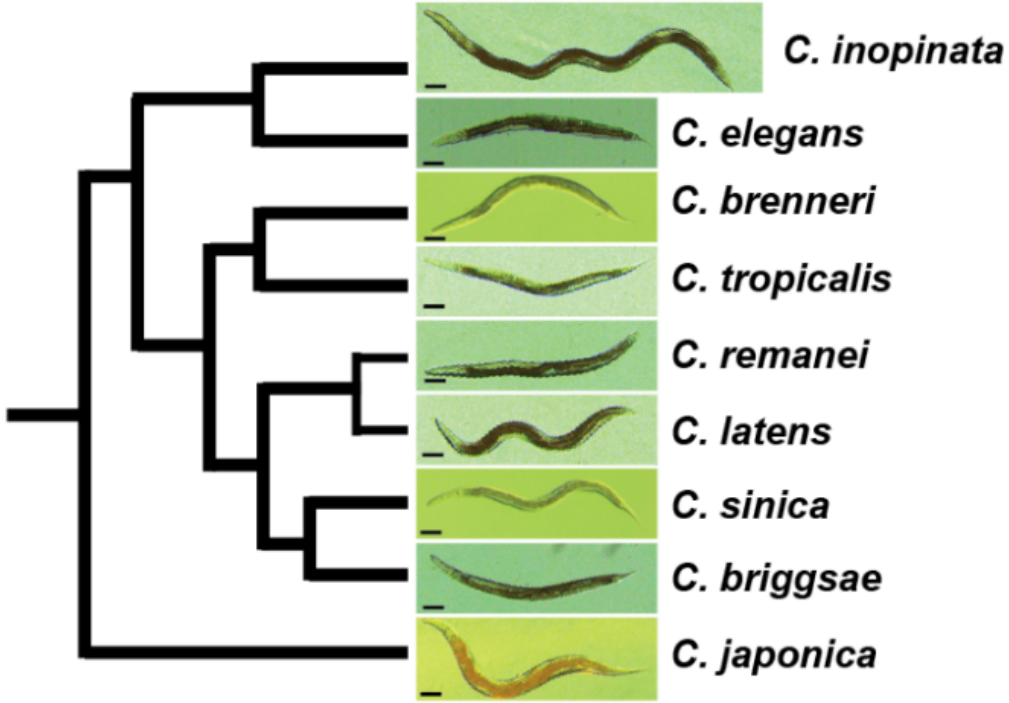
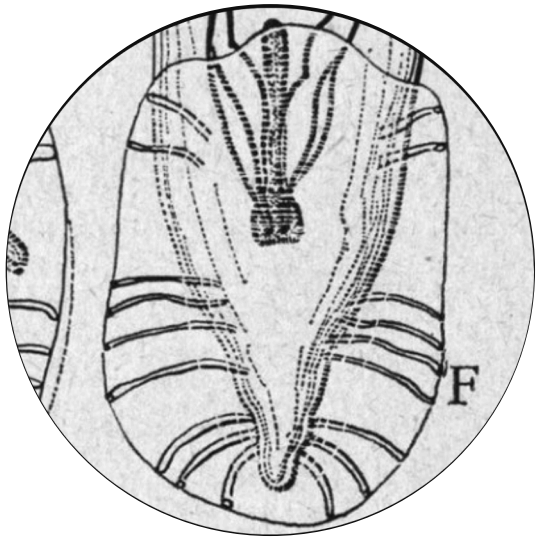
Degradation of the repetitive genomic landscape in a close relative of *C. elegans*

Gavin C. Woodruff[1,3] and Anastasia A. Teterina[1,2]

gavincw@uoregon.edu

https://gcwoodruff.github.io/





@weird_worms

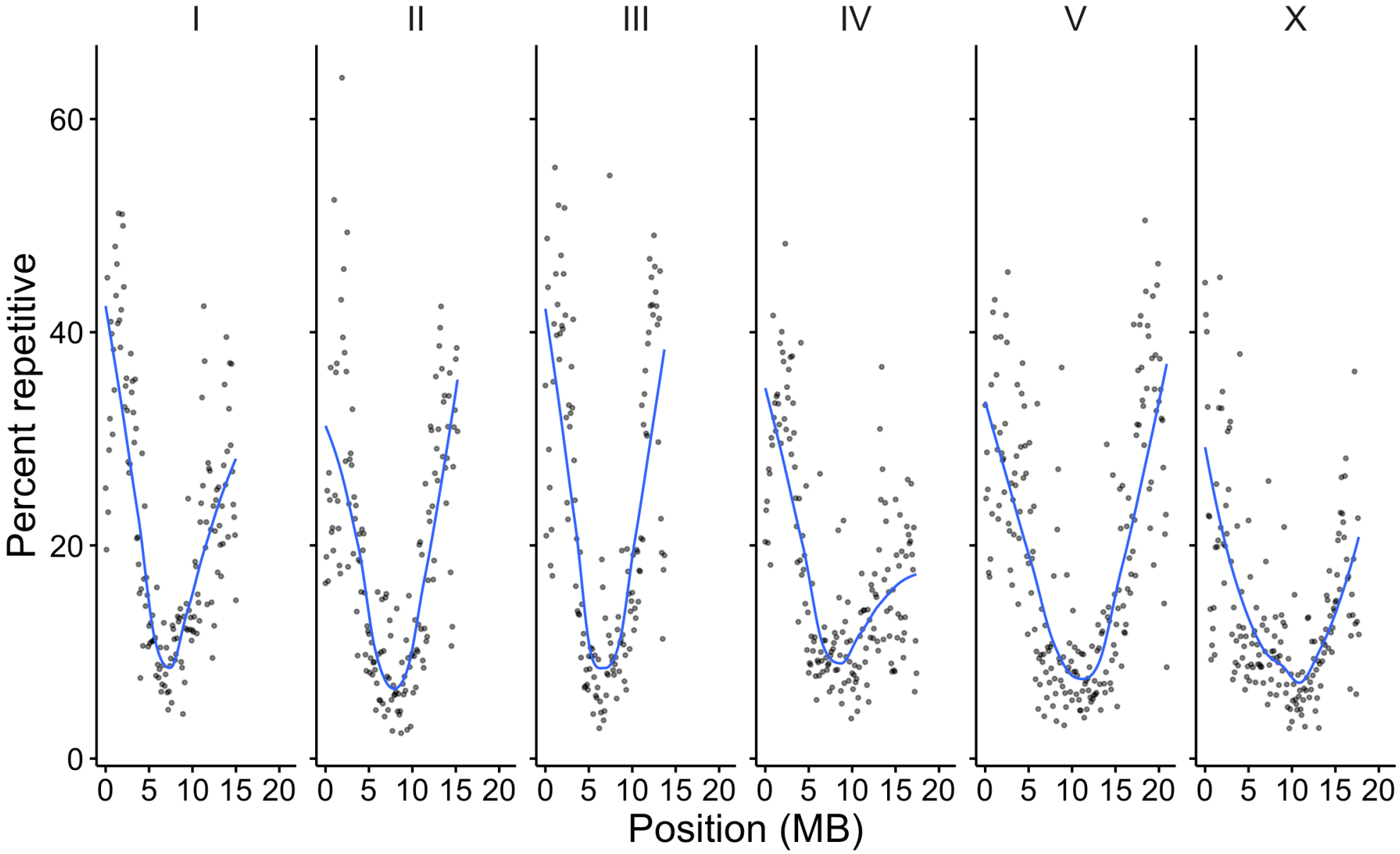
[1]Institute of Ecology and Evolution, University of Oregon
[2]Center of Parasitology, Severtsov Institute of Ecology and Evolution RAS
[3]Department of Biology, University of Oklahoma (starting as Assistant Professor August 2020!!!)

This story is on bioRxiv!

https://www.biorxiv.org/content/10.1101/797035v1.full

(also hopefully coming soon to a peer-reviewed journal near you)

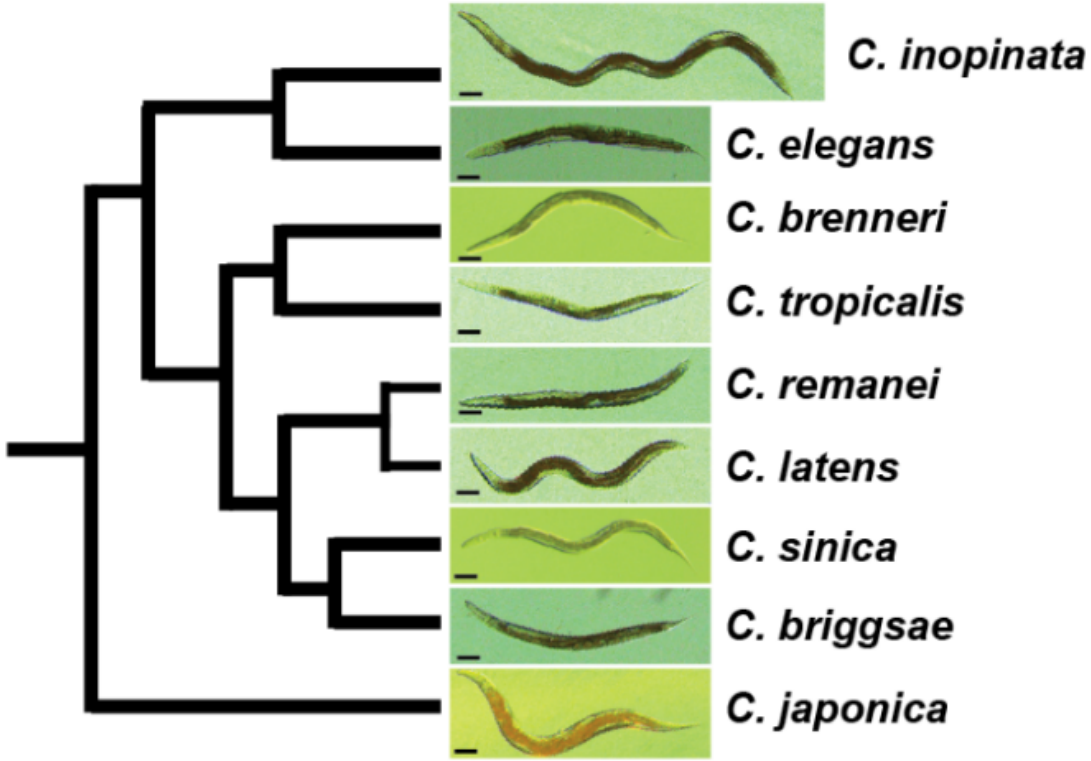
Repetitive elements covary with chromosome position in *C. elegans*



Custom figure; first reported in *C. elegans* genome consortium 1998

What causes repetitive genomic landscapes?

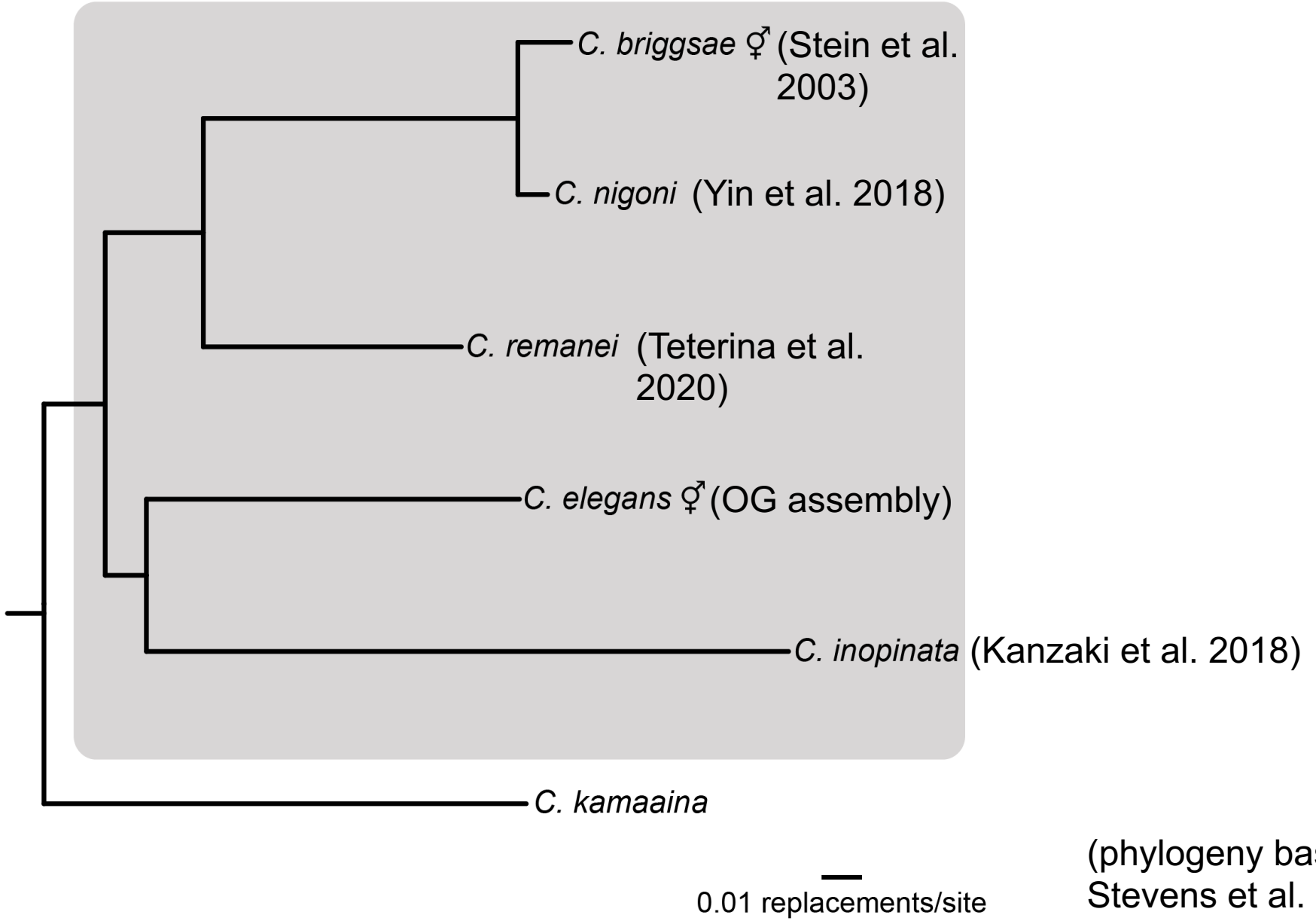
Aside: *C. inopinata* as a system for evo-devo and multidisciplinary biology



- *C. inopinata* is a close relative of *C. elegans*.
- *C. inopinata* is morphologically divergent.
- *C. inopinata* is amenable to sophisticated experimental approaches.
- *C. inopinata* is ecologically divergent.
- *C. inopinata* serendipitously co-occurs with the fig/fig wasp system.

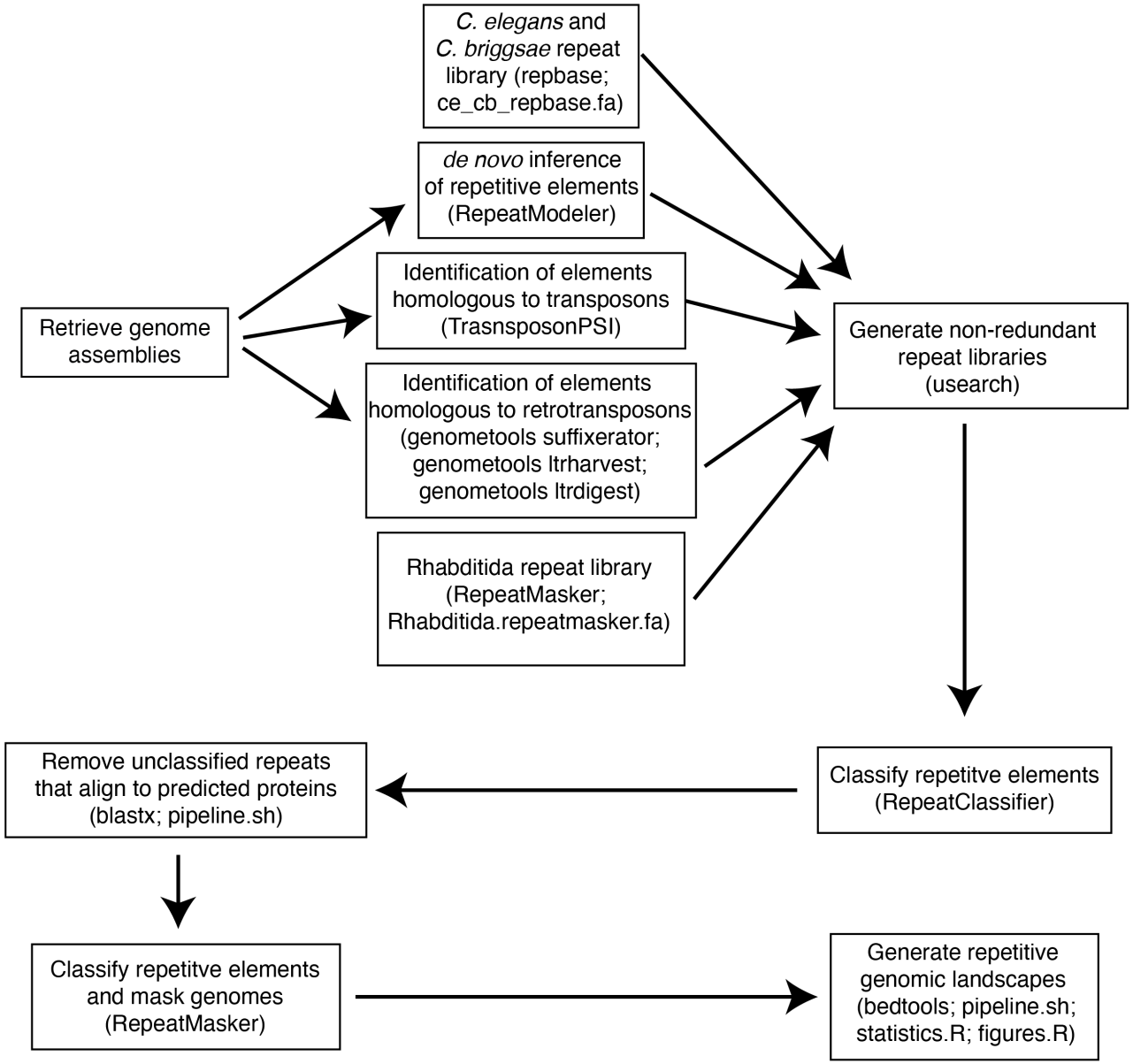
We looked at the repetitive genomic landscape in five worm species to understand how they evolve.

