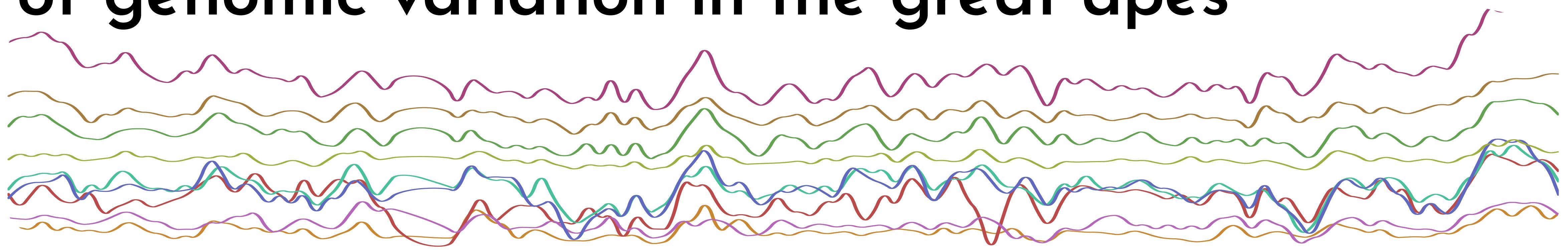


# Shared evolutionary processes shape landscapes of genomic variation in the great apes



Murillo F. Rodrigues

2023 Society for Molecular Biology & Evolution meeting

July, 2023

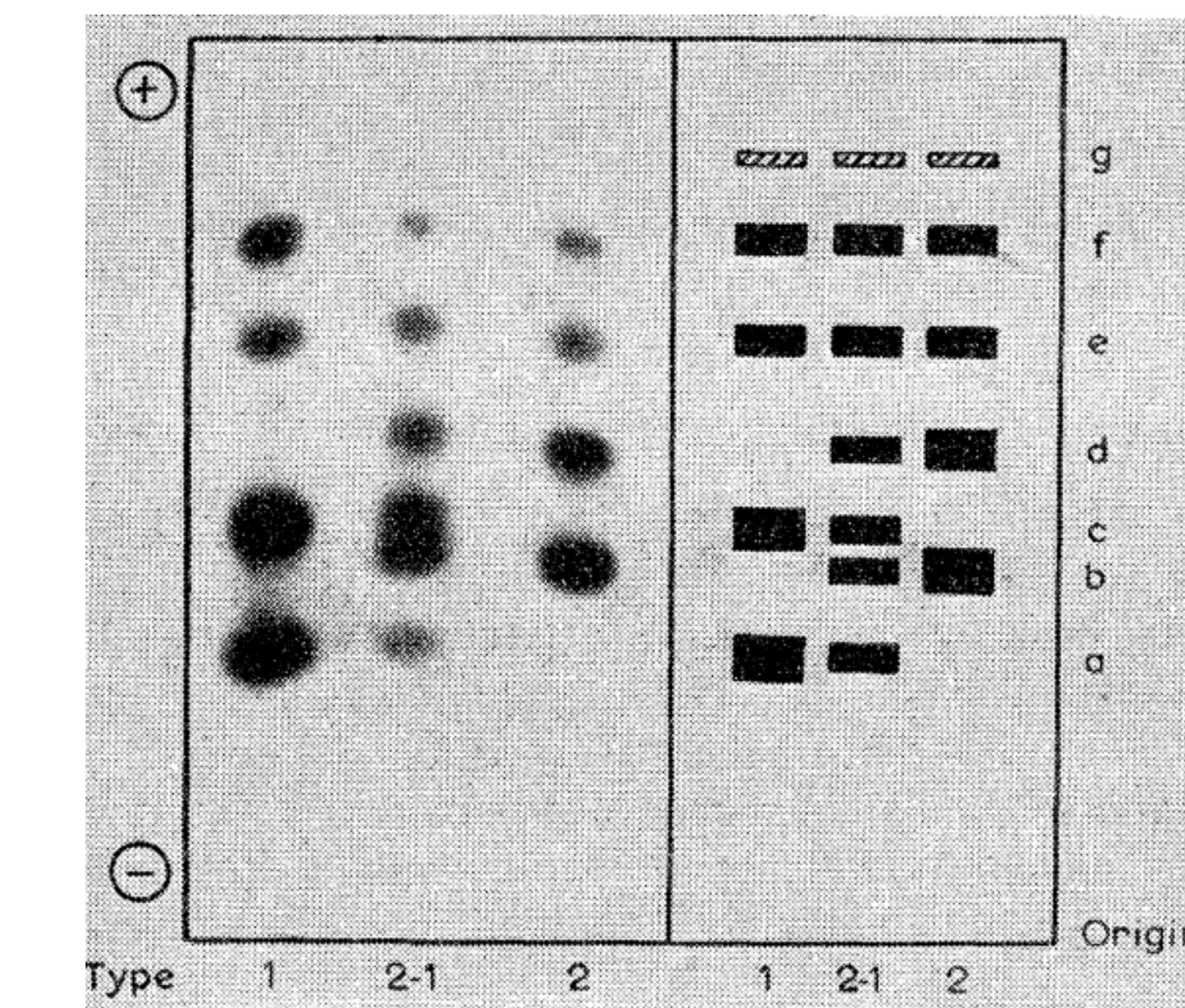
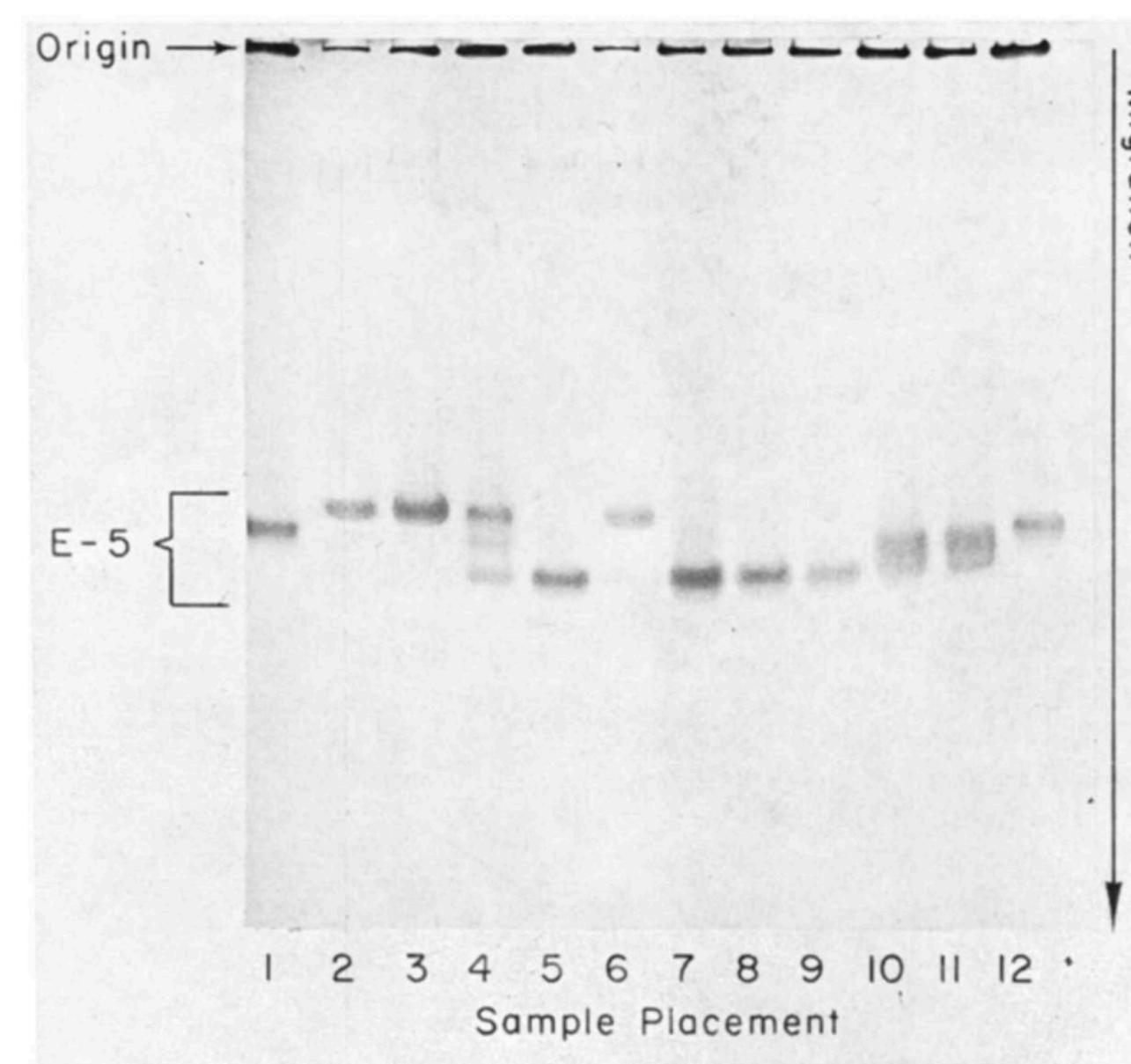
# Genetic variation

# Genetic variation

Few, independent loci carried sufficient information to infer some evolutionary processes!

582

J. L. HUBBY AND R. C. LEWONTIN

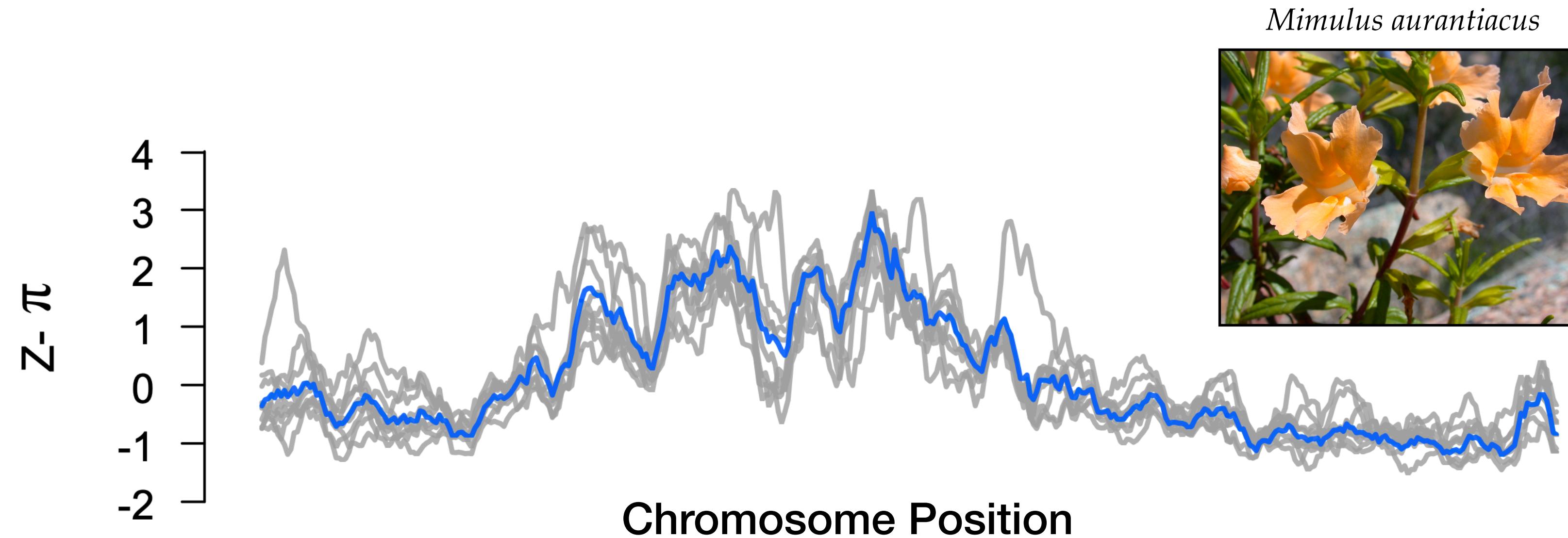


# Genomic patterns of variation within groups of species

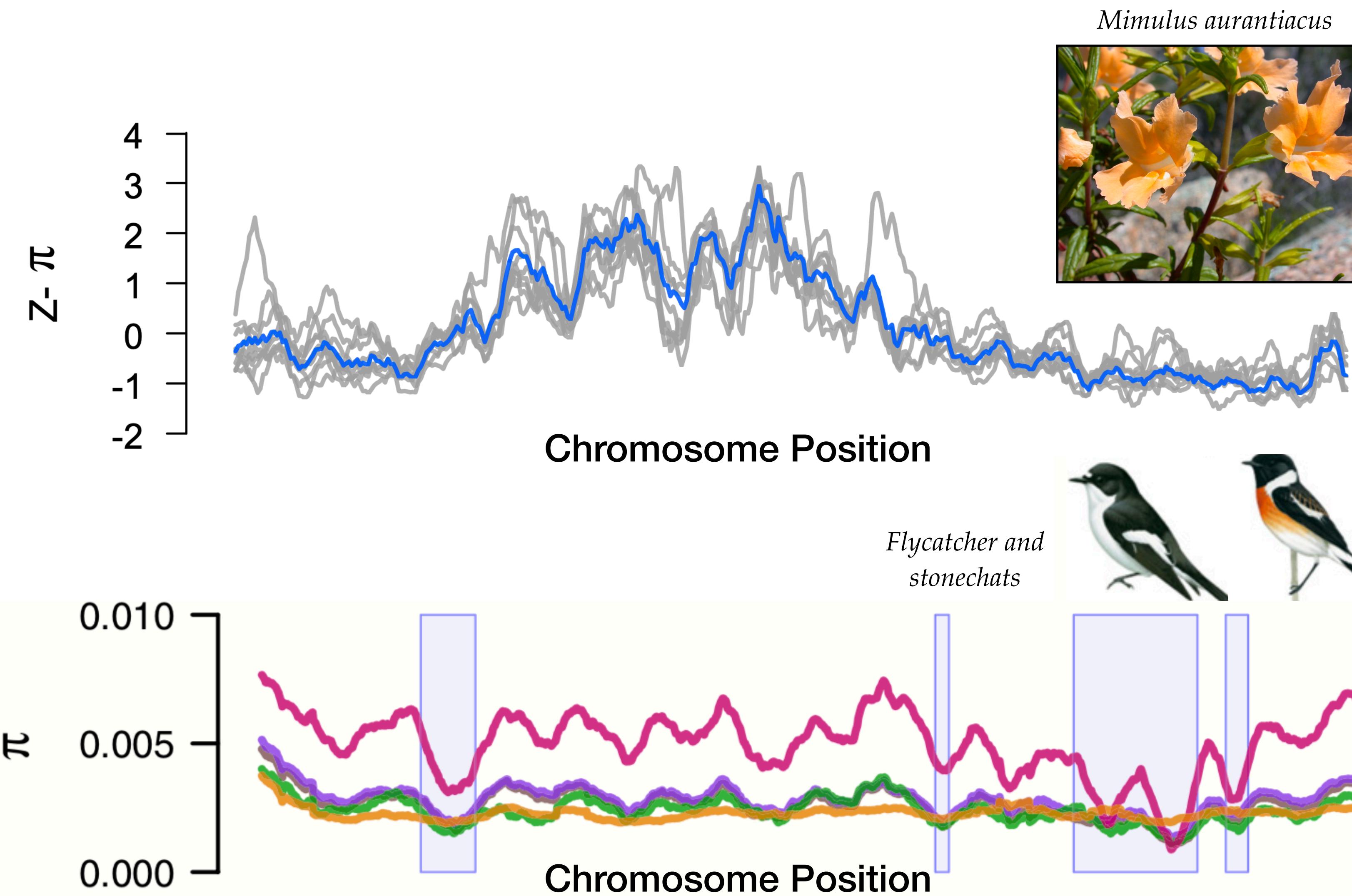
But now, we have access to chromosome-level data from multiple species, so we can also leverage linkage relations across the chromosome!

# Genomic patterns of variation within groups of species

But now, we have access to chromosome-level data from multiple species, so we can also leverage linkage relations across the chromosome!

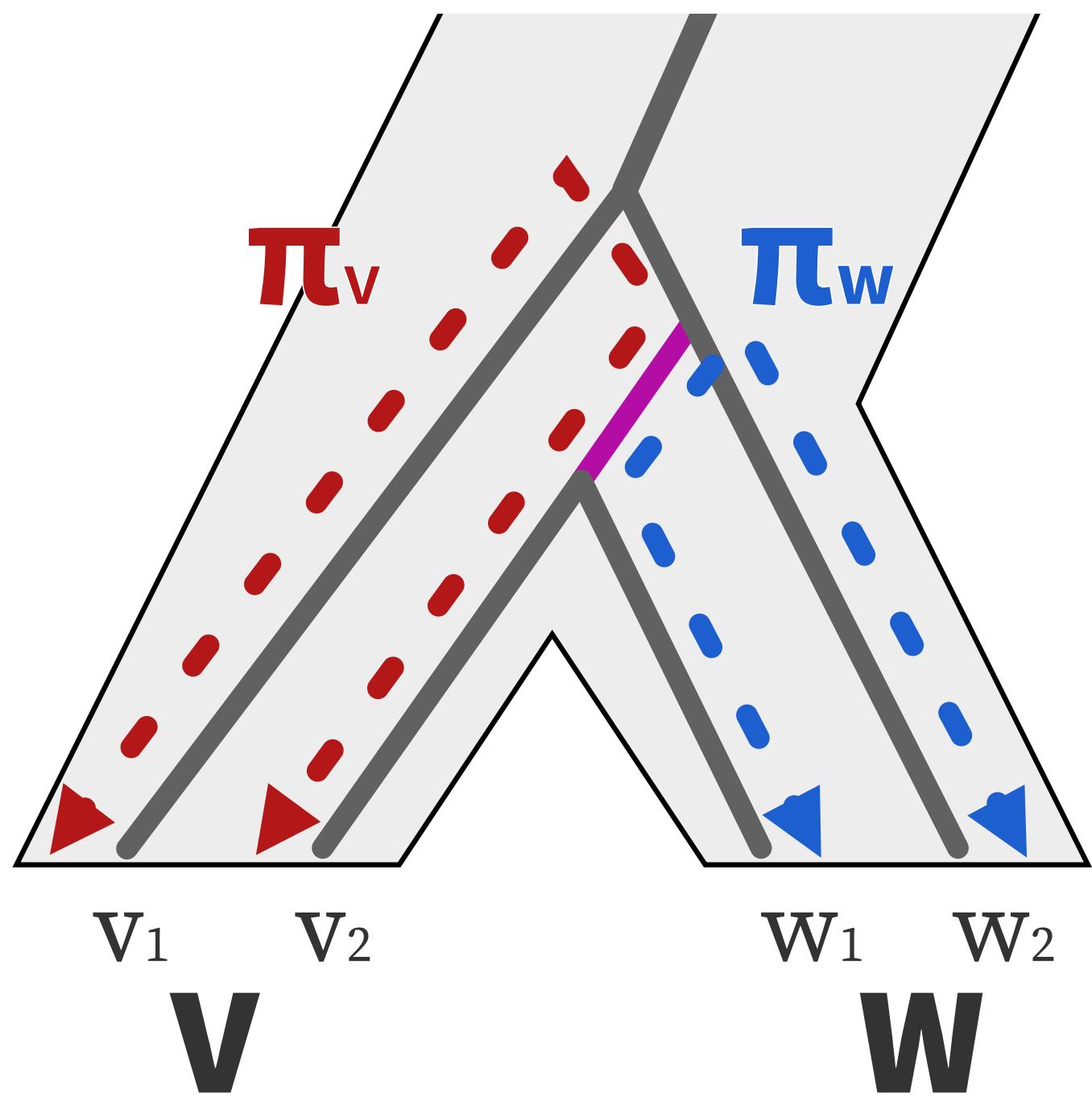


# Genomic patterns of variation within groups of species



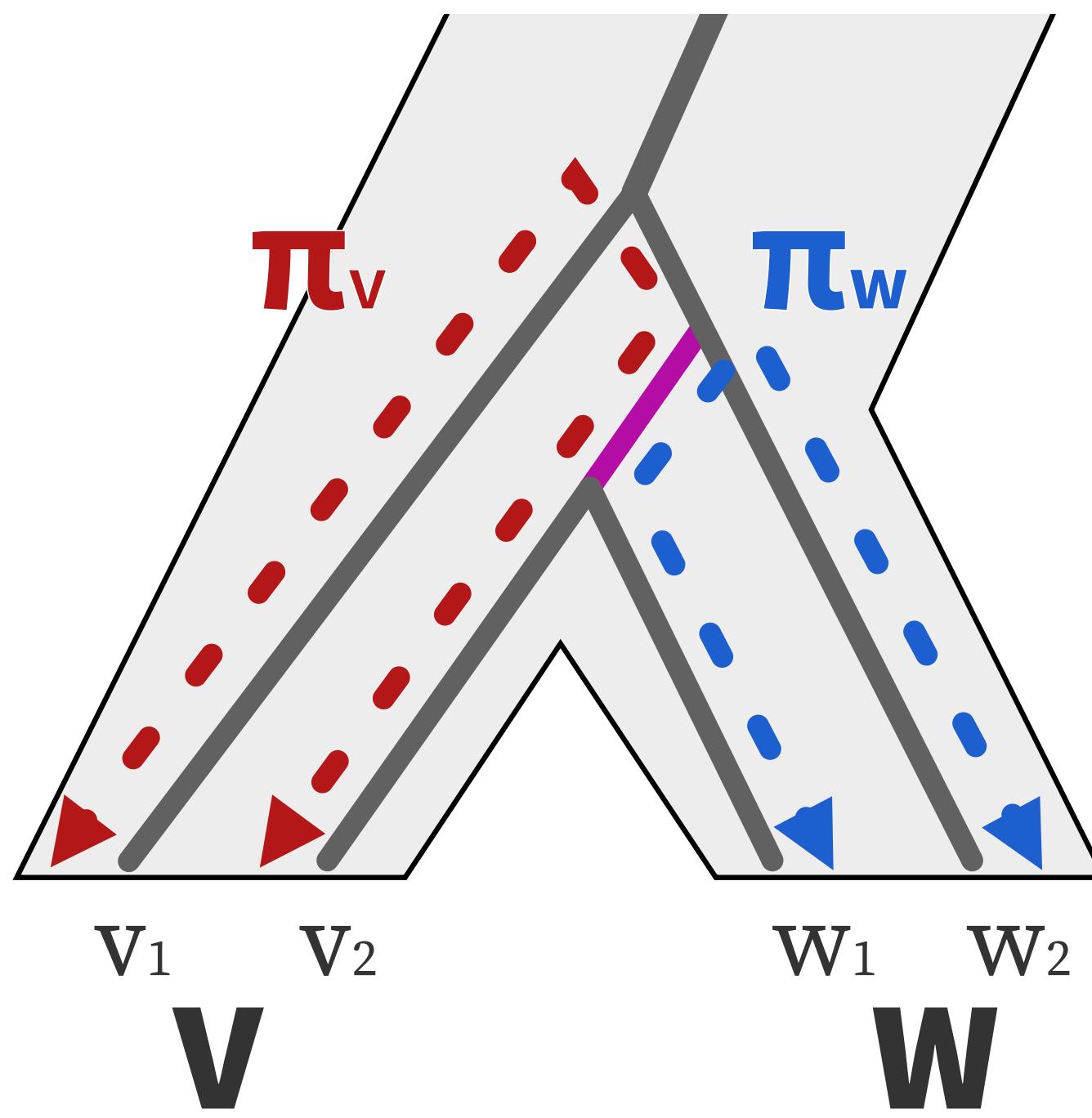
# What can cause correlated patterns of variation?

