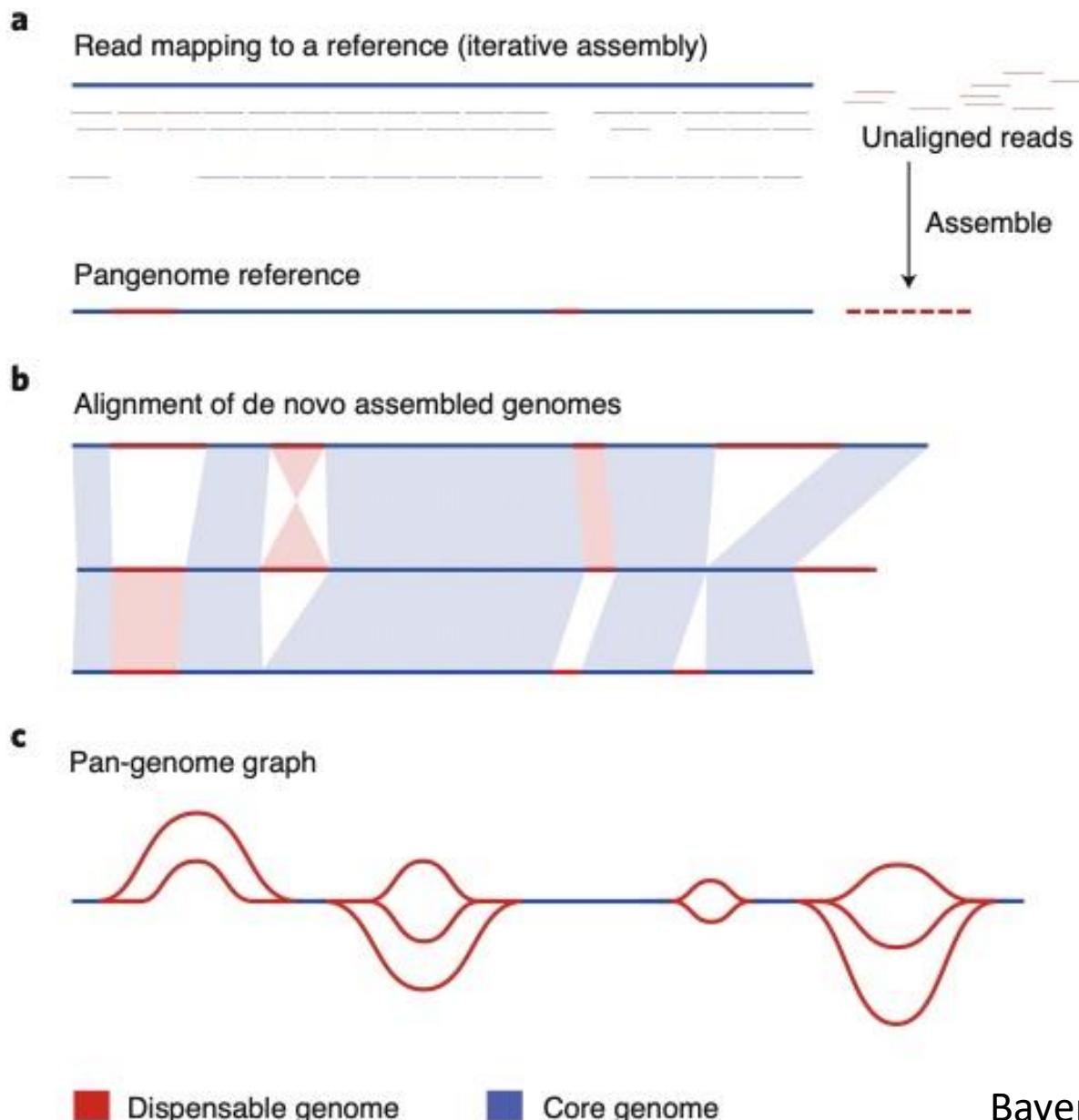


Pangenomes as a new tool for studying ecology and evolution of natural populations

Scott V. Edwards

Museum of Comparative Zoology, Harvard University, Cambridge, USA

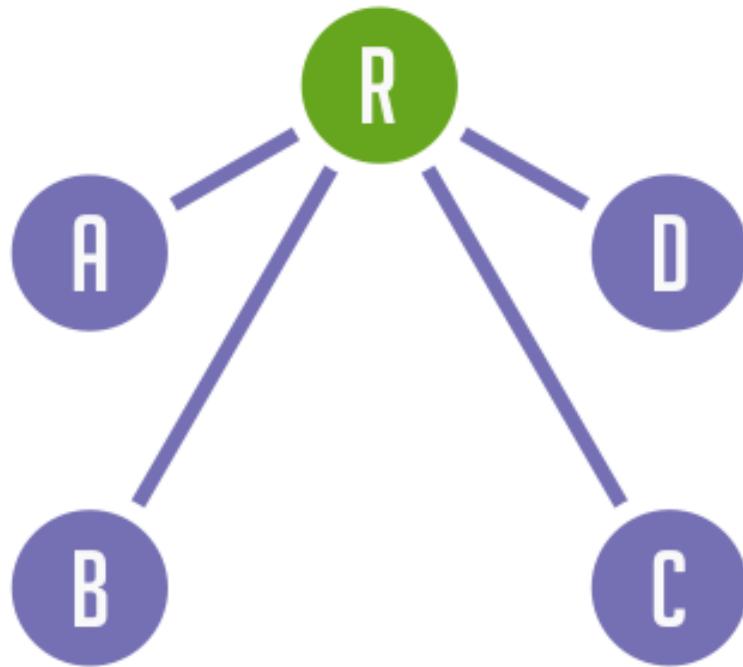
Pangenomes: moving beyond reference-based genomics