Five available *Caenorhabditis* chromosome-level genome assemblies were screened for repetitive elements with the same pipeline



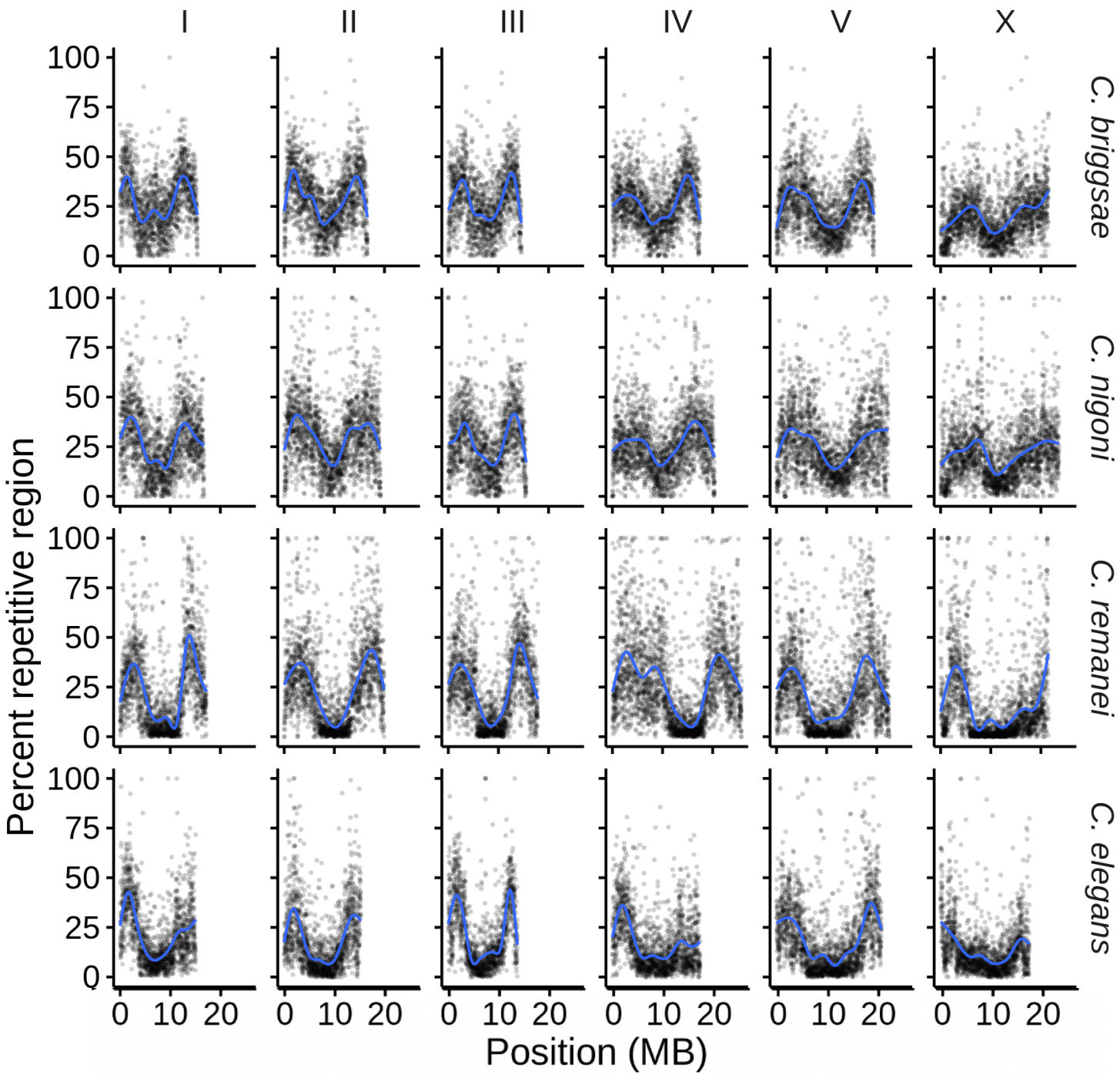
(phylogeny based on Stevens et al. 2019)

Five available *Caenorhabditis* chromosome-level genome assemblies were screened for repetitive elements with the same pipeline

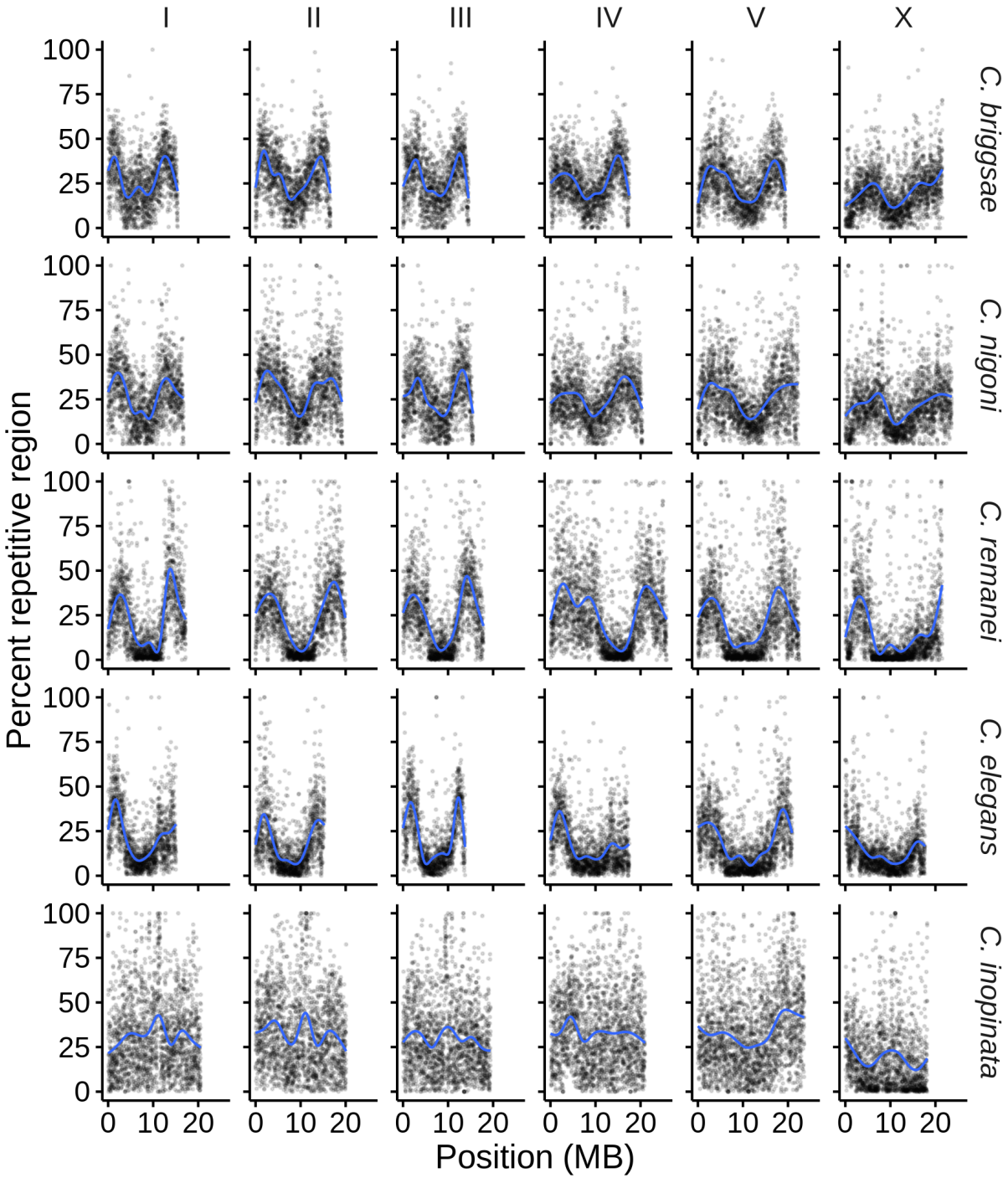


(thanks Anastasia Teterina; Coghlan et al. 2018)

Four *Caenorhabditis* species share a conserved non-random genomic distribution of repetitive elements

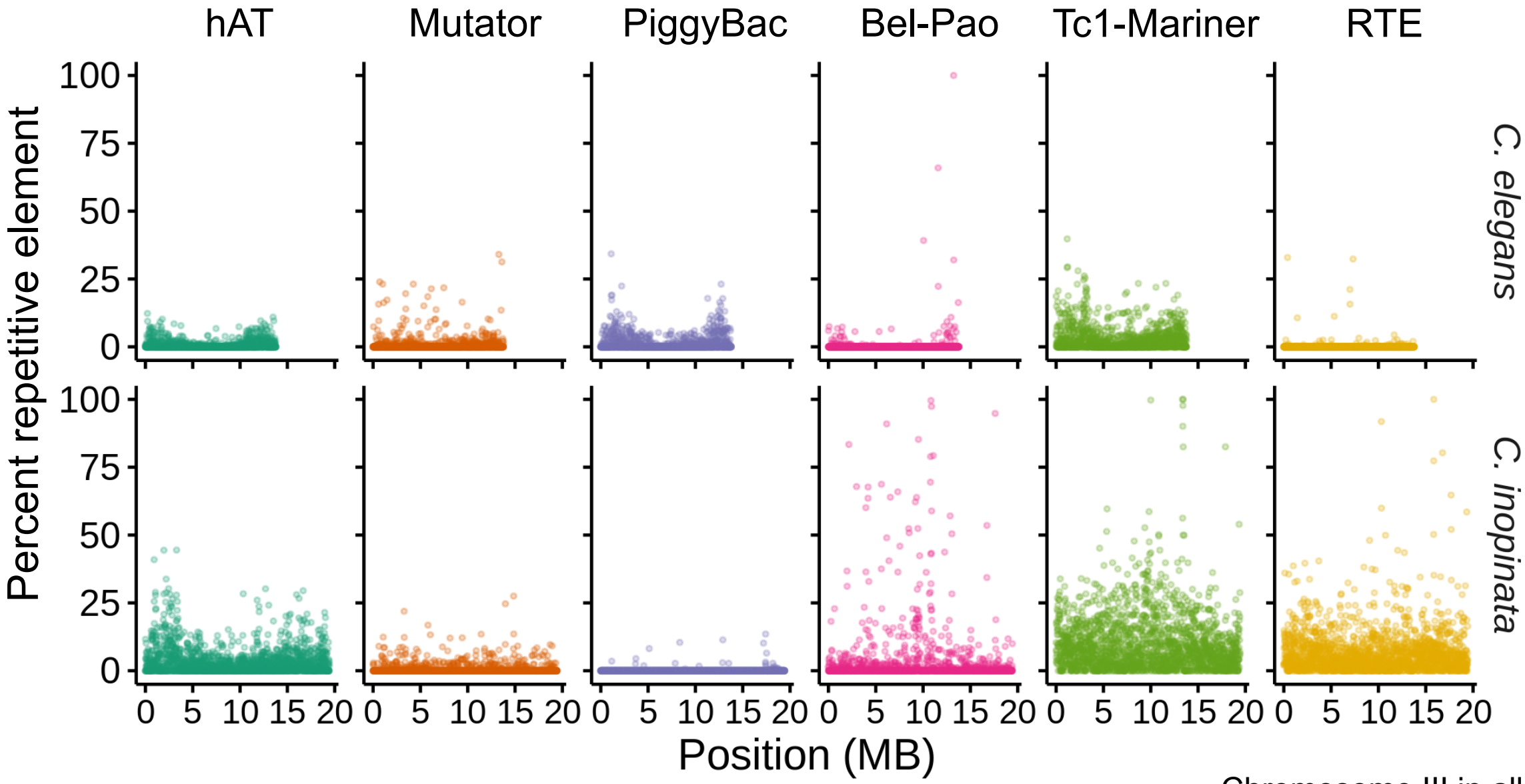


C. inopinata does not have this genomic organization of repetitive elements



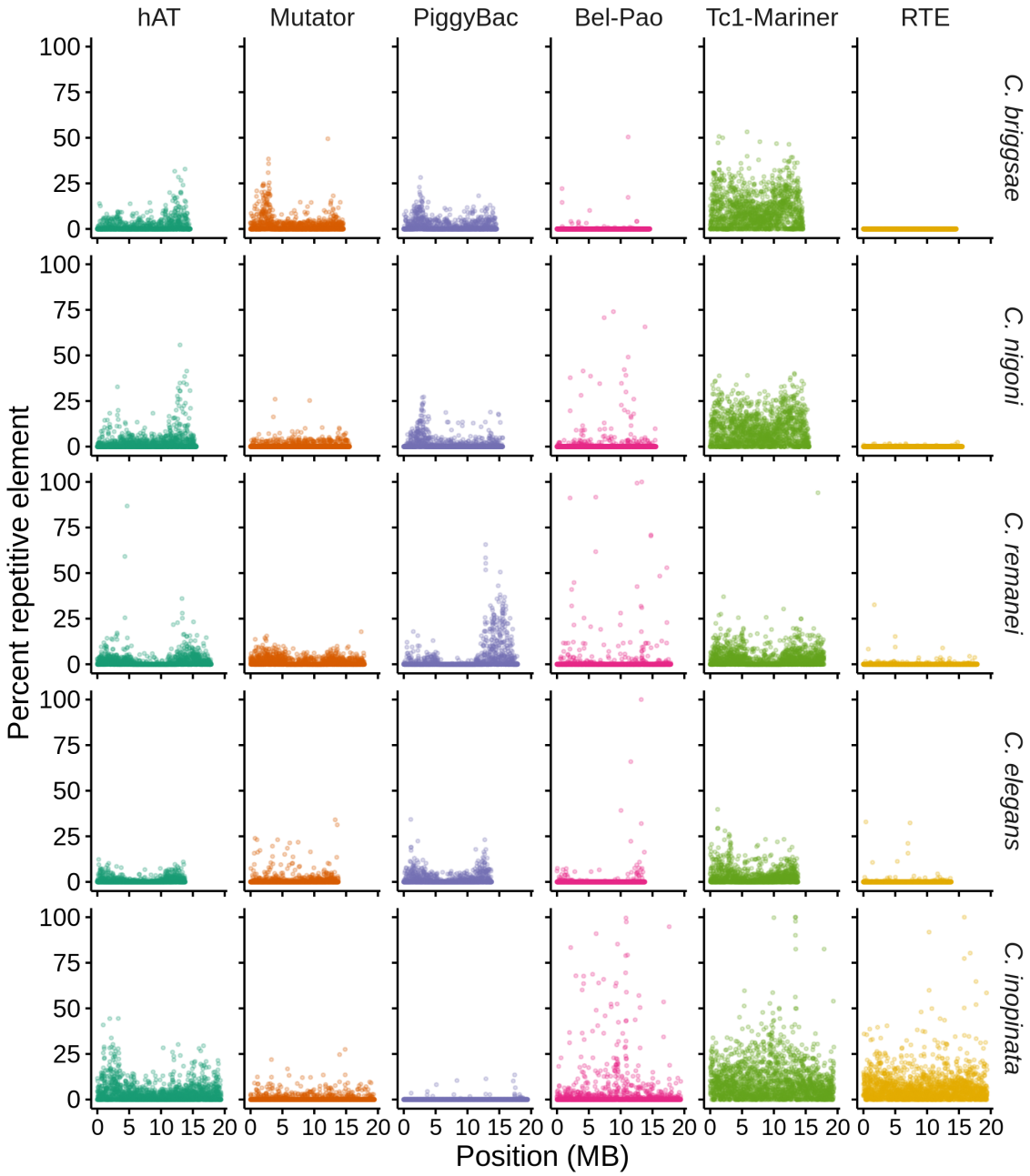
What about different kinds of repetitive elements? What do their repetitive genomic landscapes look like?