- Sorting of variation (ancestral variation and incomplete lineage sorting)

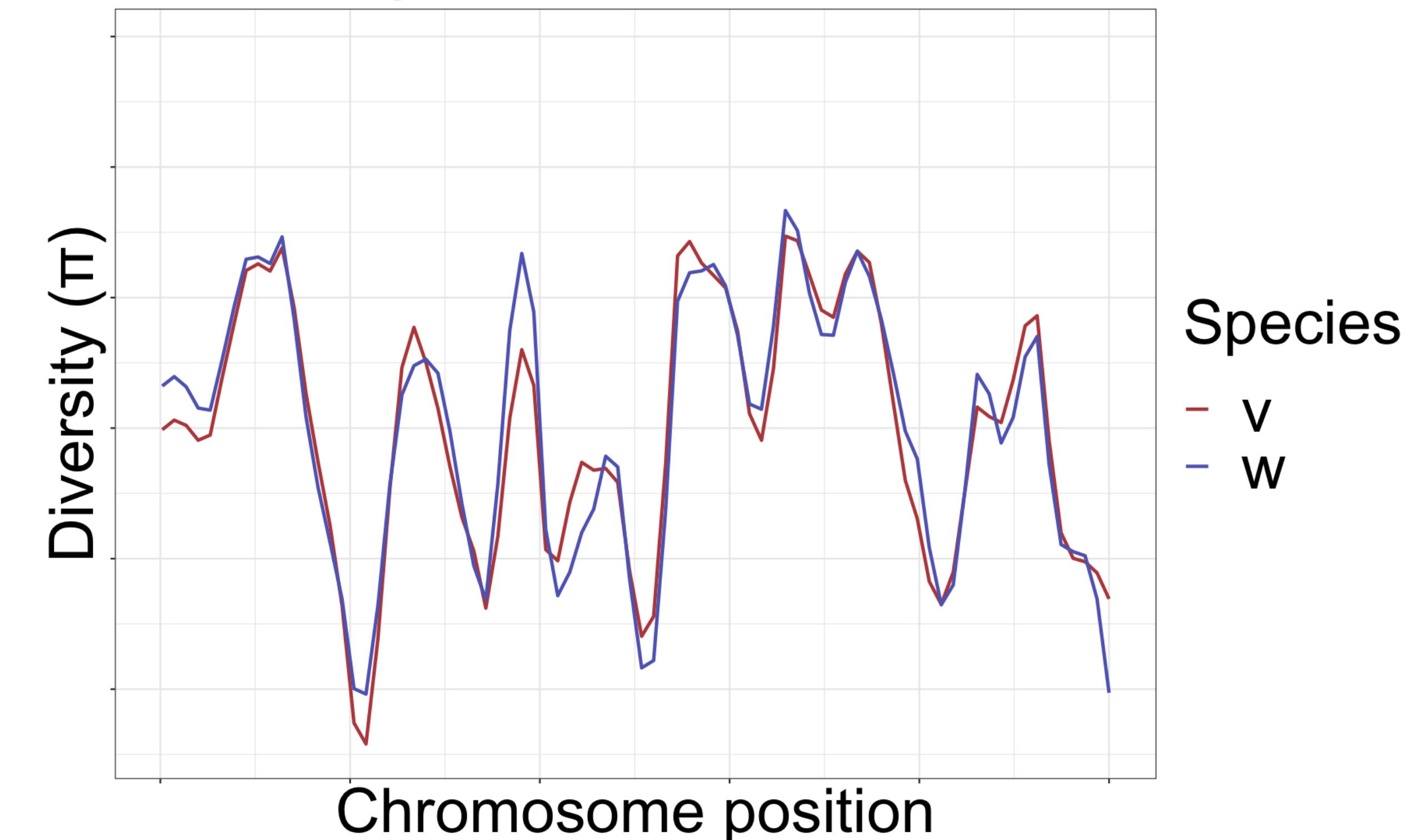


# What can cause correlated patterns of variation?

- Sorting of variation (ancestral variation and incomplete lineage sorting)



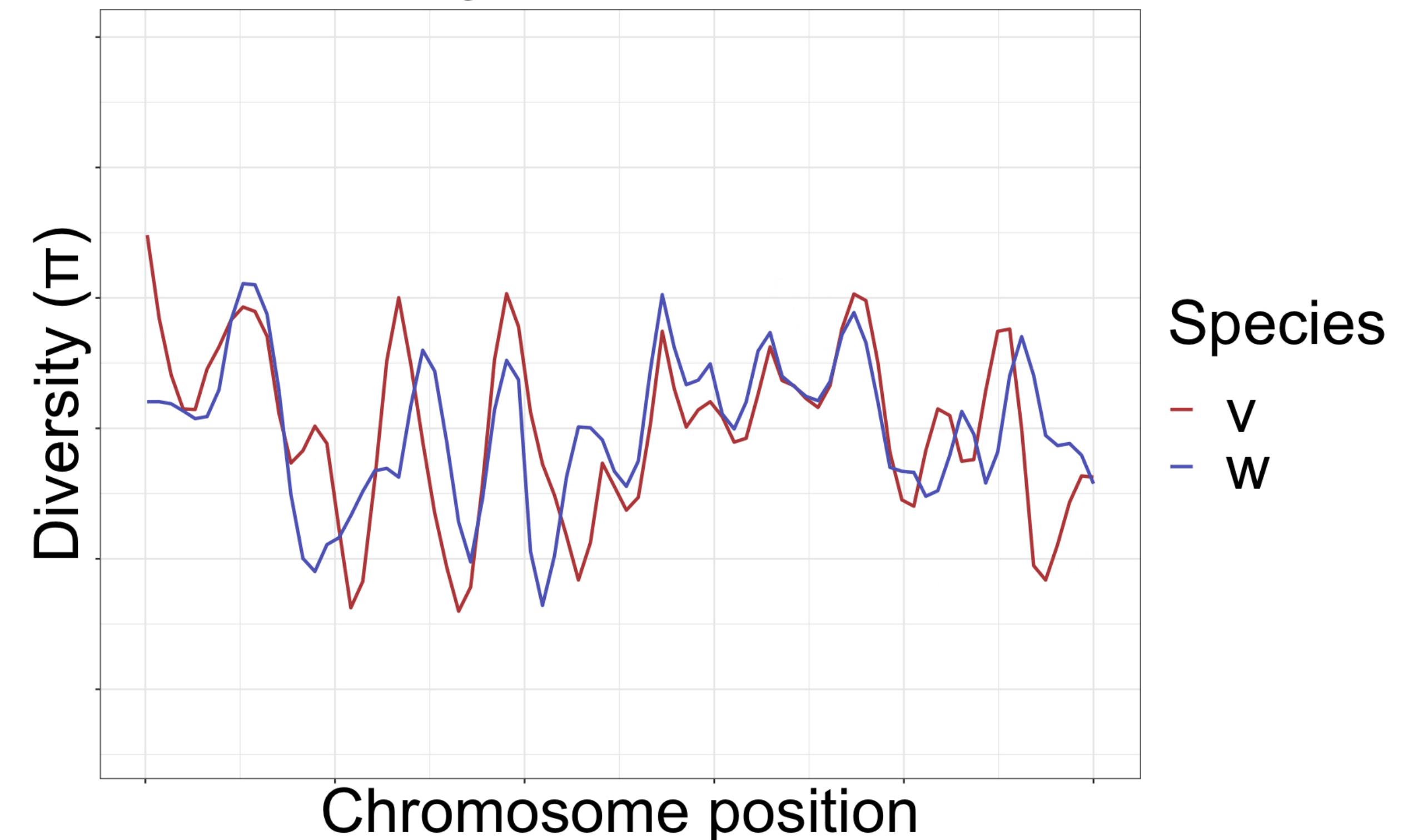
Time: 0N generations



# What can cause correlated patterns of variation?

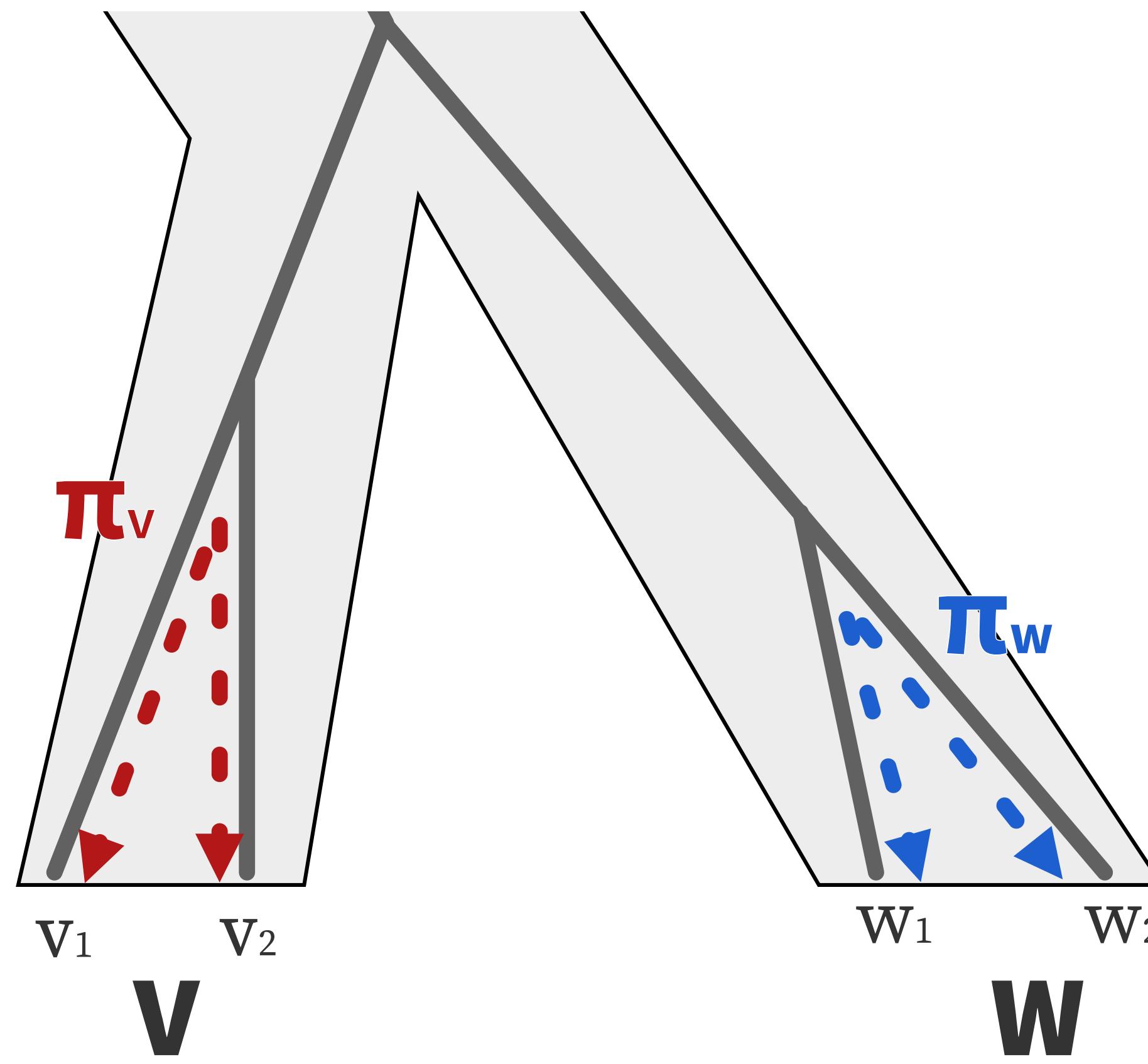
- Sorting of variation (ancestral variation and incomplete lineage sorting)

Time: 2.02N generations

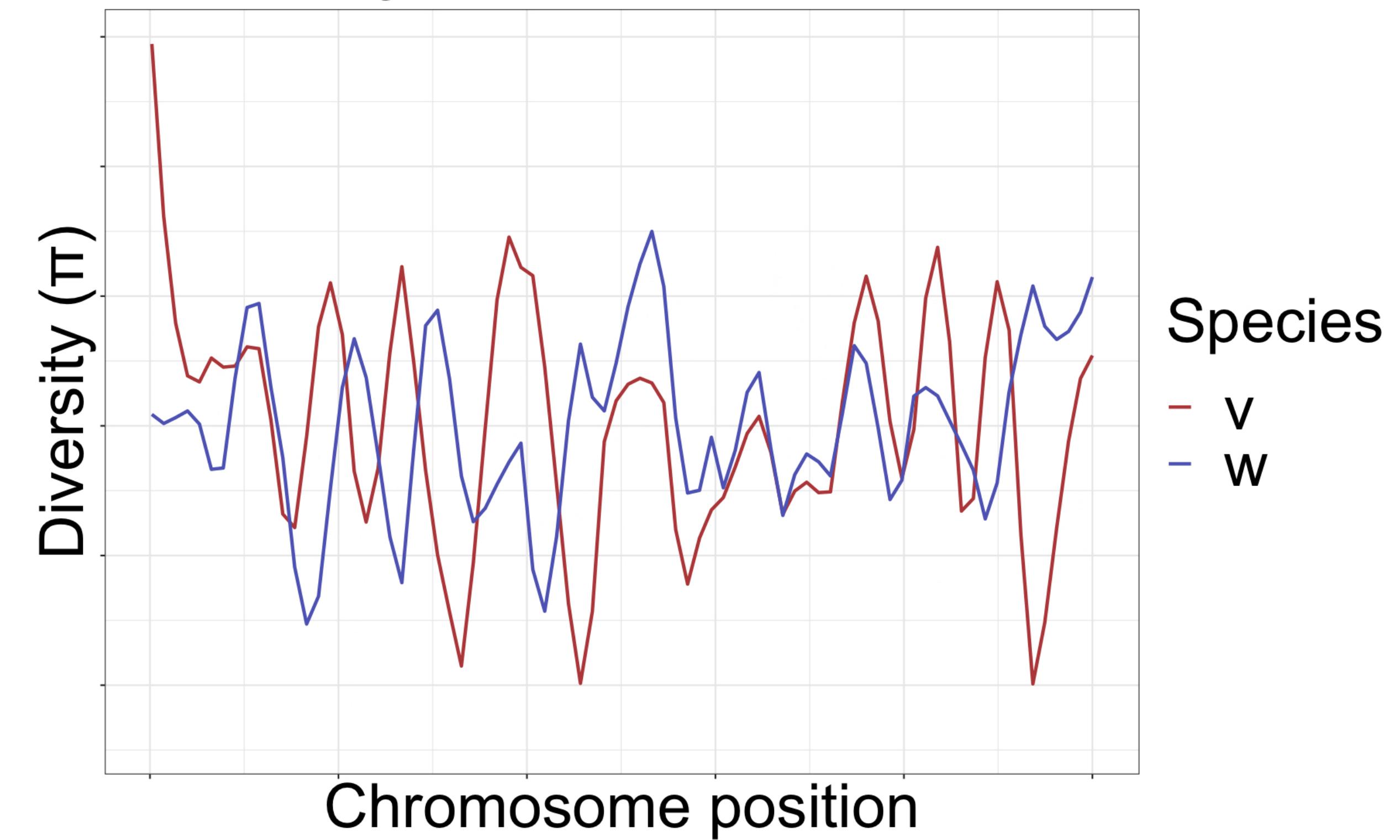


# What can cause correlated patterns of variation?

- Sorting of variation (ancestral variation and incomplete lineage sorting)

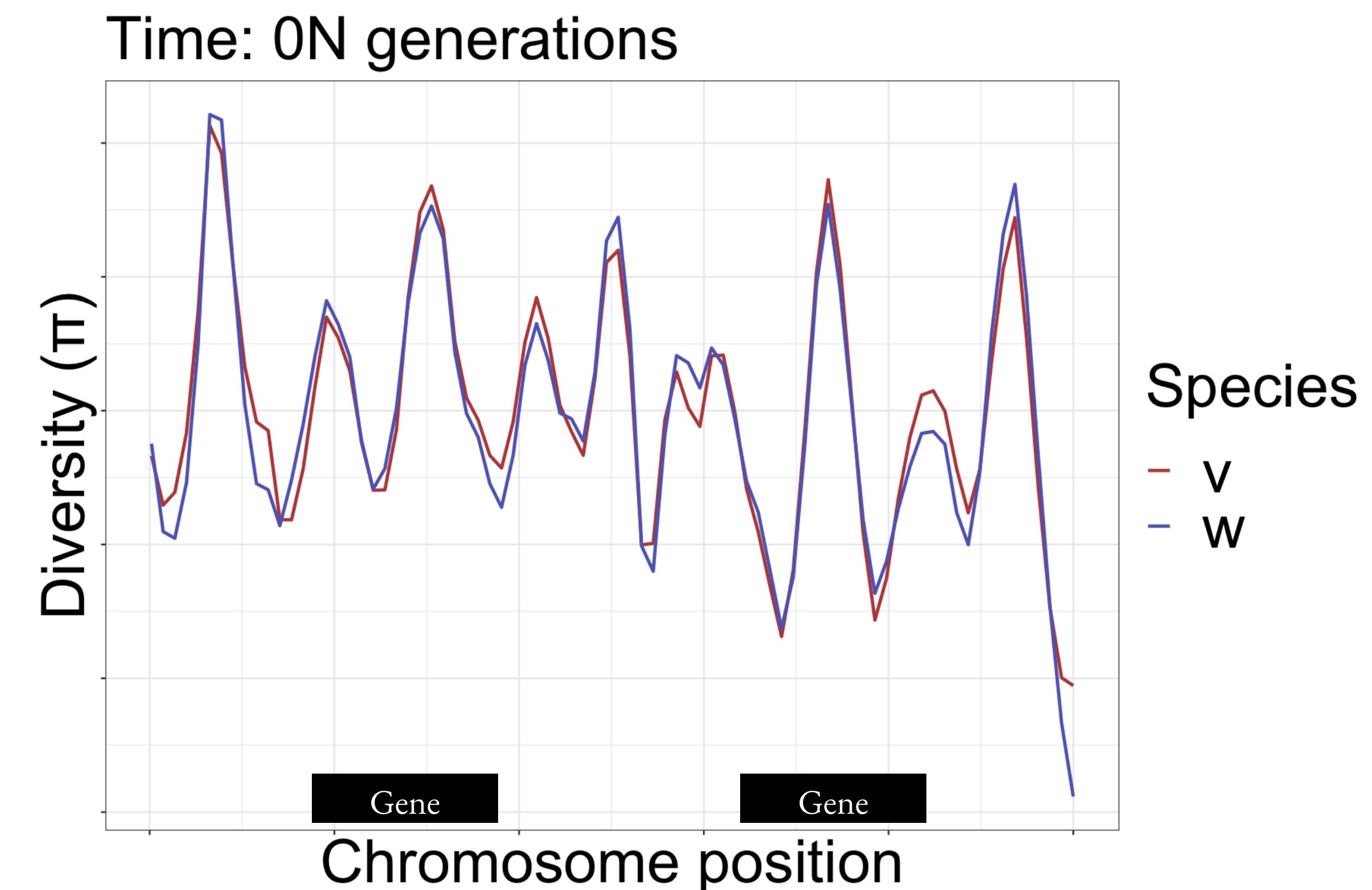


Time: 4N generations



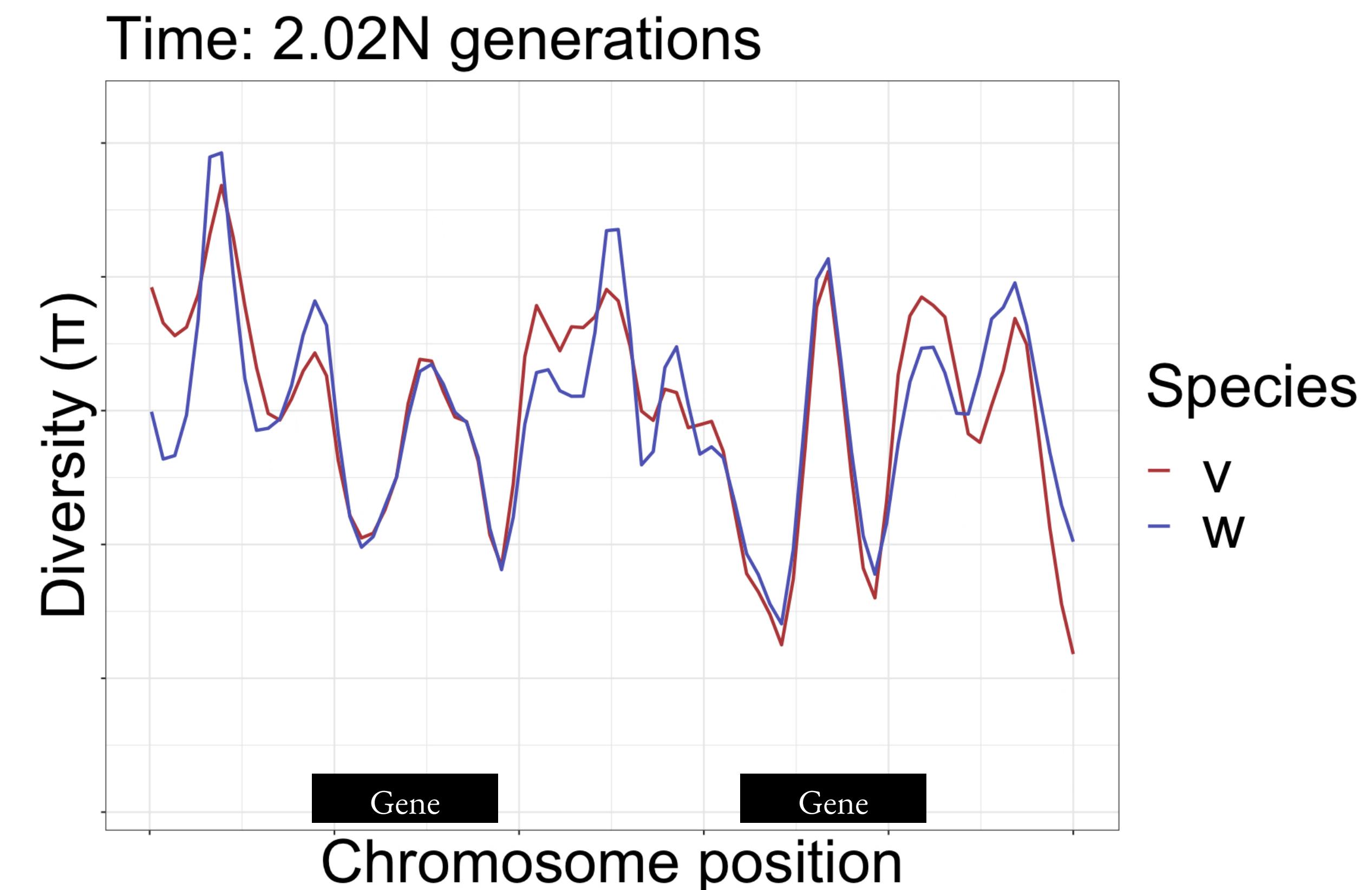
# What can cause correlated patterns of variation?

- Sorting of variation
- Shared process
  - Positive and/or negative selection
  - Mutation rate variation
  - GC-biased gene conversion



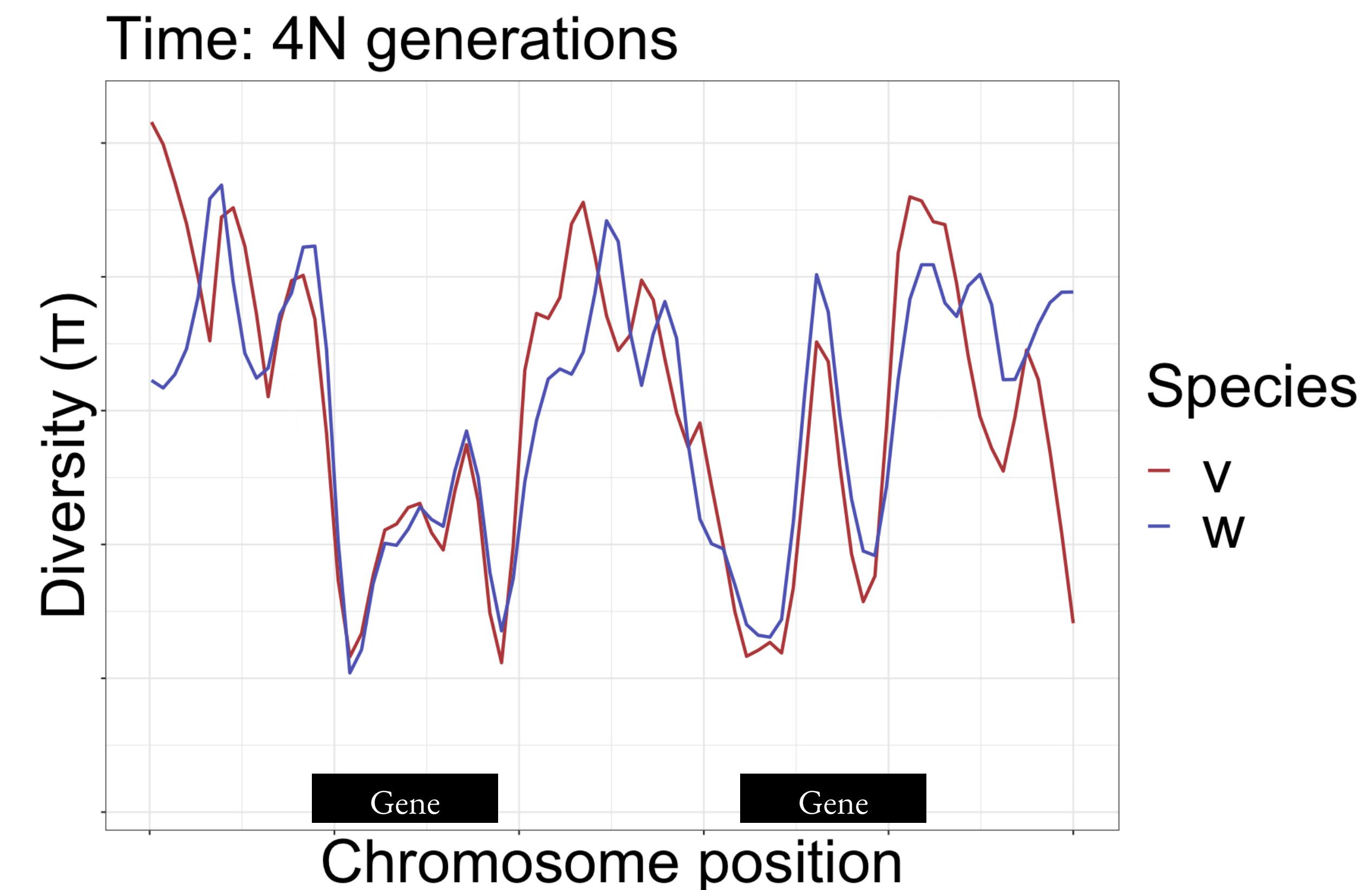
# What can cause correlated patterns of variation?

- Sorting of variation
- Shared process
  - Positive and/or negative selection
  - Mutation rate variation
  - GC-biased gene conversion



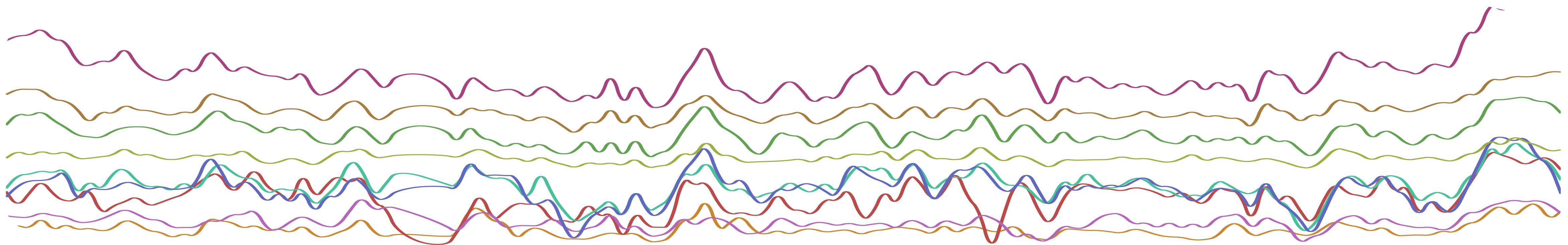
# What can cause correlated patterns of variation?