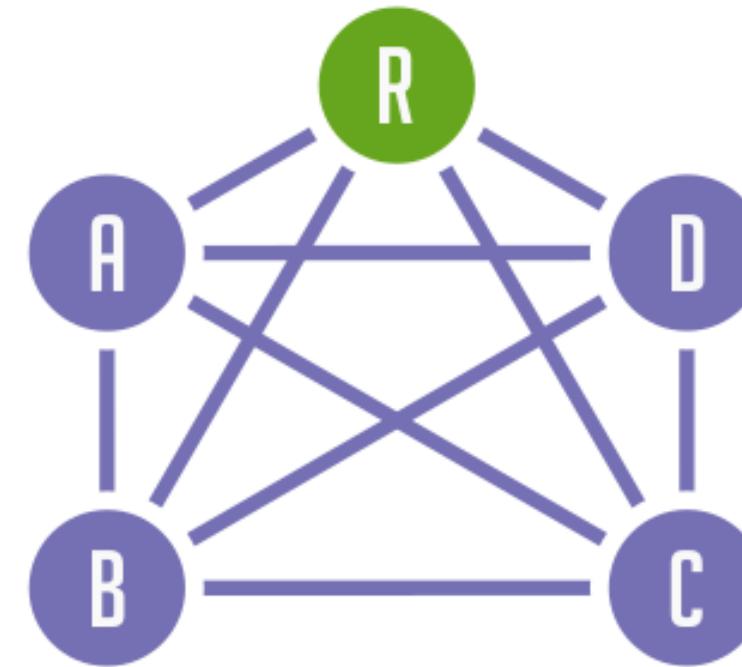
Reference-free genomics

Reference model

Genomic

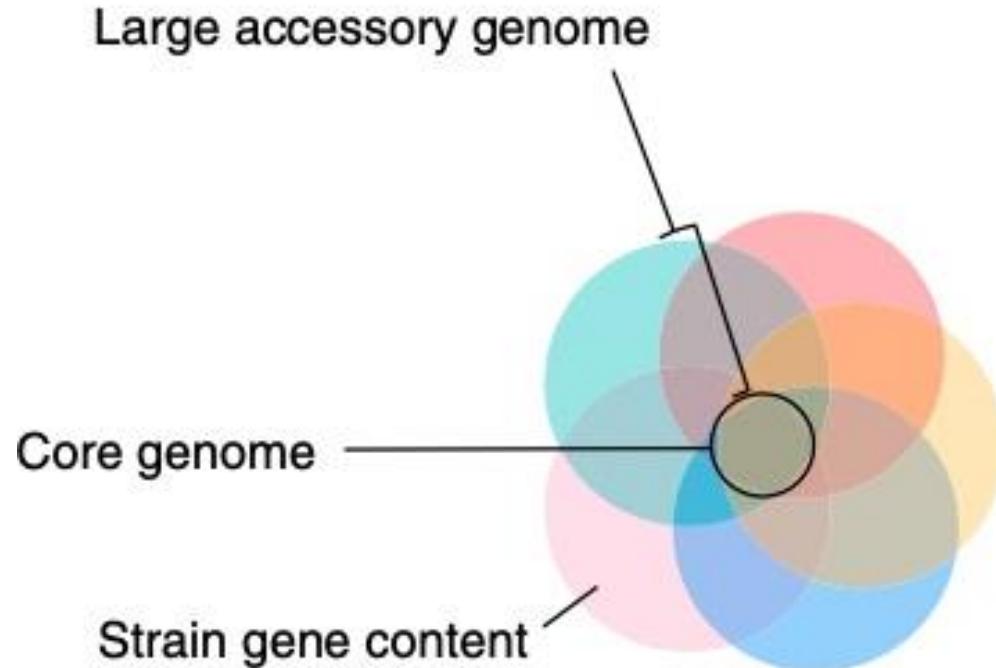


Pangenomic

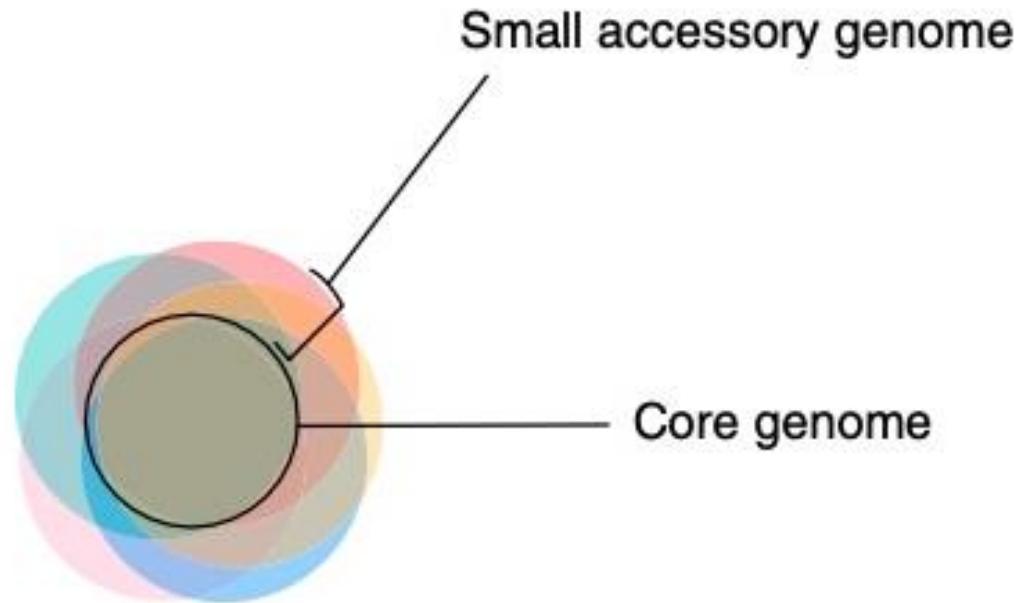


Open and closed pangenomes

Open pangenomes

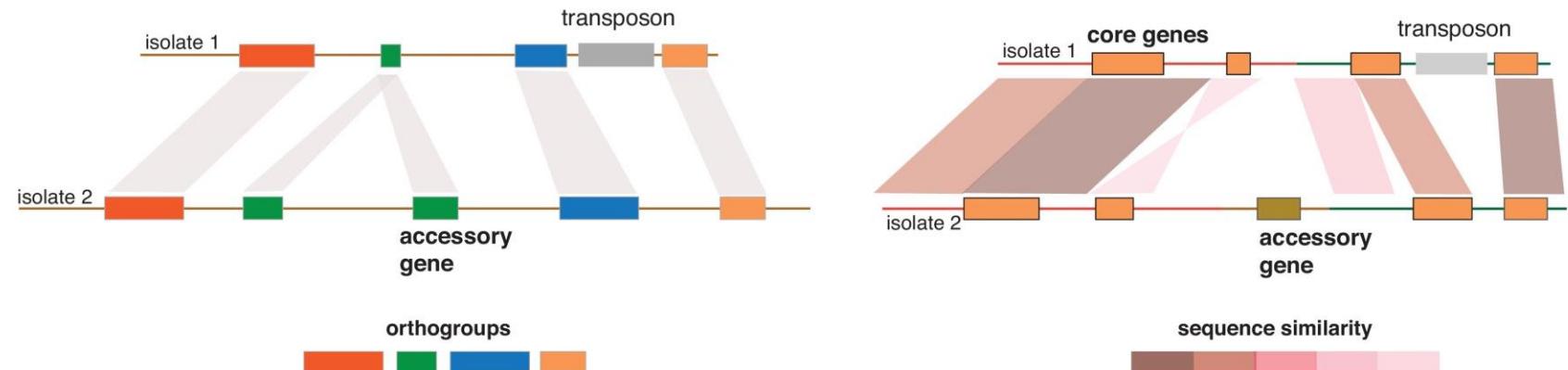
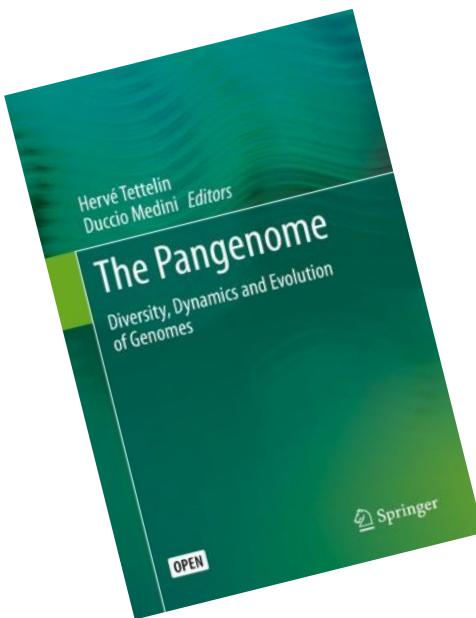


Closed pangenomes

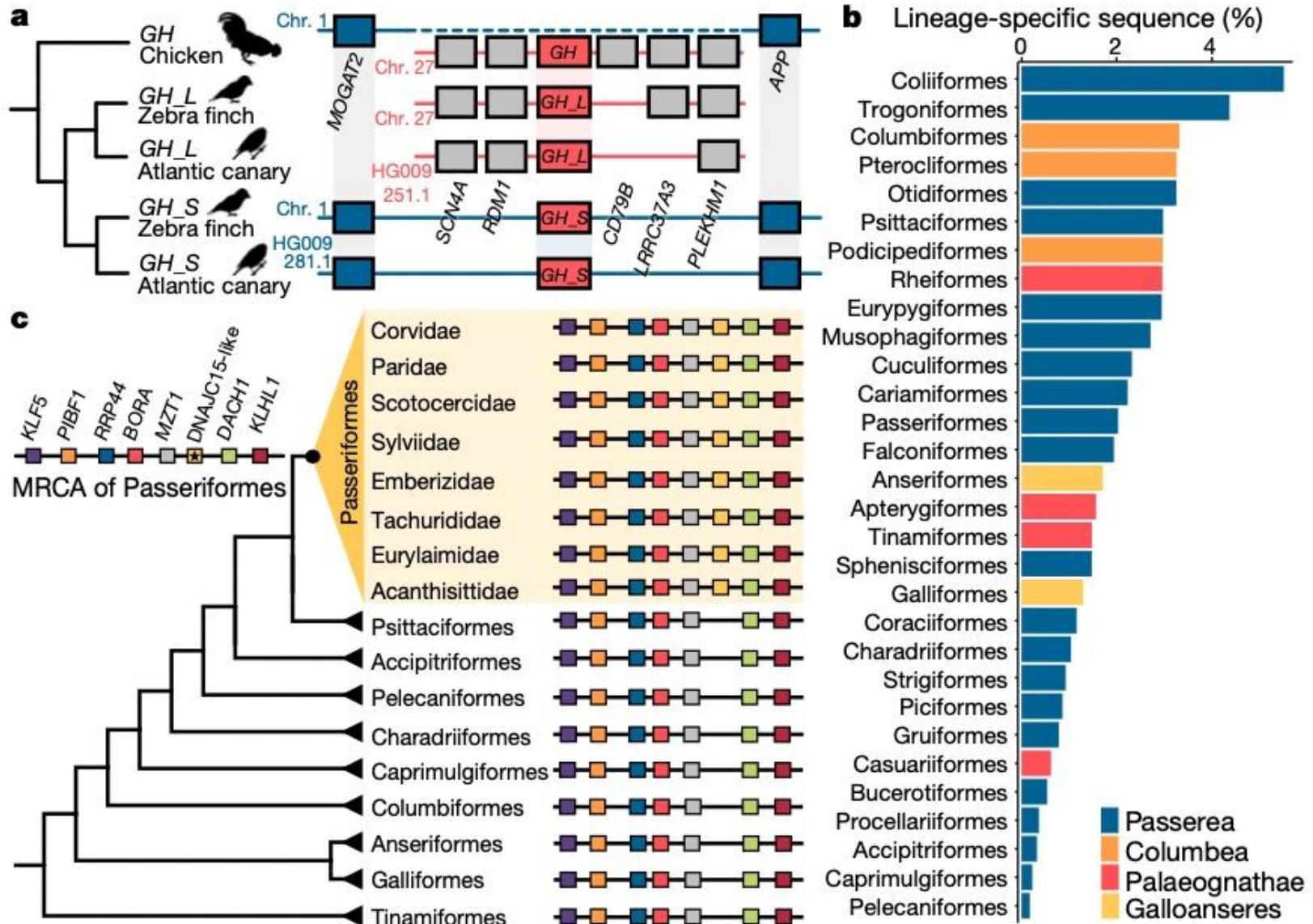


The eukaryotic pangenome

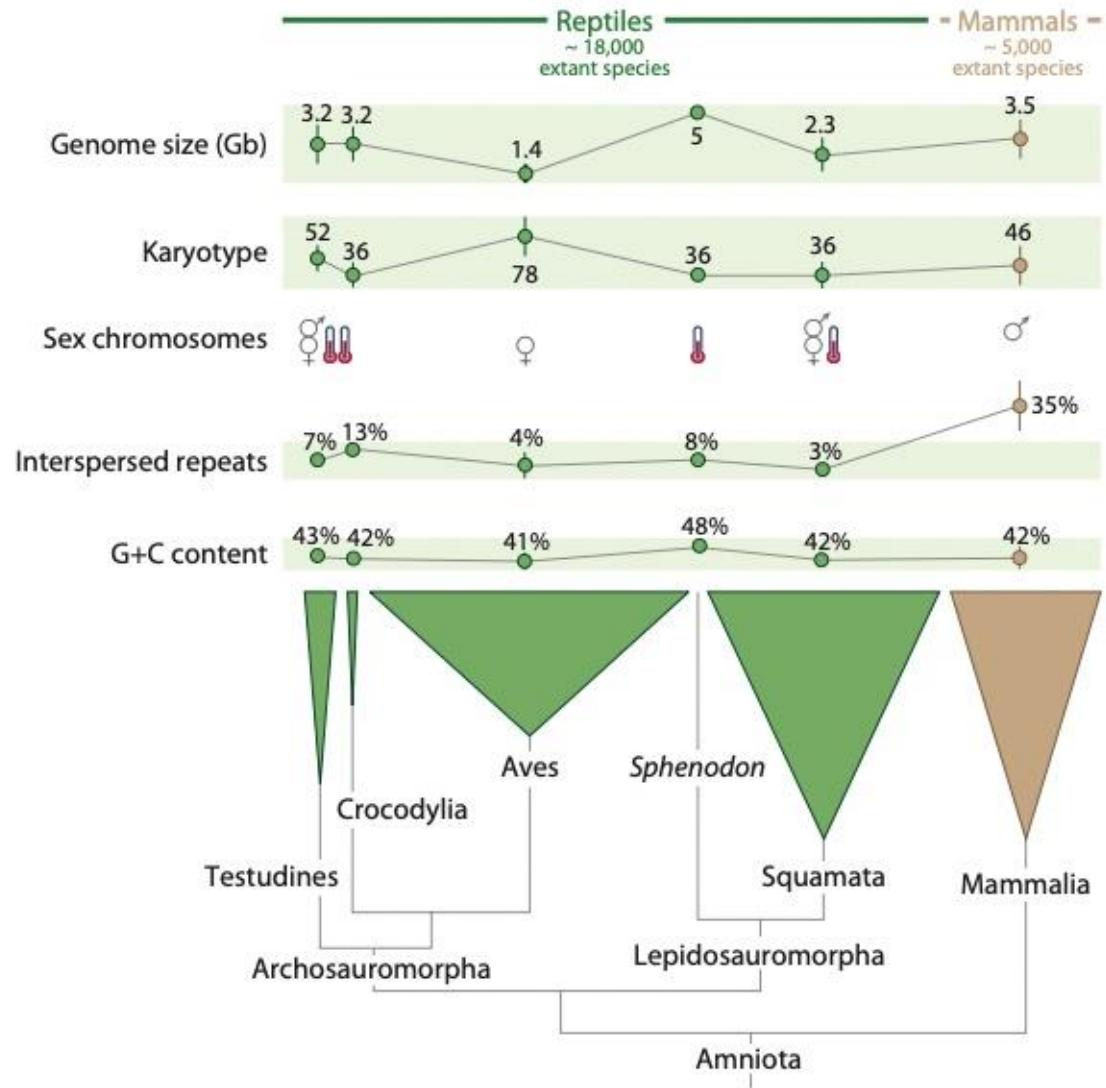
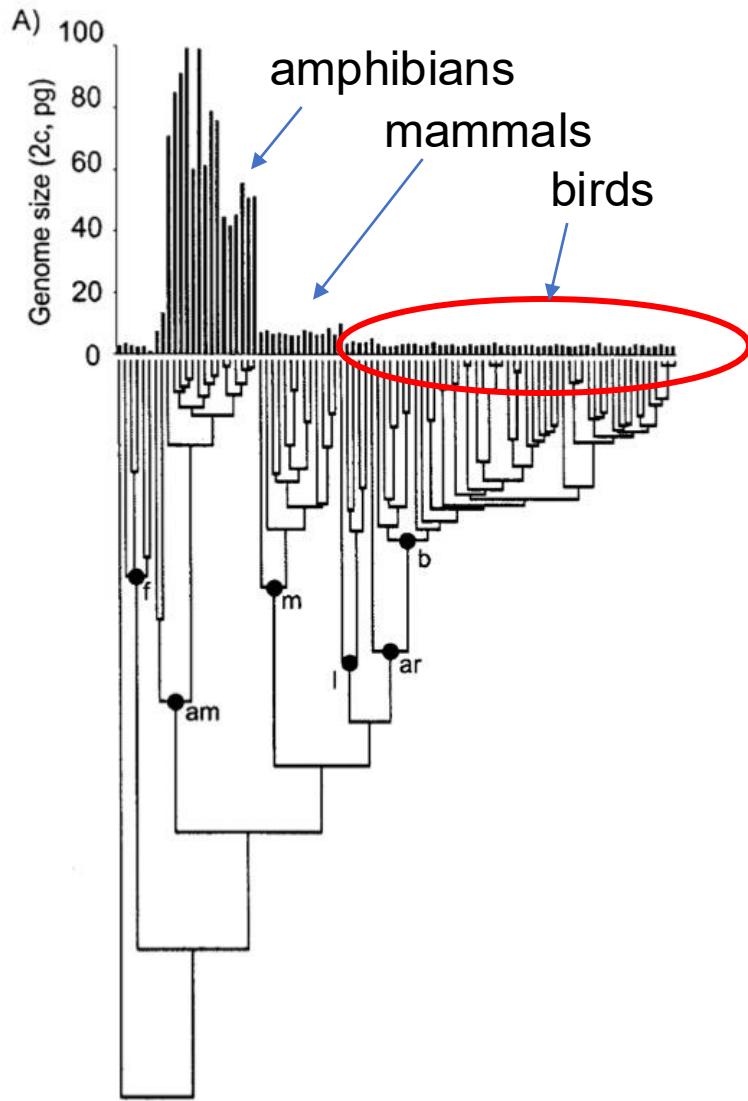
- “The existence of pangenesomes in eukaryotes is debated...Pangome studies in eukaryotes are challenging due to their more complex genome and architectures and a lack of replete genome-level sampling” (Brockhurst et al. 2019. *Current Biology*)