Different kinds of repetitive elements have different genomic landscapes



Chromosome III in all panels

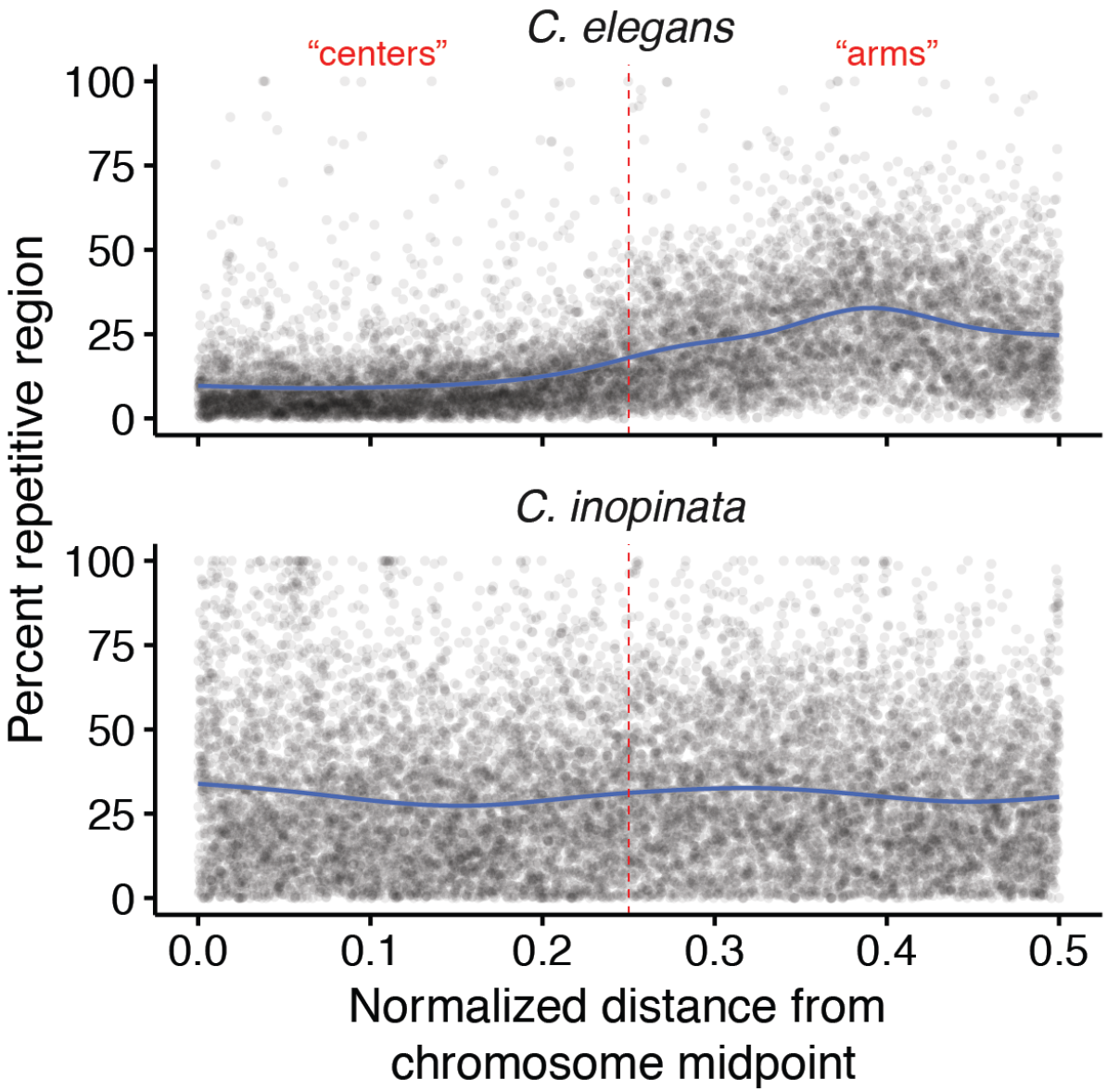
Different kinds of repetitive elements have different genomic landscapes



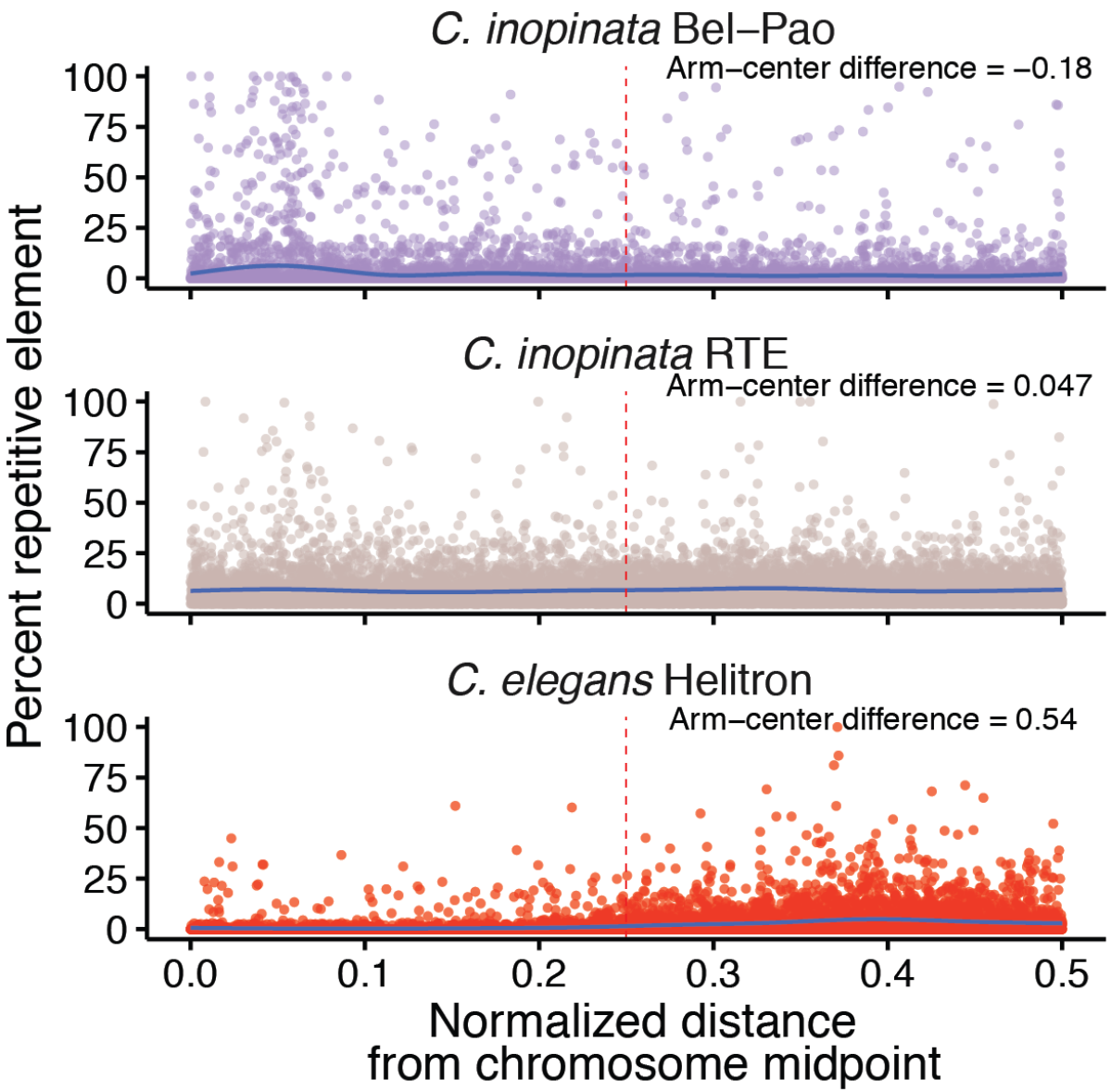
Chromosome III in all panels

Quantifying repetitive genomic landscapes

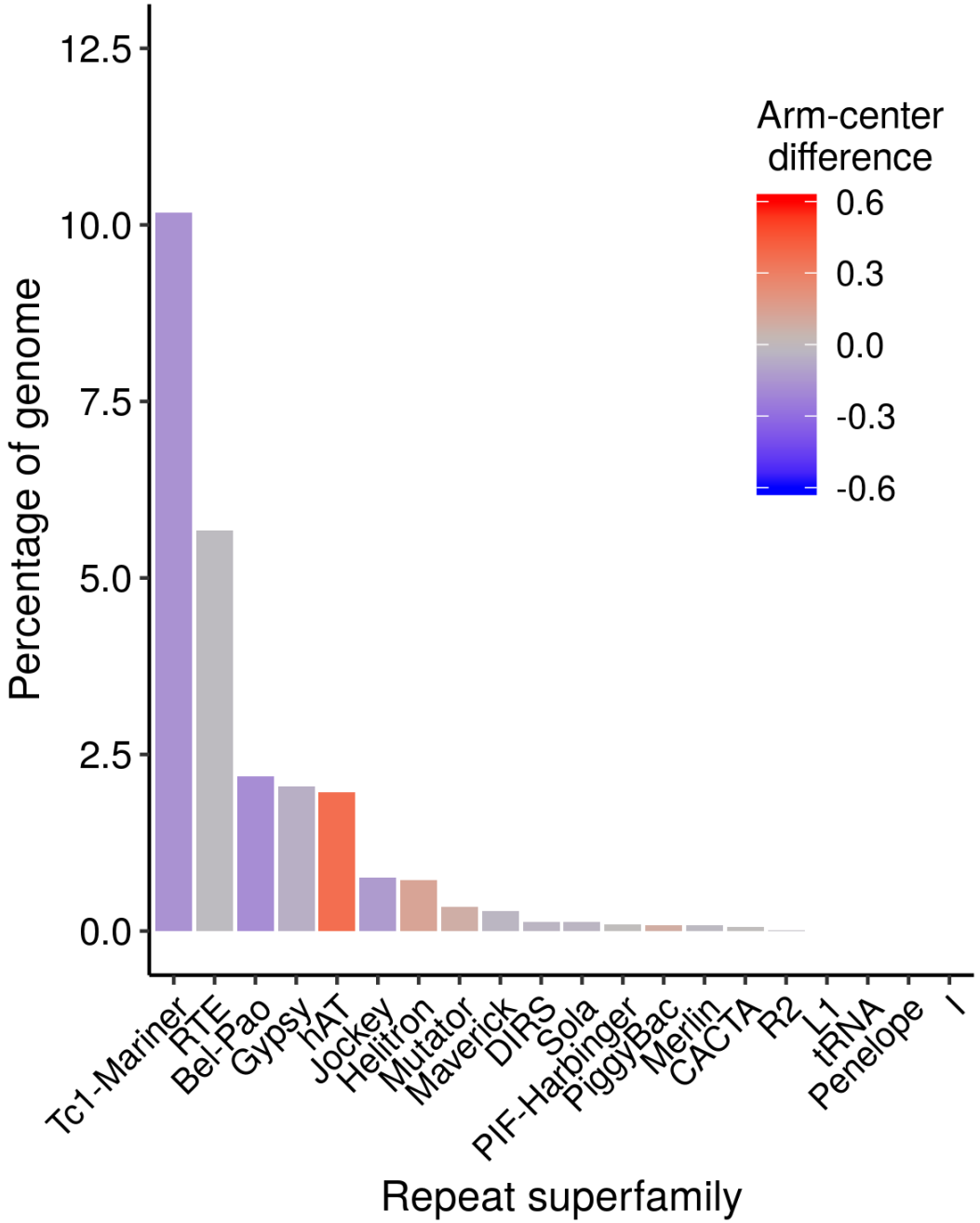
Repetitive genomic landscapes were quantified by normalizing by chromosome position



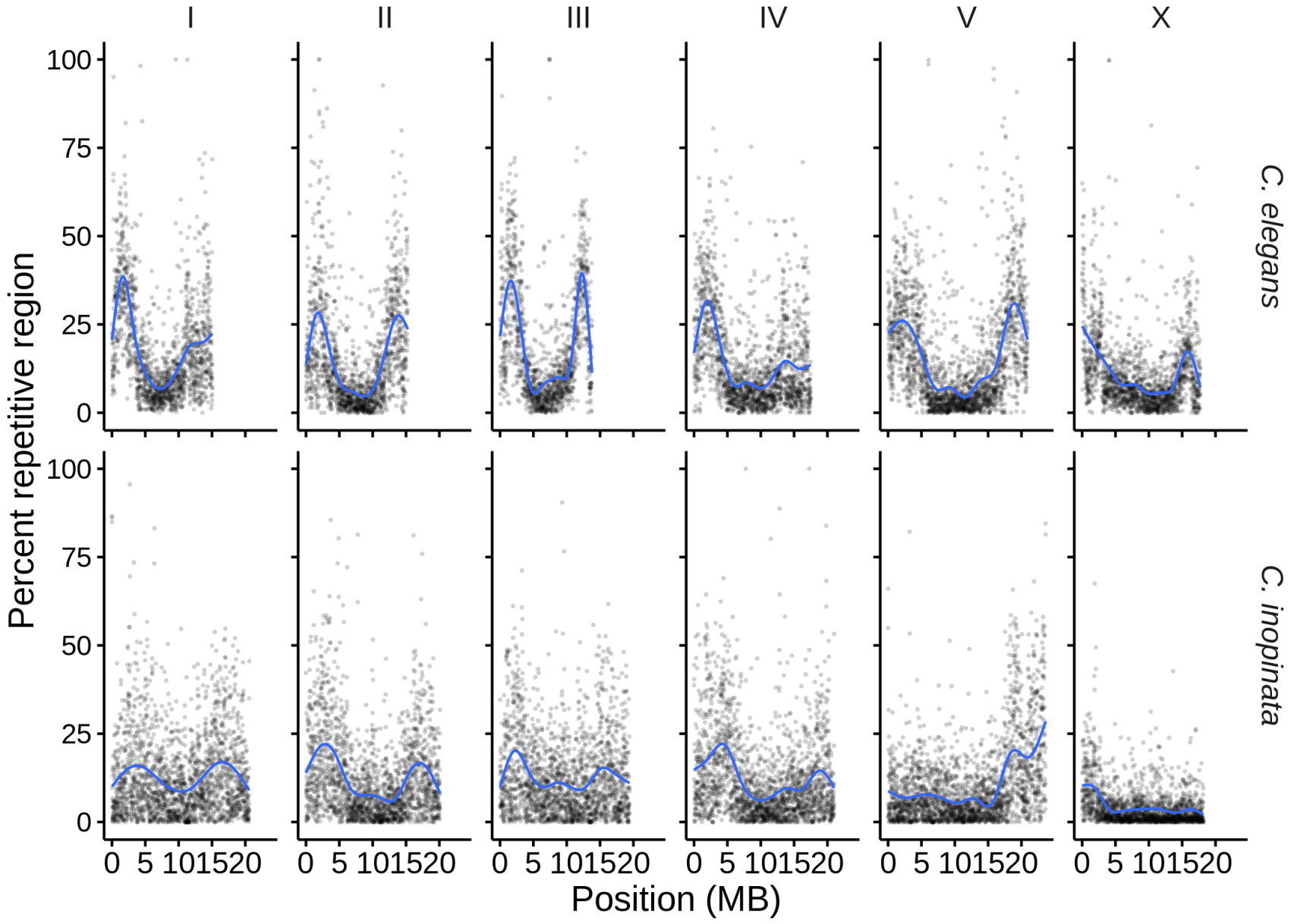
Chromosome arm-center effect sizes provide a measure of the repetitive genomic landscape



A few abundant repeat superfamilies have atypical genomic organization in *C. inopinata*



