Environmental factors have little influence on drosophilid genome size.

Samantha P. Curnow, Carl E. Hjelmen
Department of Biology, Utah Valley University

Introduction

Genome size (GS) varies widely among organisms. Within *Drosophila*, species vary from 130 million bases to almost 400 million bases (**Fig. 1**)[1].

In some organisms, genome size has been known to **correlate to** characteristics which are impacted by **temperature**[2].

We wanted to know: **is there was a relationship between genome size and climatic variables in *Drosophila* species**

We utilize **previously published data** and **comparative phylogenetic analyses** to investigate this question

Fig. 1: **Genome size** plotted as color on *Drosophila* phylogeny. Dark colors correspond to smaller genome sizes, light colors correspond to larger genome sizes

Where did we get this data?

The data for genome sizes were provided by Hjelmen et al.[1]

The data for climate factors were initially extracted from a variety of sources, including publicly accessible geographic data and temperature data from two papers by Kellermann et al.[3,4]

How did we analyze this data?

The ape[5], caper[6], and geiger[7,8] packages were used in RStudio to analyze data via PGLS (**Table 1**).

Data visualizations were produced using phytools[9], ggplot2[10], and ggpubr[11] packages in RStudio (**Fig. 2**).

What do we find?

PGLS model (y ~ x)	Standard Error	p value	R^2	Adjusted R^2
Female Genome Size ~ Critical Temperature Maximum	8.26	0.14	0.09	0.05
Female Genome Size ~ Critical Temperature Minimum	2.89	0.73	0.002	-0.01
Female Genome Size ~ Precipitation	0.01	0.37	0.01	-0.003
Female Genome Size ~ Temperature Maximum	3.07	0.64	0.003	-0.01
Female Genome Size ~ Temperature Minimum	0.97	0.52	0.006	-0.009
Female Genome Size ~ Haploid Number	8.95	0.72	0.0009	-0.006
Female Genome Size ~ Egg to adult development time (days)	2.28	0.42	0.007	-0.003
Female Genome Size ~ Geographic Area	6.06E-04	0.69	0.002	-0.01
Female Genome Size ~ Latitude Range	0.15	0.11	0.02	0.01
Female Genome Size ~ Longitude Range	0.04	0.1	0.02	0.01