- Sorting of variation
- Shared process
  - Positive and/or negative selection
  - Mutation rate variation
  - GC-biased gene conversion



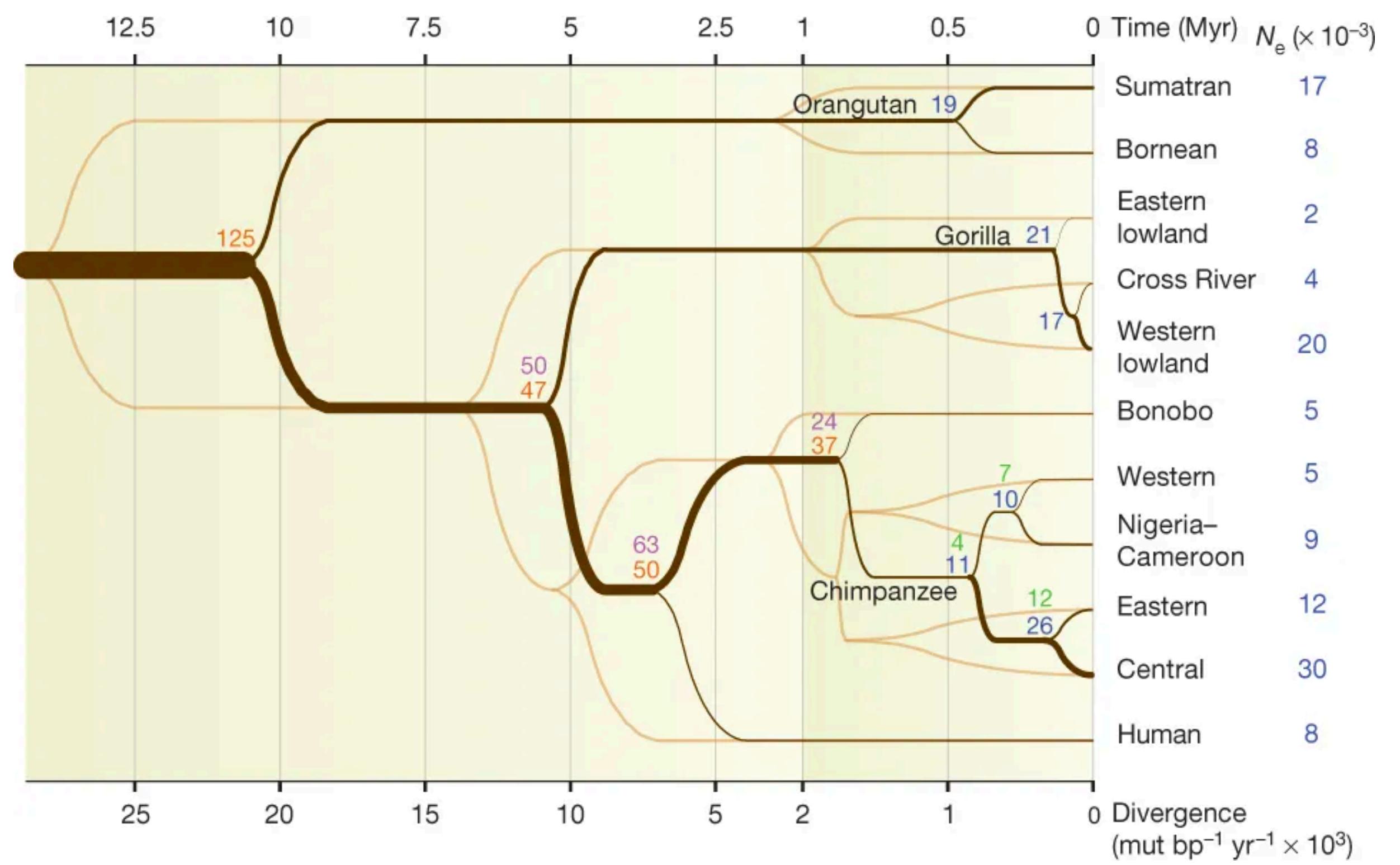
# Goals

- Describe whether and in what ways landscapes of diversity (and divergence) are correlated
- Understand how correlations between landscapes can be leveraged to tease apart the relative roles of positive and negative selection and other processes in shaping patterns of genetic variation

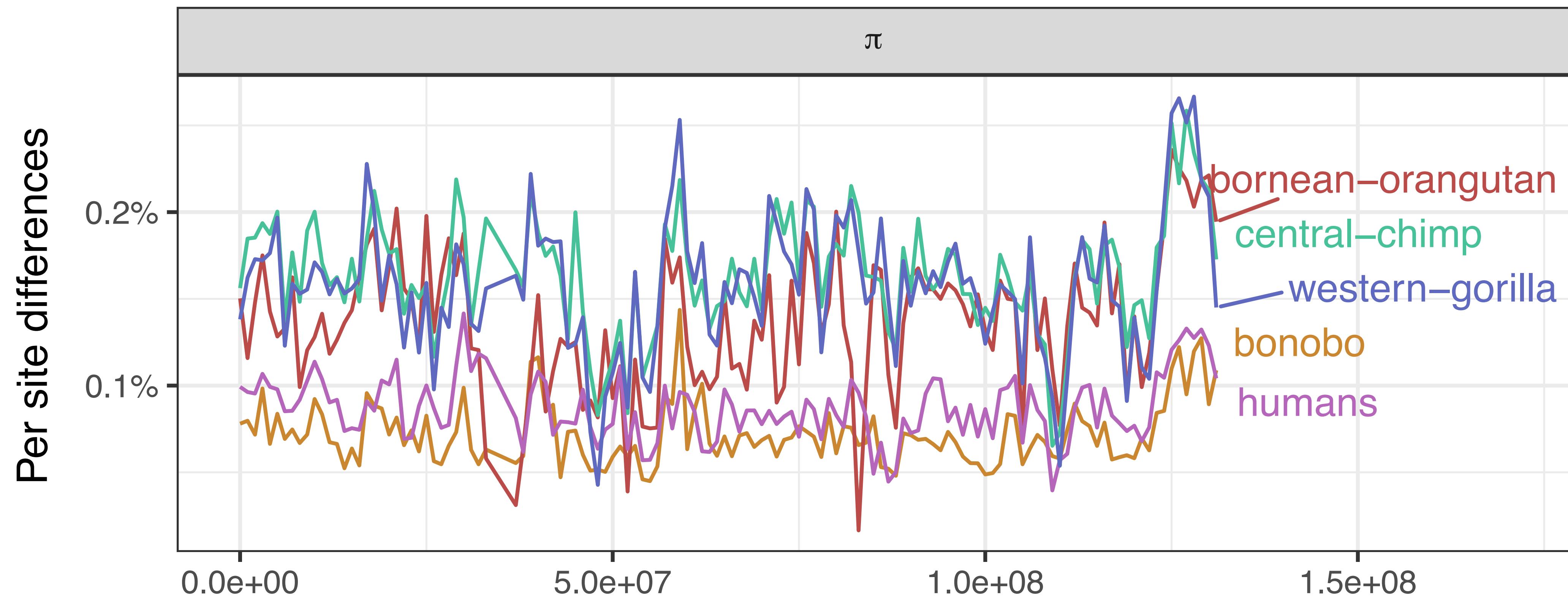


# Model system: the great apes

- High quality genomic data for 5 species
- Deep divergence, spanning over  $\sim 60N$  generations!
- Conservation of genes, recombination rate and other genomic features

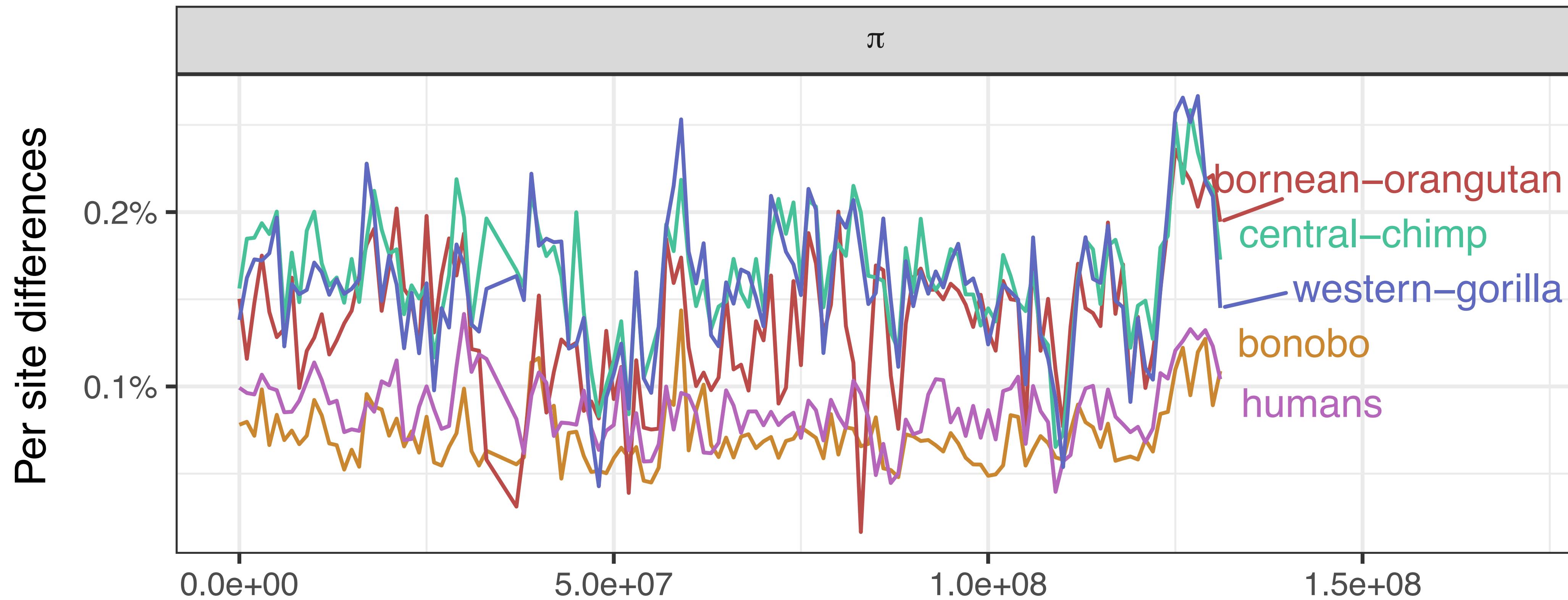


# Great apes' landscapes of genomic variation



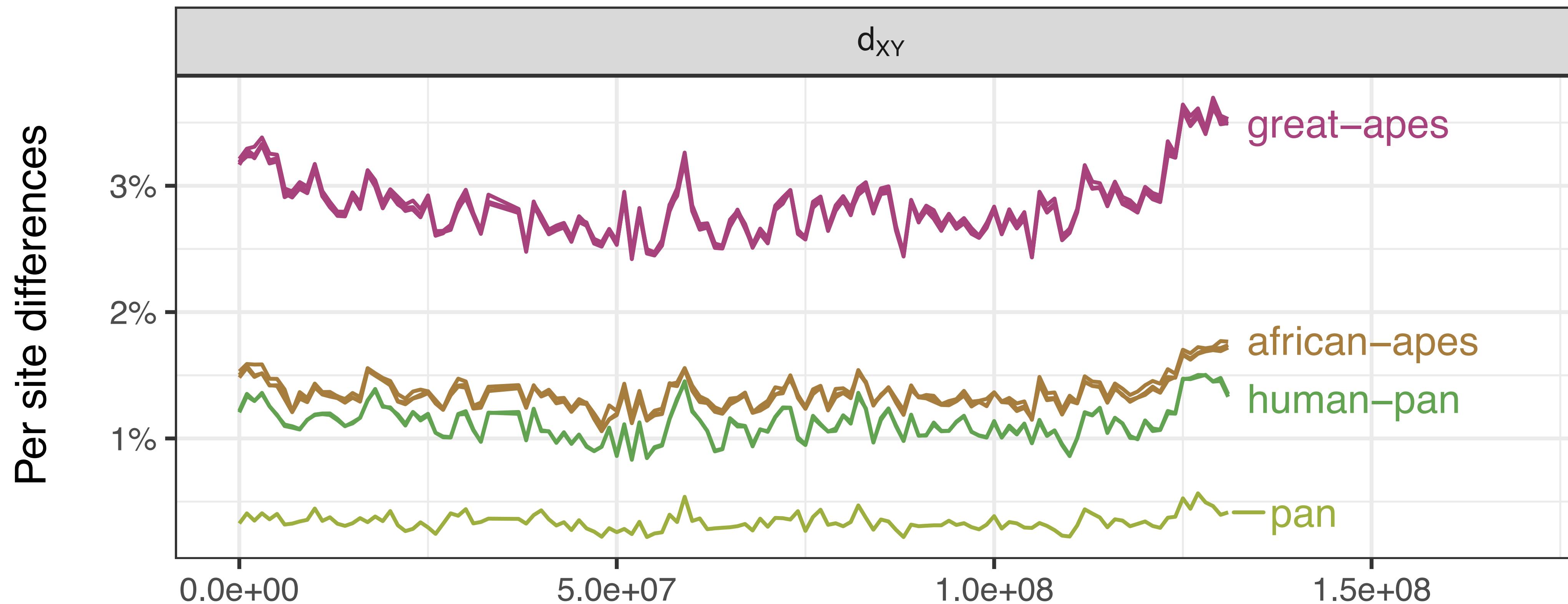
# Great apes' landscapes of genomic variation

Peaks and troughs seem to be shared across species!

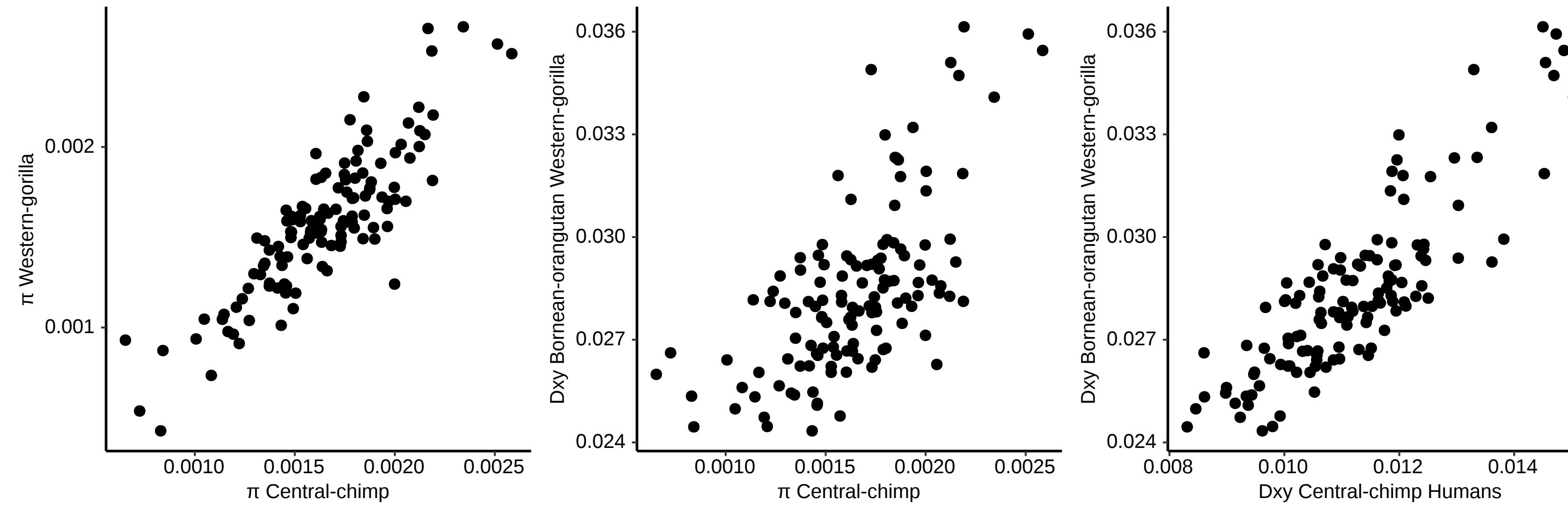


# Great apes' landscapes of genomic variation

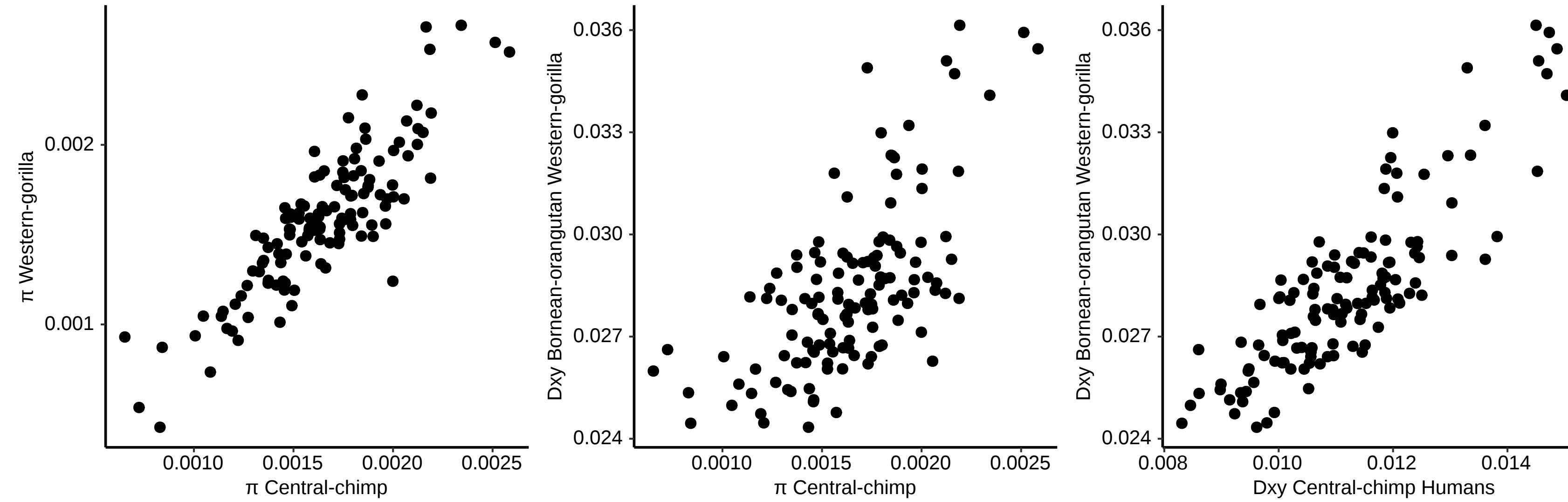
Peaks and troughs seem to be shared across species!



# Correlated landscapes of genomic variation



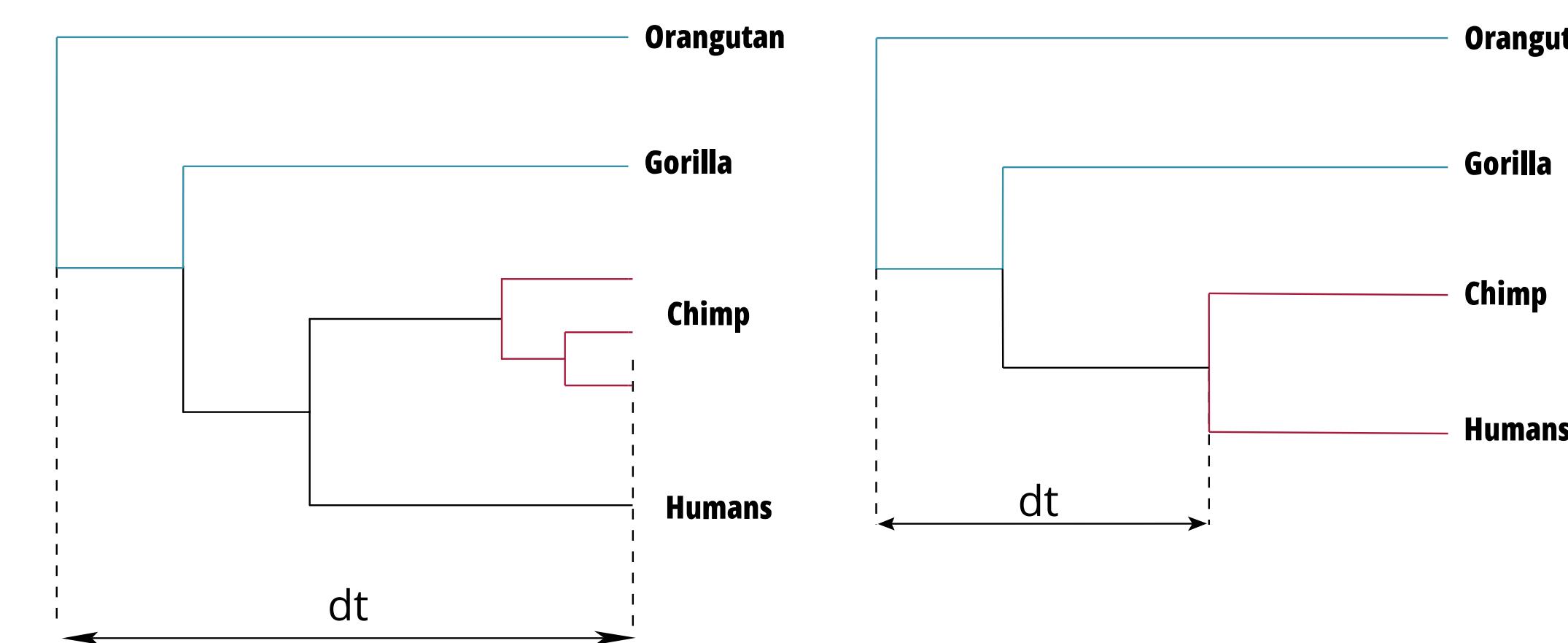
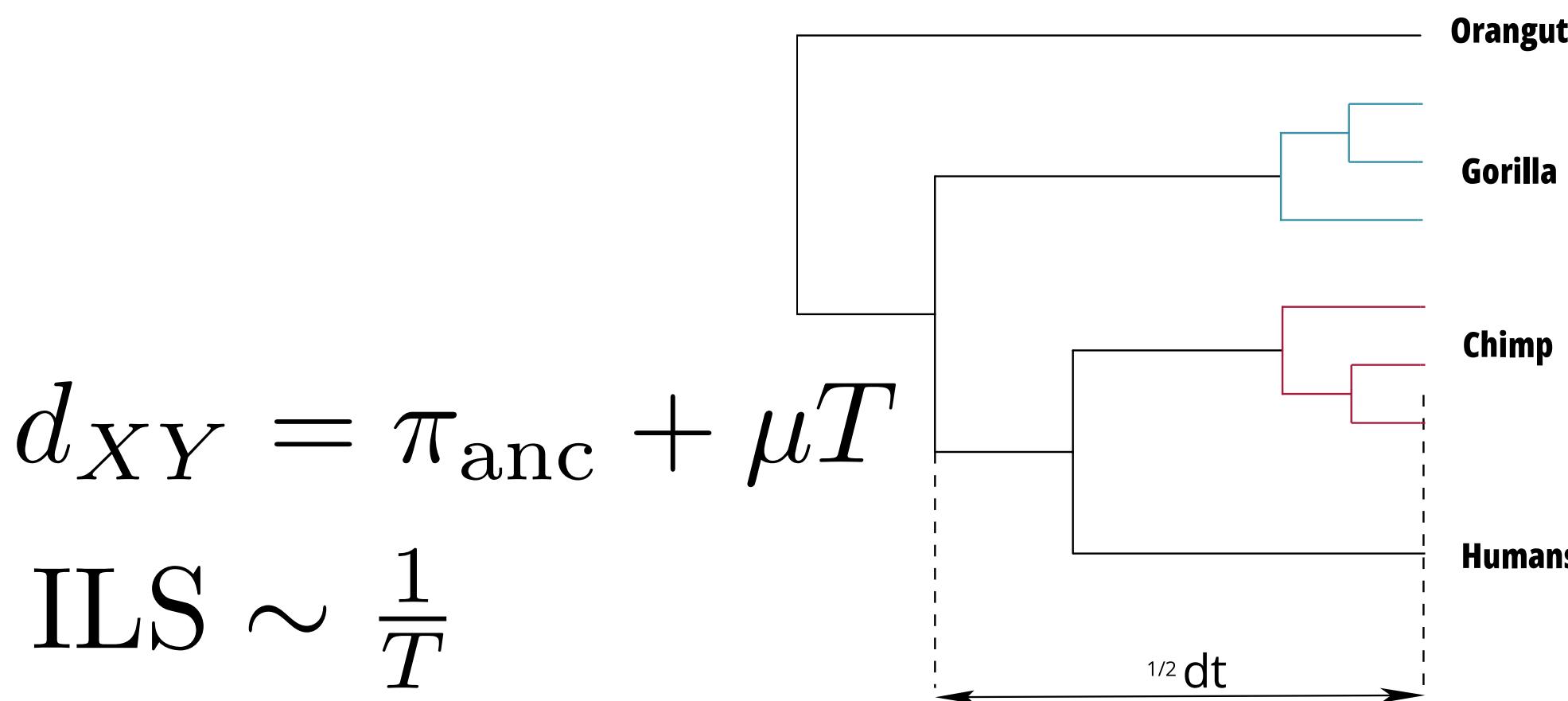
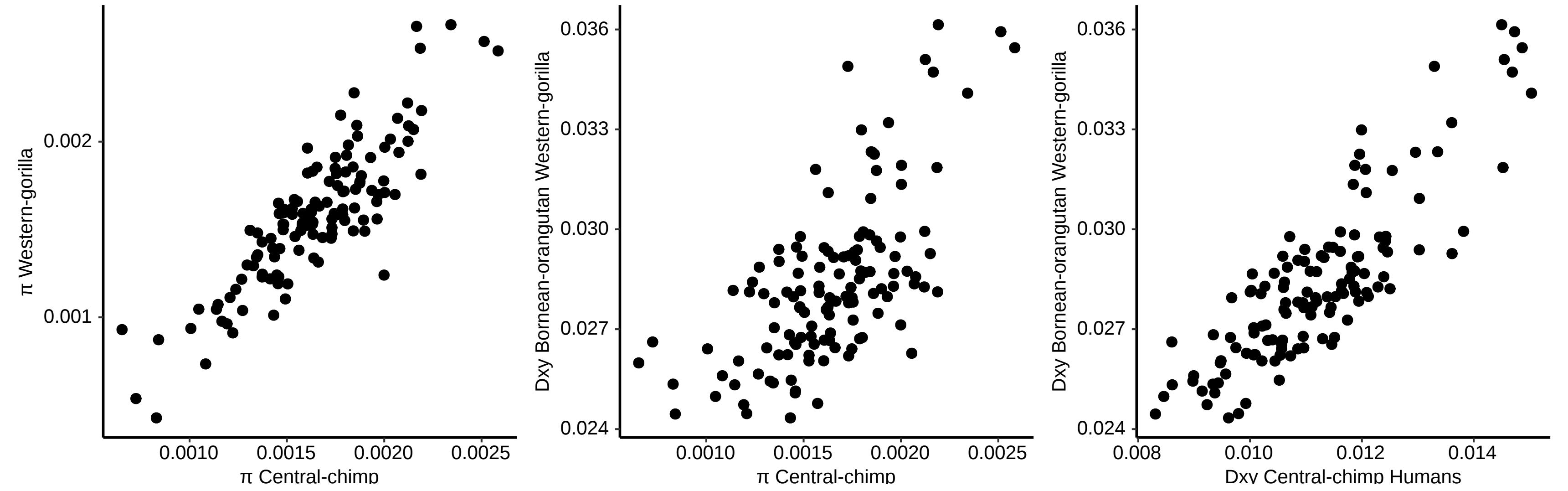
# Correlated landscapes of genomic variation