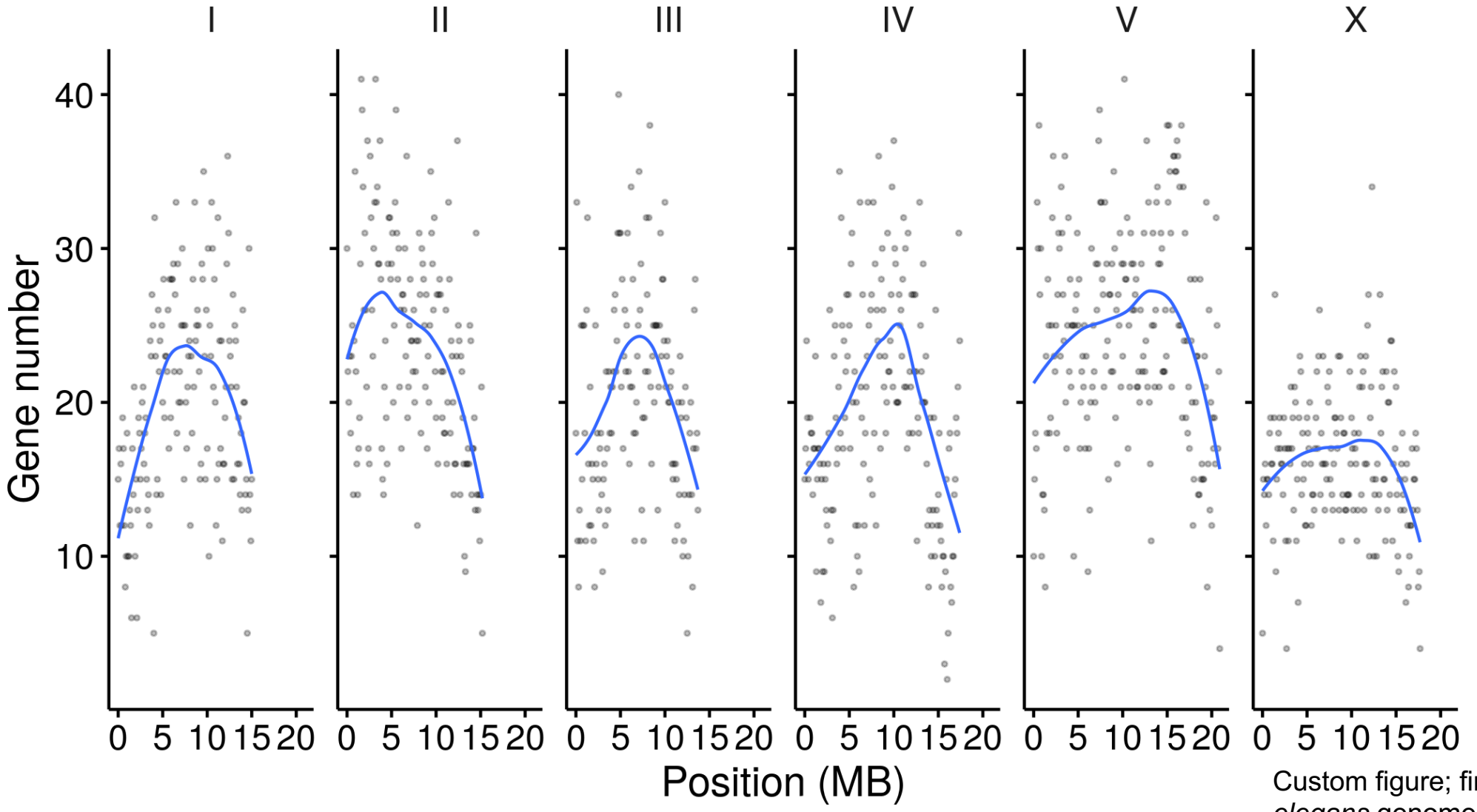
When four repeat superfamilies are removed, *C. inopinata*'s global repetitive landscape reveals a more familiar shape



Tc1-Mariner, RTE, Bel-Pao, & Gypsy removed

What is driving this superfamiliy-specific repetitive element expansion in *C. inopinata*?

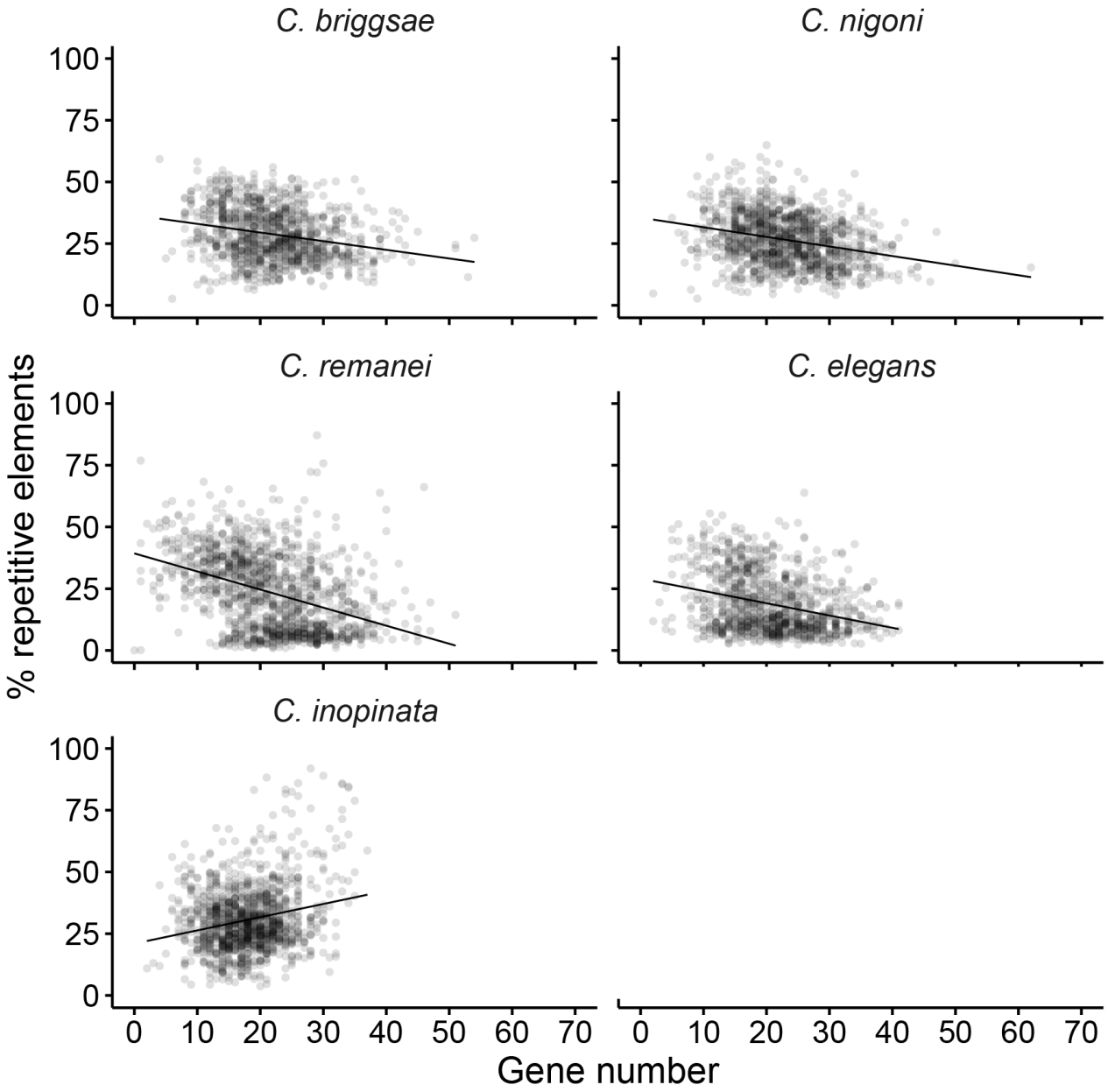
Protein-coding genes are enriched in chromosome centers in *C. elegans*



Custom figure; first reported in *C. elegans* genome consortium 1998

Is this true in *C. inopinata*?

Gene number is negatively correlated with repeat density in all *Caenorhabditis* species but *C. inopinata* (where there is a *positive* correlation!)

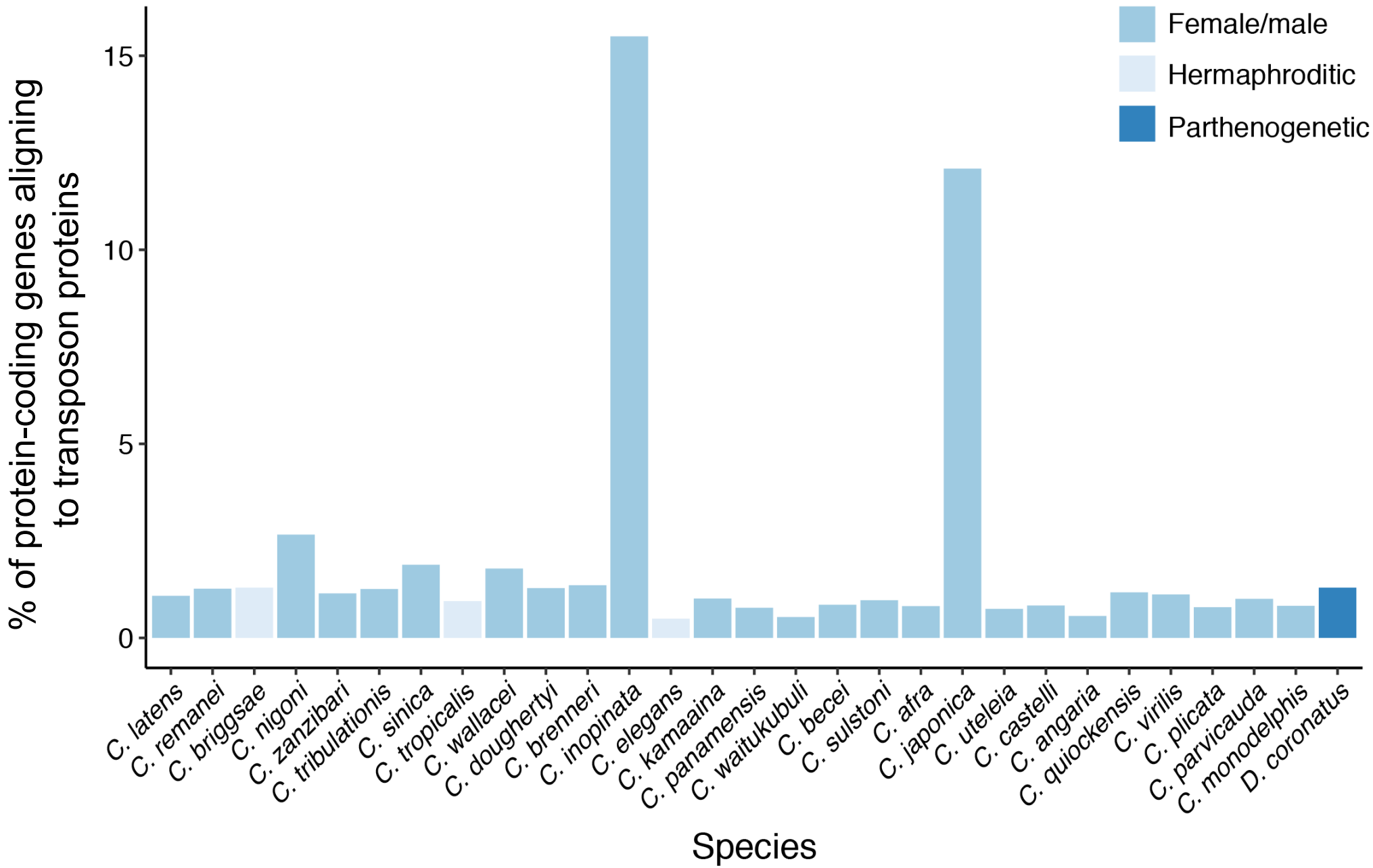


What??

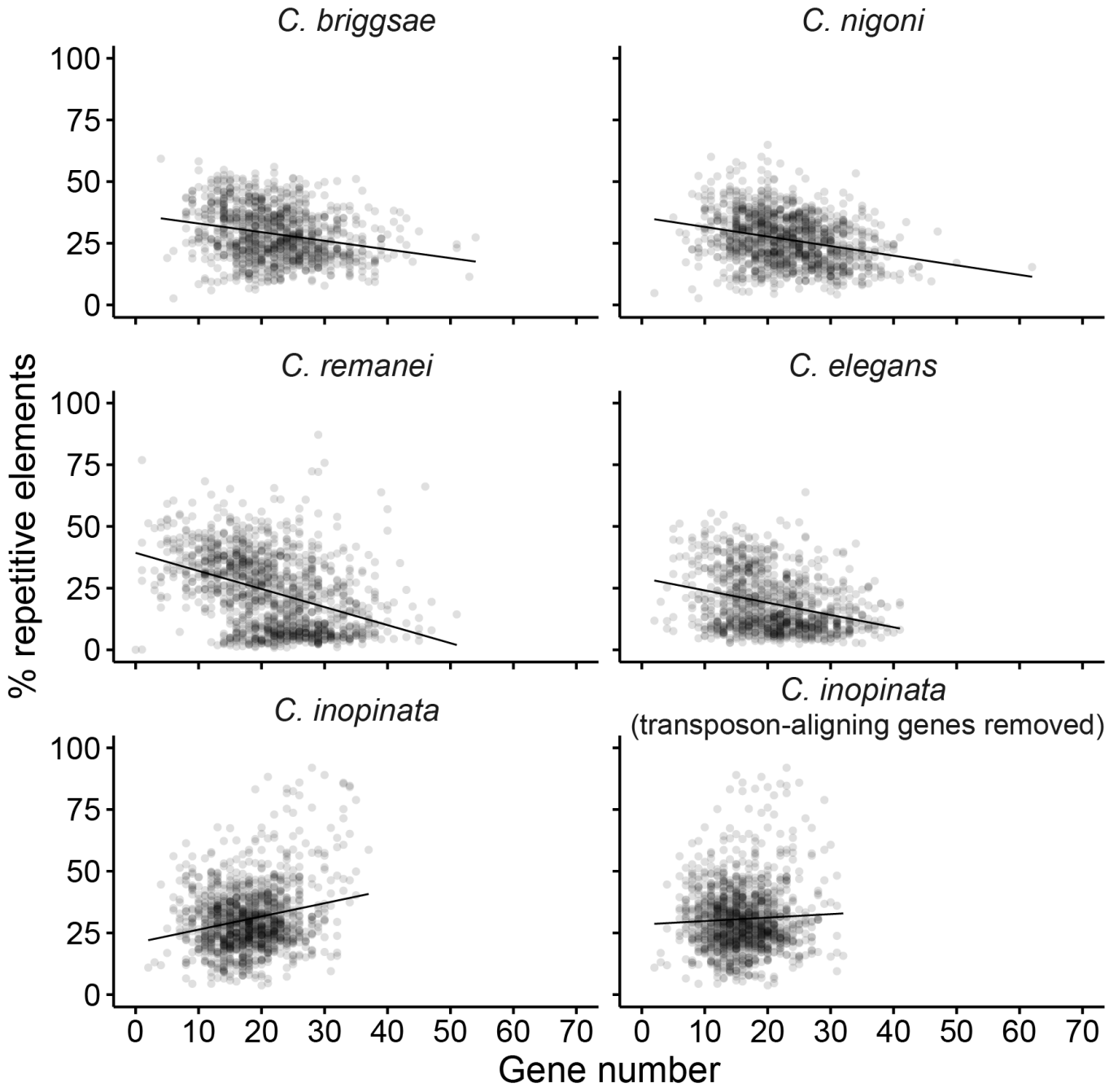
Previous analyses revealed many predicted *C. inopinata* protein-coding genes have transposon domains.

To investigate this further, I aligned all *Caenorhabditis* protein sets to a transposon database.