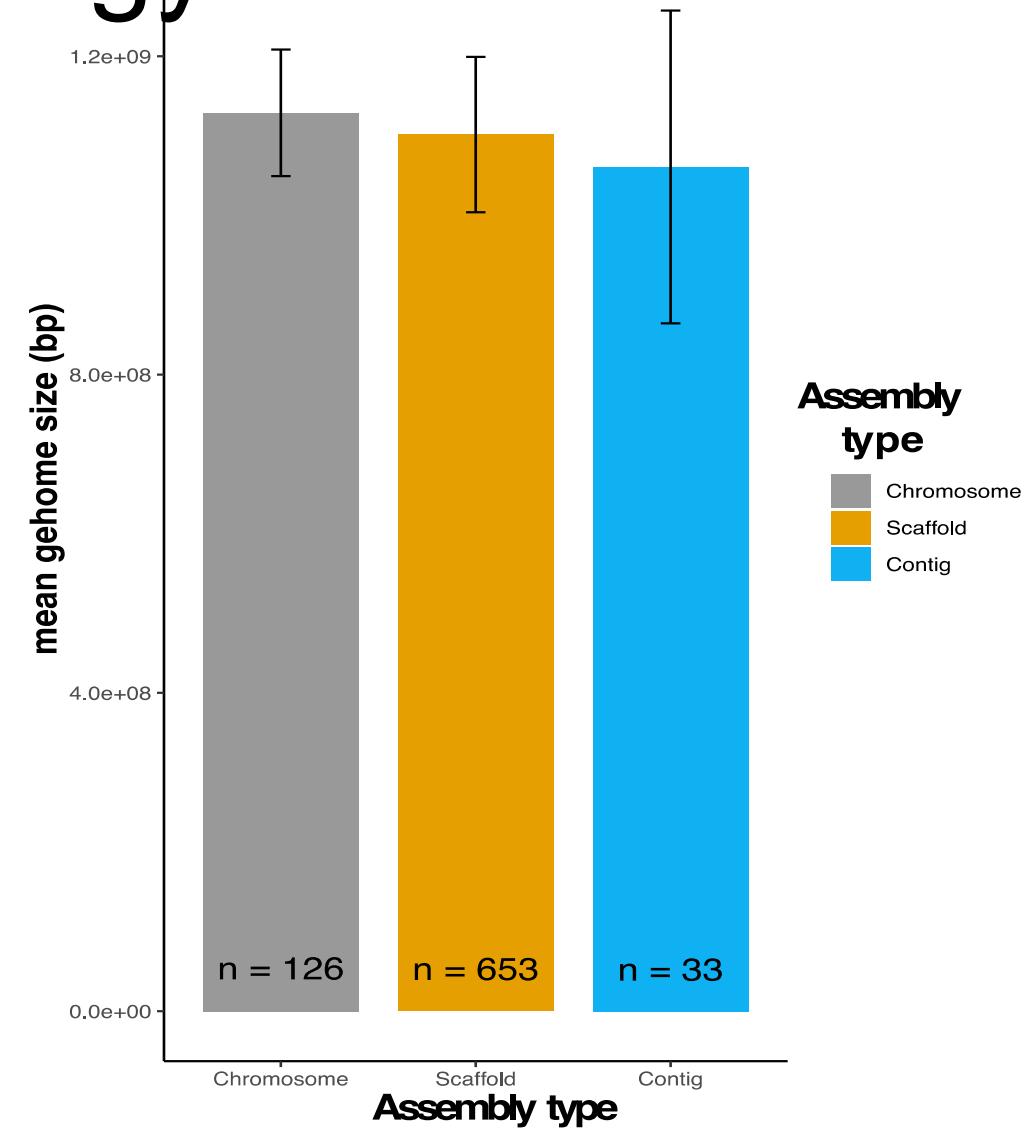
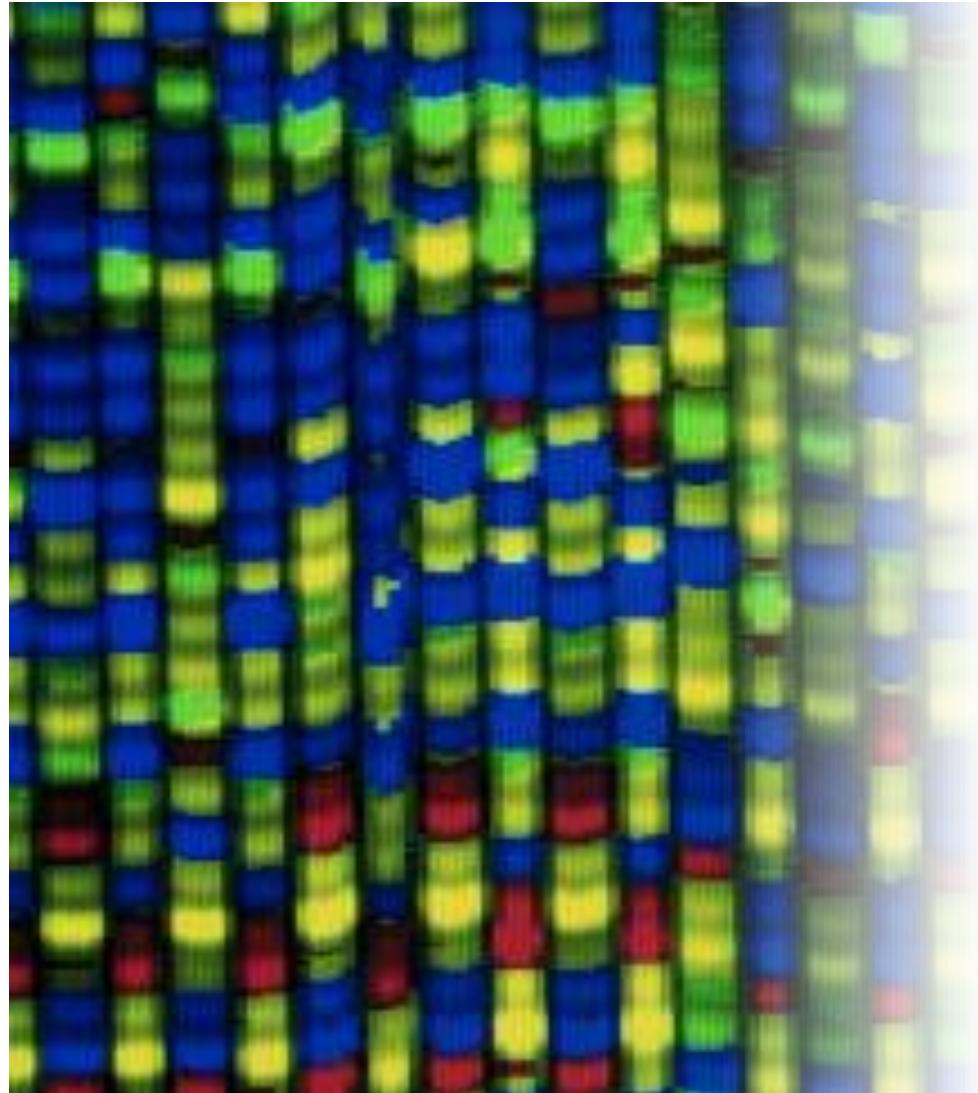
Pangenome approach to comparative genomics