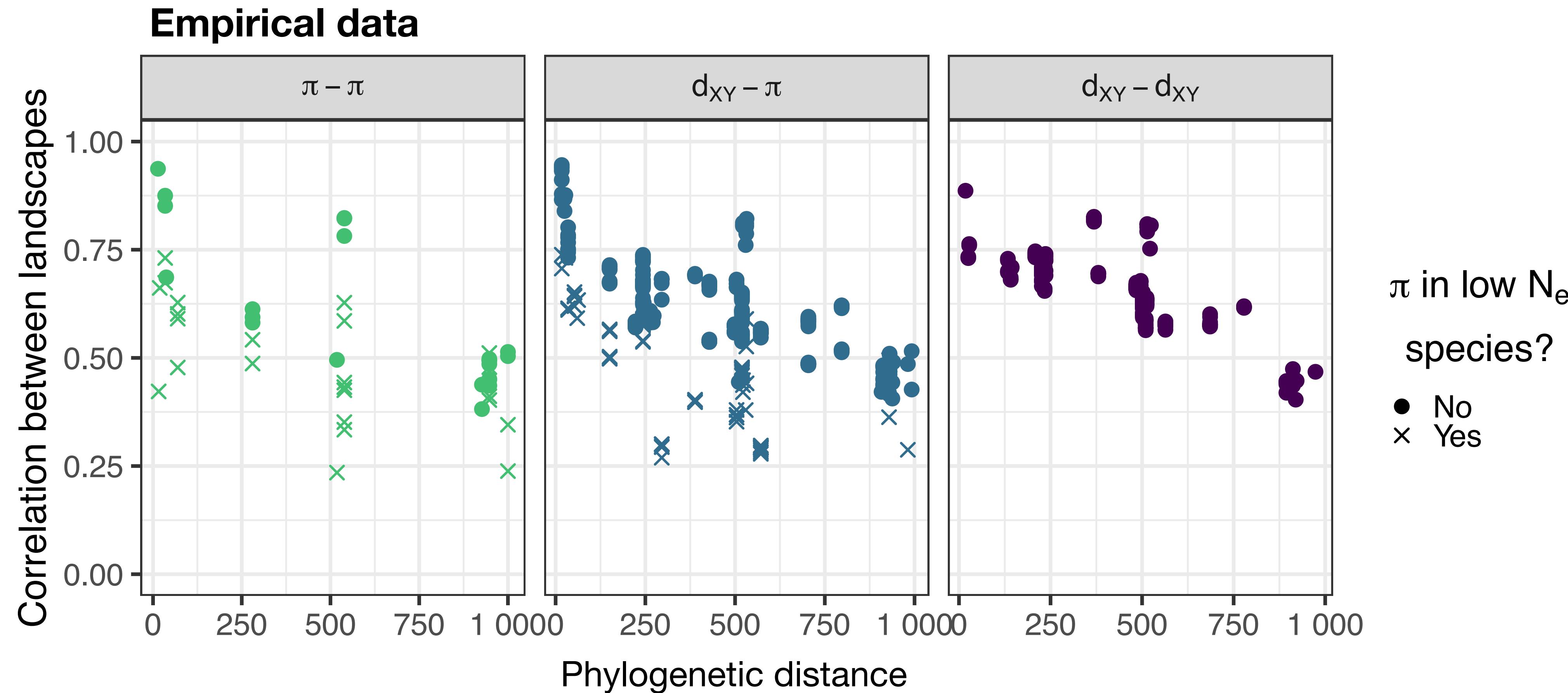
$$d_{XY} = \pi_{\text{anc}} + \mu T$$

$$\text{ILS} \sim \frac{1}{T}$$

# Correlated landscapes of genomic variation



# Correlated landscapes of genomic variation

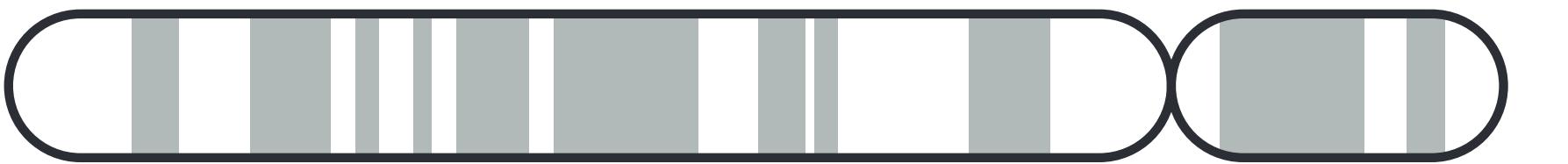


# SLiMulations

See for an explanation on how to simulate branches independently [https://tskit.dev/pyslim/docs/latest/vignette\\_parallel\\_phylo.html](https://tskit.dev/pyslim/docs/latest/vignette_parallel_phylo.html)

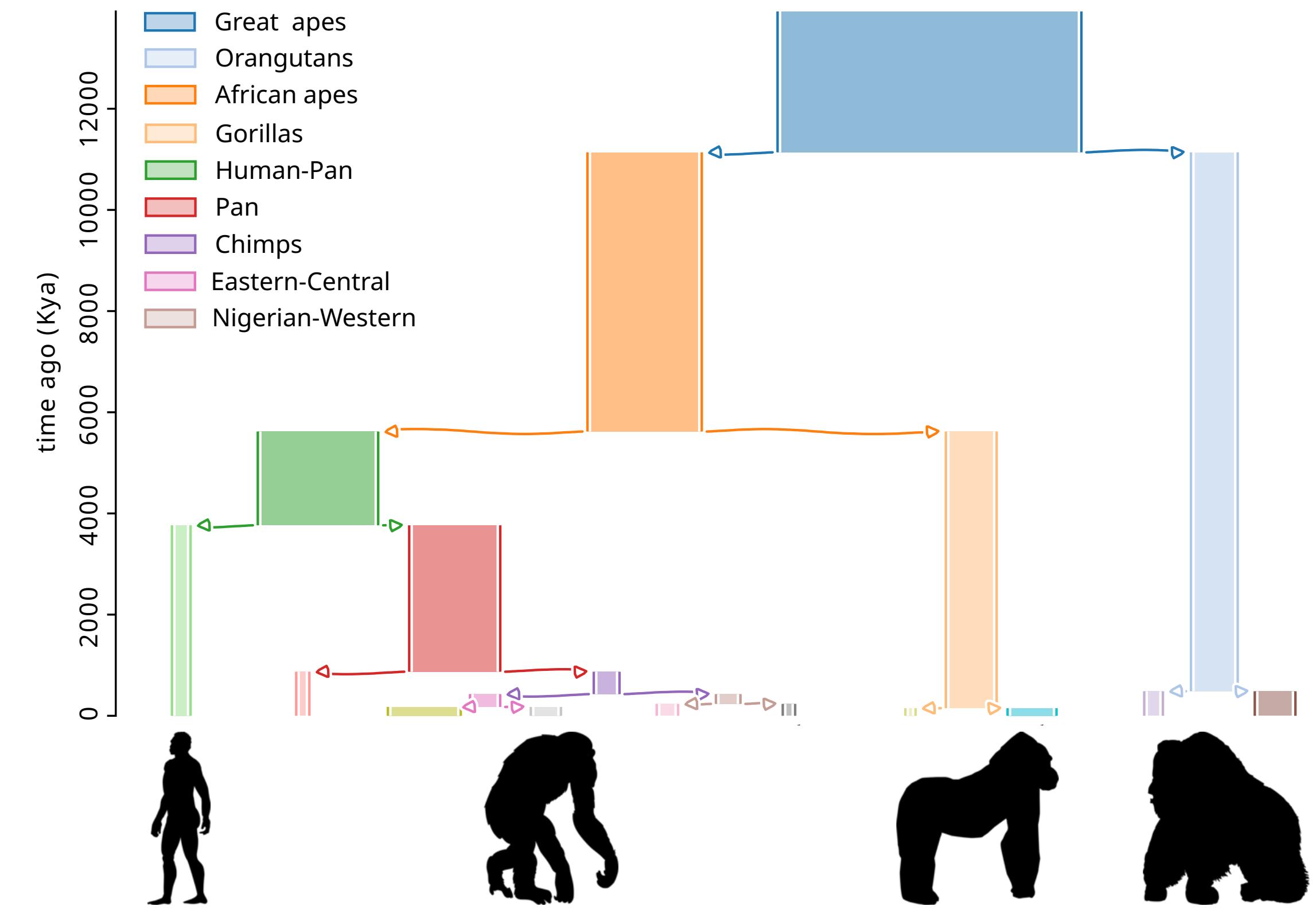
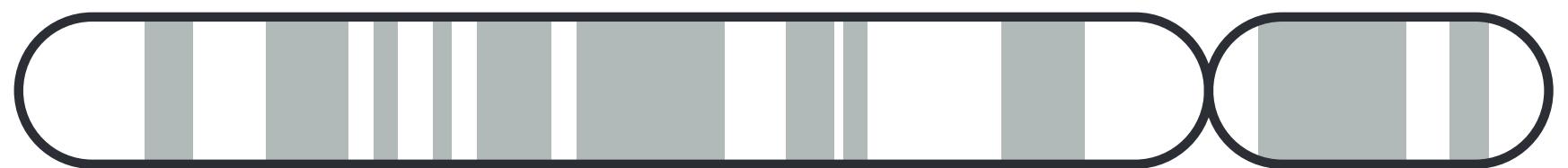
# SLiMulations

- Chromosome-scale simulation (chr12)
- Realistic — genomic annotations taken from humans

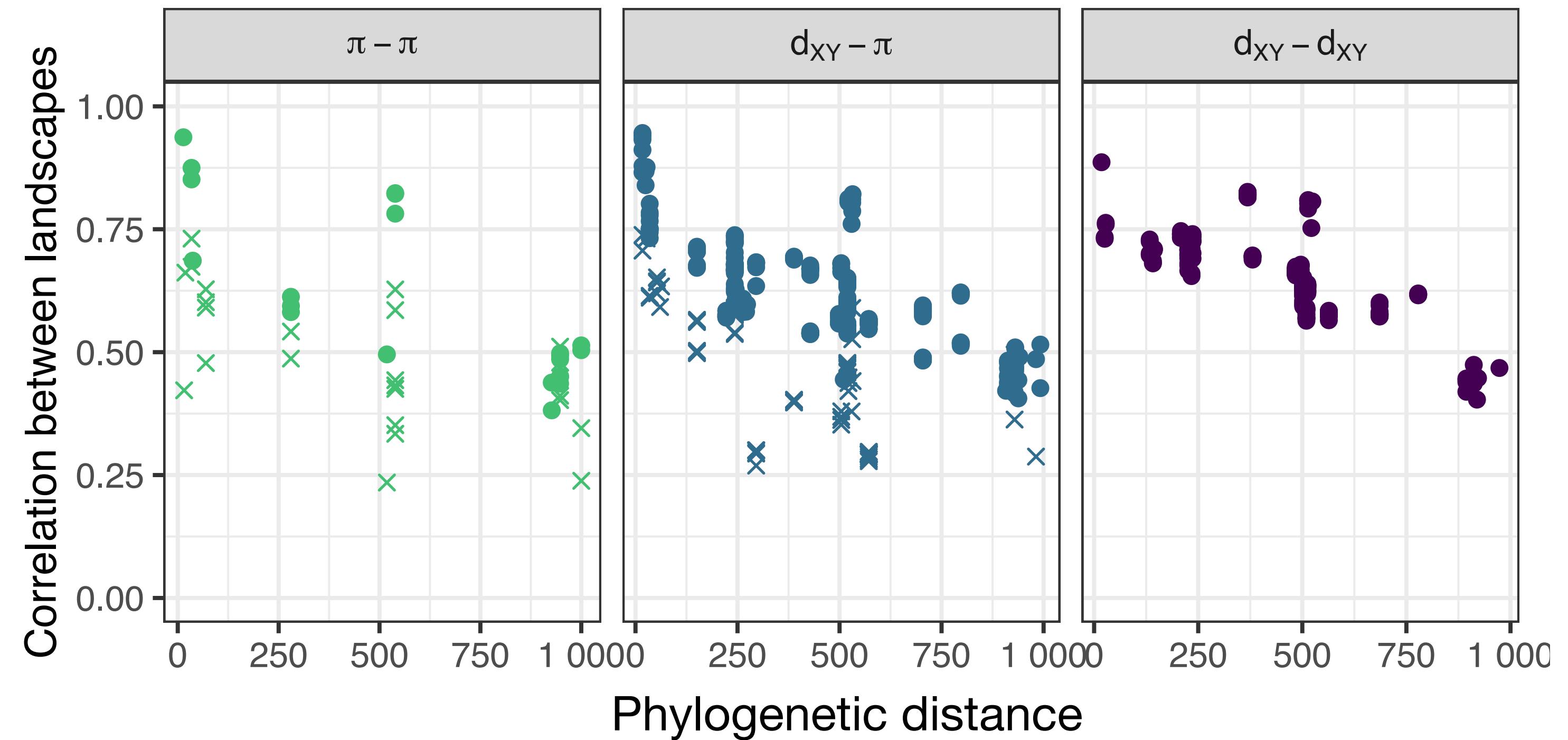


# SLiMulations

- Chromosome-scale simulation (chr12)
- Realistic — genomic annotations taken from humans
- Branches in the tree were simulated independently for computational efficiency

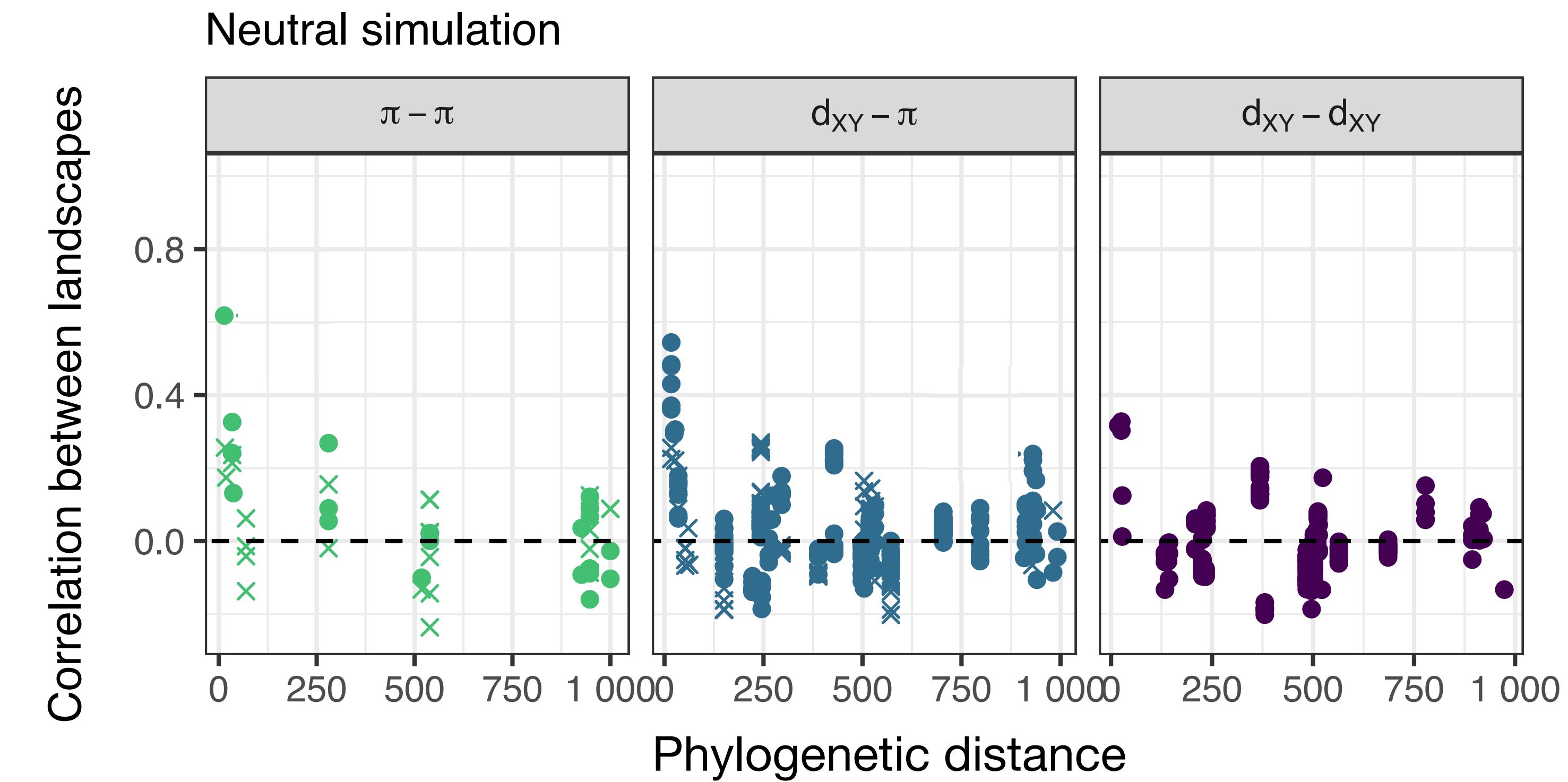


### Empirical data

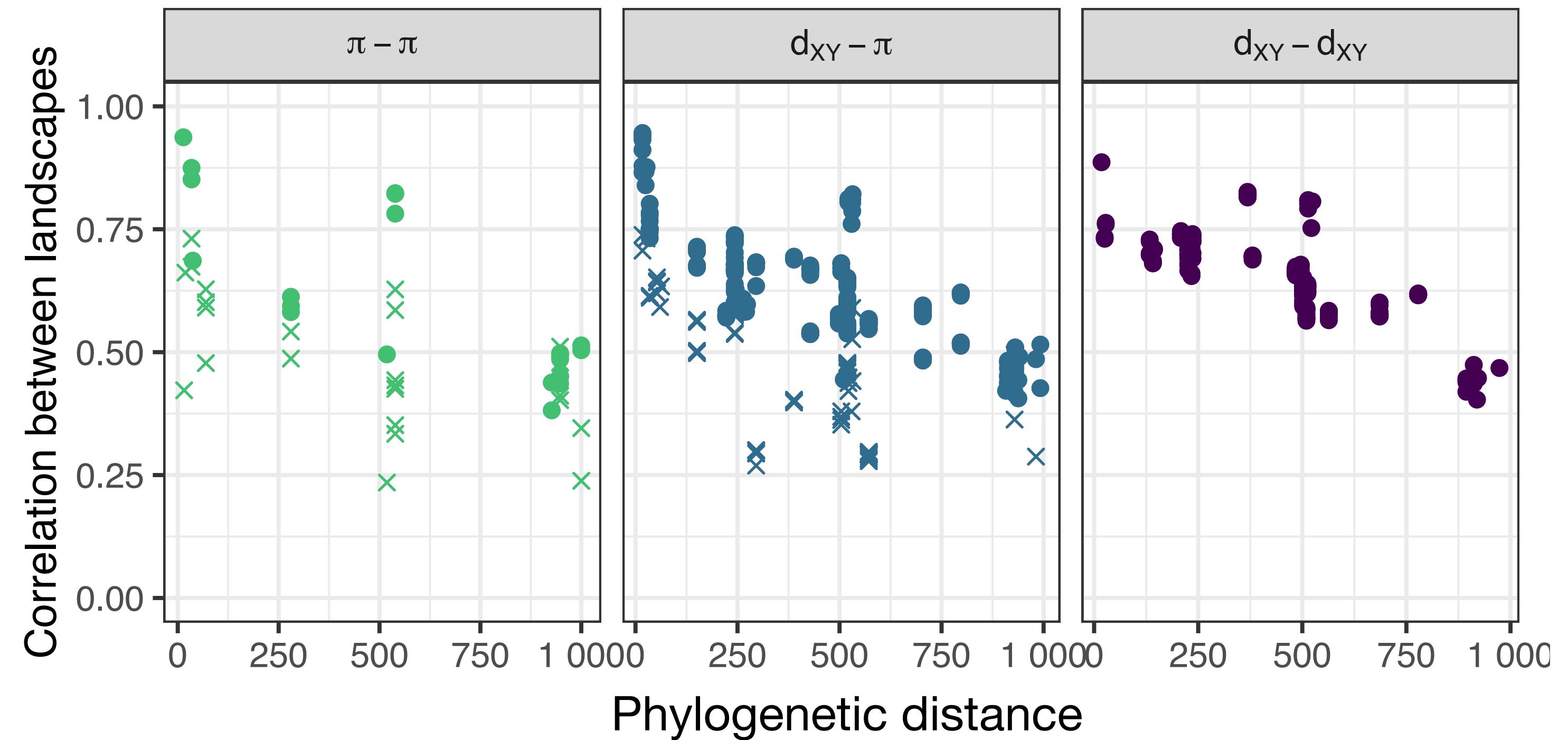


# Landscapes are not correlated in neutral simulation

- Correlations decay quickly with split time
- Recombination can create spurious correlations

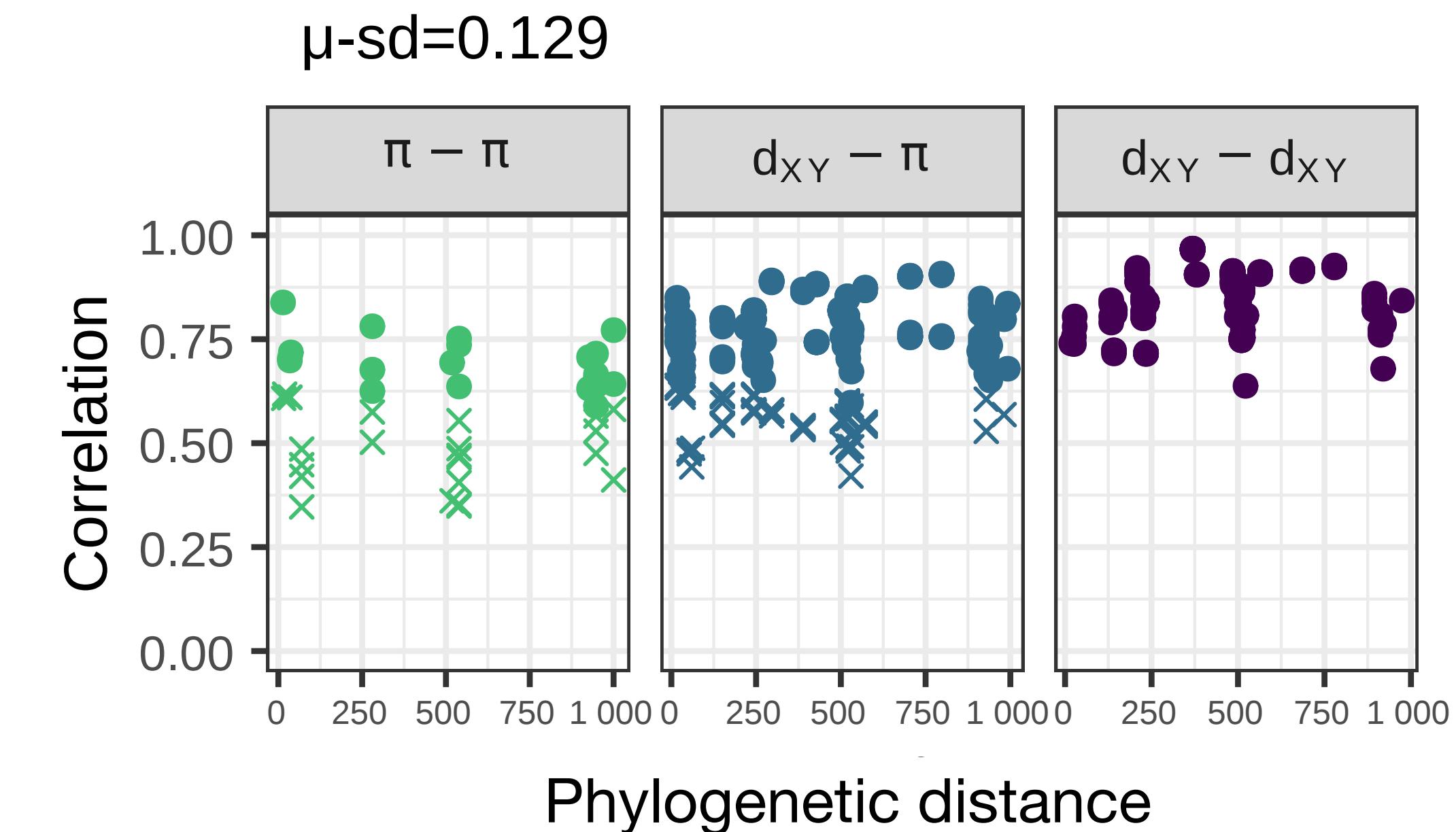
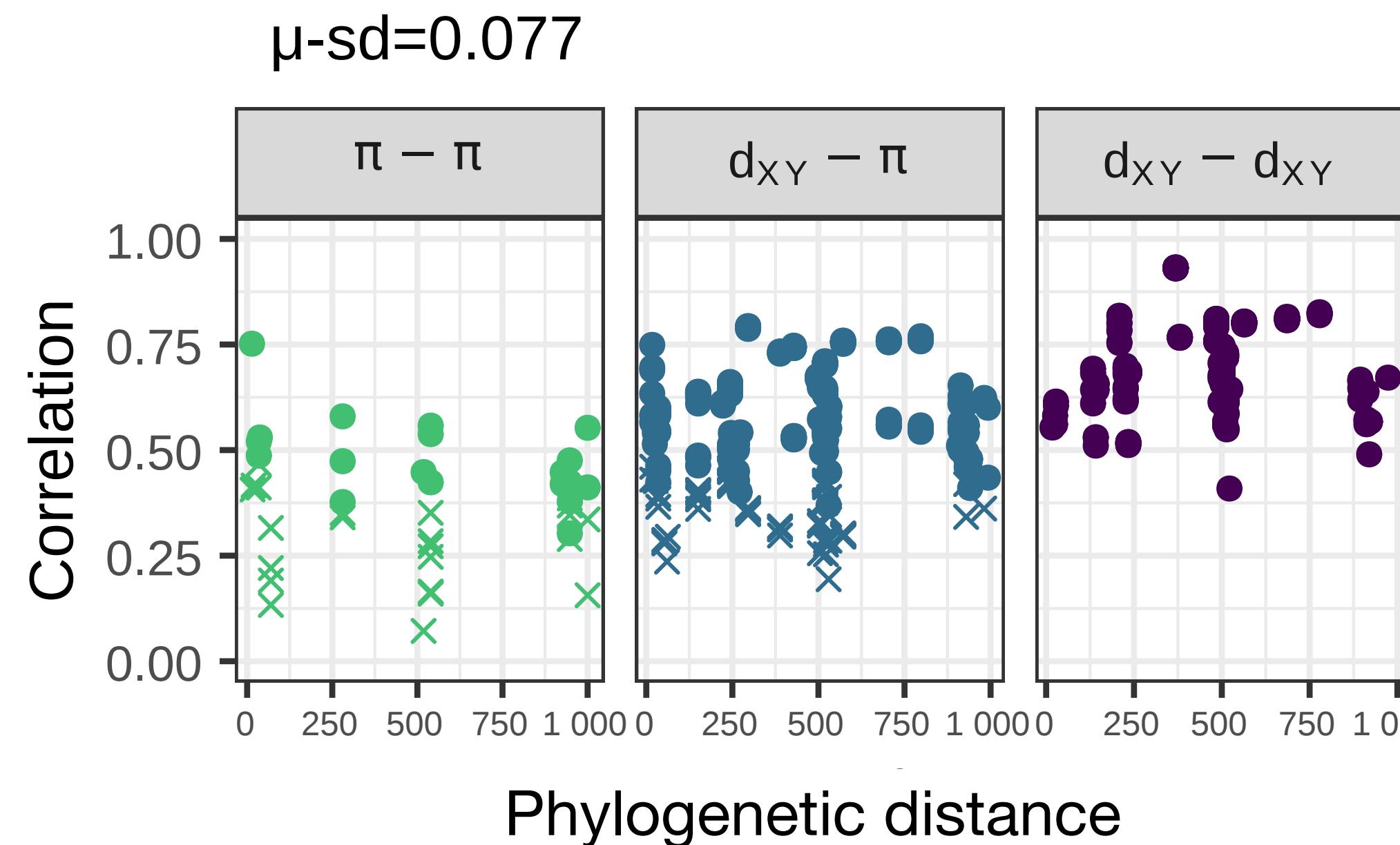