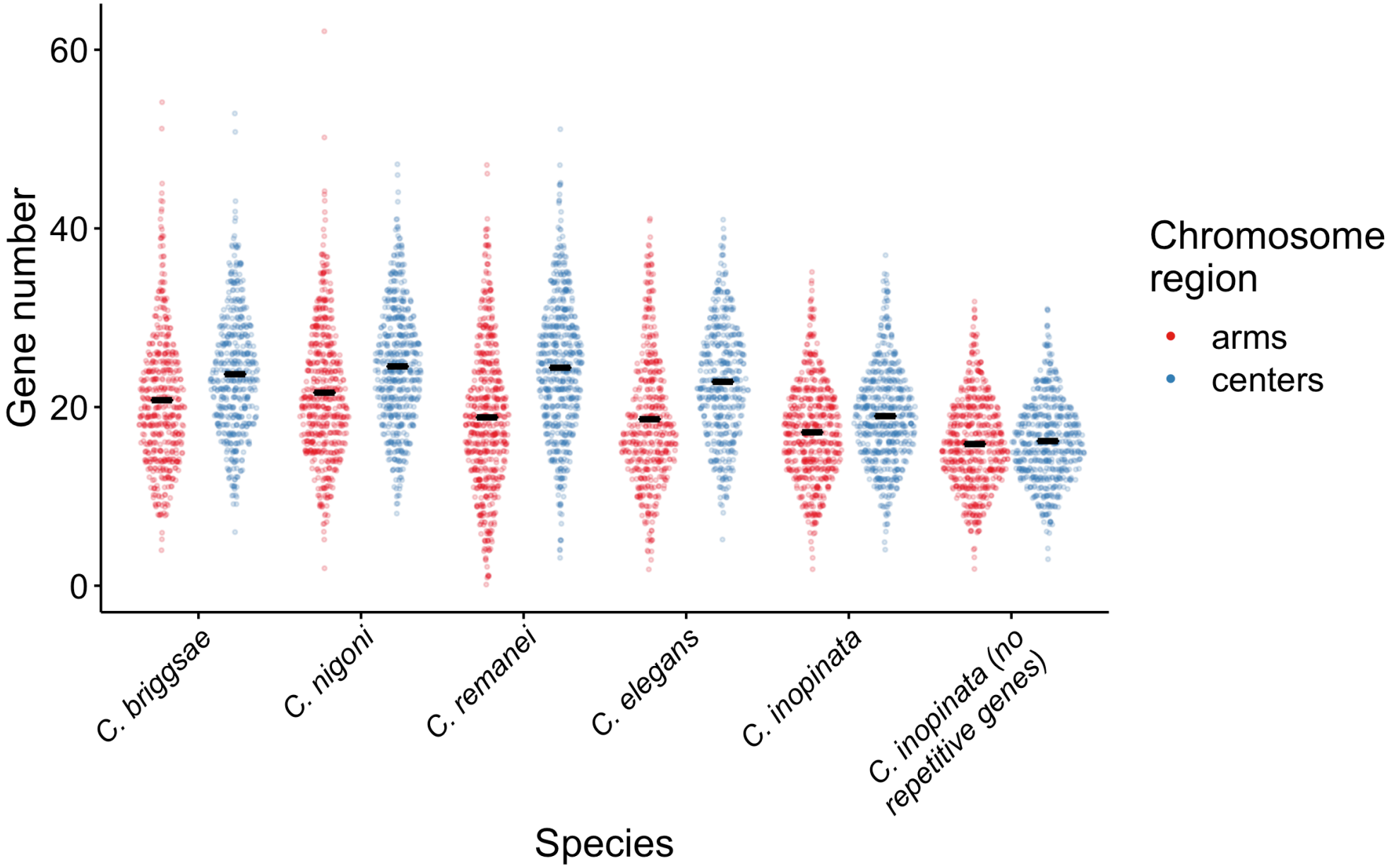
Many *C. inopinata* proteins align to transposons



When transposon-aligning genes are removed, there is no gene-repeat relationship in *C. inopinata*

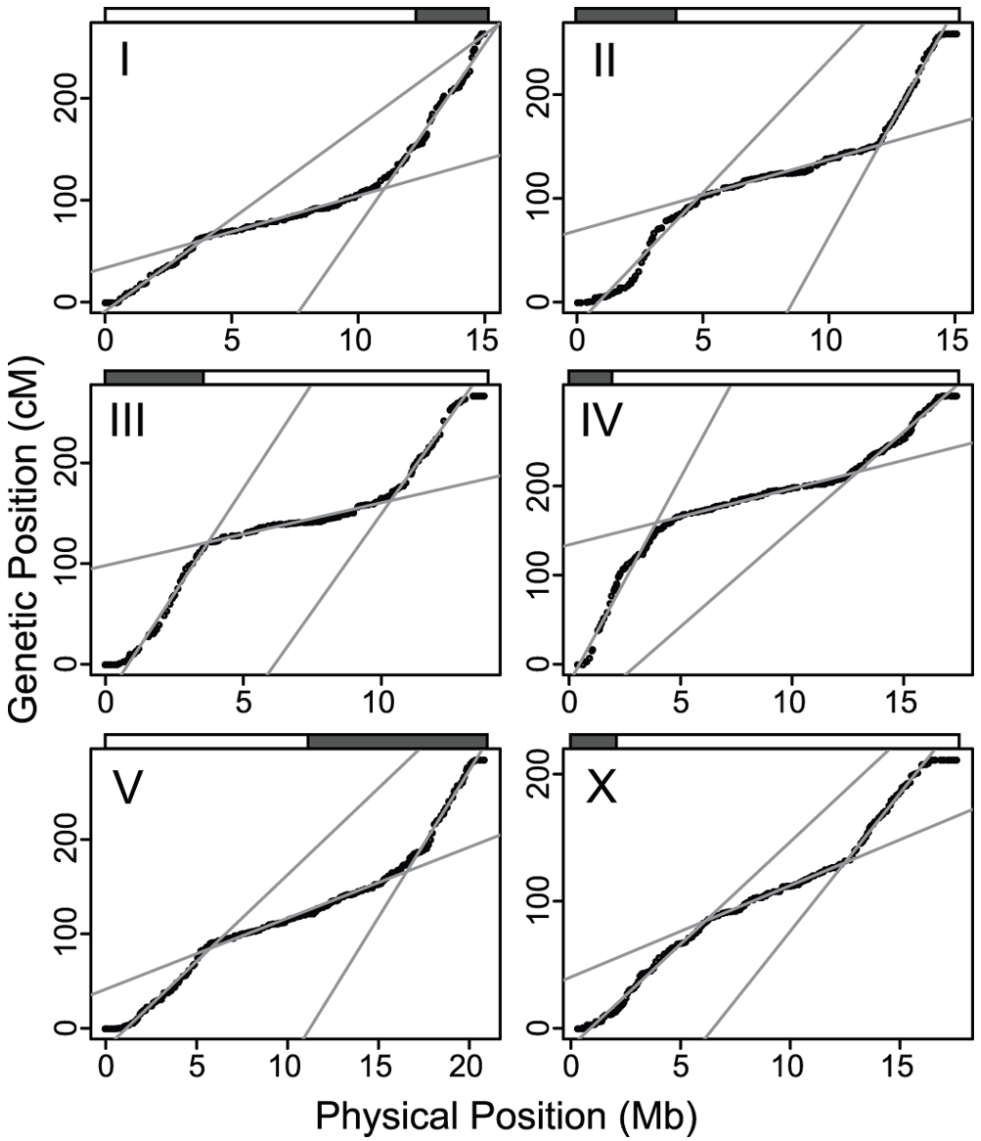


C. inopinata reveals no chromosomal clustering of protein-coding genes after removing transposon-aligning genes



But wait, recombination rate also covaries with all of these things!

Recombination rate covaries with chromosome position in *C. elegans* and *C. briggsae*



C. elegans (Rockman and Kruglyak 2009)
Also *C. briggsae* (Ross et al. 2011)

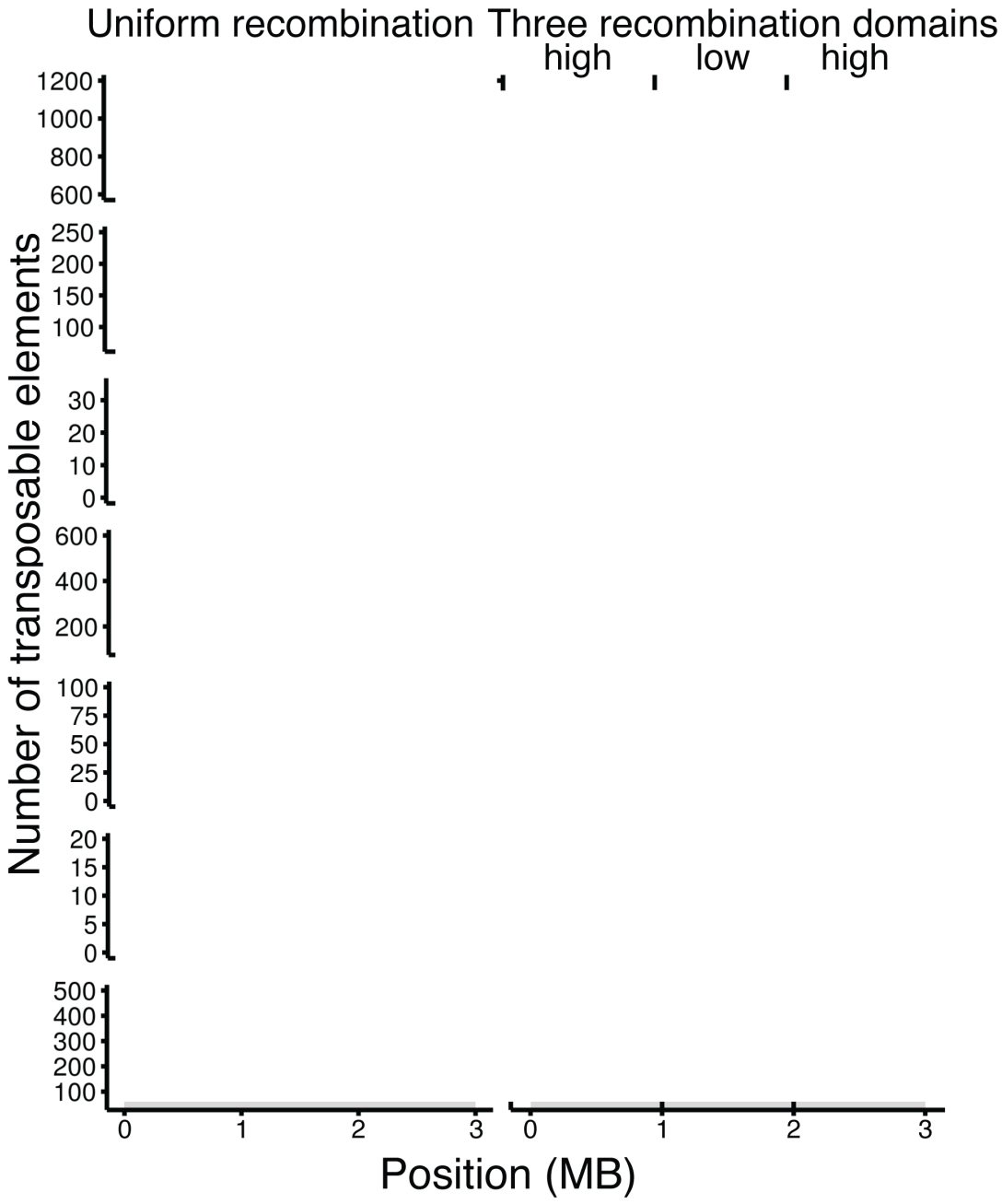
Can recombination alone explain distributions of repetitive elements?

We addressed recombination with evolutionary simulations!

- Implemented in SLiM (Haller and Messer 2019) using “copy-and-paste” TE evolution recipe.
- Population size 5,000
- One 3MB chromosome with uniform recombination rate or three domains of varying recombination (high-low-high)
- 50,000 generations
- 50 replicates per condition/scenario
- Various fitness effects of insertions across chromosome domains

(all work by Anastasia Teterina)

Genomic heterogeneity in insertion fitness effects are sufficient for generating repetitive landscapes



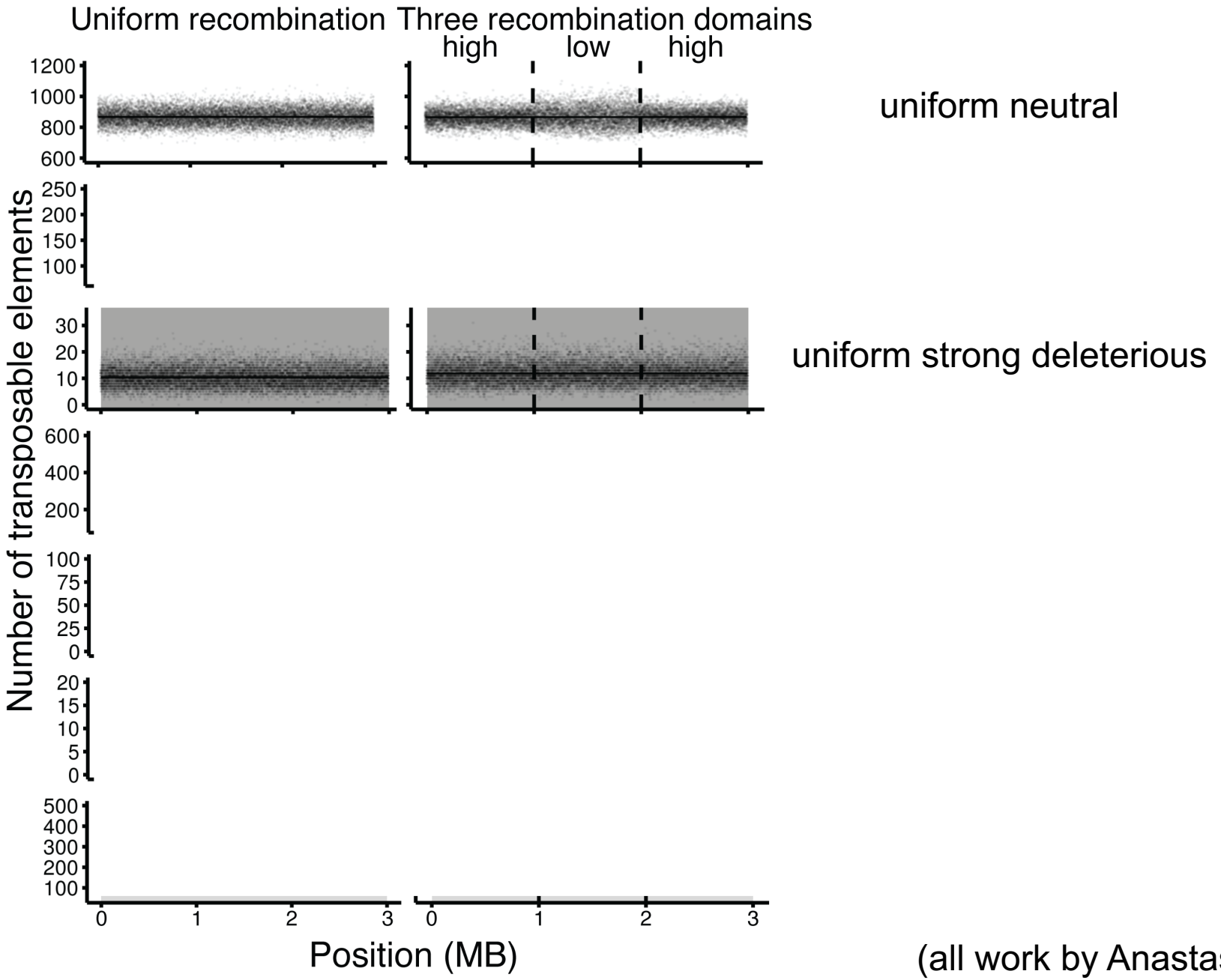
Red line: arm-center Cohen's *d* effect size > 0.2

Black line: arm-center Cohen's *d* effect size < 0.2

Background: darker shades of gray corresponds to more deleterious effects of insertion

(all work by Anastasia Teterina)

Genomic heterogeneity in insertion fitness effects are sufficient for generating repetitive landscapes