Birds have small, streamlined genomes

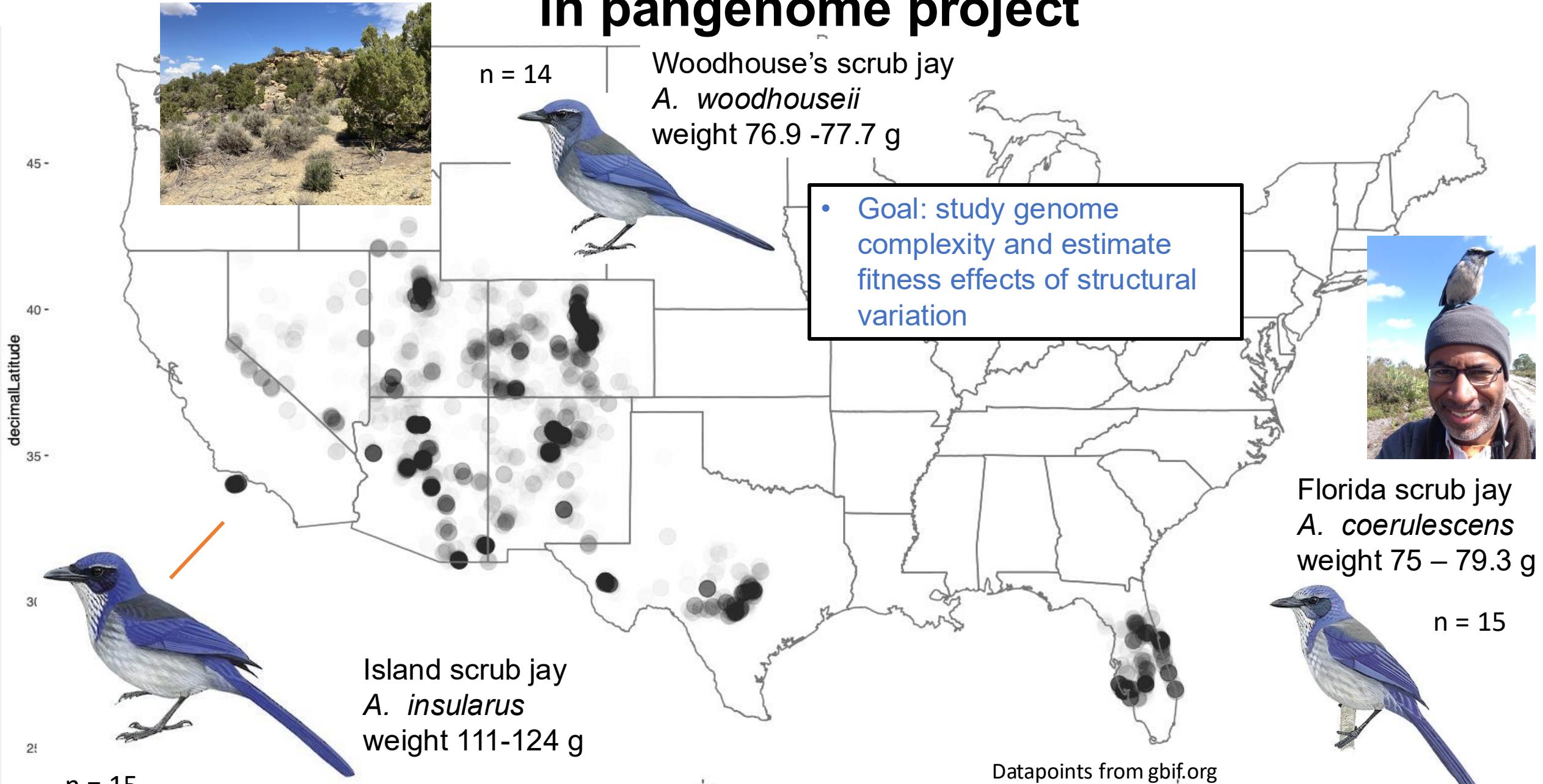


Avian genomes are growing with each new technology



Data from NCBI, accessed 13 Nov. 2021

Three scrub-jay (*Aphelocoma*) species in pangenome project



The Evolution of Comparative Phylogeography: Putting the Geography (and More) into Comparative Population Genomics

GBE

Scott V. Edwards ^{1,2,*}, V. V. Robin³, Nuno Ferrand⁴, and Craig Moritz⁵

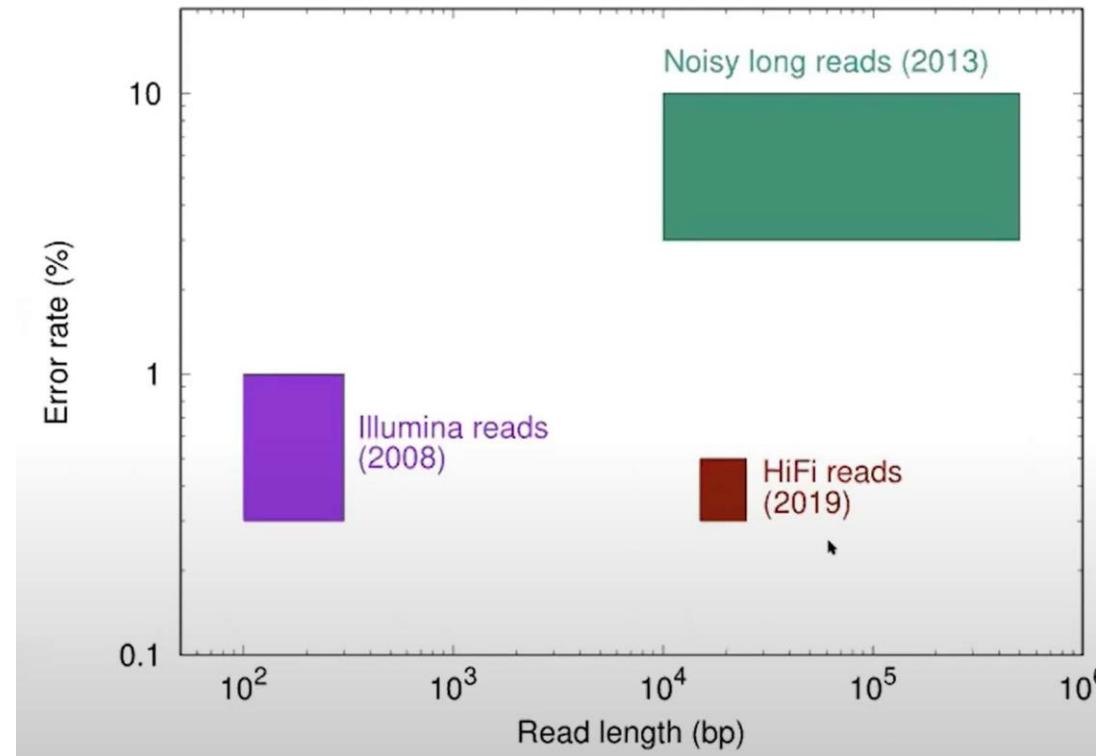
Table 1

Conceptual Relationships between the Fields of Comparative Population Genomics, Landscape Genomics, and Comparative Phylogeography

Concept/Parameter	Comparative Population Genomics	Landscape Genomics	Comparative Phylogeography
Comparative perspective	Growing	Nascent	Mature
Emphasis on space	No	Yes	Yes
Geographic scale	Random mating population	Region	Biome
Temporal scale	Arbitrary	Recent	Deep
Focus on:			
selection versus neutrality	Both	Both	Neutrality
recombination	Yes	Not yet considered	Not yet considered
geography versus environment	Nuisance parameters	Environment	Both
Future use of whole-genome sequencing	Yes	Likely	Unlikely
Growth out of museum collections community	No	No	Partial

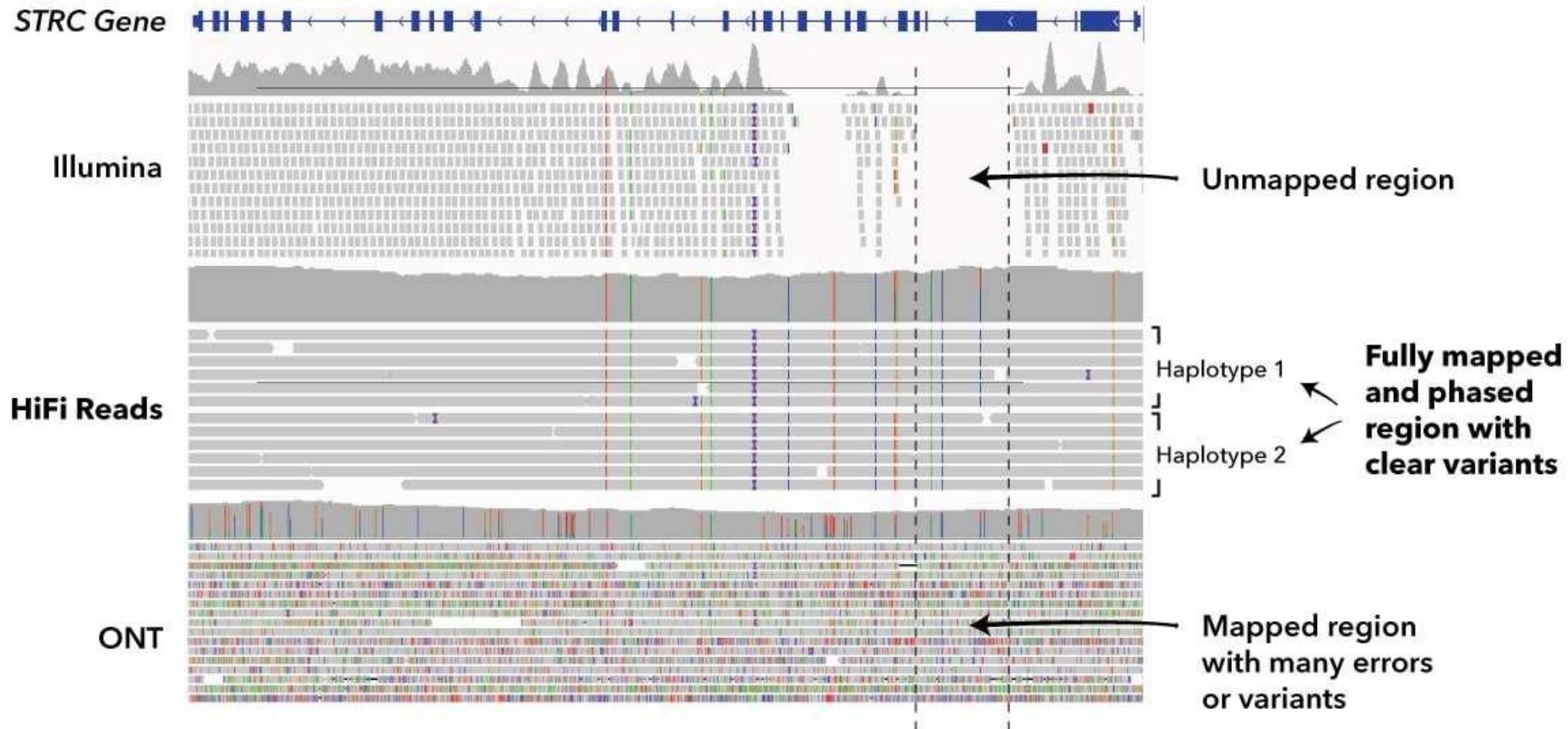
PacBio HiFi reads are long and accurate

- ▶ HiFi reads: long & accurate
- ▶ A breakthrough every ~5 years
- ▶ Most existing assemblers cannot make full use of the accuracy



