosomal genes ( $r^2 = 0.19$ ,  $p = 0.38$ ). The association between recombination rate and Tajima's D was also non-significant ( $r^2 = 0.12$ ,  $p = 0.14$ ).

## Discussion

Our efforts failed to identify statistically significant, or even suggestive, evidence that background selection (BgS) reduced a detectable amount of variation in the 43 genes studied in *Drosophila pseudoobscura*. There is a strong correlation of recombination rate to diversity in *D. pseudoobscura* and many other species (*e.g.*, Begun and Aquadro, 1992; Hellmann *et al.*, 2003; Roselius *et al.*, 2005; Smukowski and Noor, 2011; McGaugh *et al.*, 2012); however, there is persuasive evidence that much of this correlation is caused by selective sweeps or other forms of directional selection (Sella *et al.*, 2009; Slotte *et al.*, 2010; McGaugh *et al.*, 2012). Therefore, it is possible that positive selection explains most of this correlation, and the contribution of BgS may be quite small.

While we focus our analyses on *D. pseudoobscura*, its genome may be more weakly affected by BgS than other genomes because of its high recombination rate. The average recombination rate in *D. pseudoobscura* is roughly four-fold higher than that in *D. melanogaster* (Ortiz-Barrientos *et al.*, 2006), and higher still than in mammals. Since large effects of BgS are not expected in species with high recombination rates (Nordborg *et al.*, 1996), our result may not be unexpected. Further, *D. pseudoobscura* has a very large effective population size (Schaeffer, 1995), and this may minimize effects of BgS on Tajima's D or other metrics (Charlesworth *et al.*, 1993; Charlesworth *et al.*, 1995).

Conversely, our tests were conservative in many respects, and we may therefore lack power to identify BgS. First, our sample size of genes was greatly limited by restricting the analysis to those genes with a single intron and no derived fixed differences between species. Although our sample size was larger than several classic studies which detected associations of recombination rate with nucleotide diversity (Begun and Aquadro, 1992 surveyed only 20 genes), effects of BgS may be often more elusive than those of selective sweeps, except in regions of severely reduced recombination (Charlesworth *et al.*, 1993; Charlesworth *et al.*, 1995). In addition to a small number of genes, our estimates of intron  $\pi$  and Tajima's D were based on a limited number of available sequences (McGaugh *et al.*, 2012), some of which were also short sequences, and thus the estimates may be imprecise or insufficient for detection of BgS effects (Charlesworth *et al.*, 1995).

Second, beyond reducing our sample size and statistical power, our limitation to genes with no fixed interspecies differences may slightly bias us against detecting effects of BgS. BgS can reduce the effective population size of a region, and thus potentially facilitate weakly disadvantageous variants going to fixation. As such, we may have excluded some genes because they bore fixed differences, when those differences spread in part as a result of BgS.

Had our approach detected a signature consistent with BgS, we would have needed to address the possibility that partial sweeps contributed to the results observed. Recent examinations have found evidence that "soft" or "partial" sweeps may be quite common (Messer and Petrov, 2013). Similarly, we cannot exclude possible selective sweeps outside the genes but near them, though LD decays rapidly in *Drosophila*, even in regions of low recombination (Langley *et al.*, 2000). While we cannot rule out abundant soft or partial sweeps in *D. pseudoobscura*, partial sweeps are predicted to produce positive Tajima's D measures, whereas genome wide patterns of variation in *D. pseudoobscura* exhibit a negative Tajima's D (Moriyama and Powell, 1996; Hamblin and Aquadro, 1999; Machado *et al.*, 2002).

Although multiple instances of selective sweeps have been identified, unambiguously demonstrating effects of background selection as distinct from effects of selective sweeps remains an ongoing, yet elusive, goal. We hope that new approaches, even if conservative, may provide further insight into the potential impact of BgS on genome wide or gene-specific patterns of nucleotide variation.

**Acknowledgments:** This research was performed as part of an undergraduate independent study fellowship to AJH. We thank S. McGaugh for sharing her data and for a first-pass at identifying candidate genes bearing no fixed differences. We also thank C. Smukowski Heil for providing an estimate of pi for the autosomes, and comments on the manuscript.