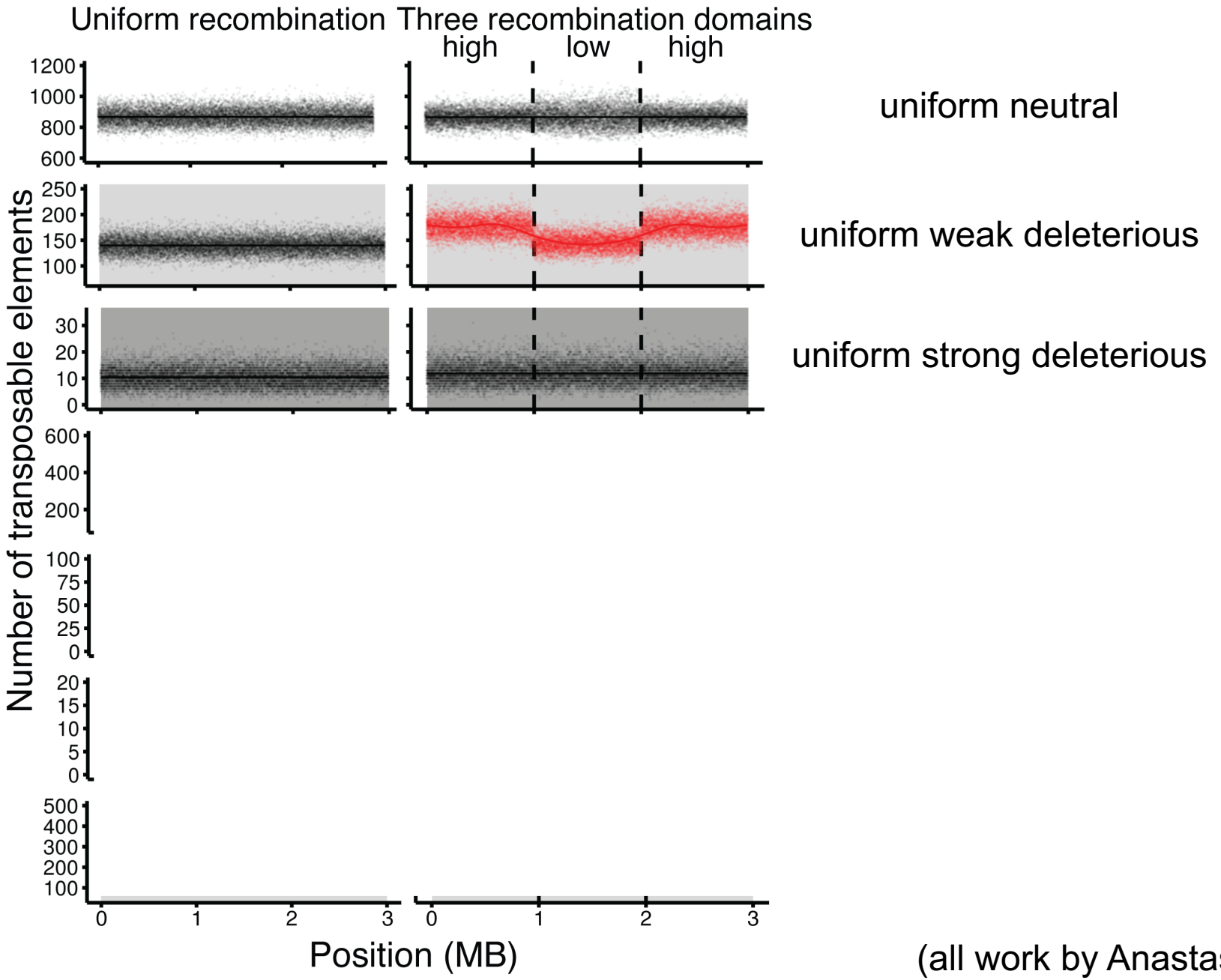
Red line: arm-center Cohen's *d* effect size > 0.2

Black line: arm-center Cohen's *d* effect size < 0.2

Background: darker shades of gray corresponds to more deleterious effects of insertion

(all work by Anastasia Teterina)

Genomic heterogeneity in insertion fitness effects are sufficient for generating repetitive landscapes



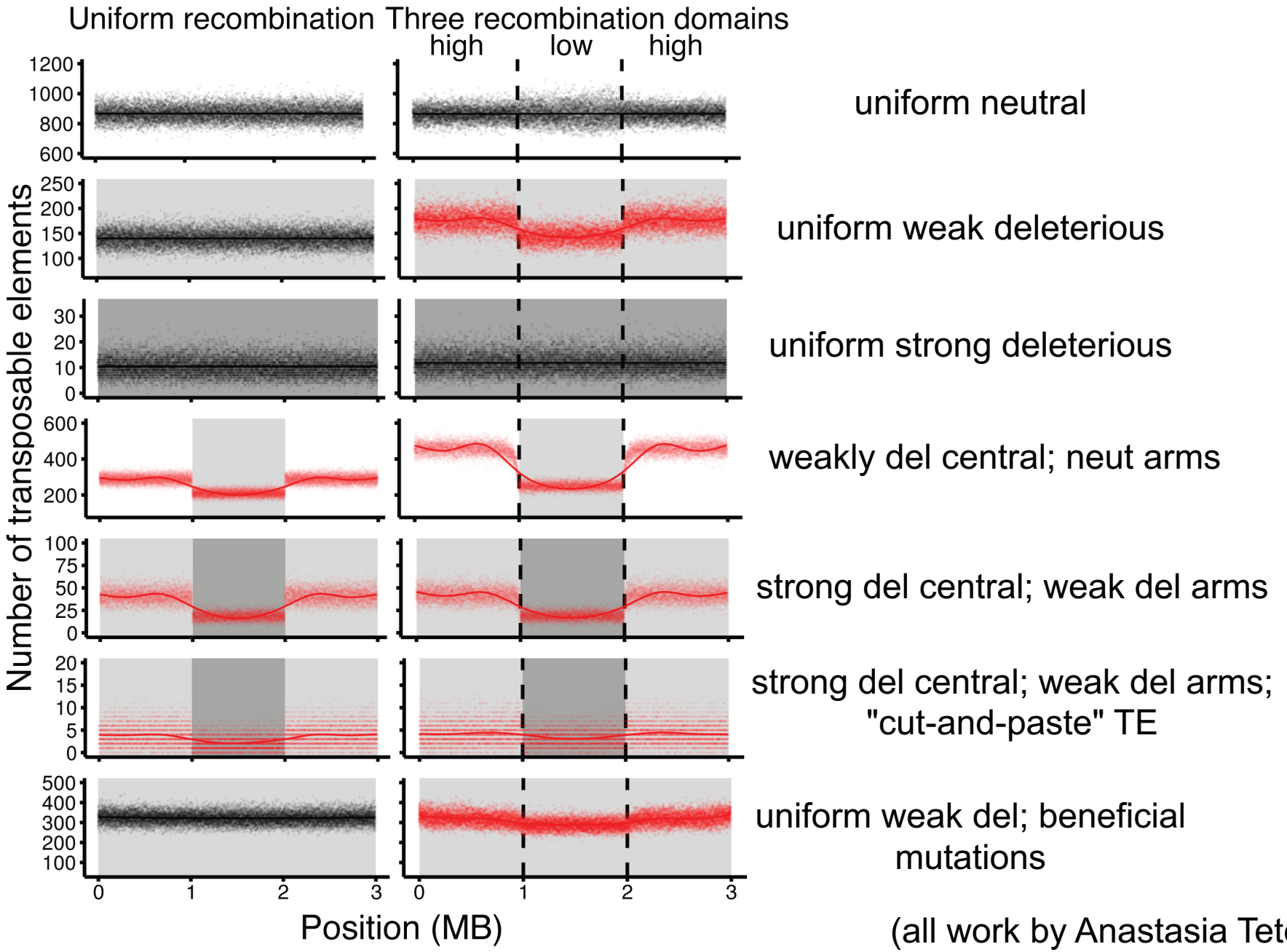
Red line: arm-center Cohen's *d* effect size > 0.2

Black line: arm-center Cohen's *d* effect size < 0.2

Background: darker shades of gray corresponds to more deleterious effects of insertion

(all work by Anastasia Teterina)

Genomic heterogeneity in insertion fitness effects are sufficient for generating repetitive landscapes



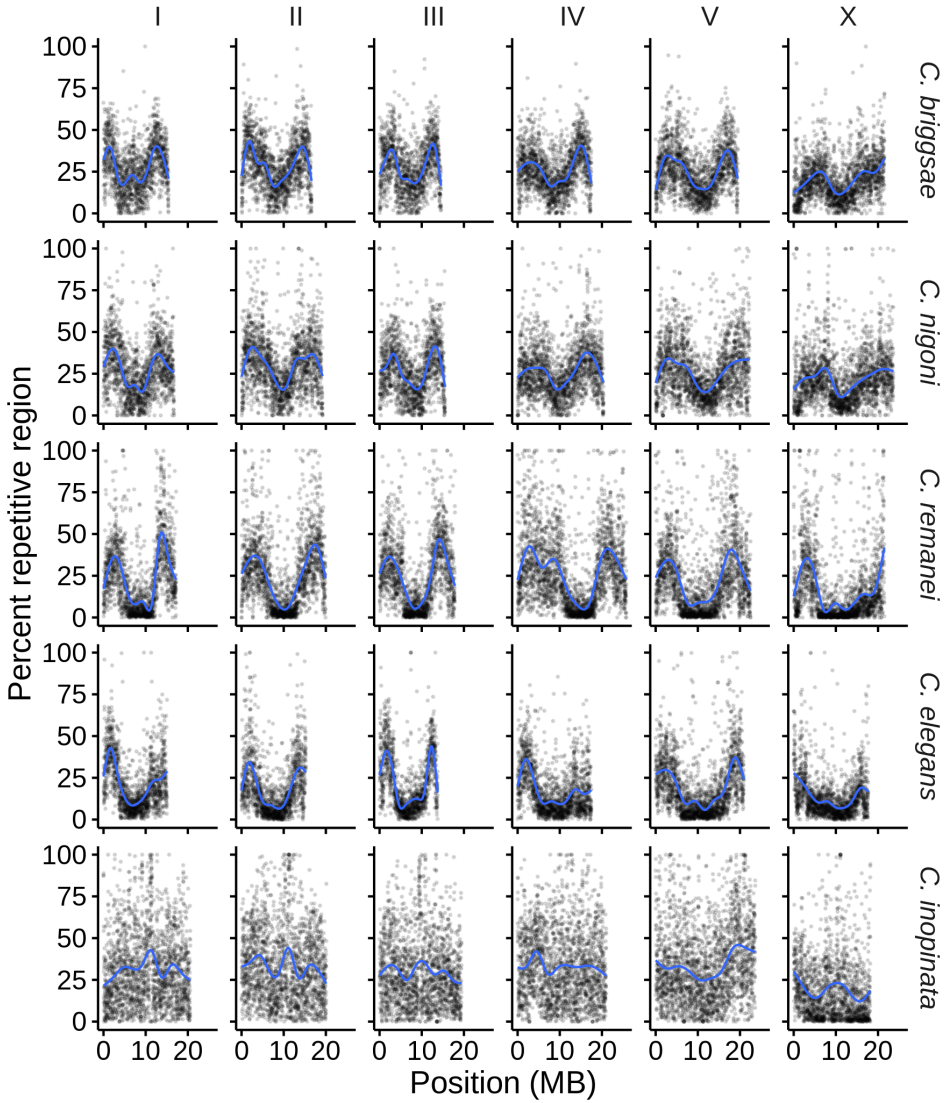
Red line: arm-center Cohen's *d* effect size > 0.2

Black line: arm-center Cohen's *d* effect size < 0.2

Background: darker shades of gray corresponds to more deleterious effects of insertion

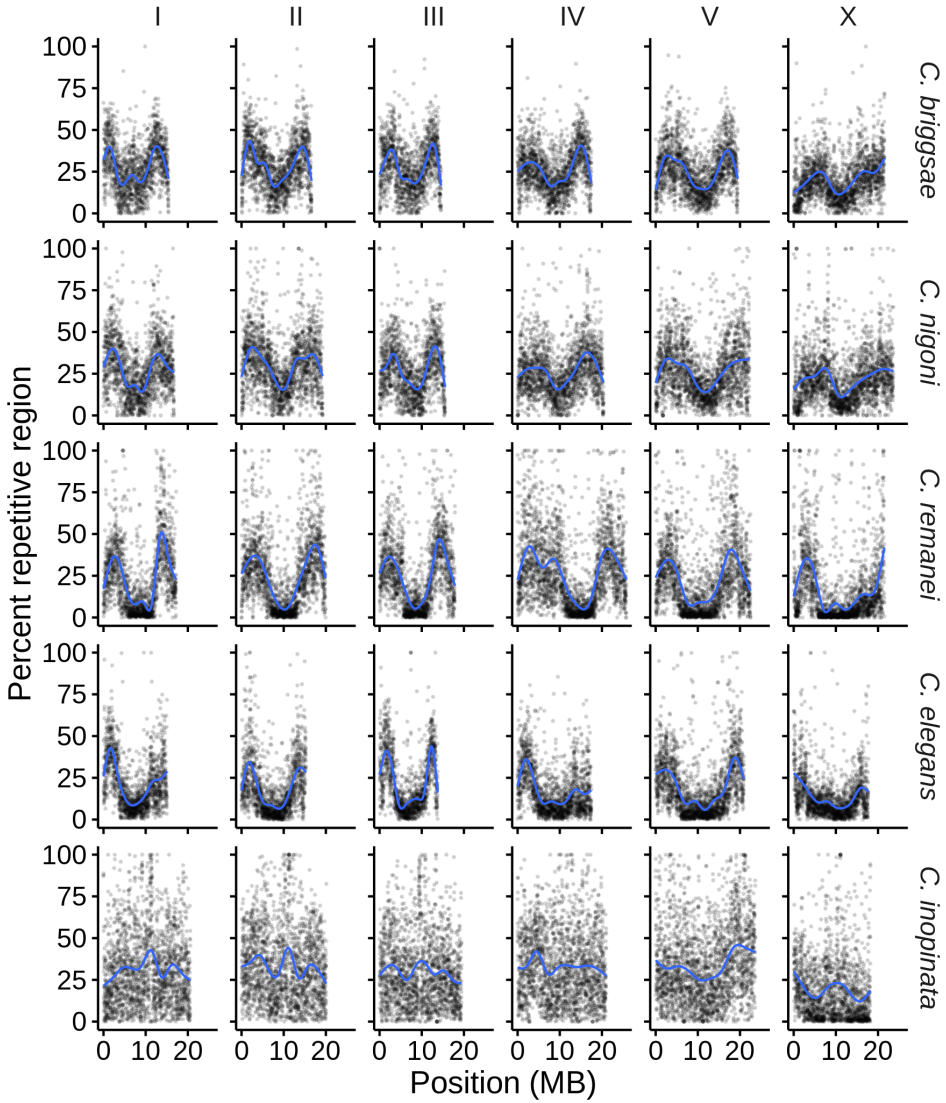
(all work by Anastasia Teterina)