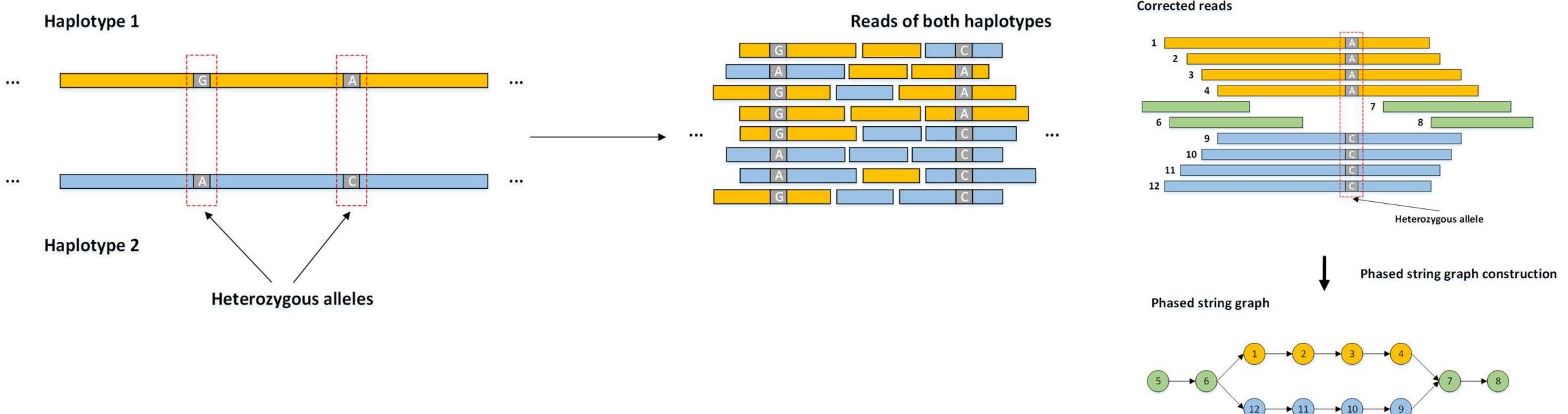
Courtesy Haoyu Cheng, Dana Farber Cancer Institute

PacBio HiFi reads are long and accurate



HG002 GRCh38 chr15:43,599,422-43,619,001 (19 kb)

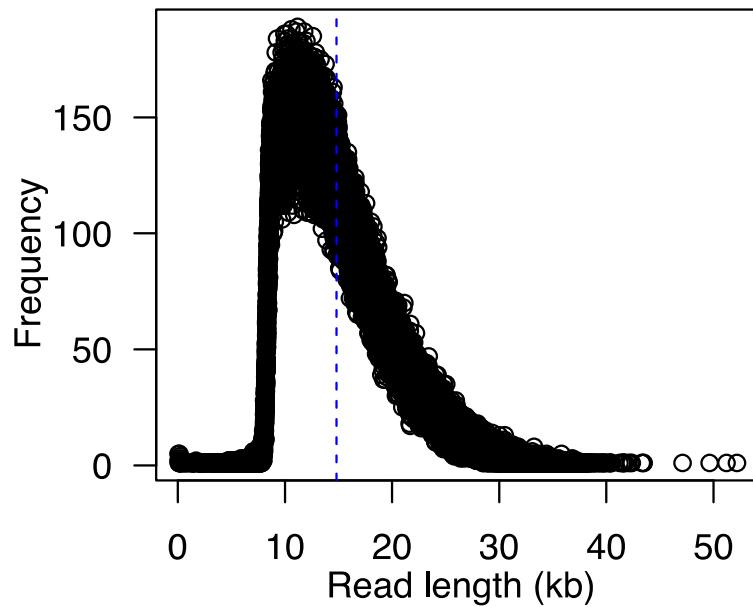
Hifiasm – a HiFi accurate read assembler that resolves haplotypes



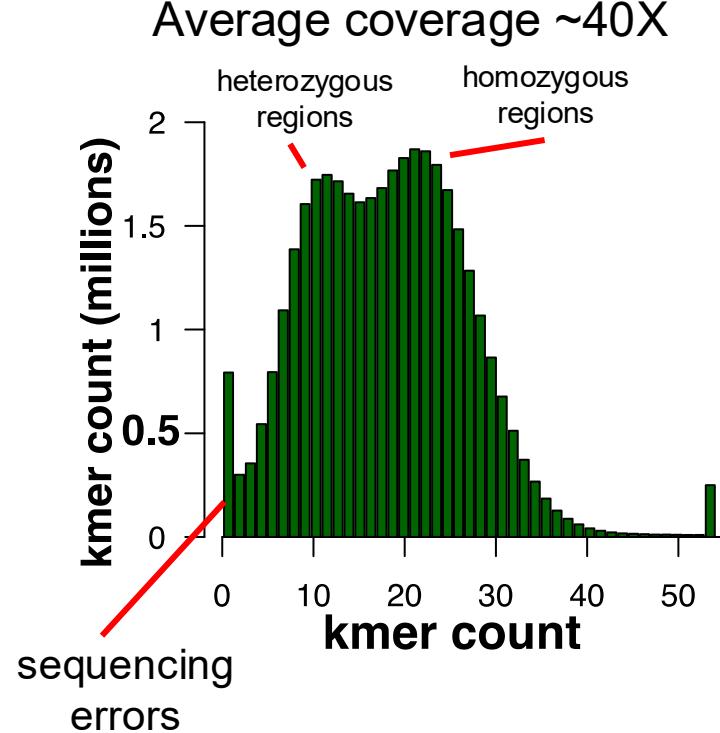
Courtesy Haoyu Cheng, Dana Farber Cancer Institute

Scrub-jay PacBio HiFi data characteristics

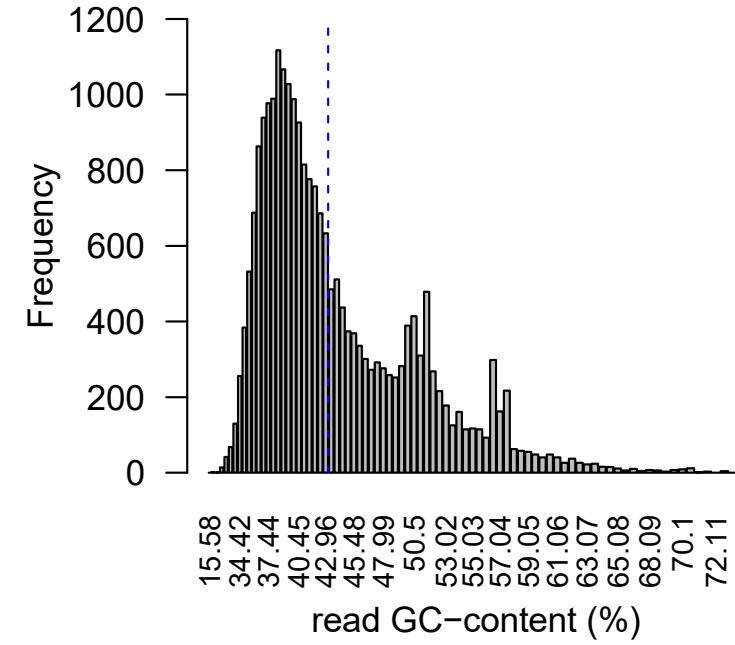
Average read length 14.8 kb



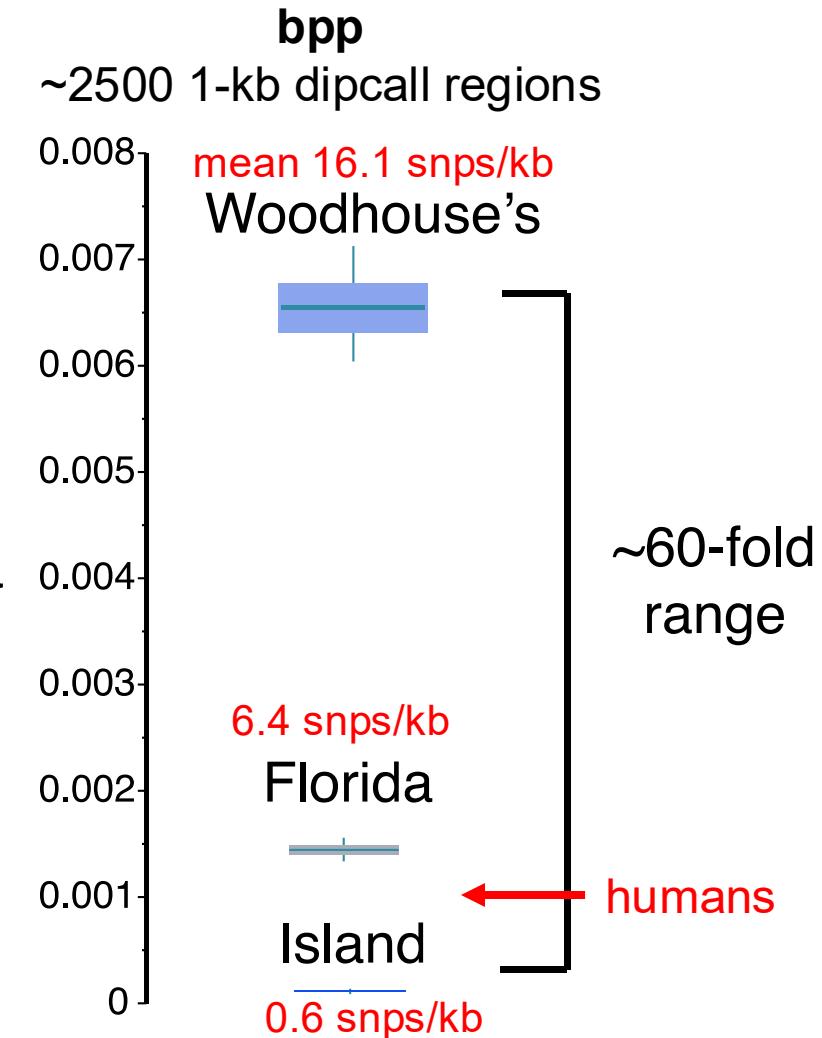
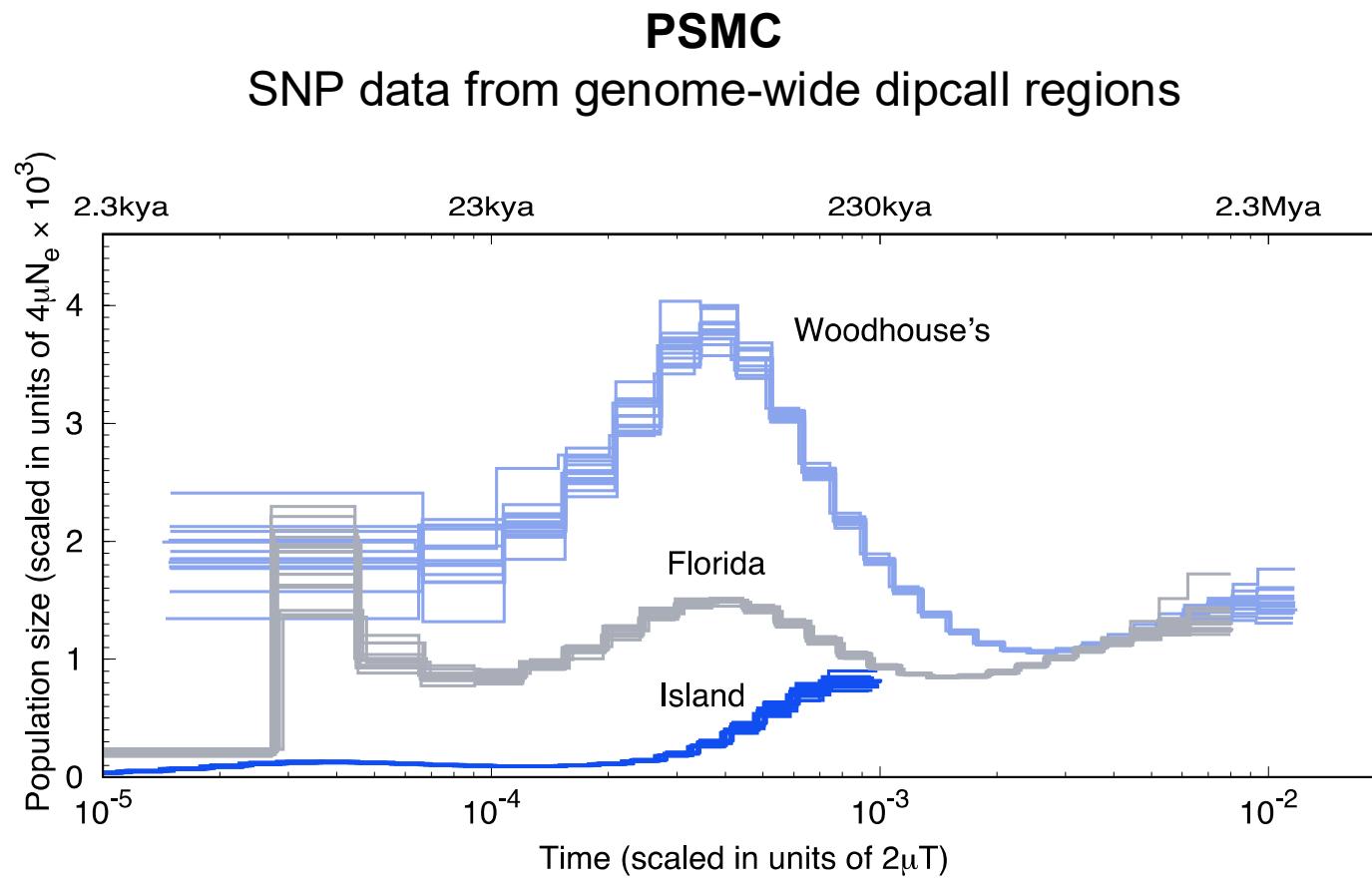
Average coverage ~40X