### Empirical data



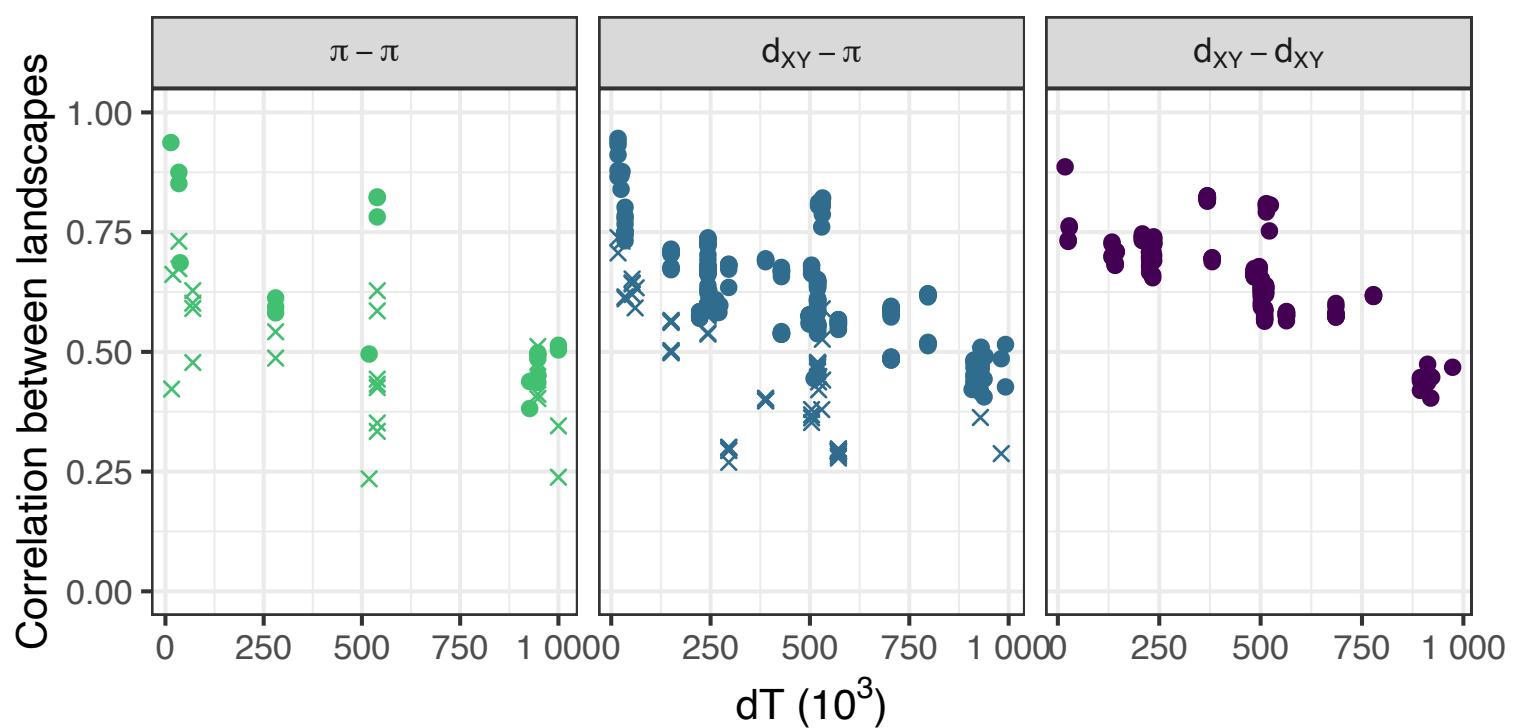
# Variation in local mutation rate can contribute to correlated landscapes

- Variation in mutation rates can maintain landscapes correlated
- But correlations do not decay over time



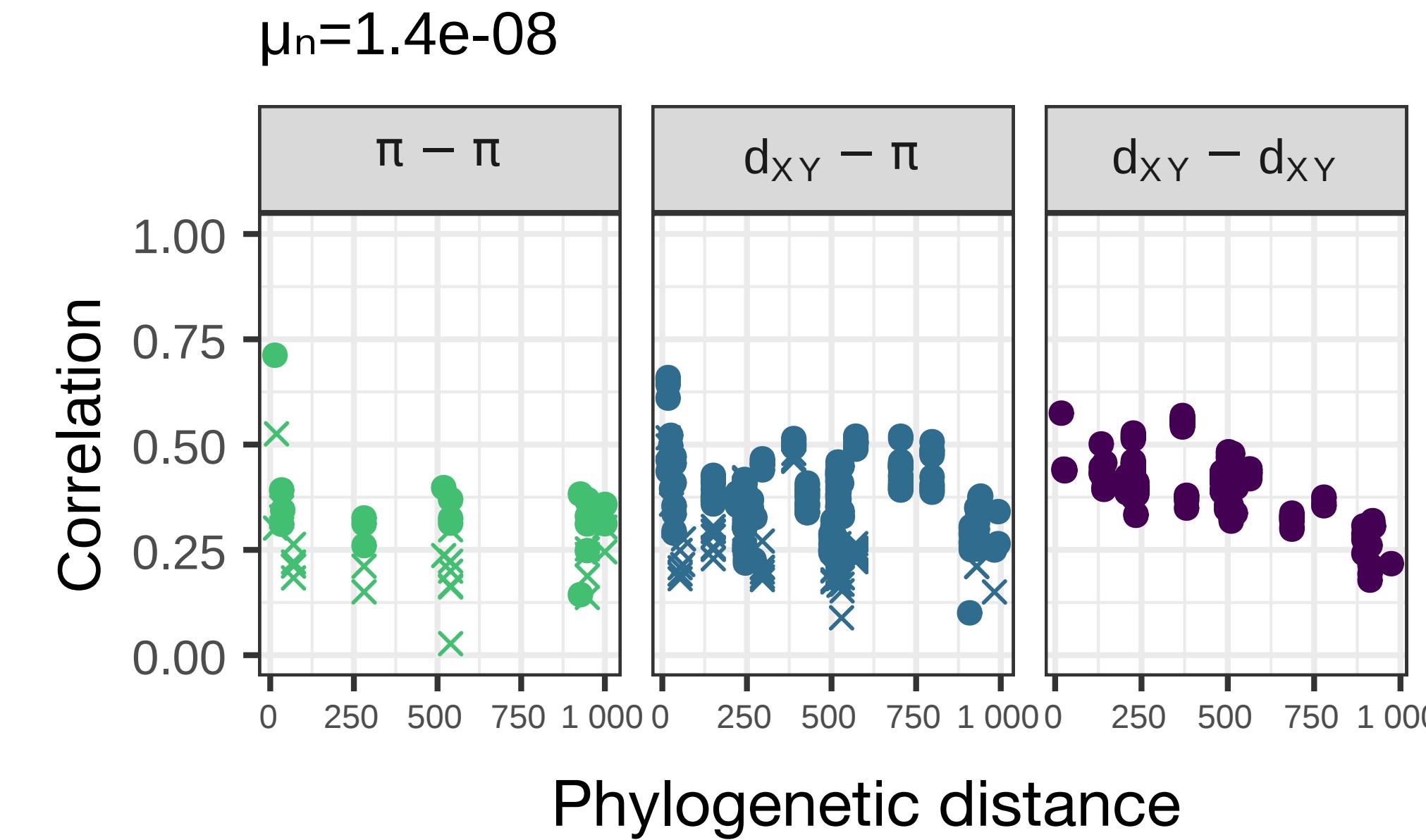
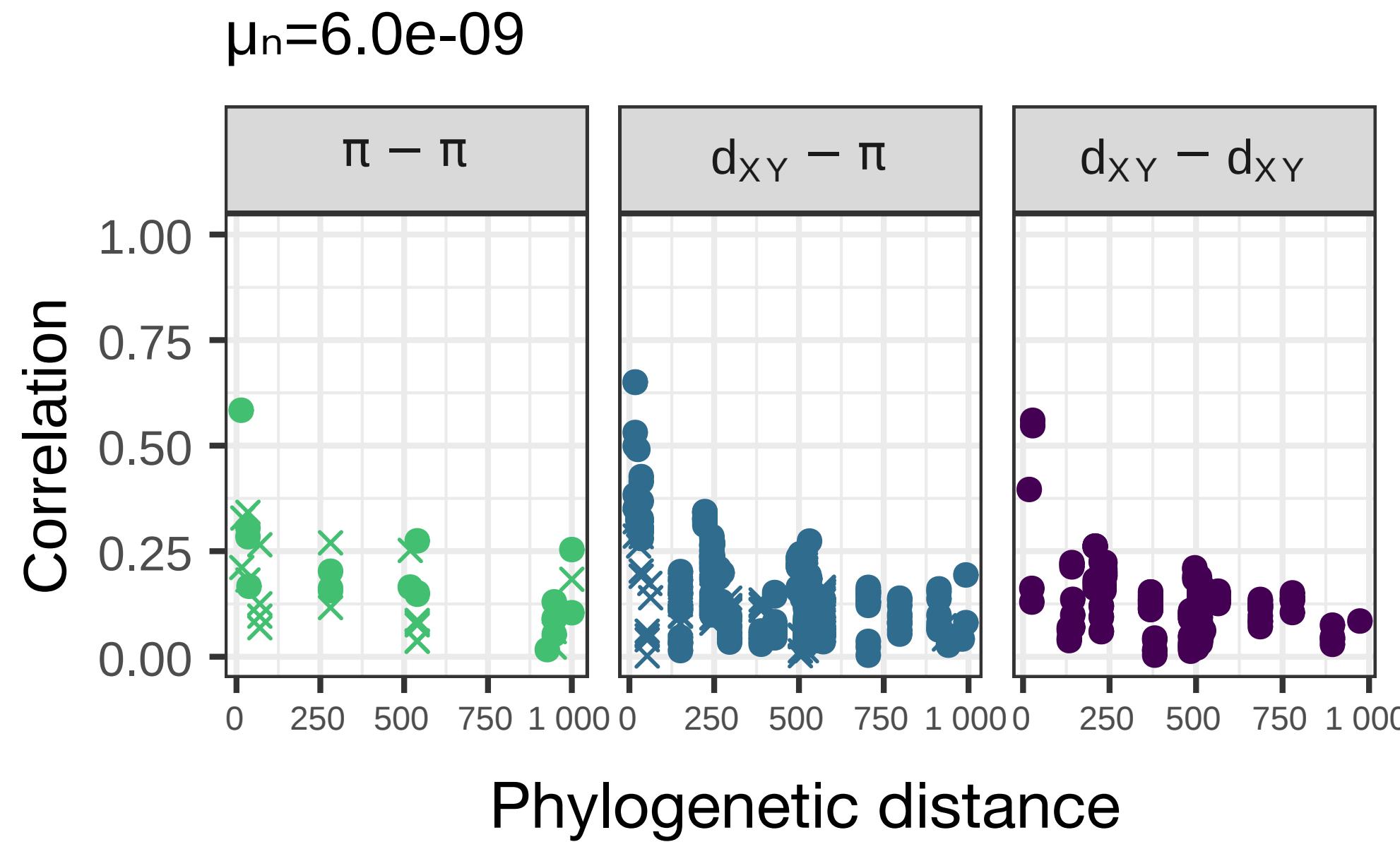
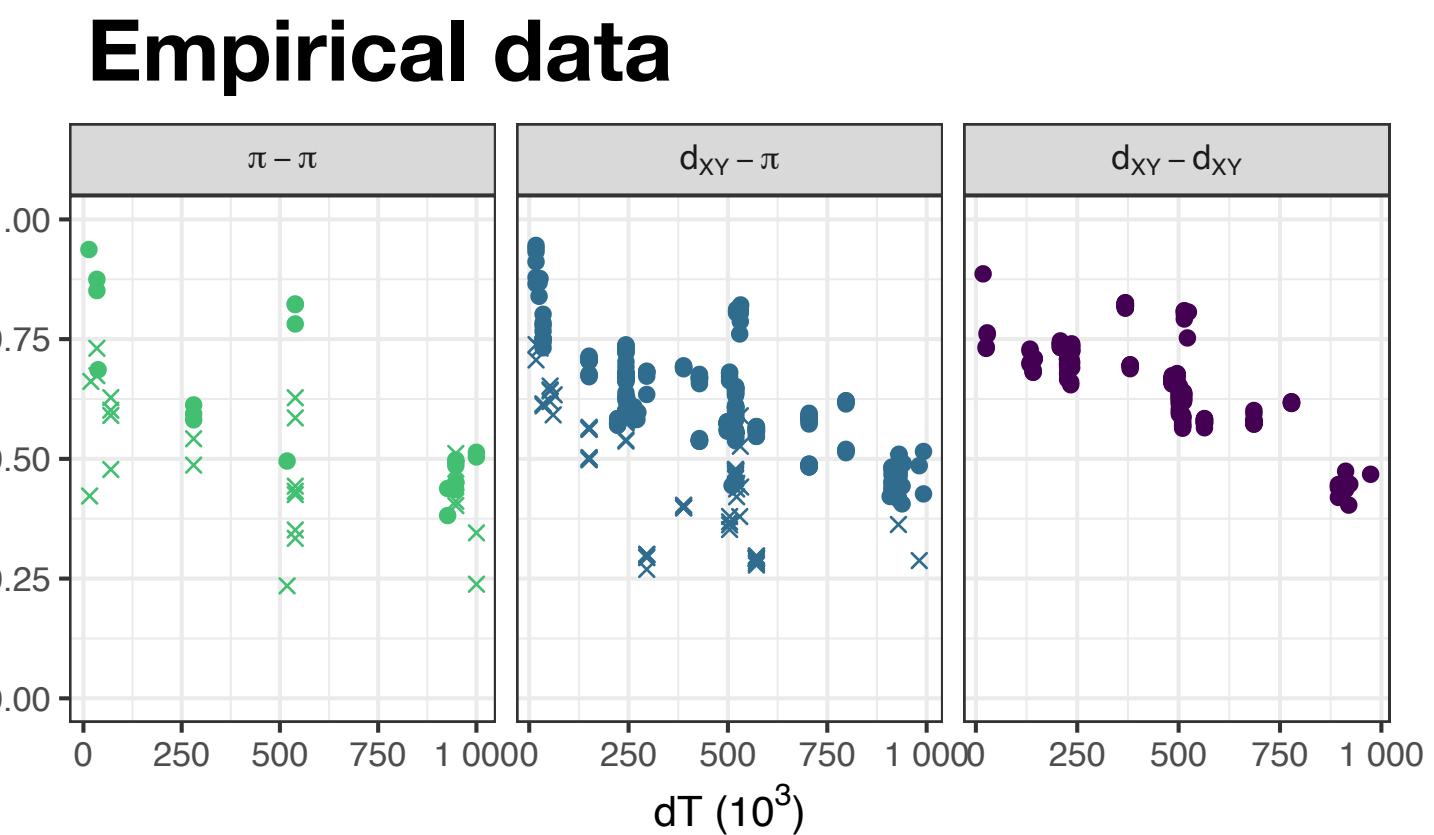
# Negative selection

## Empirical data



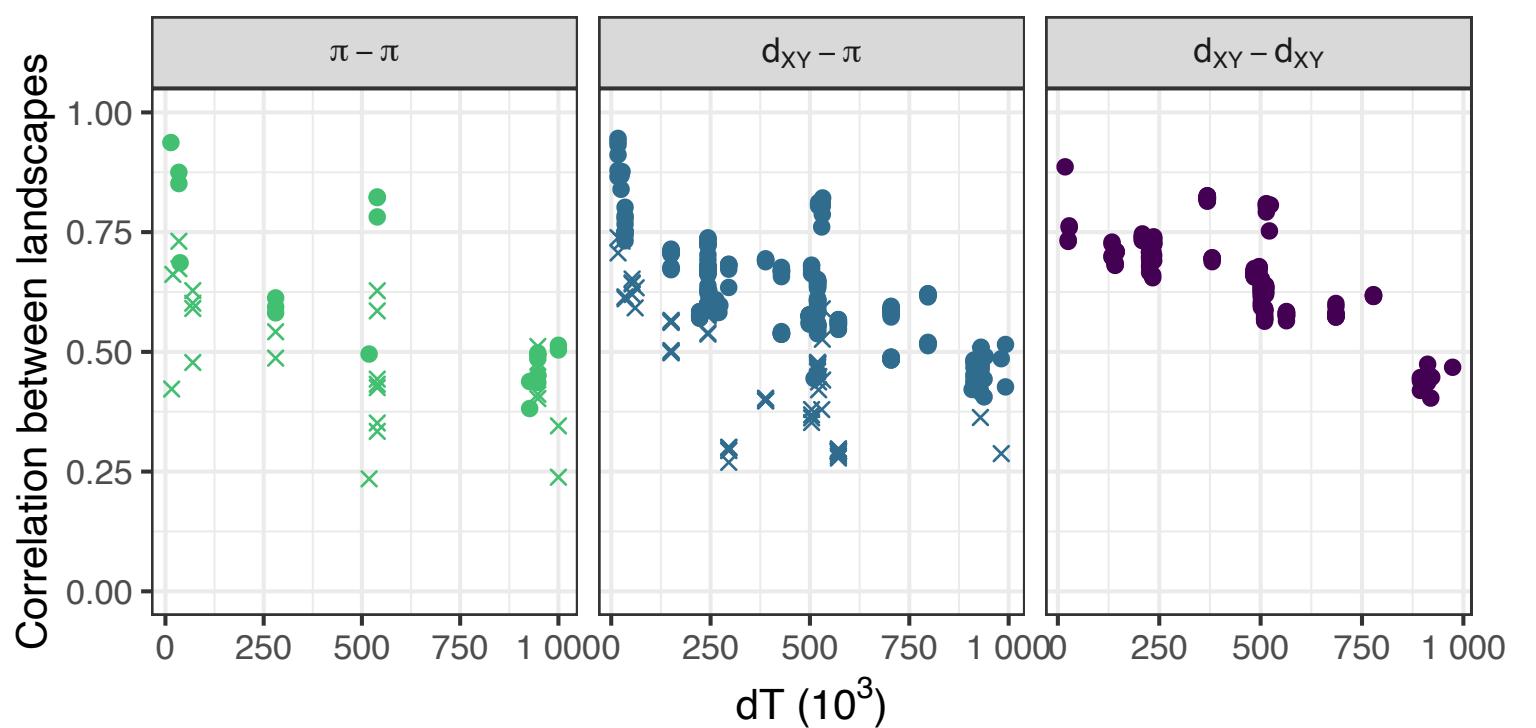
# Negative selection produces weak correlations

- Correlations can emerge with negative selection
- But still nowhere near what is seen in the real data



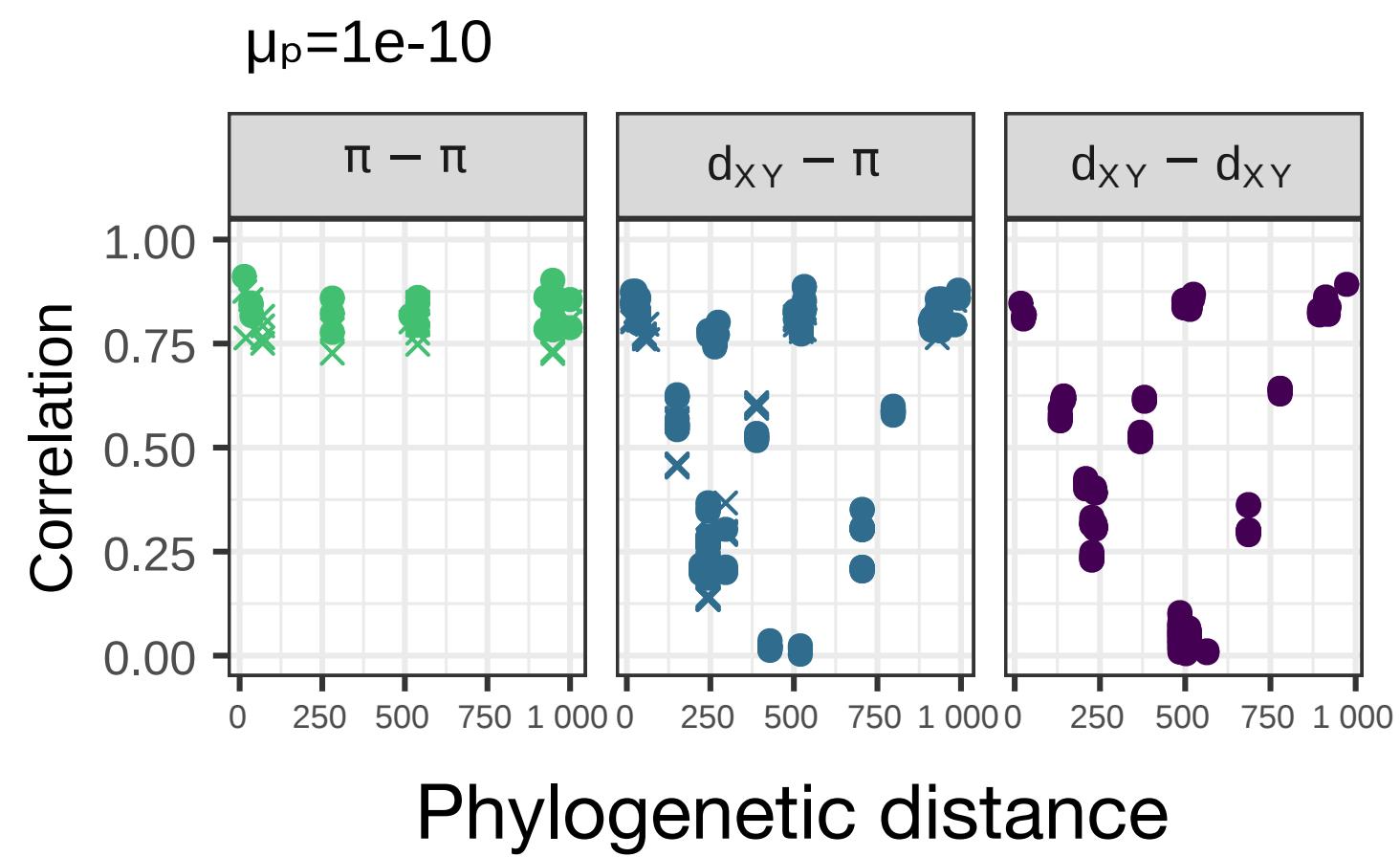
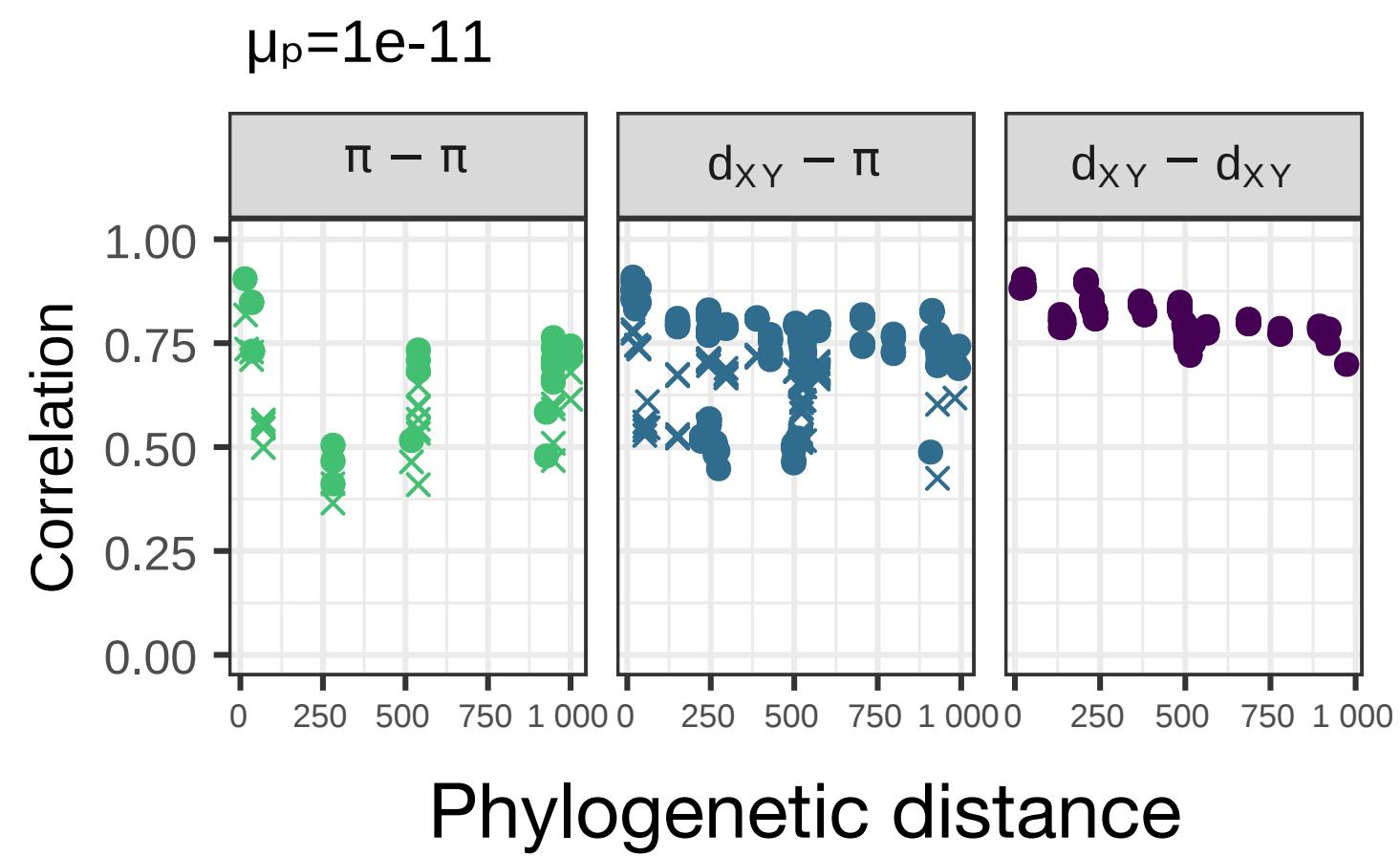
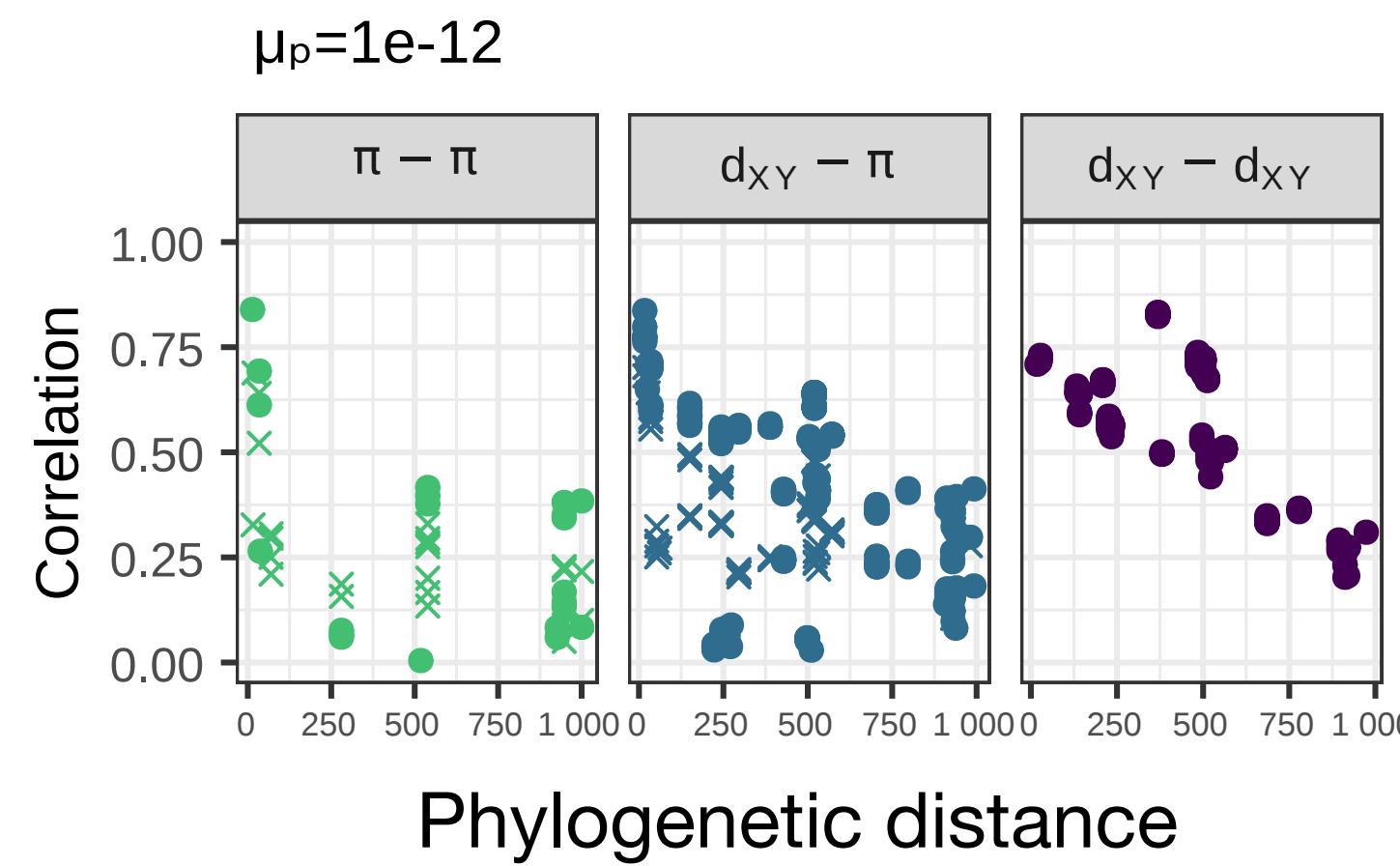
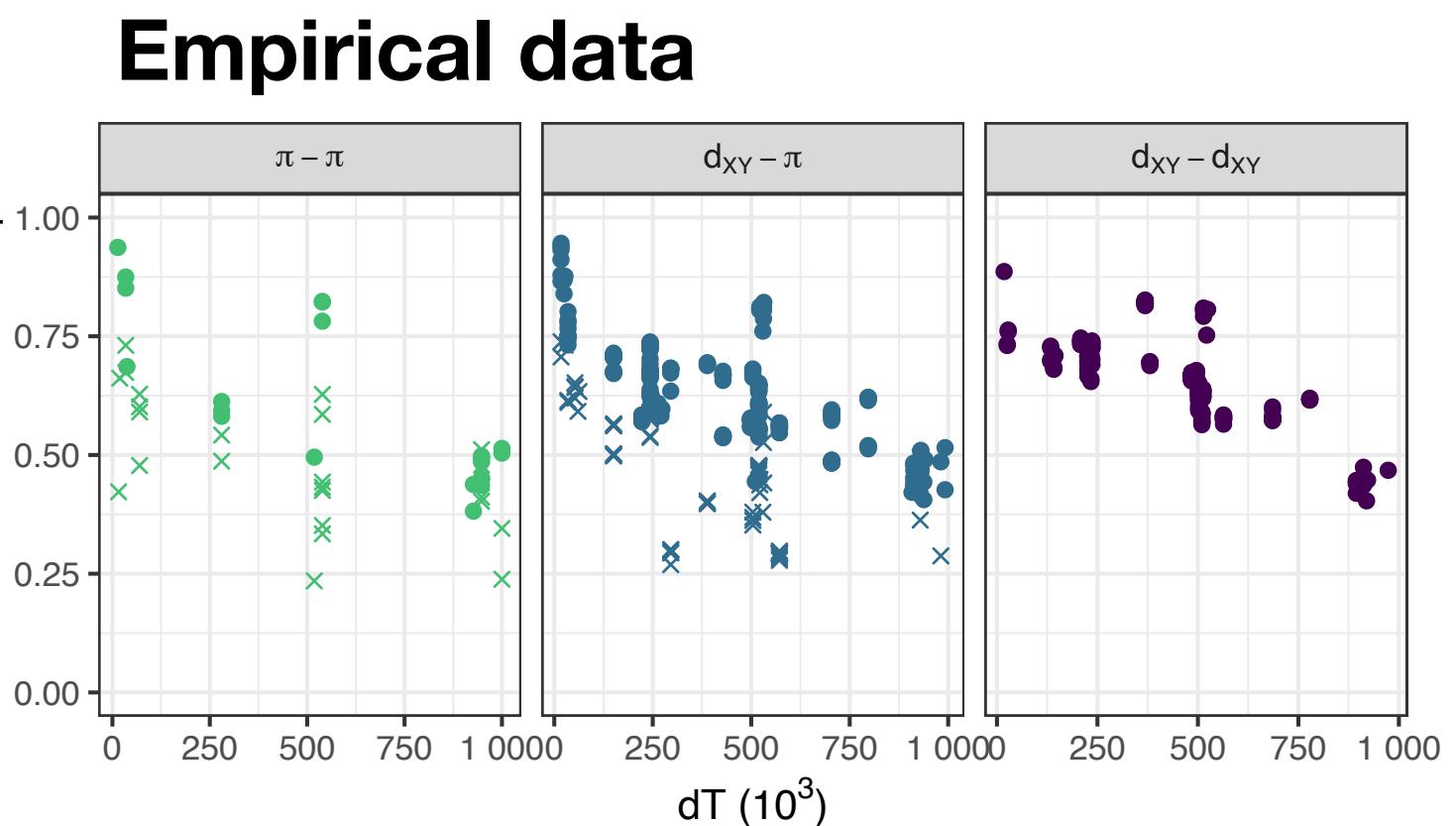
# Positive selection