Concluding Thoughts

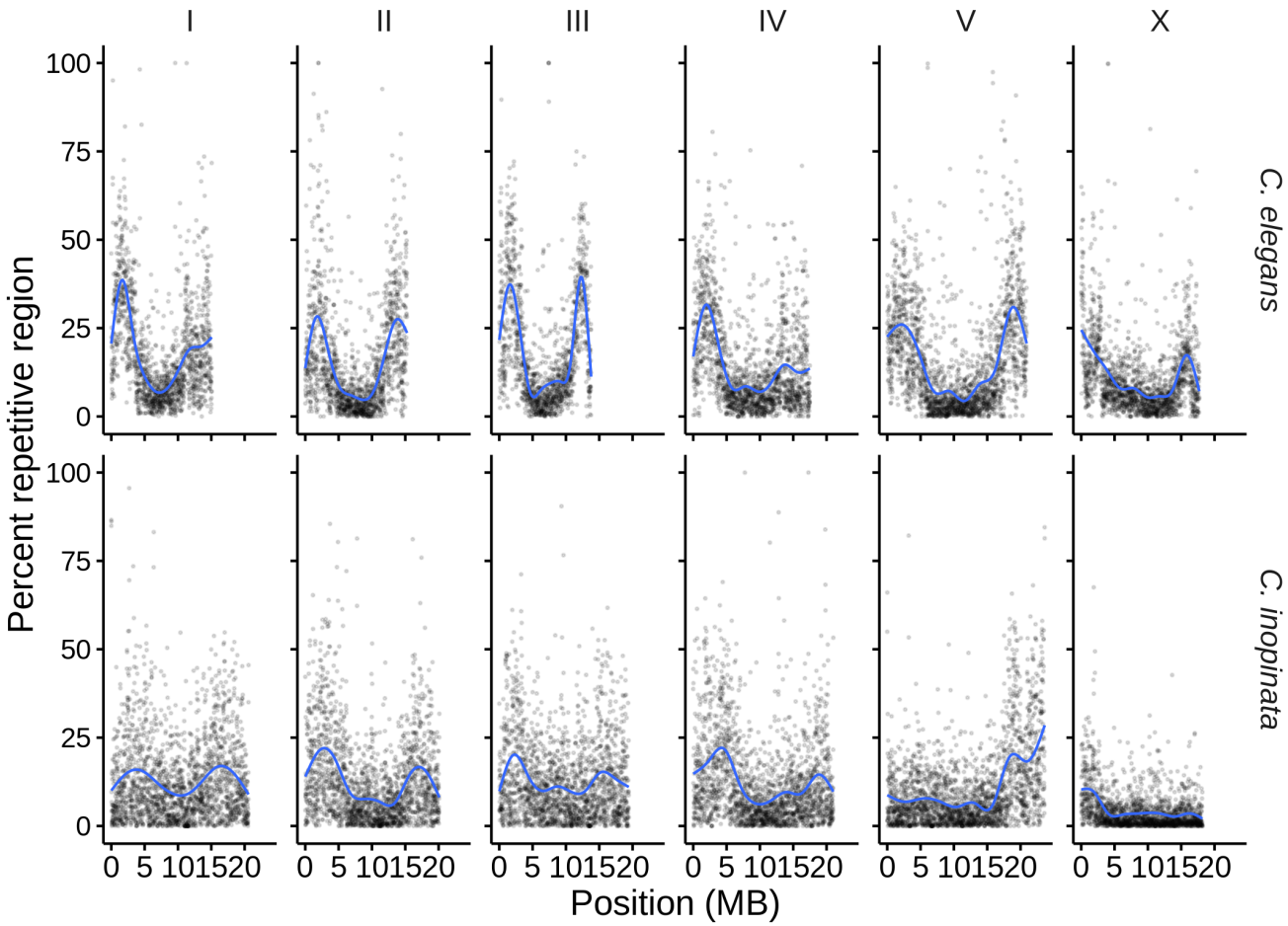


C. inopinata has a divergent repetitive genomic landscape.

Concluding Thoughts

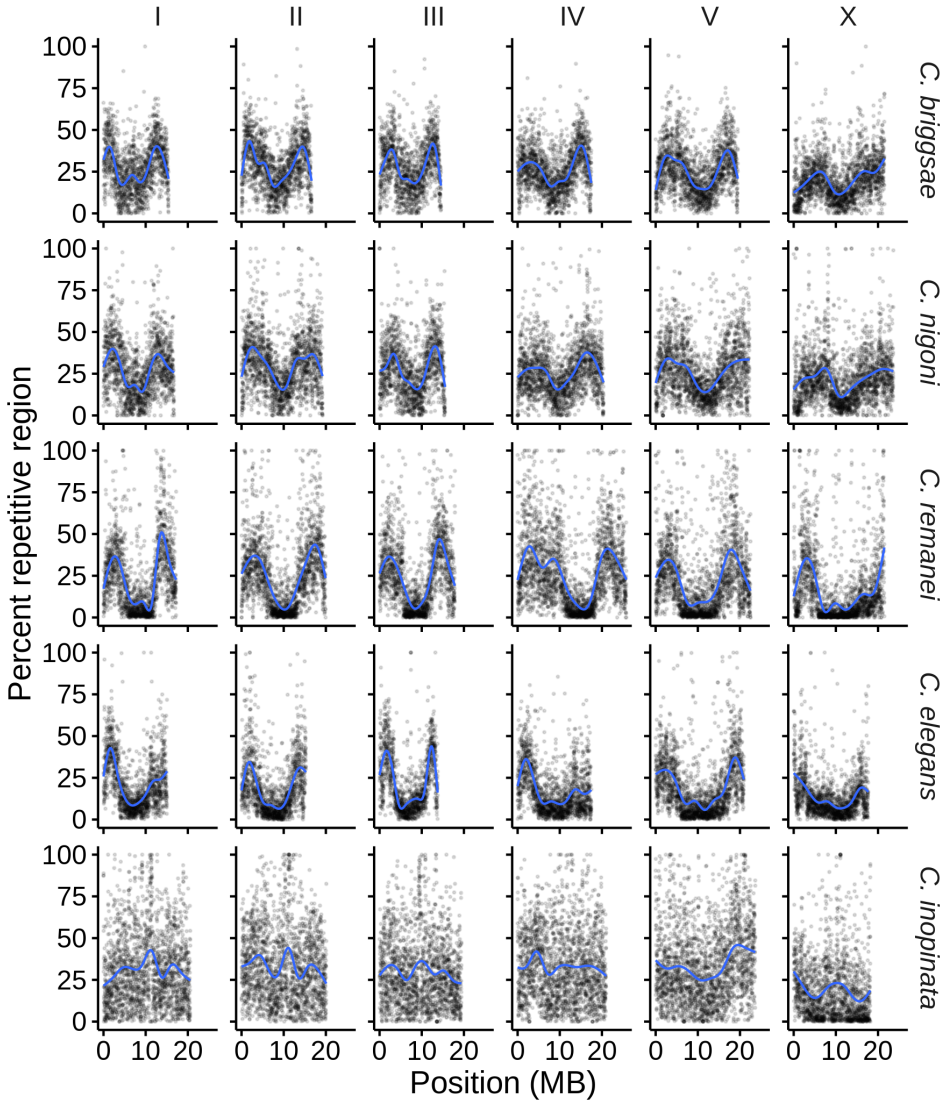


C. inopinata has a divergent repetitive genomic landscape.

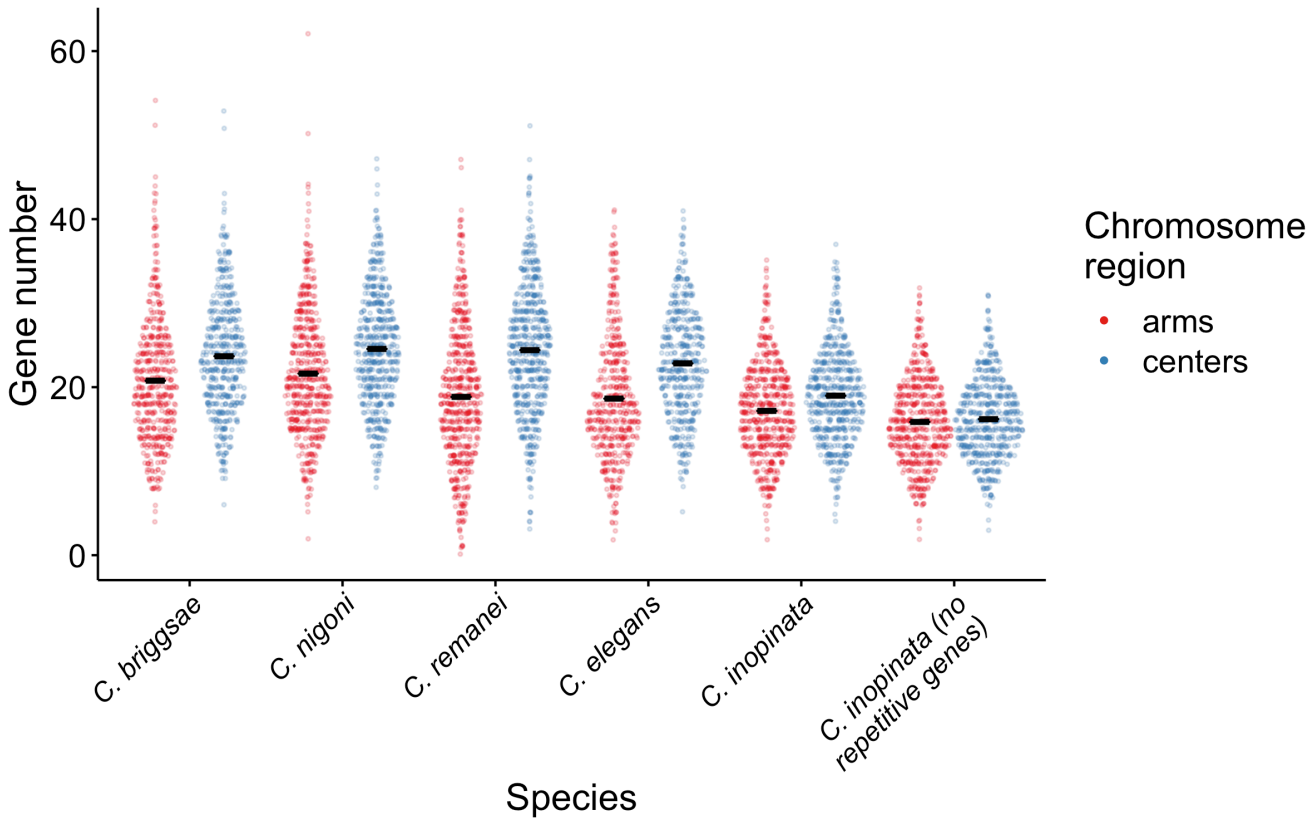


This pattern is largely driven by four divergent transposable element superfamilies.

Concluding Thoughts

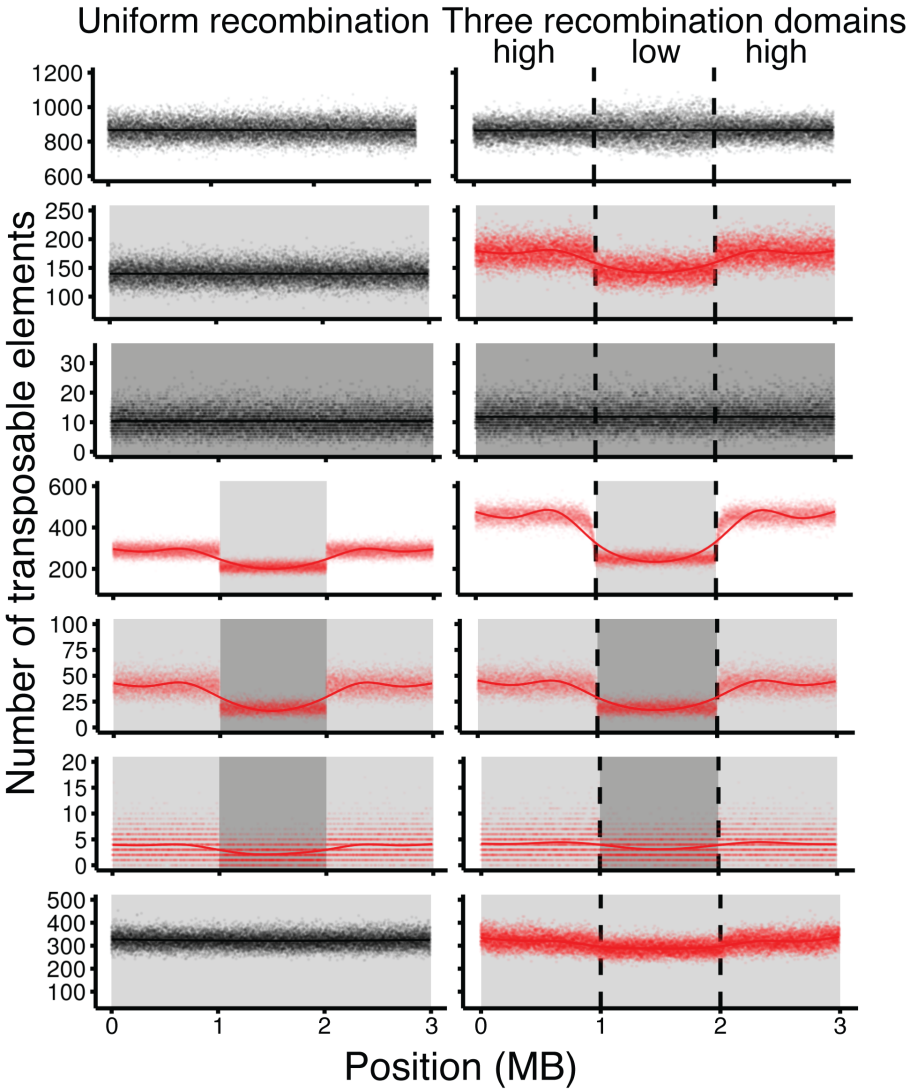


C. inopinata has a divergent repetitive genomic landscape.



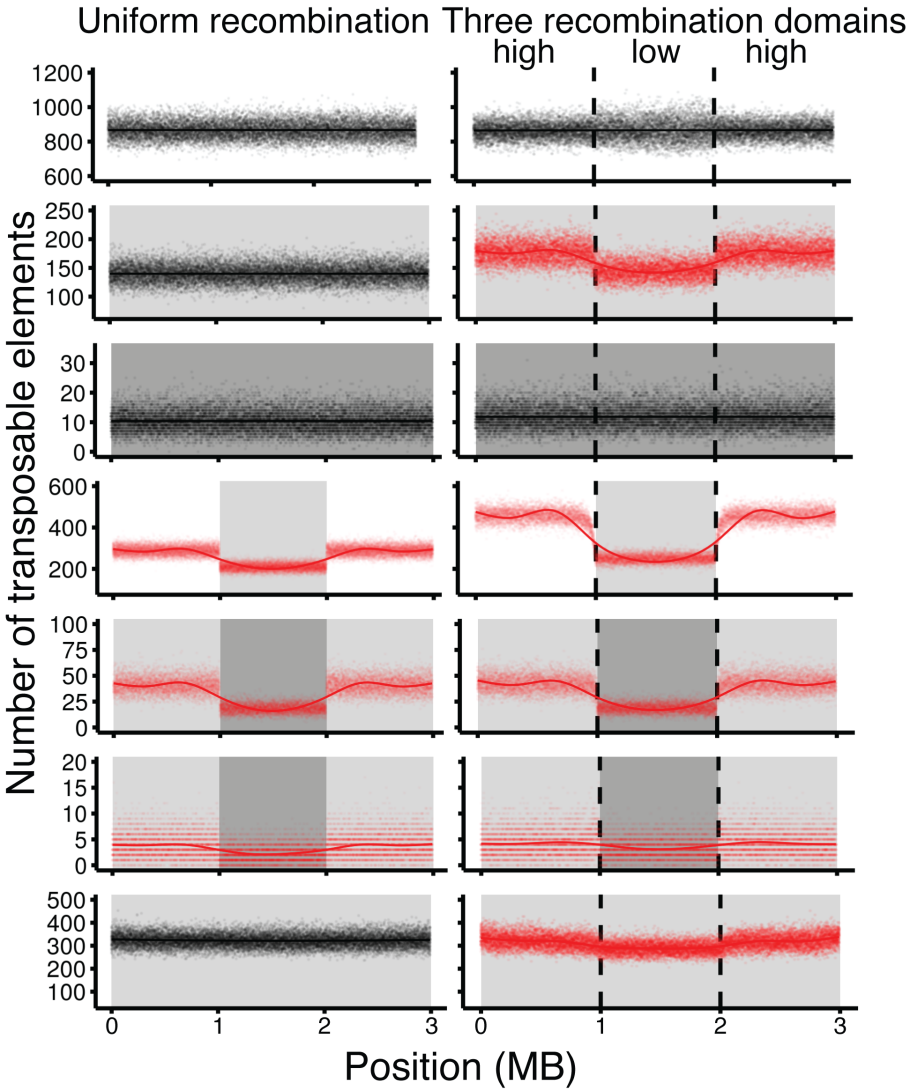
C. inopinata has lost an ancestral genomic organization of protein-coding genes.

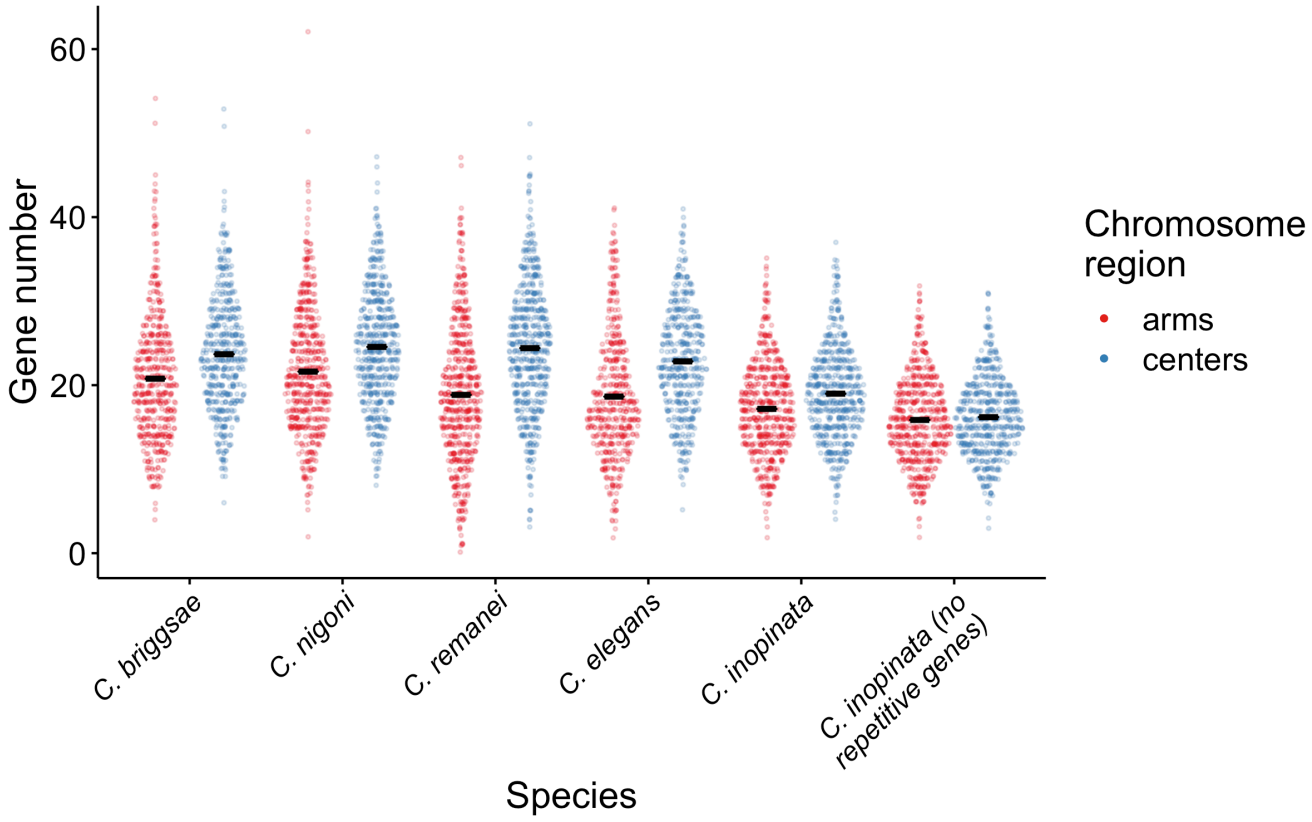
Concluding Thoughts



Genomic clusters of insertion fitness effects are sufficient for generating repetitive landscapes, although domains of varying recombination can cause such landscapes when selection is weak.