Table 1: **Phylogenetic generalized least squares** (PGLS) data for genome size of female drosophilid data analyzed with environmental factors. Based on these analyses, **no significant relationship** between climatic variables and genome size.
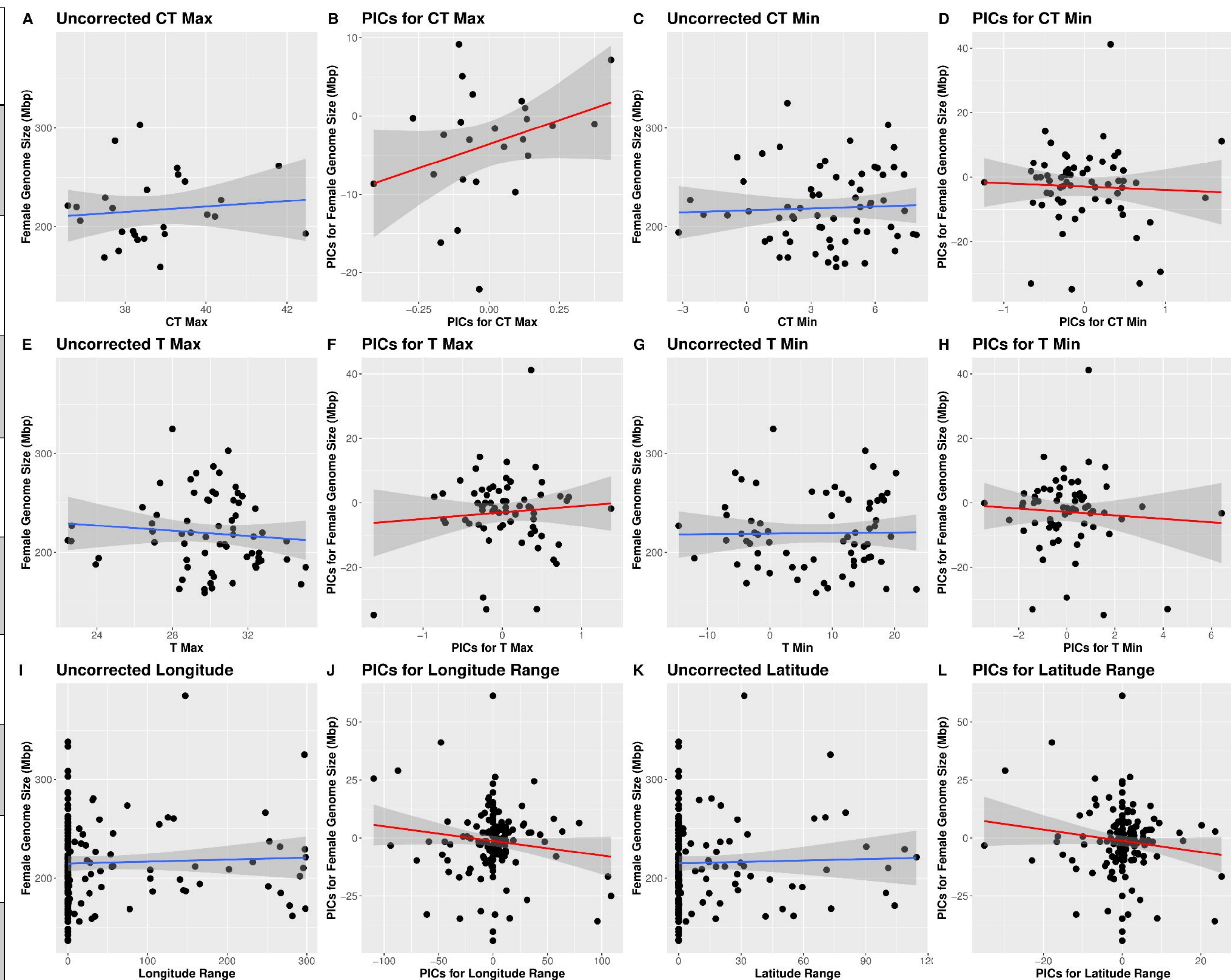

Fig 2: **Data visualizations** for the analyses for female genome size and **critical temperature** (maximum and minimum), **extremes of temperature** values (maximum and minimum), longitude range, and latitude range.

What does this mean?

Given our data in *Drosophila*, we find **no significant relationship** with climatic variables and GS (**Table 1 and Fig 2**)

The data used in our analyses and data visualizations is limited (for the number of species, environmental variables, and accuracy of geographic distribution data).

There may still be a correlation, but as of yet is undetected by our data and methods.

Where do we go from here?

More data may reveal a relationship between genome size and environmental factors.

Such a pattern may be elucidated by:
- Using data from **more species**
- Running similar analyses on **other groups**
- Obtaining highly **reliable geographic data**
- Testing genomes of a region against **climate data over decades**

References:
1. Hjelmen *et al.* (2019) *G3*, 9: 3167-3179
2. Hessen *et al.* (2013) *Biological Reviews* 88: 476-489;
3. Kellermann *et al.* (2012) *Proceedings of the National Academy of Sciences of the United States of America*, 109(40), 16228-16233
4. Kellermann *et al.* (2012) *Evolution: International Journal of Organic Evolution*, 106(11), 3377-3389
5. Paradis & Schliep (2019) Bioinformatics 35: 526-528
6. Orme *et al. (2018)*. R package version 1.0.1.
7. Harmon *et al.* (2008). Bioinformatics 24: 129-131
8. Pennell *et al.* 2014. Bioinformatics 30: 2216-2218
9. Revell (2012). *Methods in Ecology and Evolution*. 3: 217-223
10. Wickham (2016). Springer-Verlaug, New York
11. Kassambara (2020). R Package version 0.4.0