## Empirical data



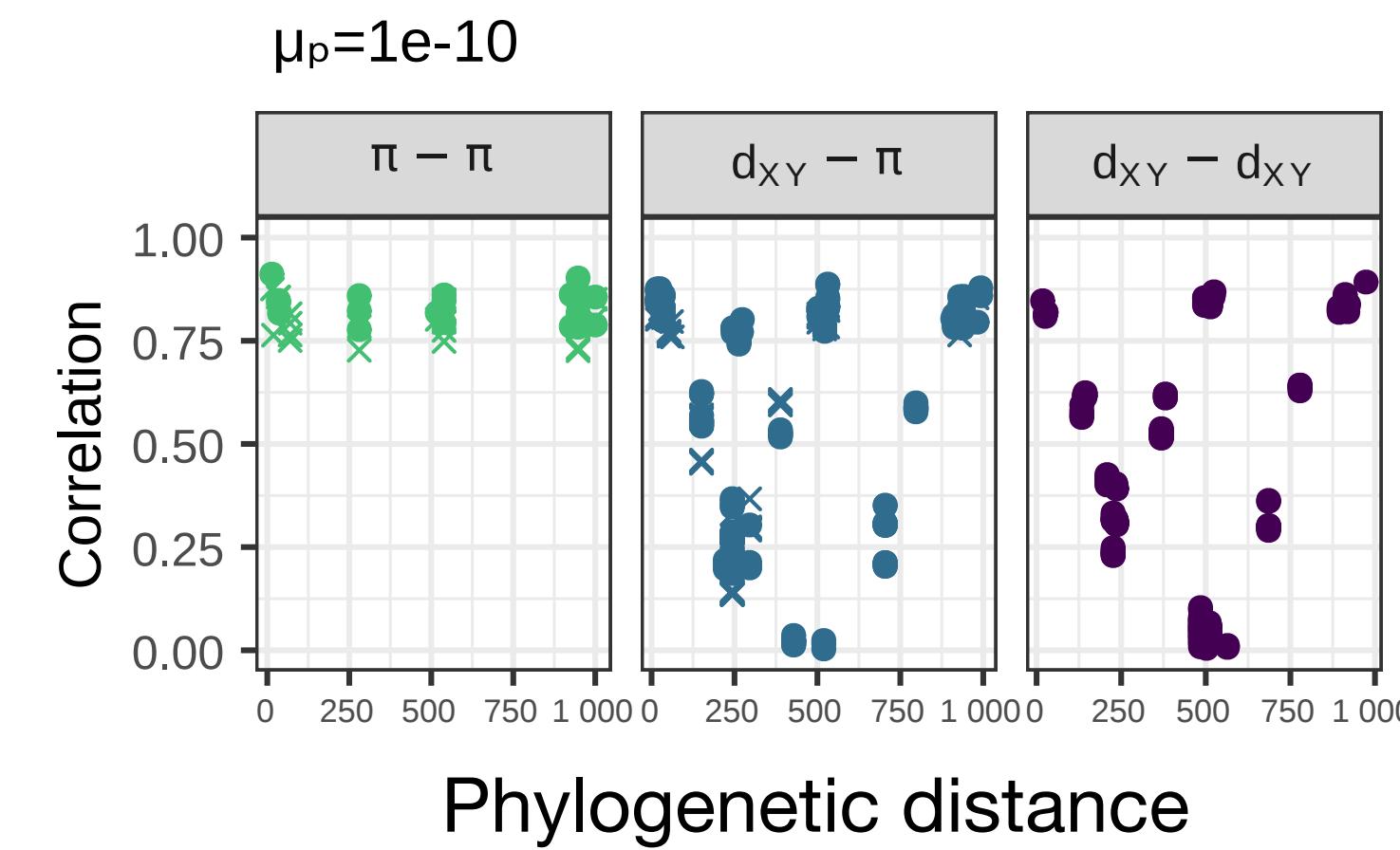
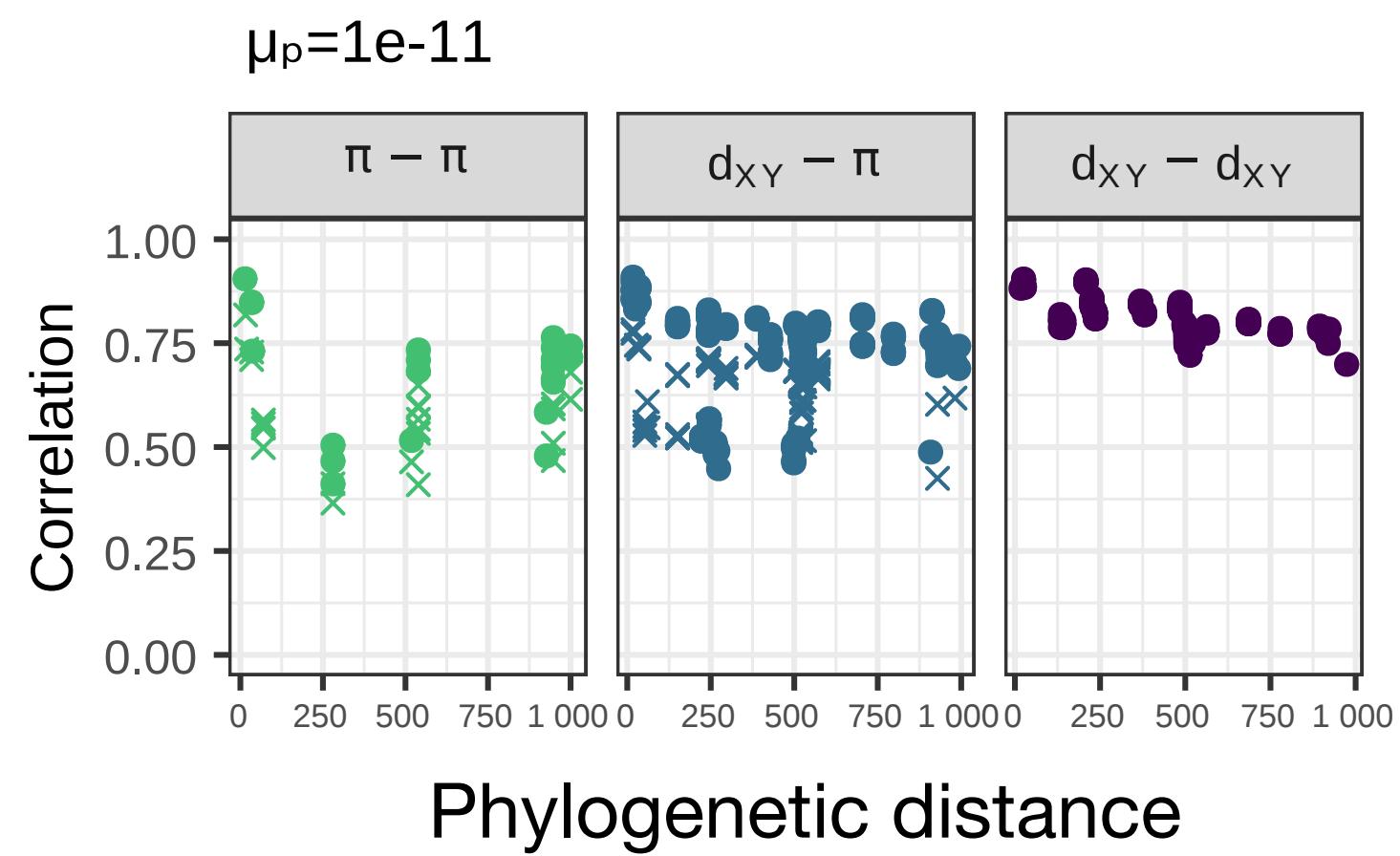
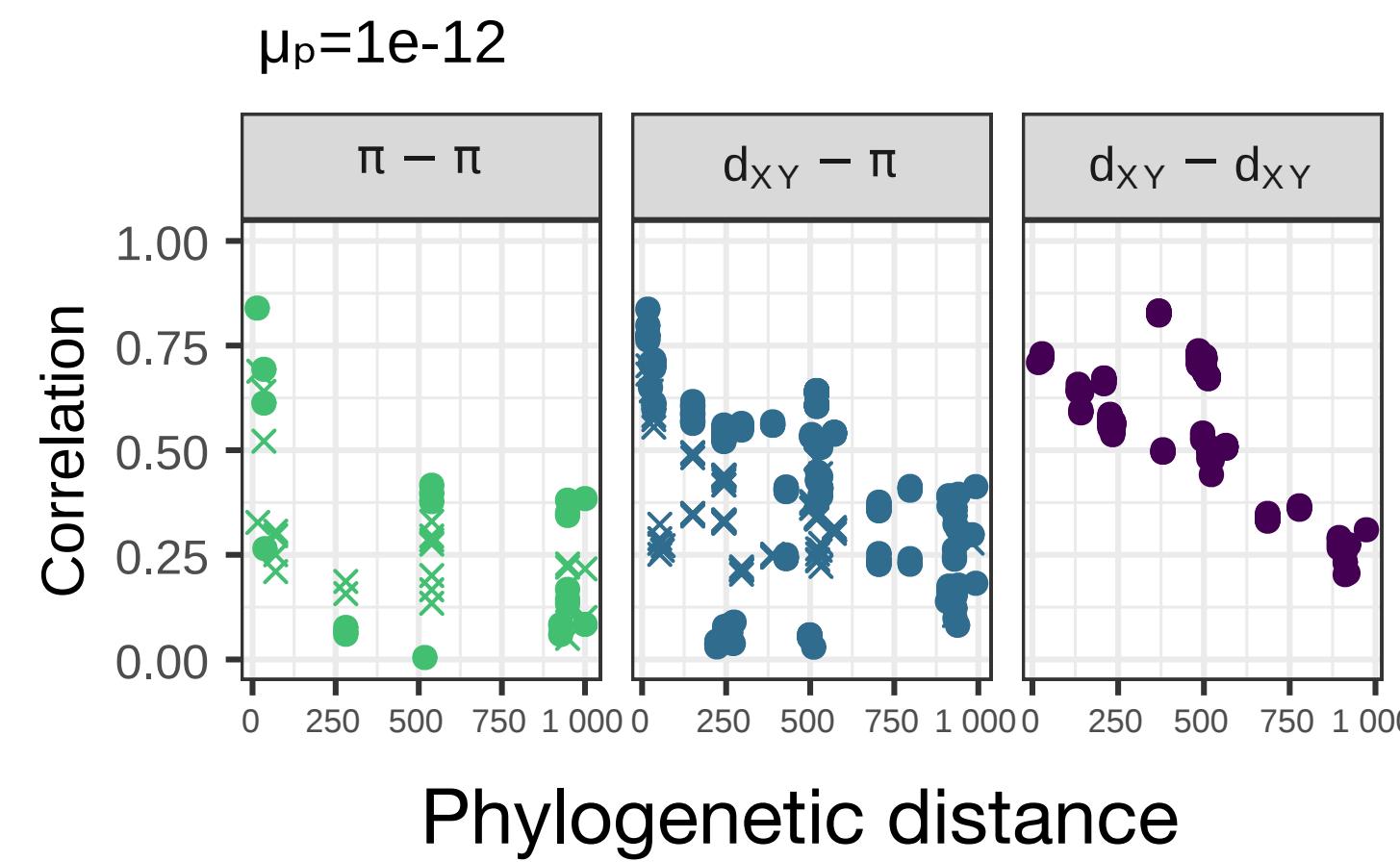
# Positive selection produces strong correlations

- Positive selection creates correlations
- With too much positive selection, correlations can start to break down

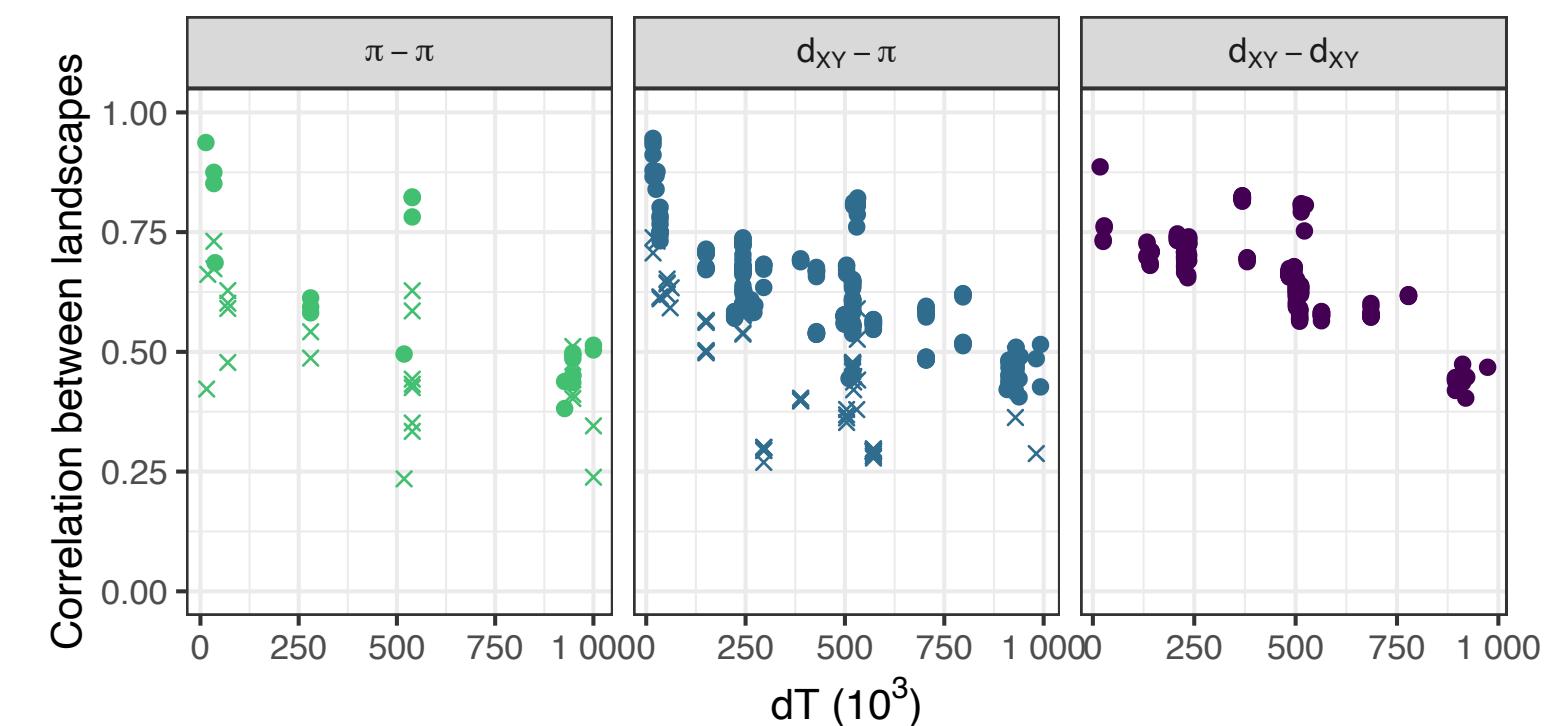


# Positive selection produces strong correlations

- Positive selection creates correlations
- With too much positive selection, correlations can start to break down

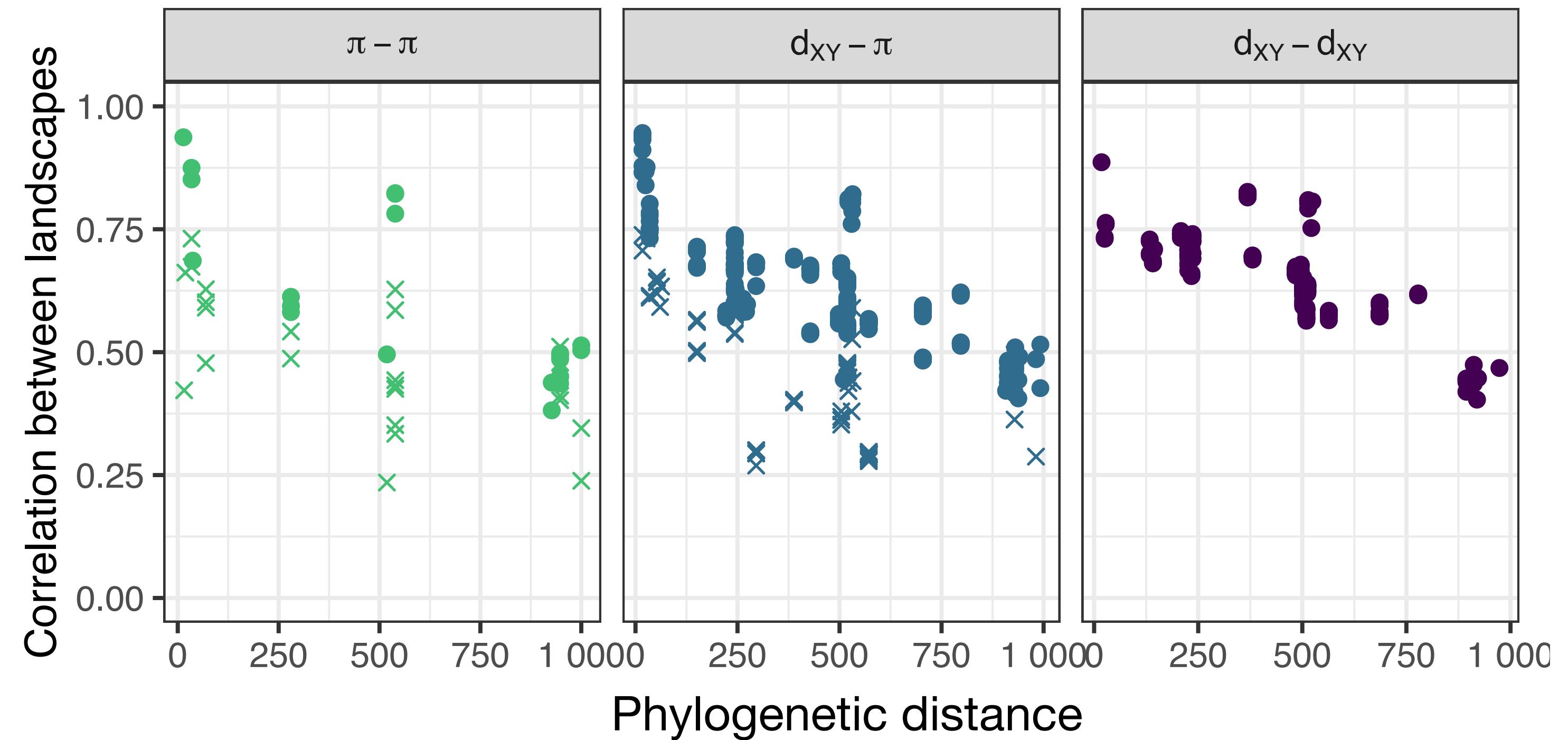


**Empirical data**



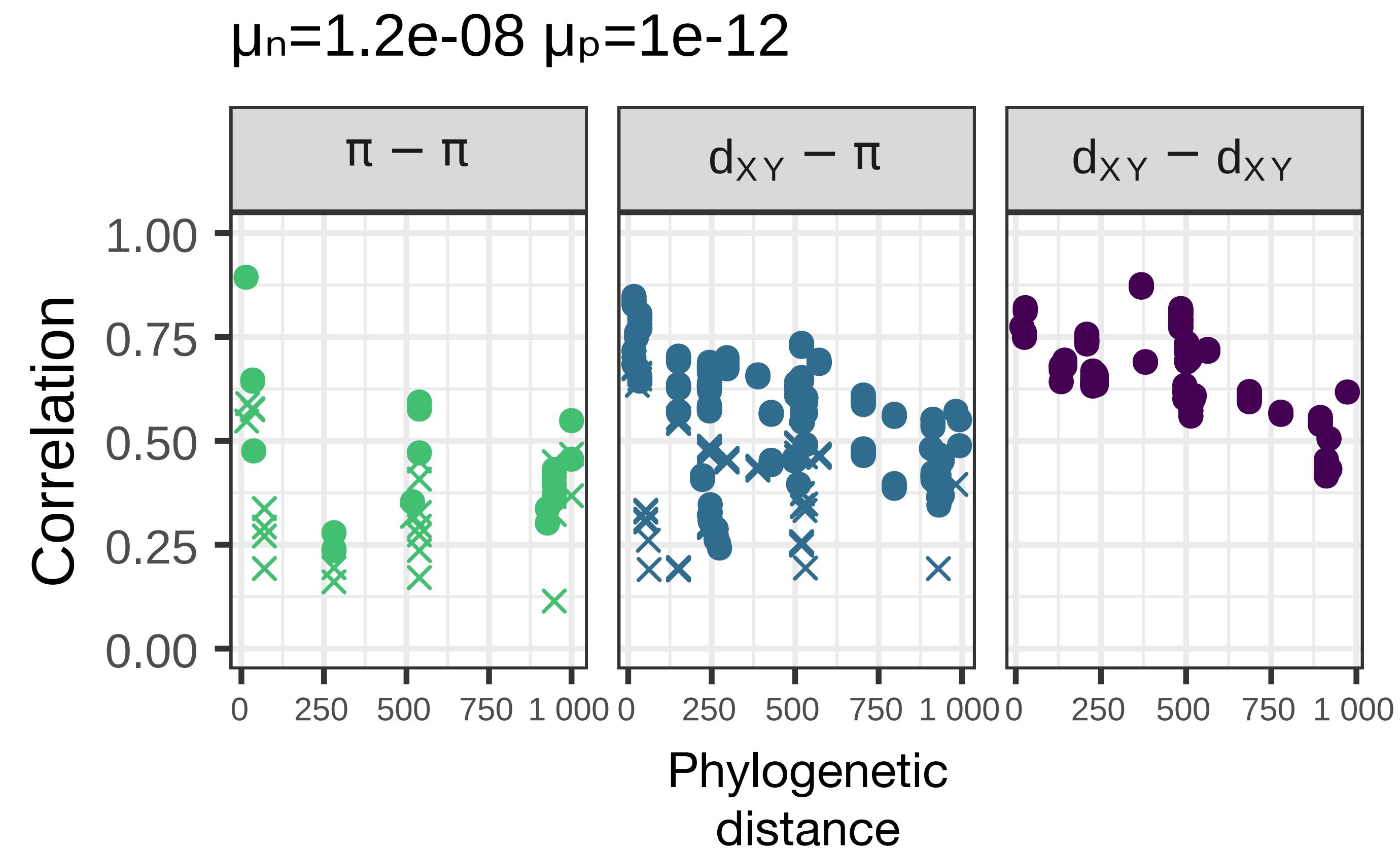
$$d_{XY} = \downarrow\pi_{anc} + \uparrow rT$$