Concluding Thoughts

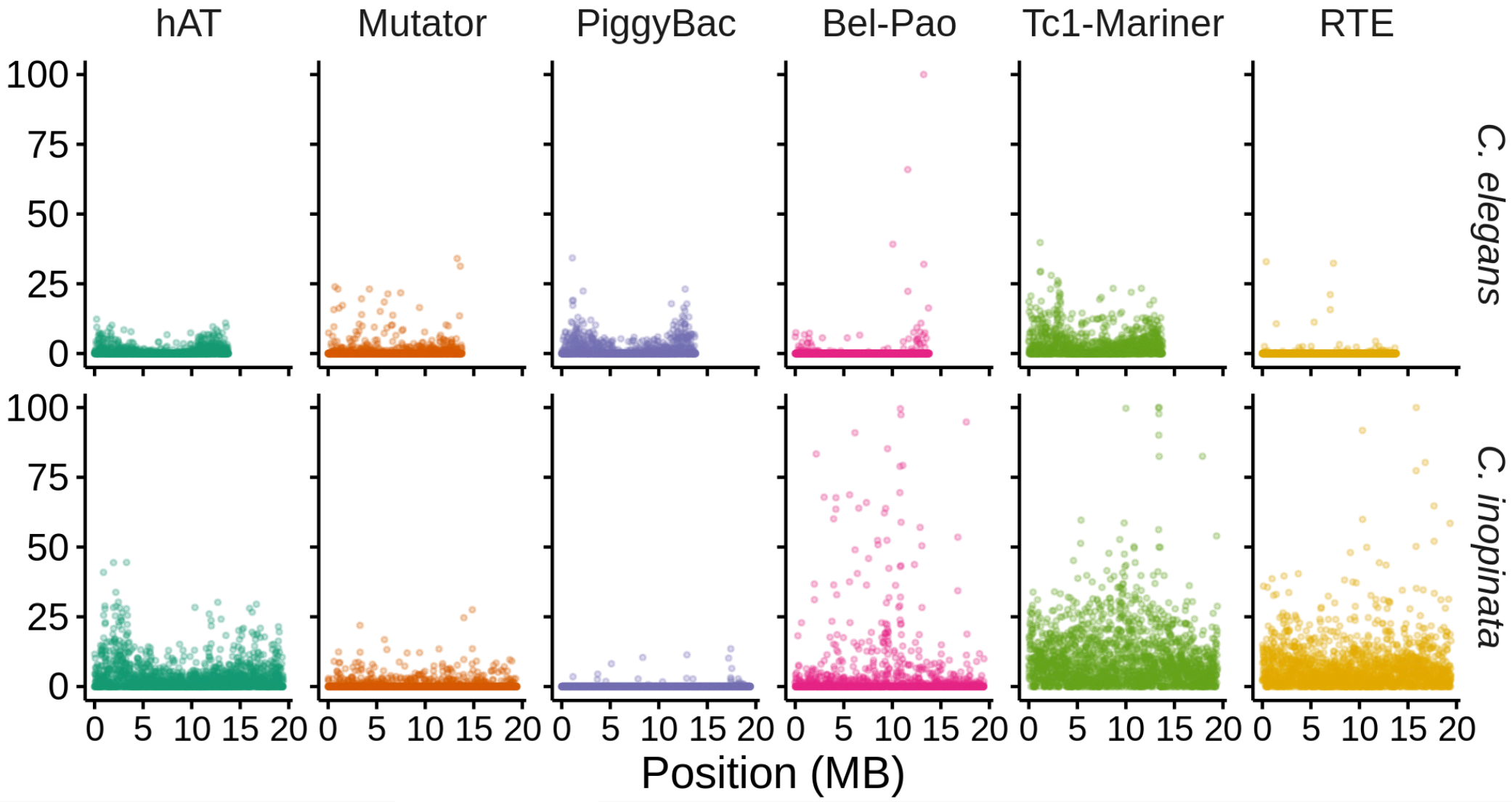




Genomic clusters of insertion fitness effects are sufficient for generating repetitive landscapes, although domains of varying recombination can cause such landscapes when selection is weak.

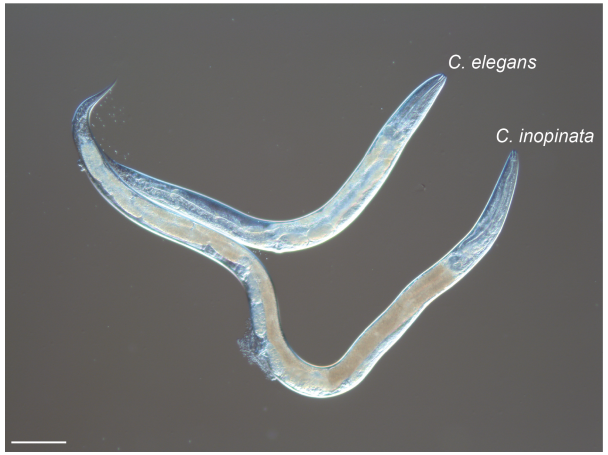
Relaxed selection on certain repeat types due to the loss of genetic clusters can explain the genomic repetitive landscape of *C. inopinata*…

Concluding Thoughts

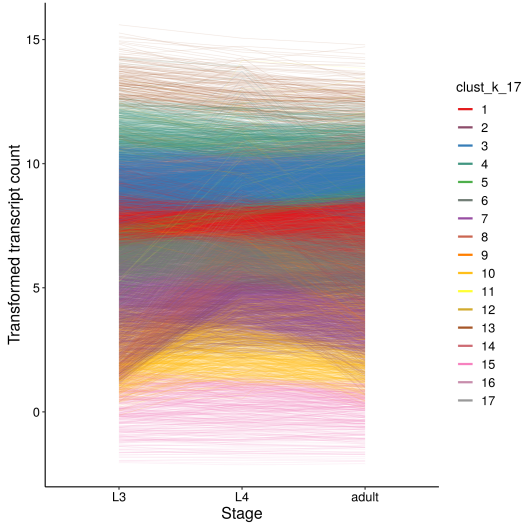


…but other historical or genomic factors are needed to explain the idiosyncrasy of genomic organization of various transposable element taxa within *C. inopinata.*

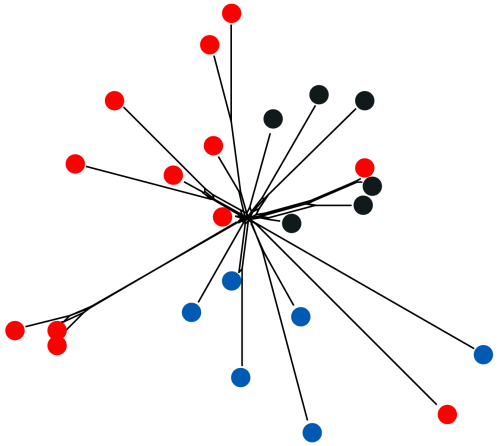
Ongoing projects in *C. inopinata* biology



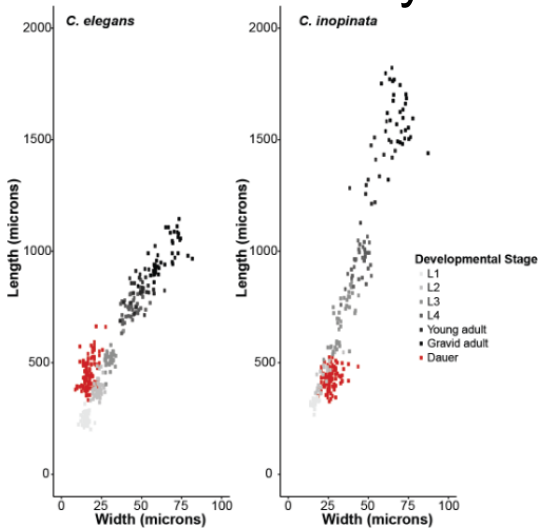
Evo-devo of body size & heterochrony



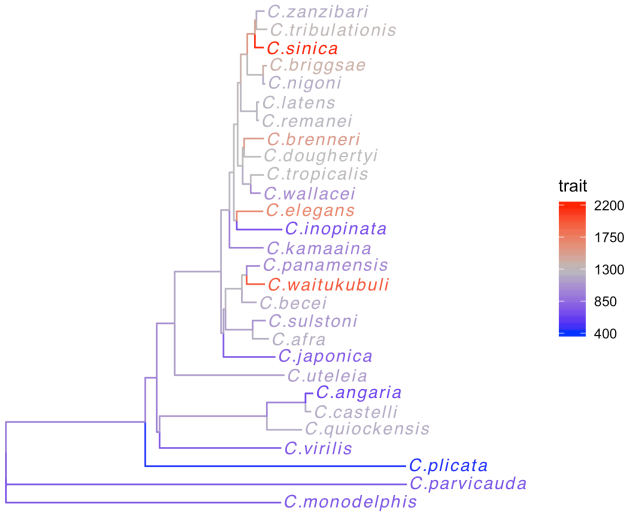
Evolution of gene expression



Population genomics



Phenotypic plasticity

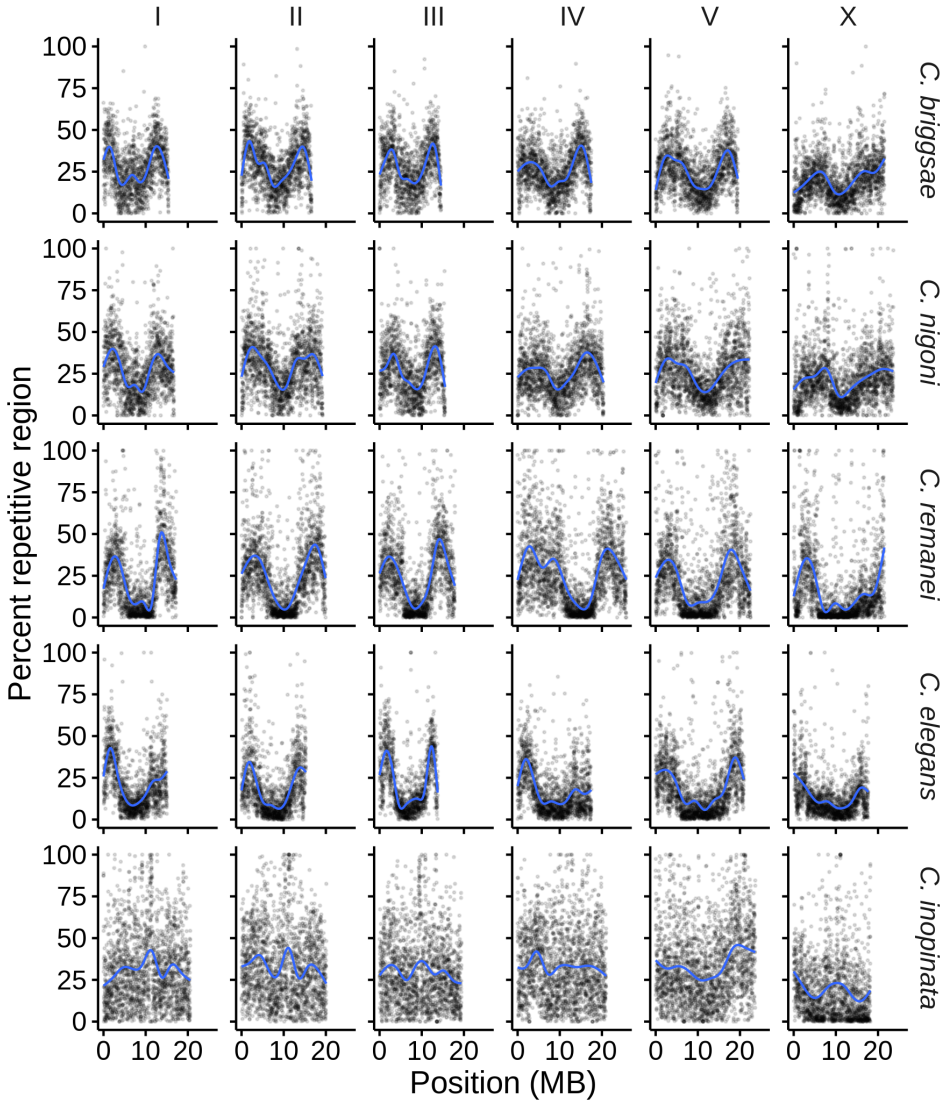


Phylogenetic comparative genomics

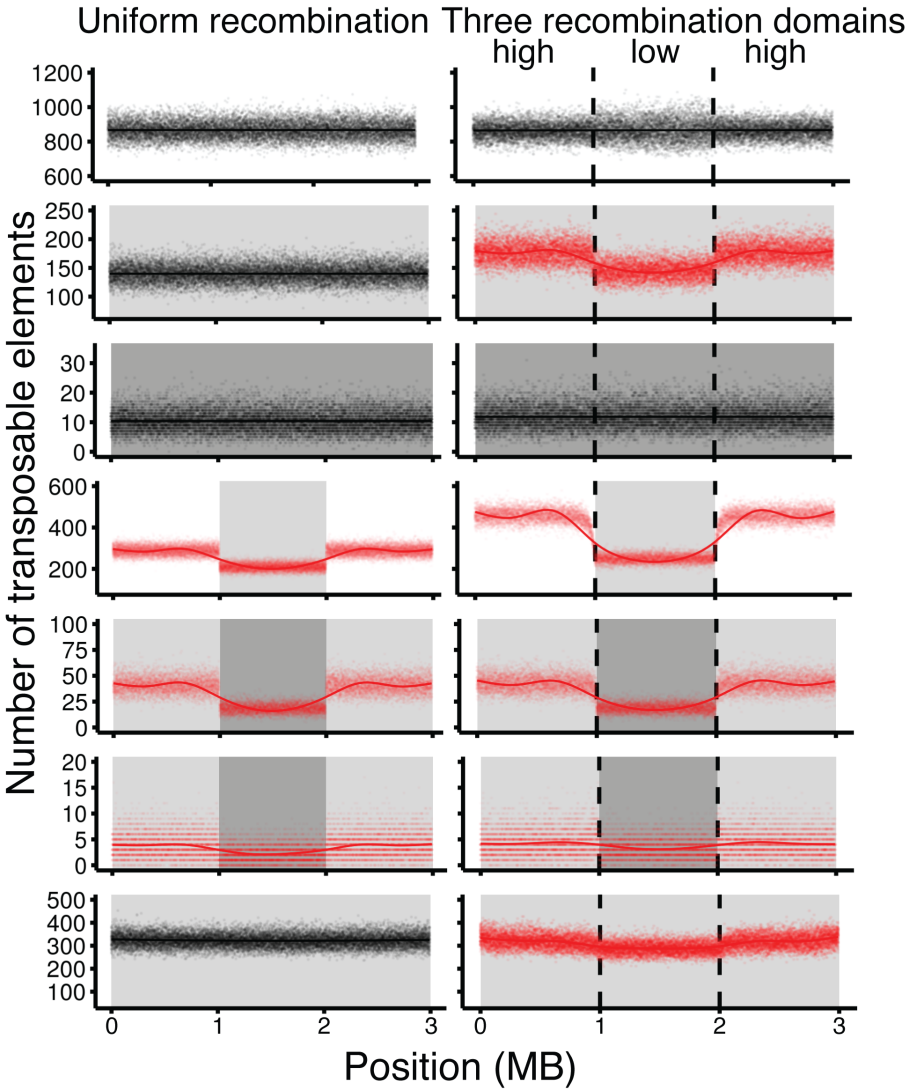
Microbial evolutionary ecology

Please contact me if you are interested in these sorts of projects!!

Thanks for listening!



C. inopinata has a highly repetitive genomic landscape.



Genomic clusters of insertion fitness effects are sufficient for generating repetitive landscapes, although domains of varying recombination can cause such landscapes when selection is weak.

Thanks

- WormBase, caenorhabditis.org, and the worm community
- Patrick Phillips
- Jeff Adrion
- Peter Ralph and his lab members
- Andy Kern and his lab members