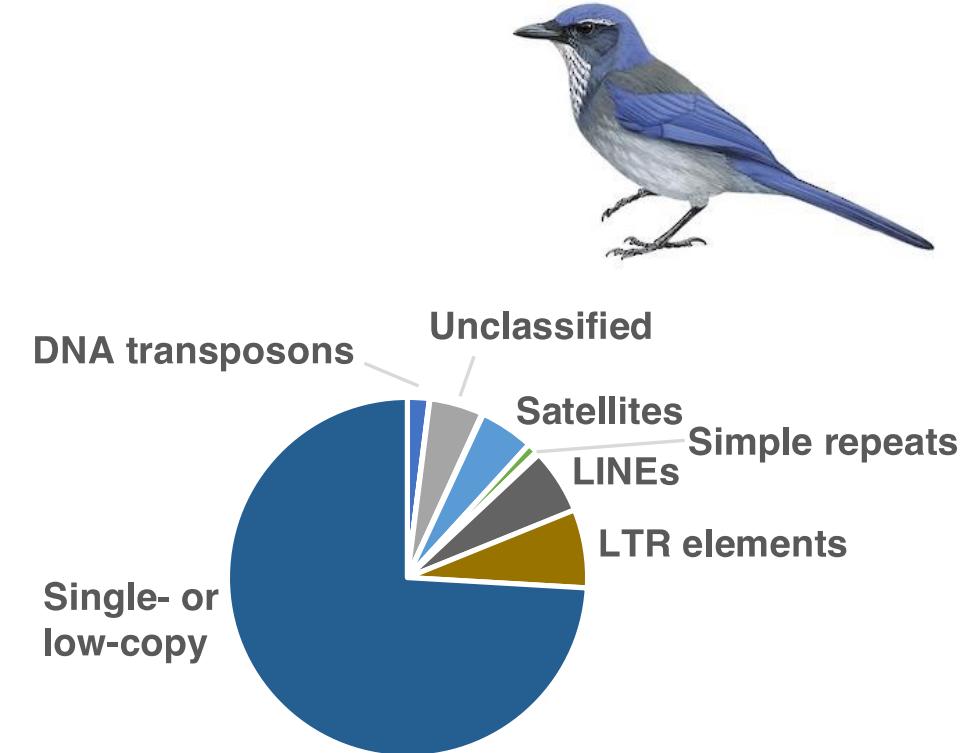
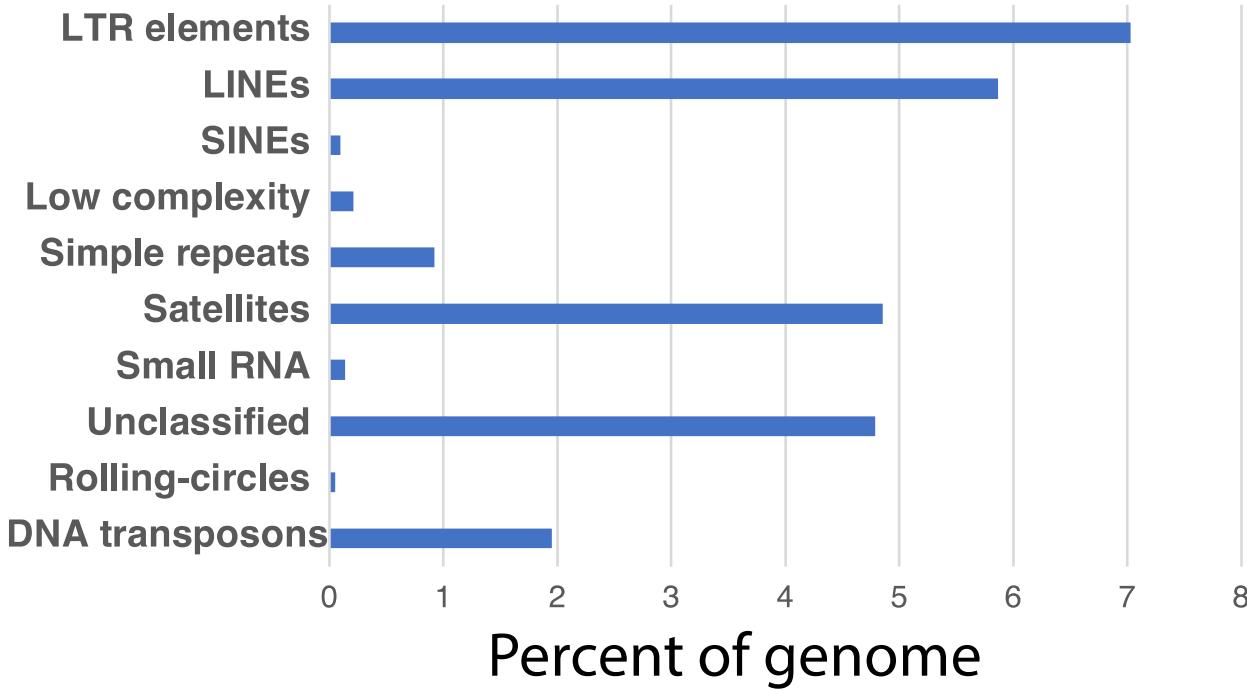
GC-content