### Empirical data



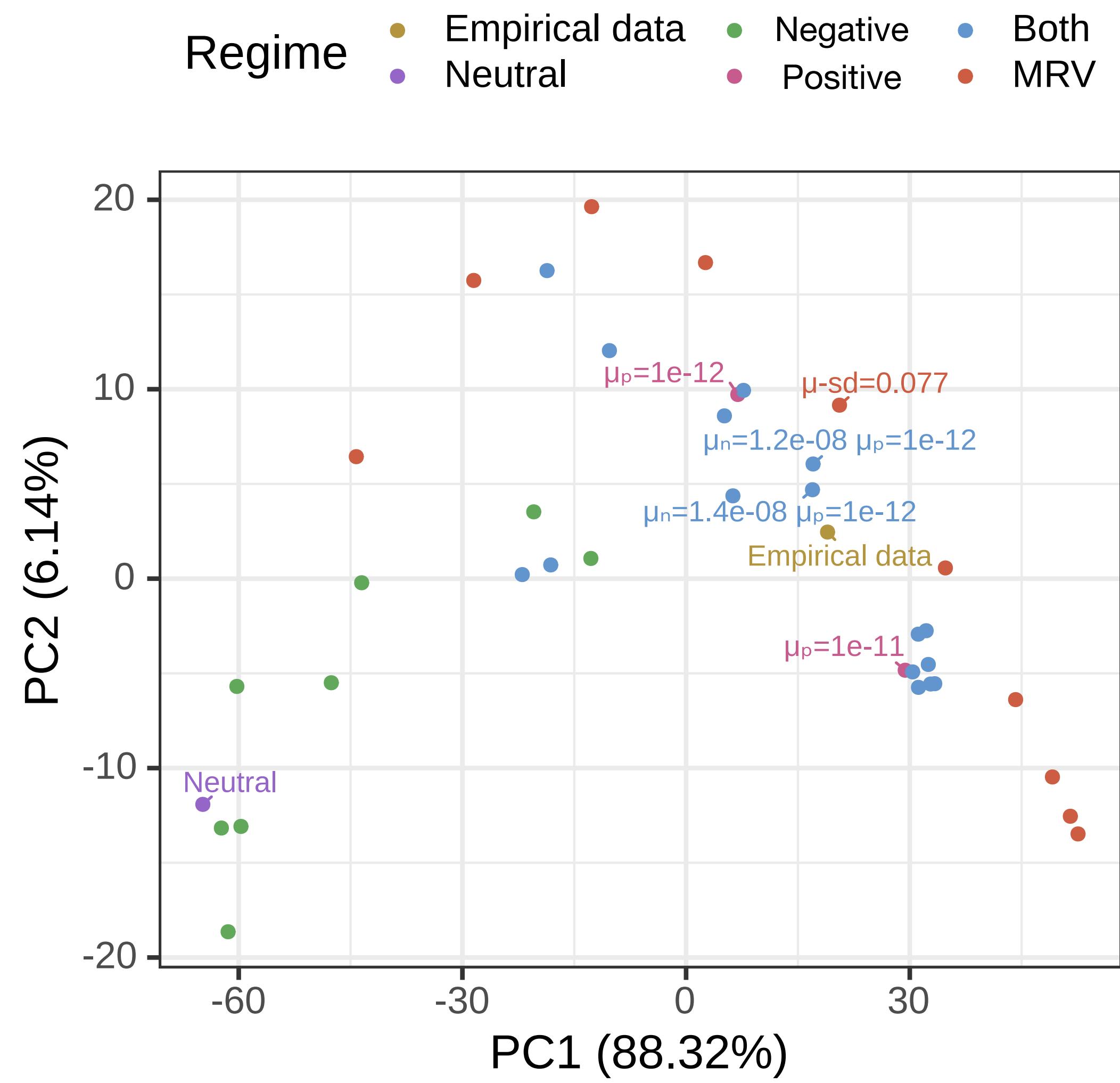
# Positive and negative selection result in the best fit

- Good fit!
- Differentiates better between low and high  $N_e$  species
- Some of the variance, especially in diversity is not recapitulated with the simulations



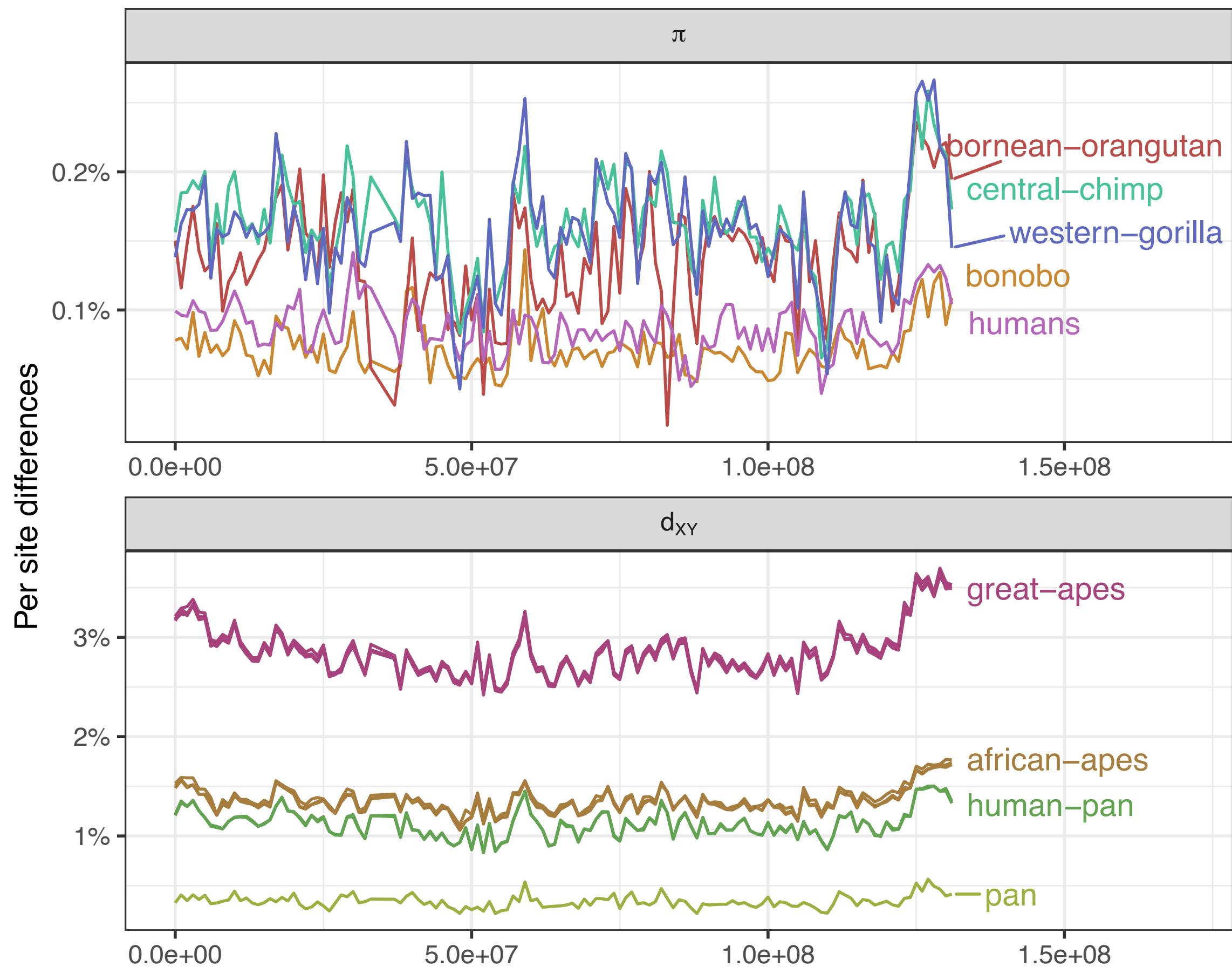
# Similarity between simulations and data

- PCA on the matrix of all 1260 pairwise correlations and the 36 simulations + data
- Simulations with both negative and positive selection fit better!
- In our best fitting simulation, around 10% of fixations in exons along the human lineage were due to positive selection



# Conclusions

- Landscapes of diversity and divergence carry a lot of information about underlying processes
- Modeling variation in multiple species at once can be incredibly powerful
- For a proper statistical model fitting, we need theory and/or more computational efficiency





THE PREPRINT SERVER FOR BIOLOGY

## **Shared evolutionary processes shape landscapes of genomic variation in the great apes**

Murillo F. Rodrigues, Andrew D. Kern, Peter L. Ralph

**doi:** <https://doi.org/10.1101/2023.02.07.527547>

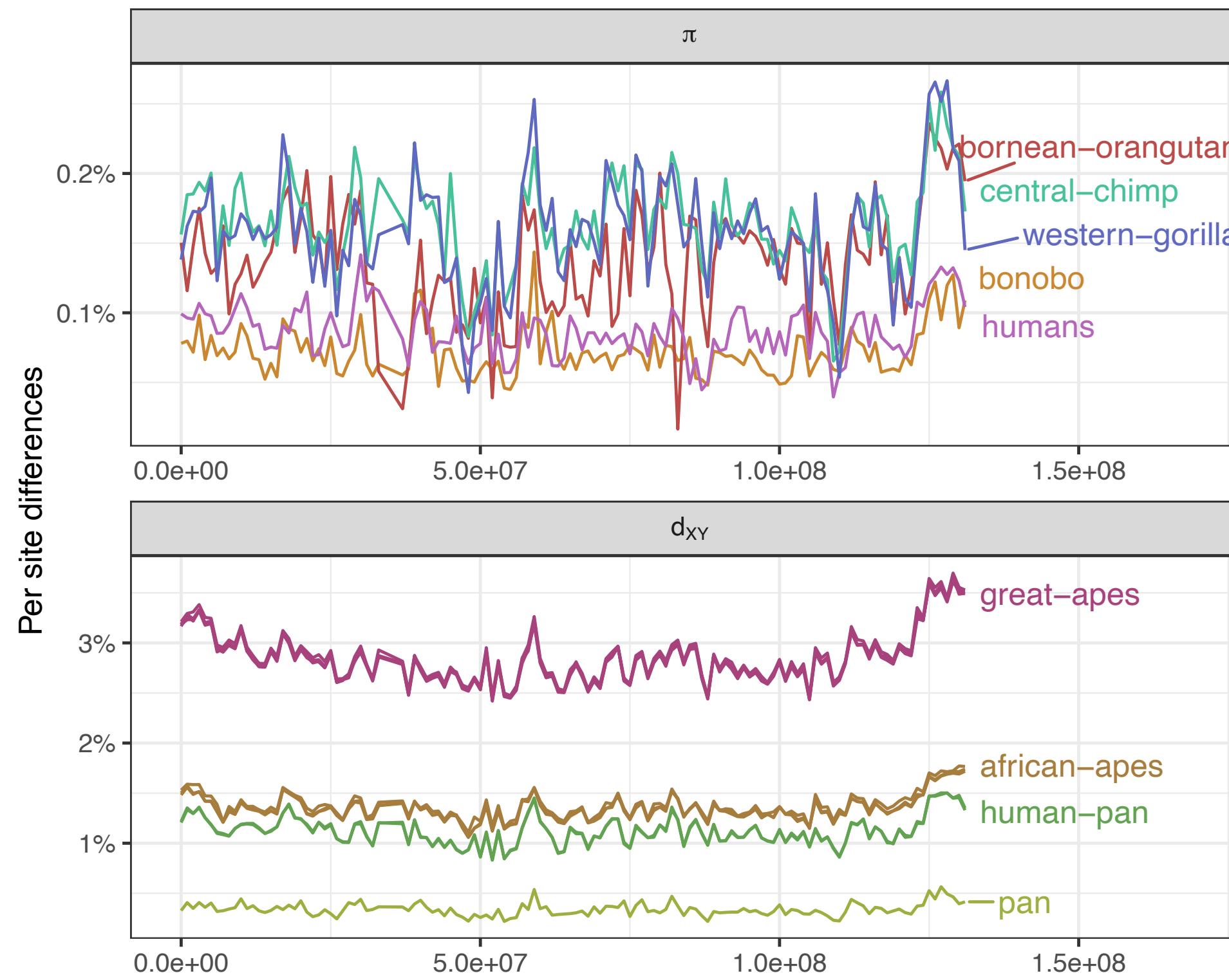
# Acknowledgements

- Peter Ralph
- Andrew Kern
- Kern & Ralph CoLab

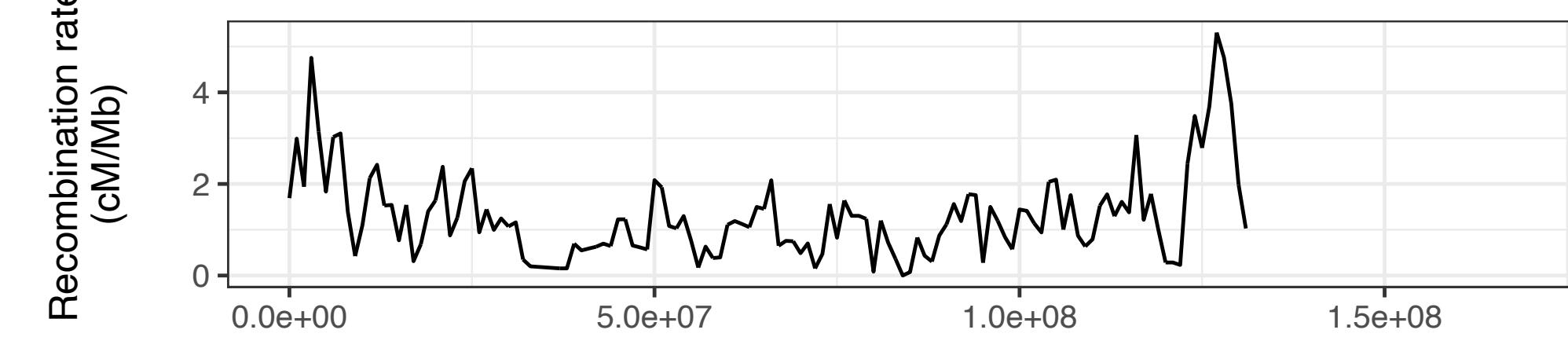




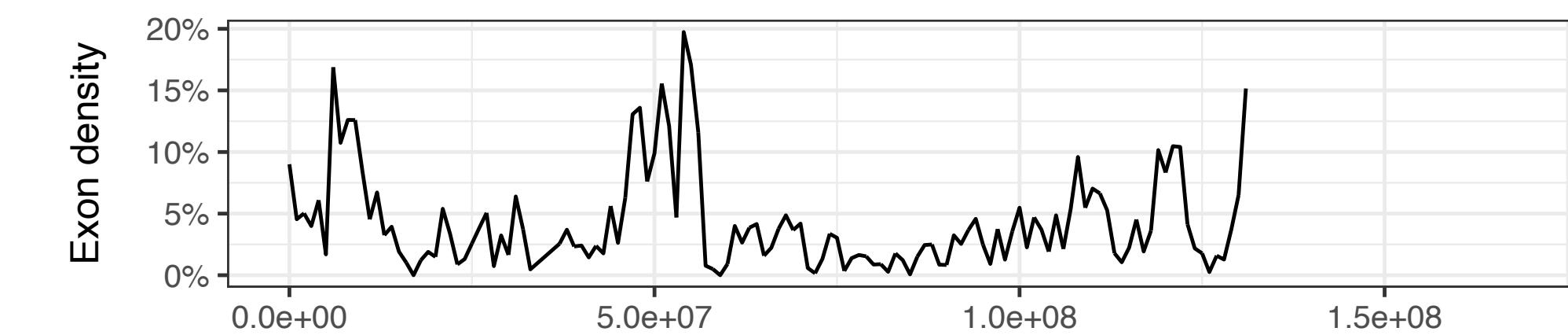
A



B

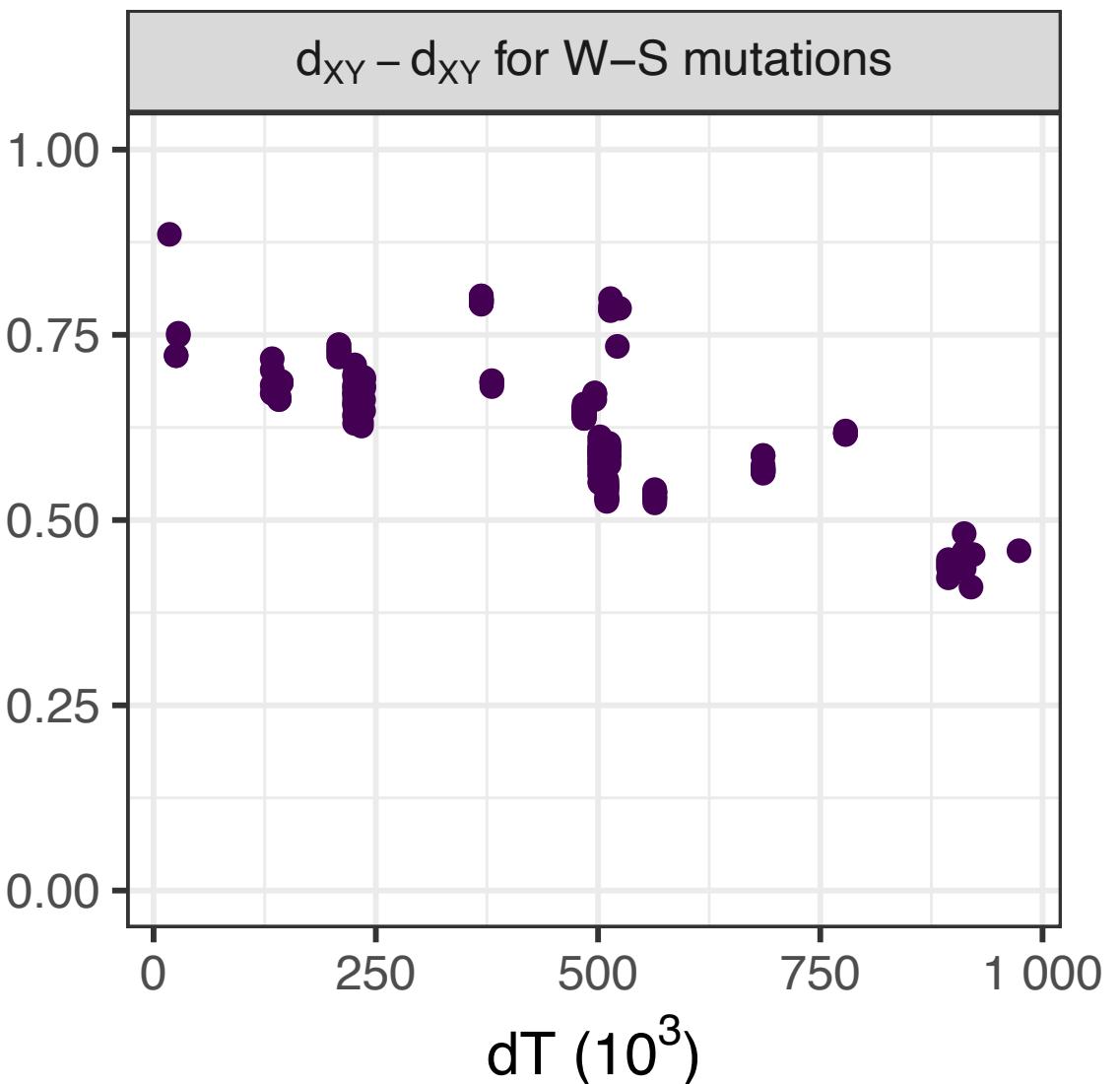
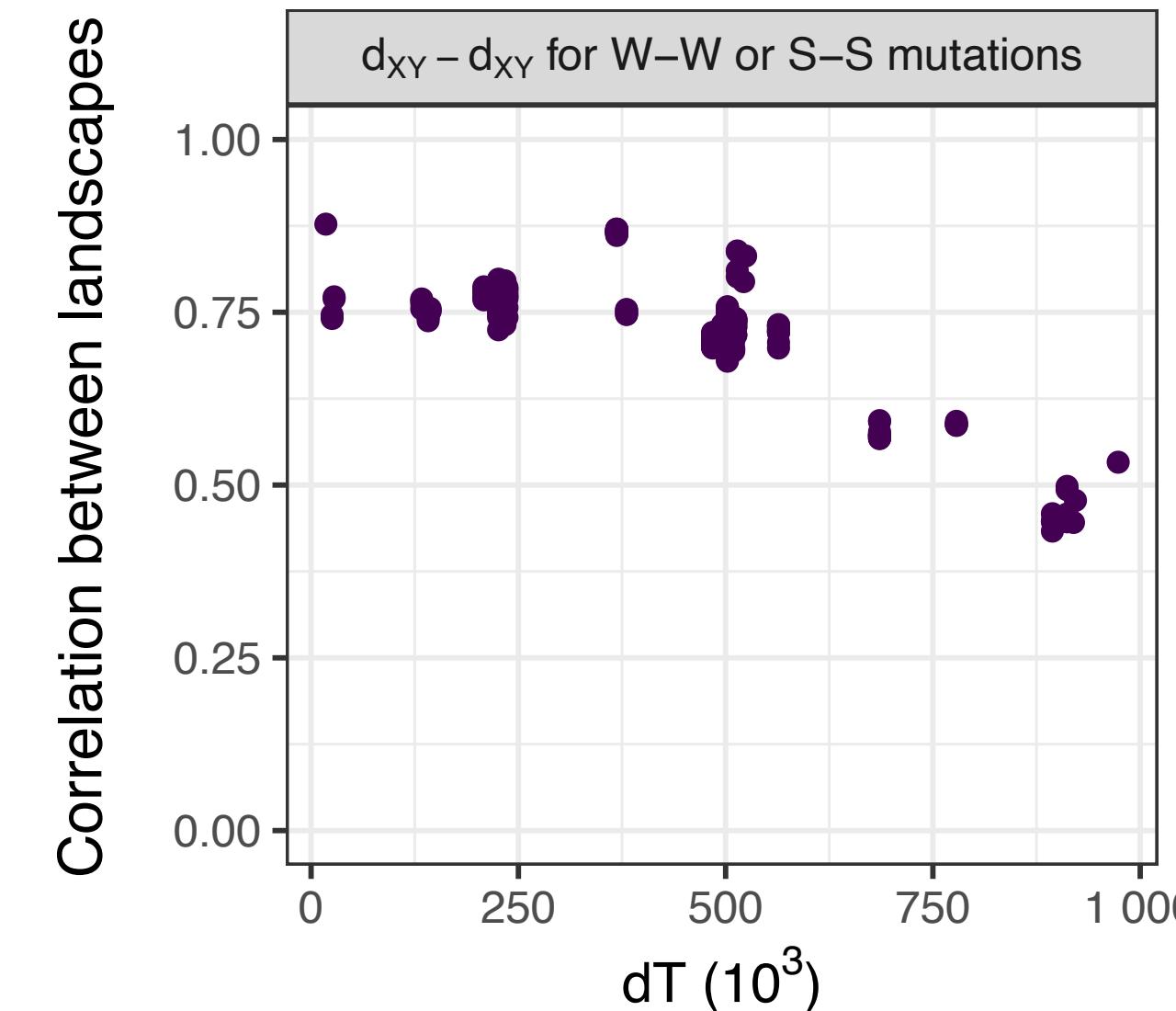
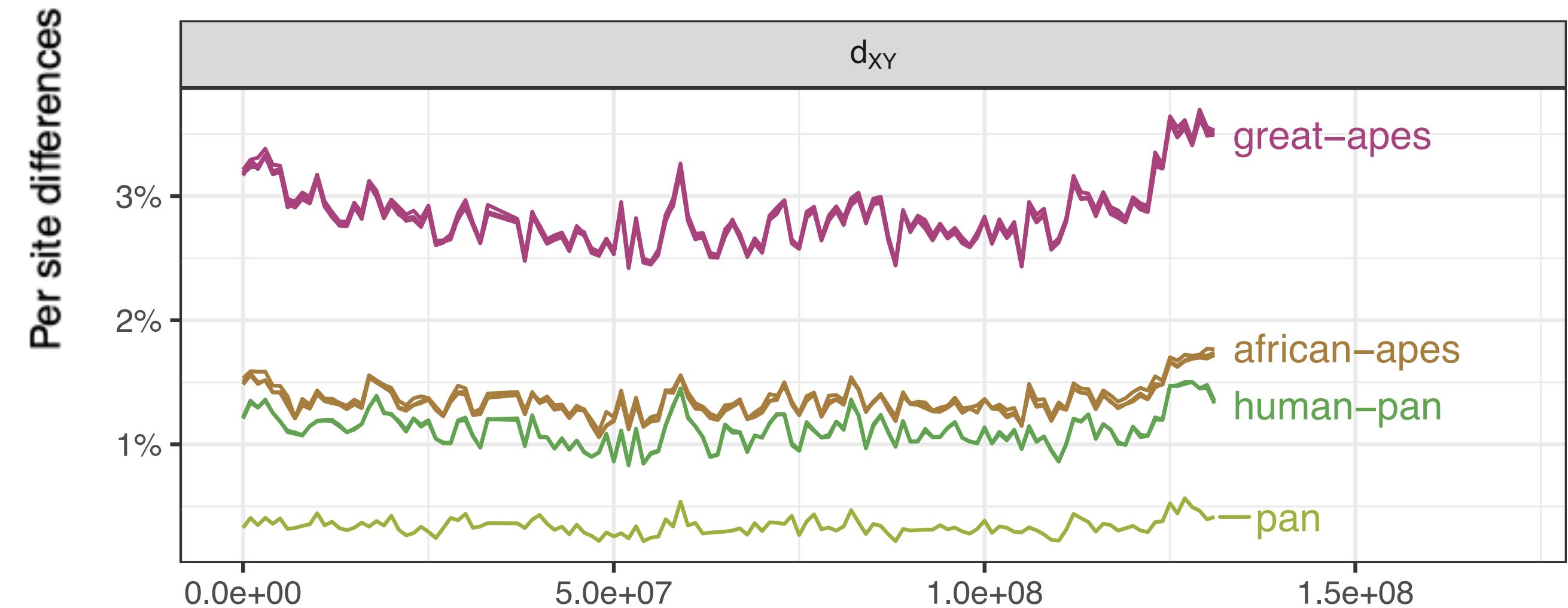


