GENOME SIZE VARIATION IN *D. MELANOGASTER*

A Senior Scholar Thesis

by

BEN ALFREJD

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

April 2011

Major: Biomedical Science

GENOME SIZE VARIATION IN *D. MELANOGASTER*

A Senior Scholar Thesis

by

BEN ALFREJD

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by:

Research Advisor: J. Spencer Johnston
Director for Honors and Undergraduate Research: Sumana Datta

April 2011

Major: Biomedical Science

ABSTRACT

Genome Size Variation in *D. Melanogaster*.
(April 2011)

Ben Alfrejd
Biomedical Science
Texas A&M University

Research Advisor: Dr. J. Spencer Johnston
Department of Entomology

As yet significant portions of the genetic variation in complex traits have not been explained with genome wide association experiments; and this has led to the search for the "missing heritability". Our data support the hypothesis that variation in genome size may account for some of the missing heritability. We measured female genome sizes for 34 *Drosophila melanogaster* inbred strains that derived from isofemale lines established from a natural population in Raleigh, NC, in addition to a group of 40 strains artificially selected for increased and decreased body size. We provide the first evidence that significant intraspecific genome size variation exists among these *Drosophila melanogaster* lines and that selection has a downsizing effect on the extent of variation.

DEDICATION

To Dr. J Spencer Johnston, without whom this would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Microevolutionary theory, according to Allen Templeton, states that selection, natural or artificial, cannot act upon an organism without genetic variation among the individuals (Templeton 2006). Such required variation can be due to differently encoded genes, gene regulation, or as we hypothesize here may take the form of genomic size differences among populations. But does variation in genome size exist within a *Drosophila* species?

We know that the genomes of two distinct *Drosophila* species can vary widely in size (Gregory 2011). Within the Genus *Drosophila*, genome size can vary up to three-fold (Bosco et al. 2008); and even among the most closely related species in the Drosophila genus, significant genome size variation exists (Gregory & Johnston 2008; Bosco et al 2007; Vieira et al. 2002), such that even the very closely related Drosophila in the melanogaster subgroup show a two-fold variation of genome size (J. S. Johnston, unpublished,).

This thesis follows Molecular Biology and Evolution.

We also know what genetic changes produce genome size change. A major contributor to a significant portion of the genome size variation is the transposable element content in each species (Petrov et al. 1995). With the availability of genomic data from *Drosophila* species, tables have been prepared that give the types of transposable elements in different *Drosophila* species and the relative number of copies of each element differ between species (Drosophila 12 genomes consortium, 2007). There are differences, but these are not the whole story of genome size variation. Variation also exists for satellite DNA (Bosco et al. 2007), regulatory regions (Kim et al., 2009) and microRNA's (Grun et al. 2005) and within and between specific genes (Stage and Eickbush, 2007).

What we don't know is the role of selection in production of this variation. What is not observed is the intraspecific variation upon which selection can act to produce these differences (Bennett et al. 2008).

We hypothesize that selection has constrained genome size within *D. melanogaster*. To test this we compare inbred lines produced by very different means in two populations and accordingly, we determine the extent of conspecific genome size variation. One of these lines was established as 34 different isolines (single female isolate) where any differences should be due to chance and another line consists of 40 inbred strains strongly selected for large and small body size, where any differences are those

associated with selection. We find decreased variation among size selected populations as opposed to the unselected inbred isolines.

CHAPTER II

METHODS

We examined 34 *D. melanogaster* strains obtained from the Drosophila Population Genomics Project (http://www.dpgp.org/) (R208, R303, R304, R306, R307, R313, R315, R324, R335, R357, R358, R360, R365, R375, R379, R380, R399, R427, R437, R486, R517, R555, R639, R707, R712, R714, R732, R765, R774, R799, R786, R820, R852, and R859). Stocks were maintained at room temperature on Bloomington's standard medium (The Bloomington Drosophila Stock Center, Indiana University).

Additionally, we examined 40 *D. melanogaster* strains selected for size (L1-B L1-FF, L1-G, L1-H, L1-L, L1-MM, L1-P, L1-SS, L1-V, L1-Y, L2-C, L2-E, L2-F, L2-GG, L2-H, L2-K, L2-L, L2-LL, L2-O, L2-TT, S1-AA, S1-BB, S1-G, S1-M, S1-OO, S1-TT, S1-VV, S1-W, S1-X, S1-YY, S2-E, S2-EE, S2-F, S2-FF, S2-J, S2-NN, S2-OO, S2-Q, S2-R, S2-V).