60-fold range in effective population size across species

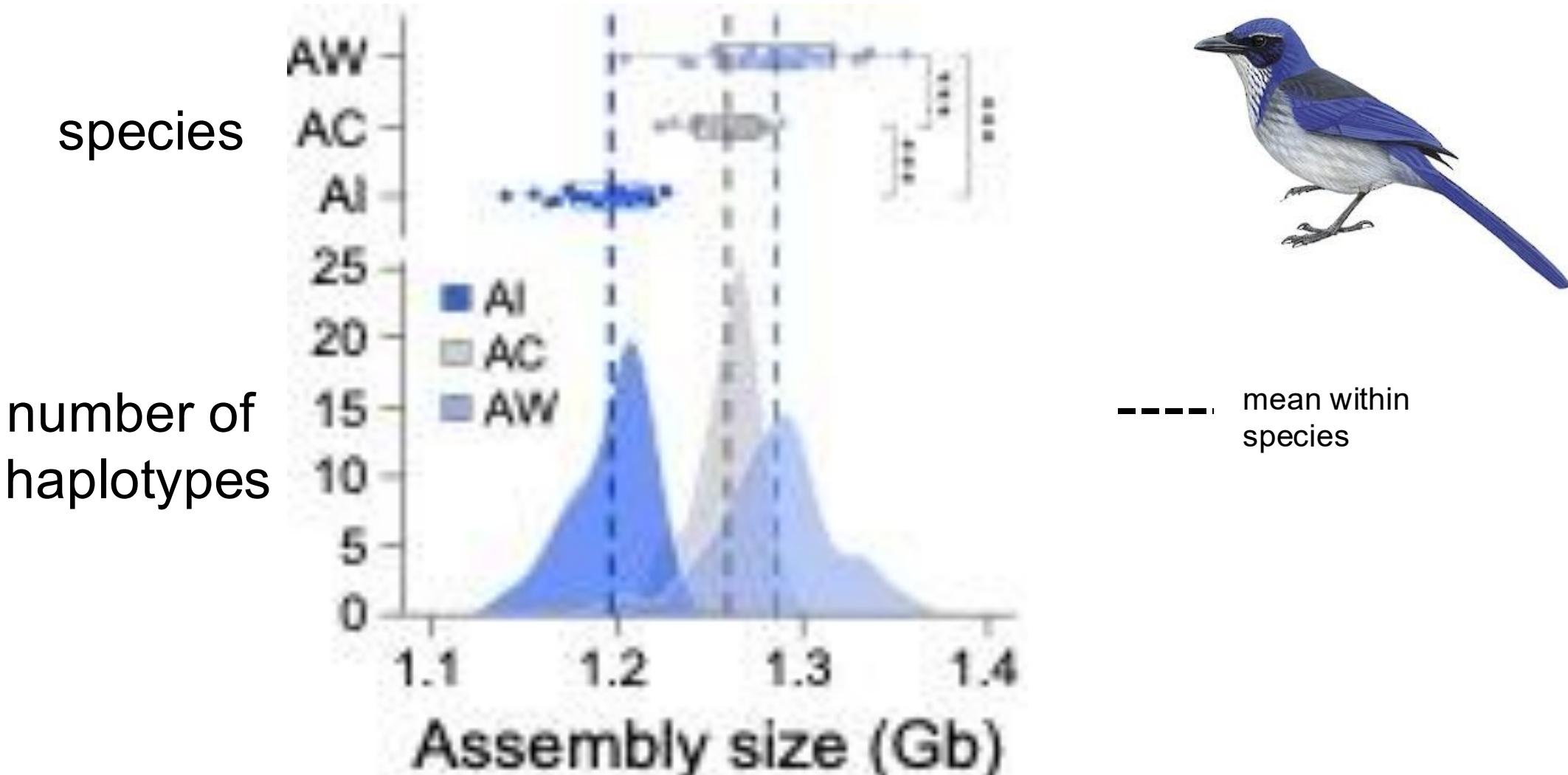


RepeatMasker analysis suggests over 25% repeats and transposable elements

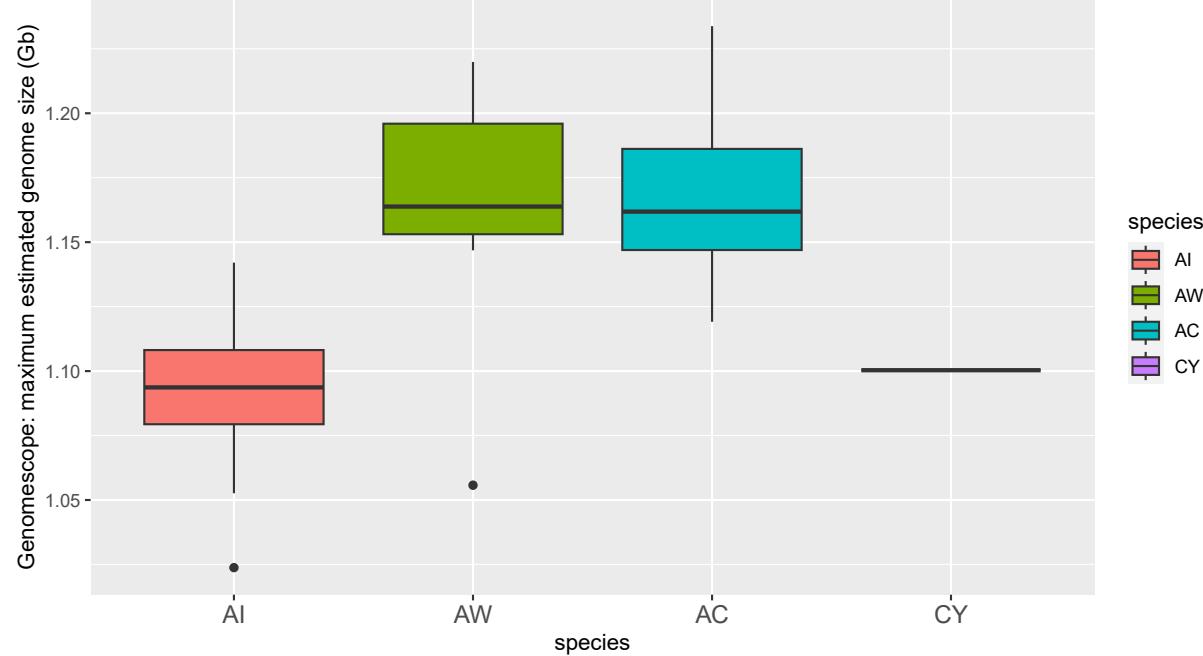
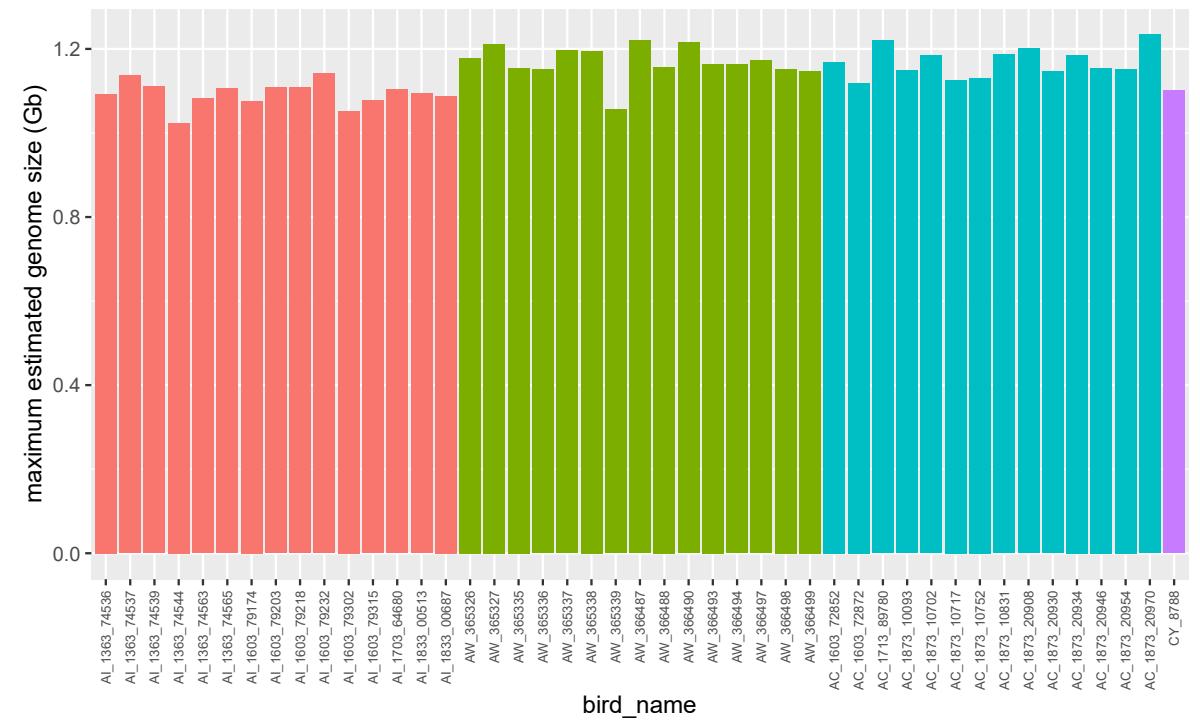


Assemblies of Island Scrub Jays are ~100 Mb smaller than Woodhouse's Scrub Jay

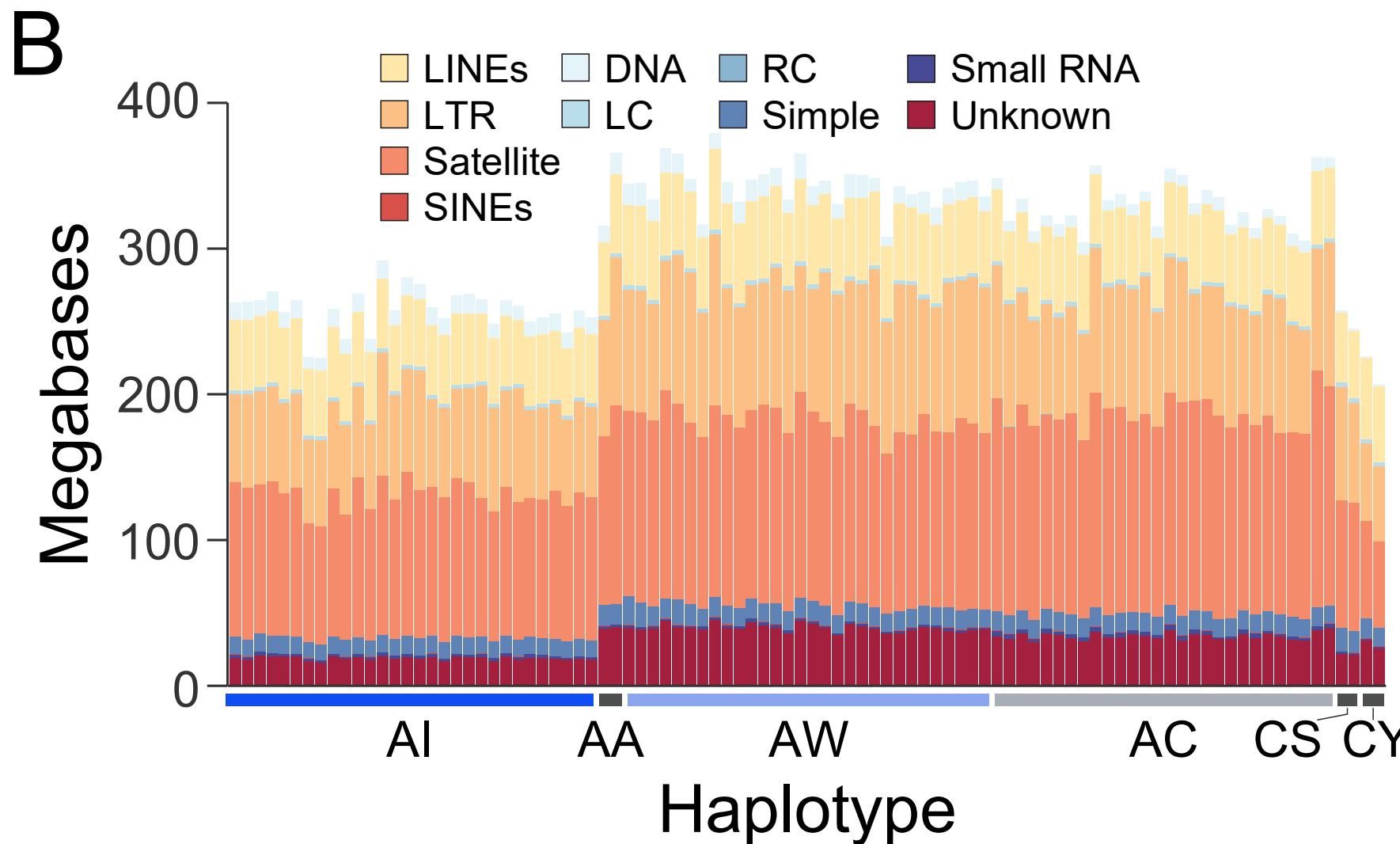
N = 30, 30 and 28 haplotypes (AI, AW, AC)