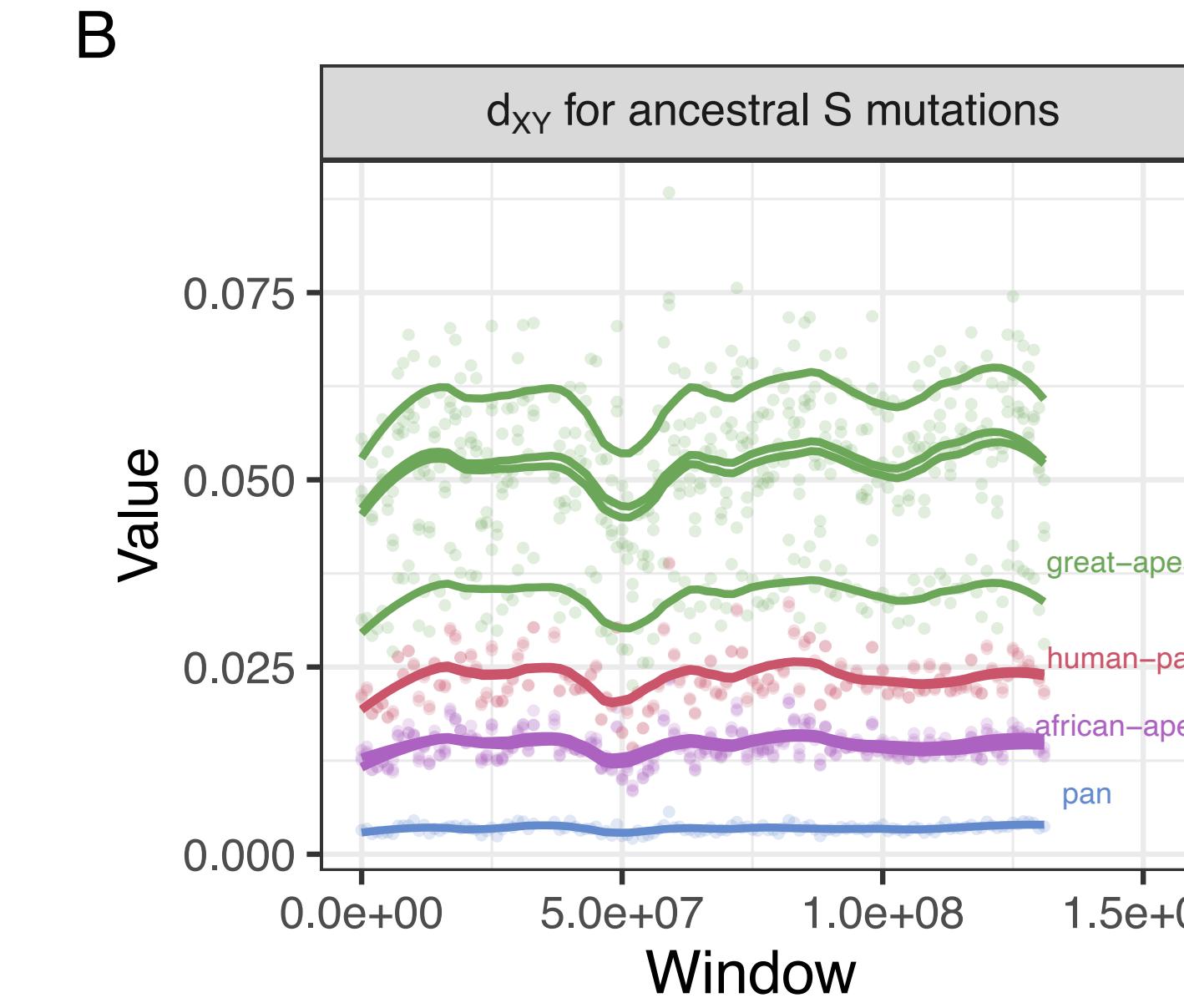
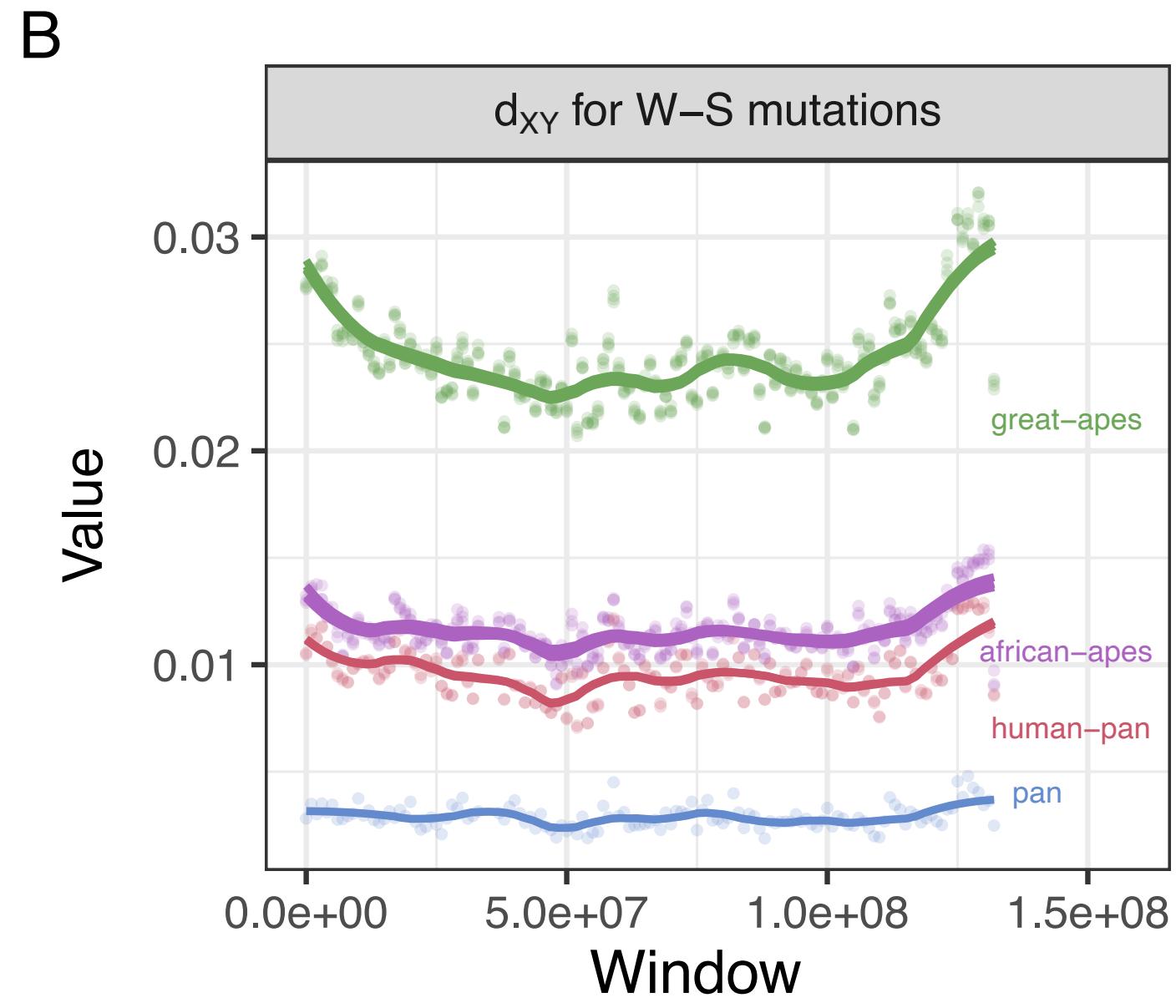
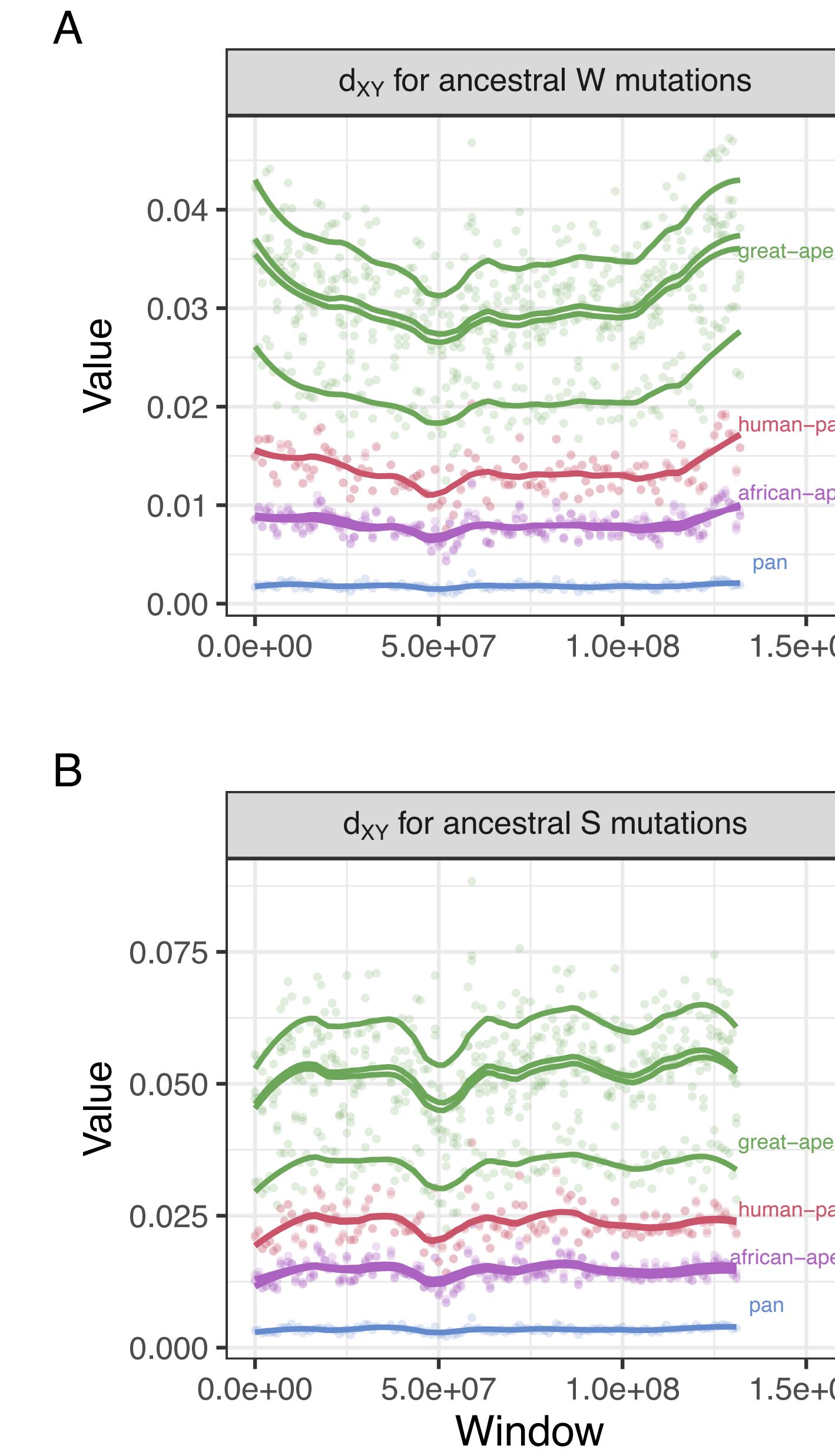
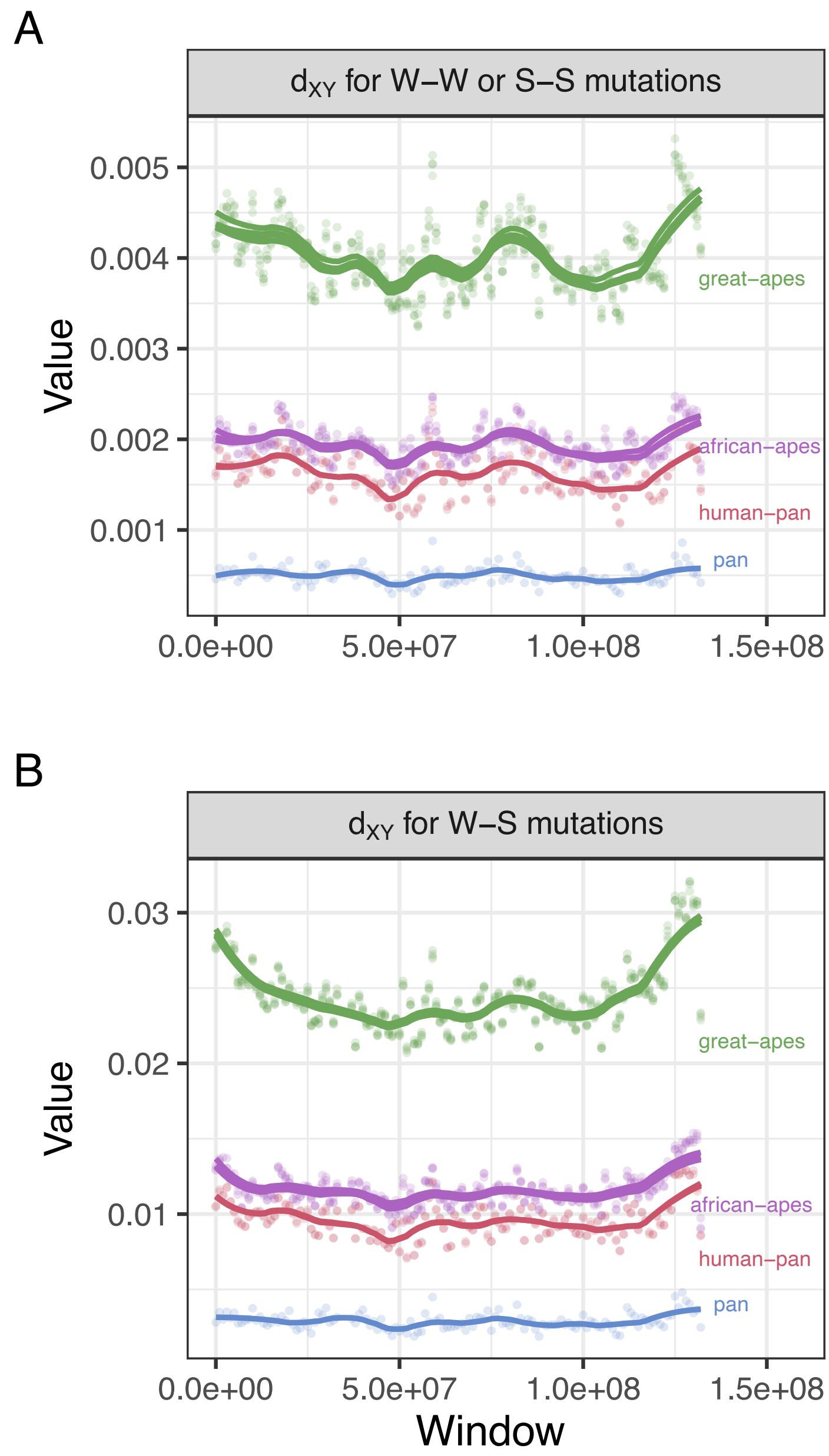
C



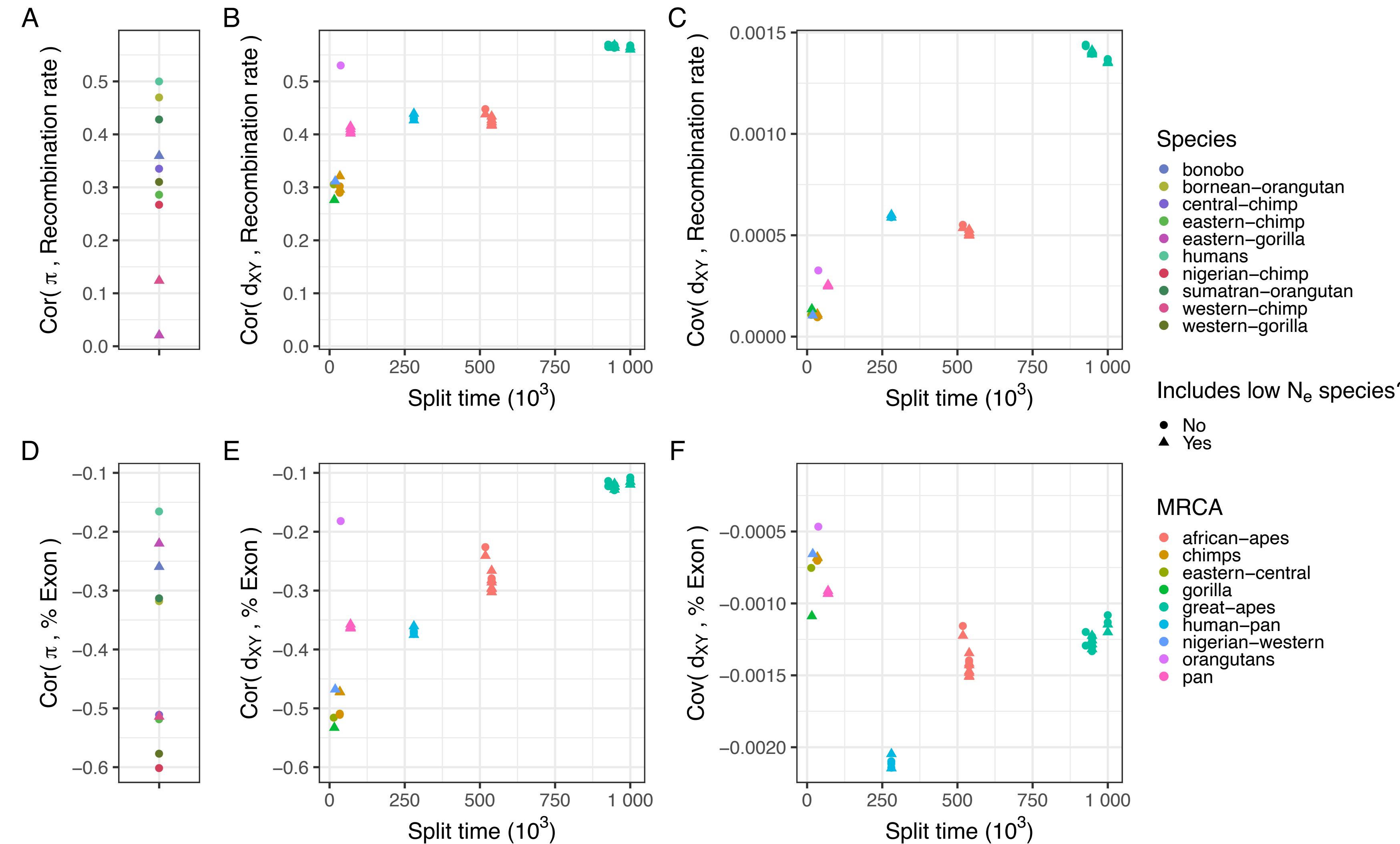
# GC-biased gene conversion

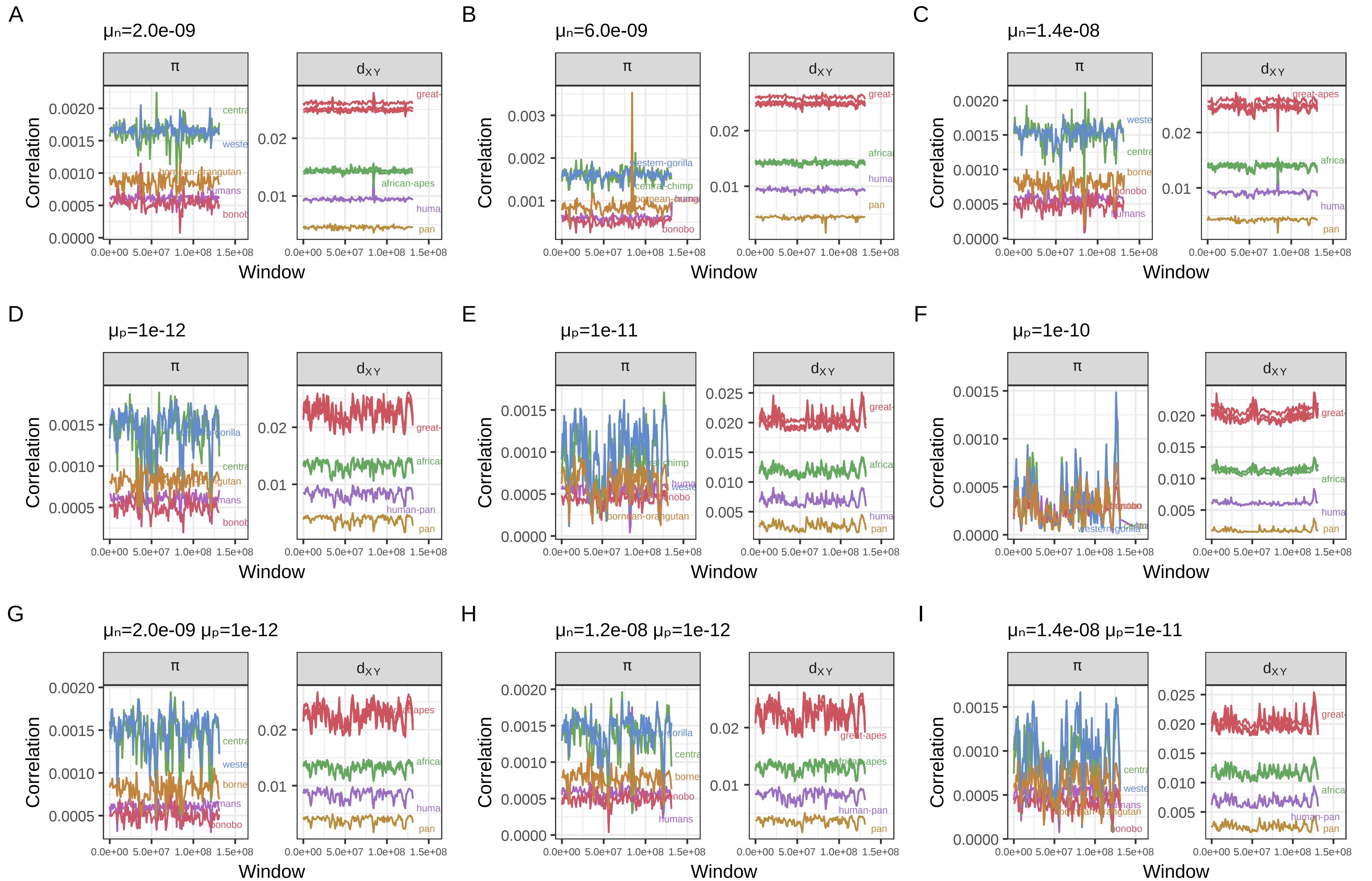
- gBGC can affect landscapes of variation
- It does not seem to account for the correlations

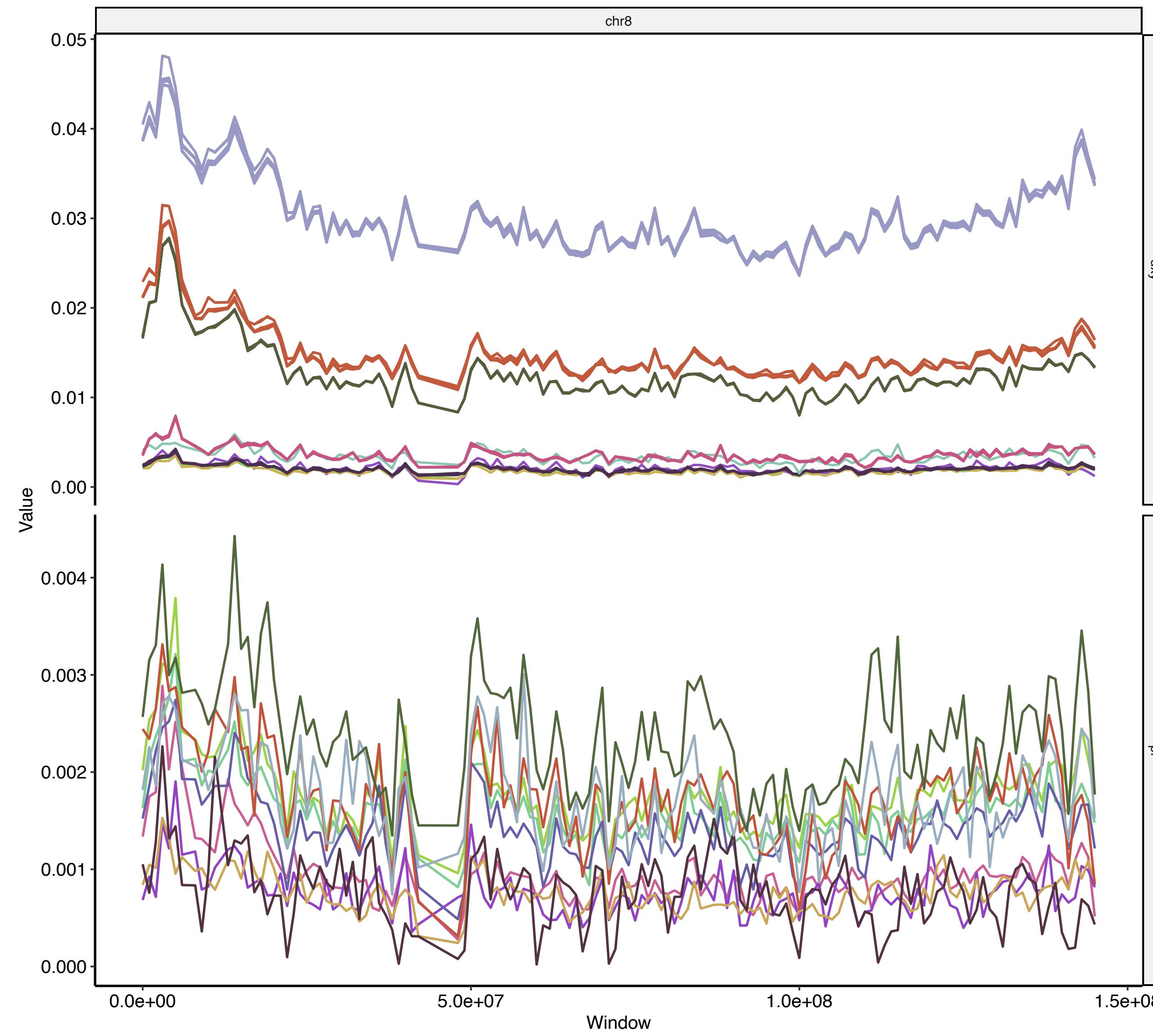




# Correlations between genomic variation and features







Species

humans	nigerian-chimp	bornean-orangutan	nigerian-western	human-pan
central-chimp	bonobo	sumatran-orangutan	chimps	african-apes
western-chimp	eastern-gorilla	eastern-central	orangutans	great-apes
eastern-chimp	western-gorilla	gorilla	pan	

- Recombination and exon map were shared among all species
- Within exons, mutations can be neutral, beneficial or deleterious
- Elsewhere, mutations were neutral
- Simulations with varying rates of deleterious and beneficial mutations
- Also varied the mutation rate along the genome (without selection)