Flow cytometric geneome size determination. *D. melanogaster* strains were compared against a *D. virilis* standard strain (1C = 328 Mb), by collecting the cephalic nuclei of both in 1 mL of Galbraiths buffer at approximately 2 C^o and 200 µl probidium idodide. The samples were then analyzed by a flow cytometer to 7,000 events each (Viera et al 2002). Flow cytometry, which is commonly used in the medical field and in plant biology, provides an accurate determination of differences in genome size (Viera et al

2002). The means were determined by fluorescence from each diploid fly nuclei sample analyzed and used to calculate the *Drosophila* genome size of each sample (Viera et al 2002).

CHAPTER III

RESULTS

The 34 strains of *Drosophila melanogaster* scored for genome size showed significant intraspecific female genome size variation. The 5 strains with the largest female genomes were R306, R380, R427, R517, R774; the 5 strains with the smallest female genomes were R208, R313, R335, R379, R786 (Table 1, Figure 1). The range for the means of the female isolines is 11.35 Megabases with the smallest genome measuring 170.3576 Megabases and the largest measuring 181.7114 Megabases. The range for these strains is further underscored by the tight fitted regression line, R^2=.8731, although the two strains of R517 and R208 appear to be moderate large and small outliers respectively (Figure 1).

Table 1. Intraspecific & sexually dimorphic genome size variation in the 10 selected strains

Strain	N (female)	Avg Female Genome Size (Mb)	s.e.	Significance Group(s) (female)	N (male)	Avg Genome Size - 175	T-test (to 175) p-value
R517	11	181.64	0.33	A	5	6.711	0.00000002
R774	3	178.14	0.09	B	3	3.137	0.0008
R380	6	177.10	0.79	BCD	7	2.124	0.003
R427	8	177.00	0.22	BCD	8	1.885	0.00002
R306	4	175.18	0.15	BCD	3	2.143	0.32
R379	3	172.45	0.19	GHIJKL	3	-2.545	0.006
R335	3	172.31	0.41	HIJKL	3	-2.688	0.02
R786	5	172.22	0.48	HIJKL	7	-2.782	0.004
R313	3	172.19	0.72	HIJKL	3	-3.844	0.06
R208	8	169.65	0.38	L	4	-4.642	0.003

Of the 40 body sized *D. melanogaster* strains, the 5 largest female genomes were S1-VV, S2-F, S2-J, S2-FF, L1-FF; the 5 smallest genomes were S1-G, S2-OO, S2-E, L1-Y, L1-MM (Figure 2). The range of means for the body size selected females is smaller at 8.8471 Megabases, and in fact the entire range of means is significantly higher overall with the smallest genome being 175.5728 Megabases and the largest genome 184.4199 Megabases. The linear regression line for this range is even more tightly fitted, R=.9538 with no apparent outliers but 2 and 3 values at the low and high ends, respectively (Figure 2).

Figure 1. Linear regression of mean genome size estimates of 34 inbred isolines.

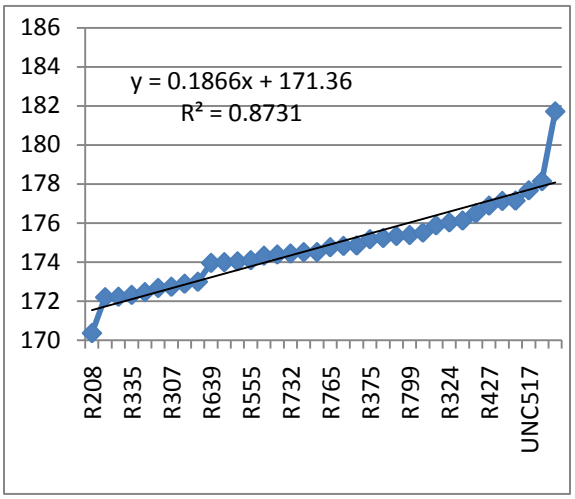
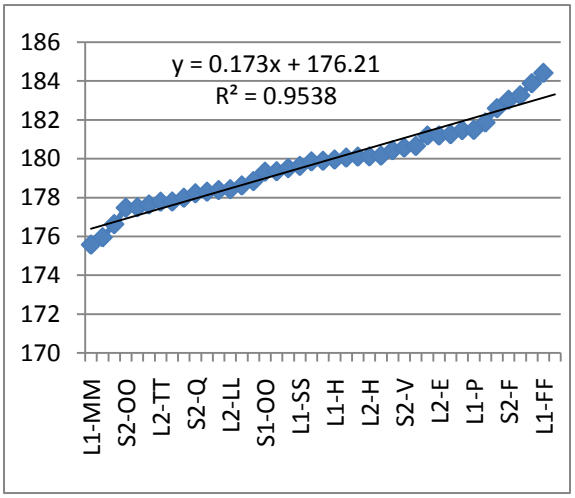


Figure 2. Linear regression of mean genome size estimates of size selected *melanogaster*.