**References:** Aguade, M., N. Miyashita, and C.H. Langley 1989, *Genetics* 122: 607-615; Begun, D.J., and C.F. Aquadro 1992, *Nature* 356: 519-520; Begun, D.J., *et al.*, 2007, *PLoS Biology* 5: e310; Charlesworth, B., 2013, *Journal of Heredity* 104: 161-171; Charlesworth, B., M.T. Morgan, and D. Charlesworth 1993, *Genetics* 134: 1289-1303; Charlesworth, D., B. Charlesworth, and M.T. Morgan 1995, *Genetics* 141: 1619-1632; Hamblin, M.T., and C.F. Aquadro 1999, *Genetics* 153: 859-869; Hellmann, I., *et al.*, 2003, *American Journal of Human Genetics* 72: 1527-1535; Innan, H., and W. Stephan 2003, *Genetics* 165: 2307-2312; Kaiser, V.B., and B. Charlesworth 2009, *Trends in Genetics* 25: 9-12; Kaplan, N.L., R.R. Hudson, and C.H. Langley 1989, *Genetics* 123: 887-899; Kulathinal, R.J., S.M. Bennett, C.L. Fitzpatrick, and M.A.F. Noor 2008, *Proceedings of the National Academy of Sciences USA* 105: 10051-10056; Langley, C.H., *et al.*, 2000, *Genetics* 156: 1837-1852; Librado, P., and J. Rozas 2009, *Bioinformatics* 25: 1451-1452; Loewe, L., and B. Charlesworth 2007, *Genetics* 175: 1381-1393; Machado, C.A., R.M. Kliman, J.A. Markert, and J. Hey 2002, *Mol. Biol. Evol.* 19: 472-488; Maynard Smith, J., and J. Haigh 1974, *Genet. Res.* 23: 23-35; McGaugh, S.E., *et al.*, 2012, *PLoS Biology* 10: e1001422; McGaugh, S.E., and M.A.F. Noor 2012, *Philosophical Transactions of the Royal Society of London Series B* 367: 422-429; McVicker, G., D. Gordon, C. Davis, and P. Green 2009, *PLoS Genetics* 5: e1000471; Messer, P.W., and D.A. Petrov 2013, *Trends Ecol. Evol.* 28: 659-669; Moriyama, E.N., and J.R. Powell 1996, *Mol. Biol. Evol.* 13: 261-277; Nei, M., 1987, *Molecular Evolutionary Genetics*. Columbia University Press; Nei, M., and T. Gojobori 1986, *Mol. Biol. Evol.* 3: 418-426; Nordborg, M., B. Charlesworth, and D. Charlesworth 1996, *Genet. Res., Camb.* 67: 159-174; Ortiz-Barrientos, D., A.S. Chang, and M.A.F. Noor 2006, *Genet. Res., Camb.* 87: 23-31; Pritchard, J.K., J.K. Pickrell, and G. Coop 2010, *Curr. Biol.* 20: R208-215; Roselius, K., W. Stephan, and T. Stadler 2005, *Genetics* 171: 753-763; Sabeti, P.C., *et al.*, 2006, *Science* 312: 1614-1620; Schaeffer, S., 1995, Pp. 329-352 in L. Levine, ed. *Genetics of Natural Populations: The Continuing Importance of Theodosius Dobzhansky*. Columbia University Press, New York; Sella, G., D.A. Petrov, M. Przeworski, and P. Andolfatto 2009, *PLoS Genetics* 5: e1000495; Slotte, T., J.P. Foxe, K.M. Hazzouri, and S.I. Wright 2010, *Mol. Biol. Evol.* 27: 1813-1821; Smukowski, C.S., and M.A. Noor 2011, *Heredity* 107: 496-508; Stephan, W., 2010, *Phil. Trans. Roy. Soc. Lond. B* 365: 1245-1253; Stephan, W., and C.H. Langley 1989, *Genetics* 121: 89-99; Strathern, J.N., B.K. Shafer, and C.B. McGill 1995, *Genetics* 140: 965-972; Tajima, F., 1989, *Genetics* 123: 585-595.



### **Antistress ability of *Myristica fragrans* (Japatrae) a nutmeg to detoxify reactive oxygen species in stress-induced *Drosophila melanogaster*.**

**Riya, Peter, H.B. Vasanth Patil, and B.Y. Sathish Kumar\*.** Post Graduate Department of Bio-technology, JSS College (An Autonomous College of University of Mysore), Ooty Road, Mysore 570 025, Karnataka, India; E-mail: [bysathish@gmail.com](mailto:bysathish@gmail.com).

### **Abstract**

Stress describes a positive or negative condition, which has an impact on an organism's genome, transcriptome, proteome, and phenome well-being. Origin of stress may vary but its effect is deleterious. The anti-stress property of *Myristica fragrans* (Leaf powder) in combating stress in stress-induced *D. melanogaster* a fruit fly were experimented where four groups of flies were reared simultaneously. The Control flies reared on normal media followed by media containing MTX (Second group), the third group on the media containing MTX and 0.5 gm of japatrae, finally flies on only 0.5 gm of japatrae in media. Then the flies were assayed for stress related marker enzymes like SOD, CAT, and GPx. Reduction in level of ROS by *Myristica fragrans* in stress-induced fruit flies has increased the ability to scavenge them and lowering the