CHAPTER IV

DISCUSSION AND SUMMARY

We set out to establish a relationship between selection and genetic variation. The fact that there is a range of different genome sizes in the inbred lines indicates that there is significant genetic variation within a natural population. The two extreme mean genome sizes, R517 and R208, provide obvious and noteworthy variation that extends the range of sizes. Such variation should be taken into account in future studies dealing with genomic research.

To test for any relationship between genome size and selection, we examined the nuclei of 40 *melanogaster* strains selected for size, both large and small. Reason should dictate that genome size would be affected by selection (Gregory 2003) in a directly proportional fashion because nuclear volume and thus cell size is correlated with genome size. Yet the range of variation is significantly constricted within the selected strains when compared to the inbred isolines. The amount of cells may have increased in the larger bodied flies, but not the genome.

Further, we know there are forces that change genome size, transposons, unequal crossing-over, duplication and deletions. So we don't have an answer why selection did not change genome size. We know there is variation for genome size. But we find that

it does not respond quickly to selection for a correlated character (size). That still begs the question why selection opposes genome size change.

Genome size varies across organisms by many times. Even within arthropods, the genome size may vary from 93 Mb is a 2-sposted mite to more than 16000 Mb in a grasshopper. What is surprising is that this variation is not simply related to any life history trait. There are a few general trends. Insects that are holometabolous that go through metamorphoses, such as flies that go from larvae through pupae to adults, tend to have genomes smaller than 2000 Mb. Yet those that are ametabolous developing by moults, such as grasshoppers and aphids, do not all have enormous genomes. The aphids have a genome around 600 mb, while the body louse has a genome of 100 mMB. Genome size is measured as 1C, the amount of DNA in a gamete, and the lack of relationship of life history and genome size is called the C-value enigma. It is an enigma, because we know how the genome size changes (transposons, microsatellitte DNA, et cetera) but we don not know the evolutionary forces that shape that change. We measured genome sizes of females from 34 *Drosophila melanogaster* strains and show that significant variation exists. We have observed significant variation in genome sizes among sequenced *Drosophila melanogaster* strains. These results indicate that a portion of "missing heritability" observed in genome wide association studies may be due to the failure to account for the effect of genome size when attempting to map genome size correlated phenotypes. These basic observations indicate that studies of

genome size can contribute to identifying “missing heritability” in genome wide association studies.

LITERATURE CITED

Bennett MD, Price HJ, and Johnston JS (2008) Anthocyanin inhibits propidium iodide DNA fluorescence in Euphorbia pulcherrima: Implications for genome size variation and flow cytometry. Annals of Botany **101**: 777-790.

Bosco G, Campbell P, Leiva-Neto JT, and Markow TA (2007) Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics **177**: 1277-1290.

Clark, AG. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203-218

Gregory TR (2003) Is small indel bias a determinant of genome size? TRENDS in Genetics **Vol.19** No.9

Gregory TR, Johnston JS (2008) Genome size diversity in the family Drosophilidae. *Heredity* **101**: 228-238.

Gregory TR (2011). Animal genome size database. http://www.genomesize.com.

Grün, D. , Wang, Y.L. , Langenberger, D. , Gunsalus, K.C. & Rajewsky, N. (2005) MicroRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput. Biol.* **1**, e13.

Hanrahan, Shawn J., Johnston, J. Spencer. 2011. New genome size estimates of 141 species of Arthropods. *Chromosome Research*. In Press

Johnston, J. S., K. S.Yoon, J. P. Strychatz, B. R. Pittendrigh, and J. M. Clark. 2007. Body lice and head lice, (Anopleura: Pediculidae) have the smallest genomes of any hemimetabolous insect reported to date. *Journal of Medical Entomology* **44:** 1109-1112.

Kim, J, X. He and S. Sinha, (2009). Evolution of regulatory sequences in 12 Drosophila species, *PLoS Genet* **5**, p. e1000330.

Stage, D. E. and T. H. Eickbush. Sequence variation within the rRNA gene loci of 12 *Drosophila* species. *Genome Res* 2007, **17**(12):1888-1897.

Templeton, A. 2006. Population Genetics and Microevolutionary Theory. Wiley-Liss, Hoboken, NJ. 705pp

Vieira C, Nardon C, Arpin C, Lepetit D, and Biémont C (2002) Evolution of genome size in Drosophila. Is the invader's genome being invaded by transposable elements? Mol Biol Evol **19**: 1154–1161.

CONTACT INFORMATION

Name:	Ben Alfrejd
Professional Address:	c/o Dr. Spencer Johnston Department of Entomology 412 Minnie Belle Heep Bldg. 2475 TAMU College Station, TX 77843
Email Address:	b_alfrejd@msn.com
Education:	B.S., Biomedical Science, Texas A&M University, May 2011 Undergraduate Research Scholar