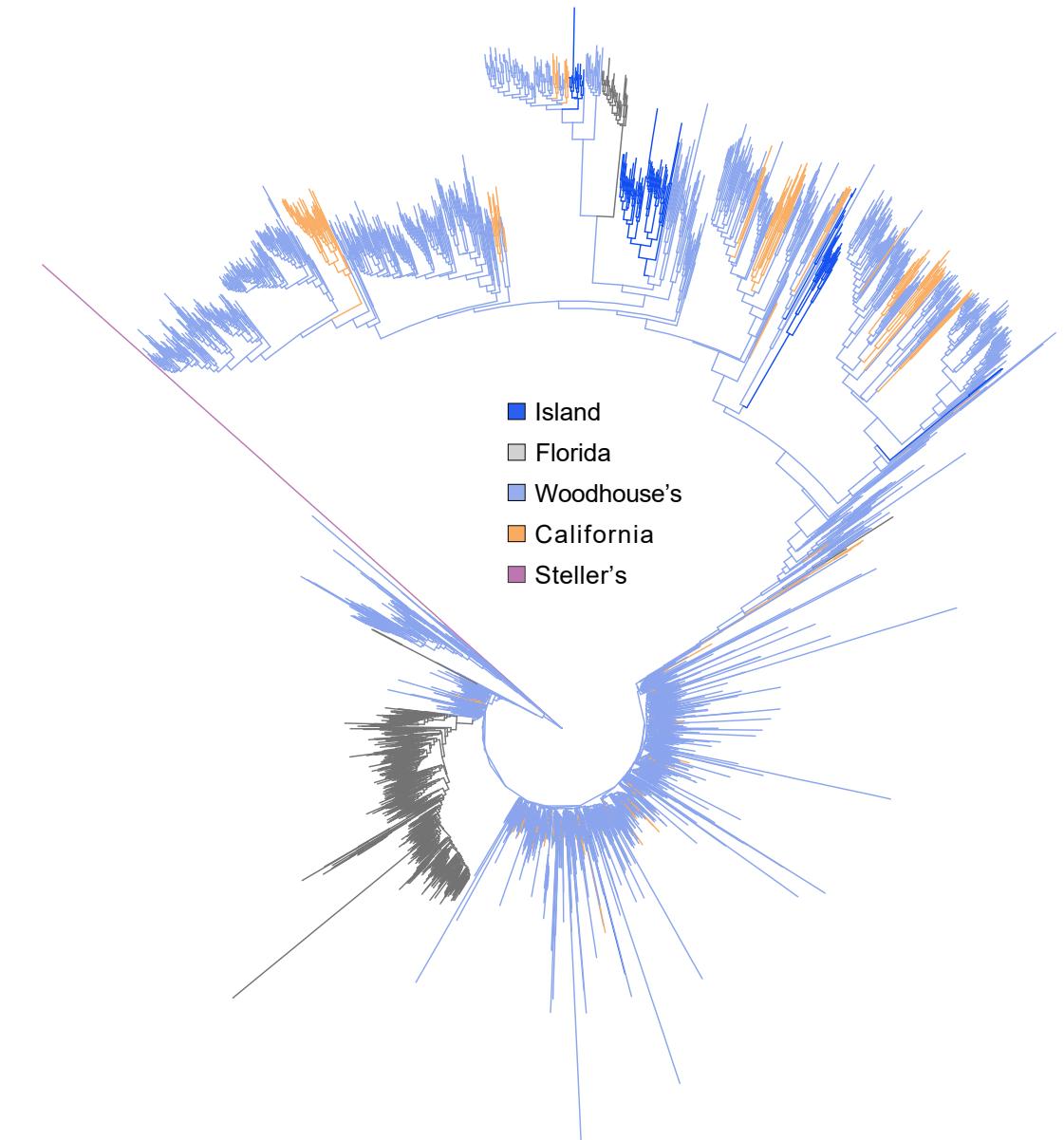
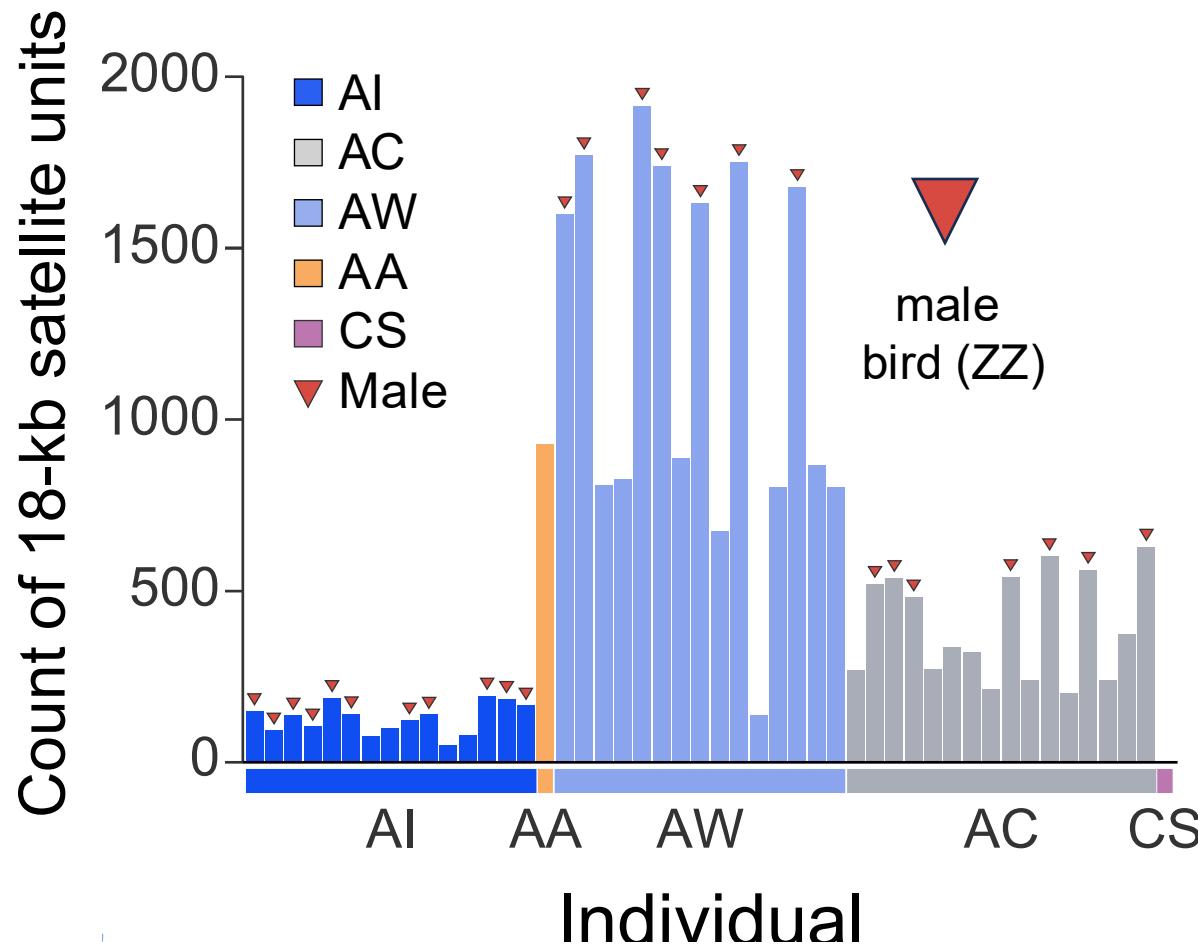
Estimates of max genome size from Genomescope using k-mers



Interspecific variation in repeat content...



...and satellites



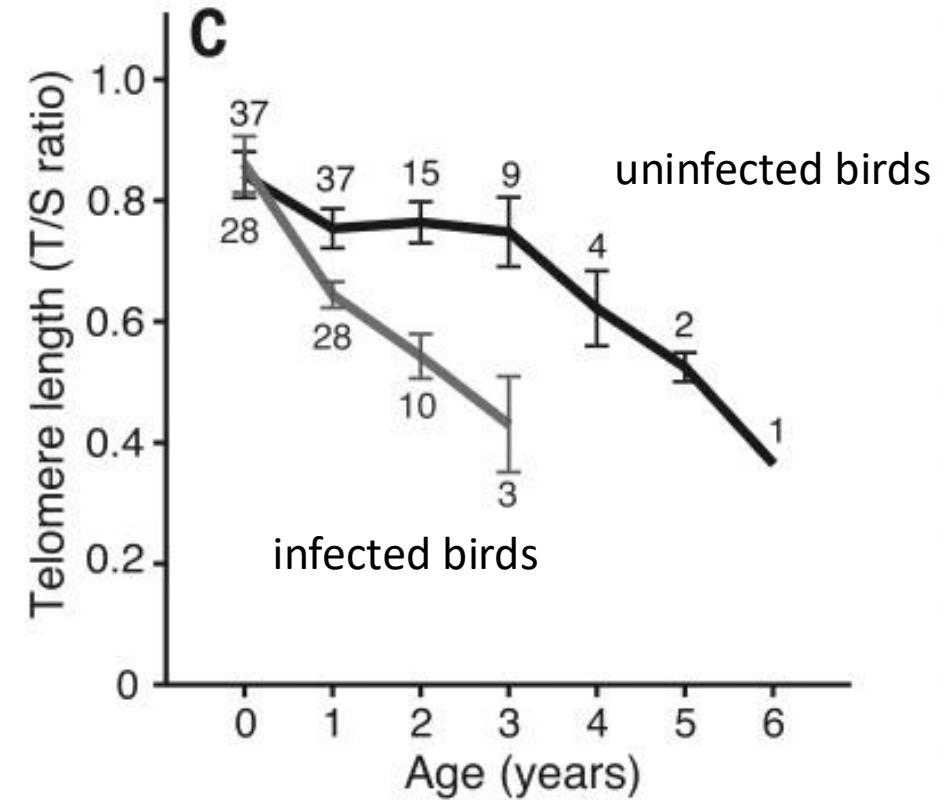
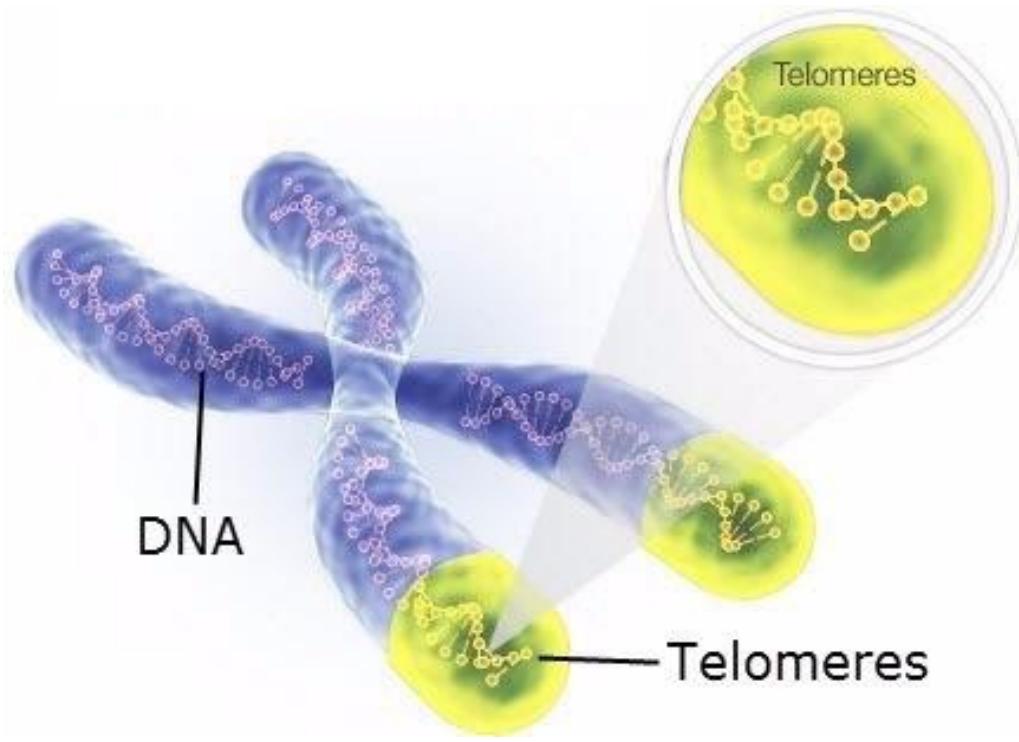
IQ-TREE v. 1.6.12, 3500 18-kb satellite sequences

Telomeres – barometers of age and stress in birds

RESEARCH | REPORTS

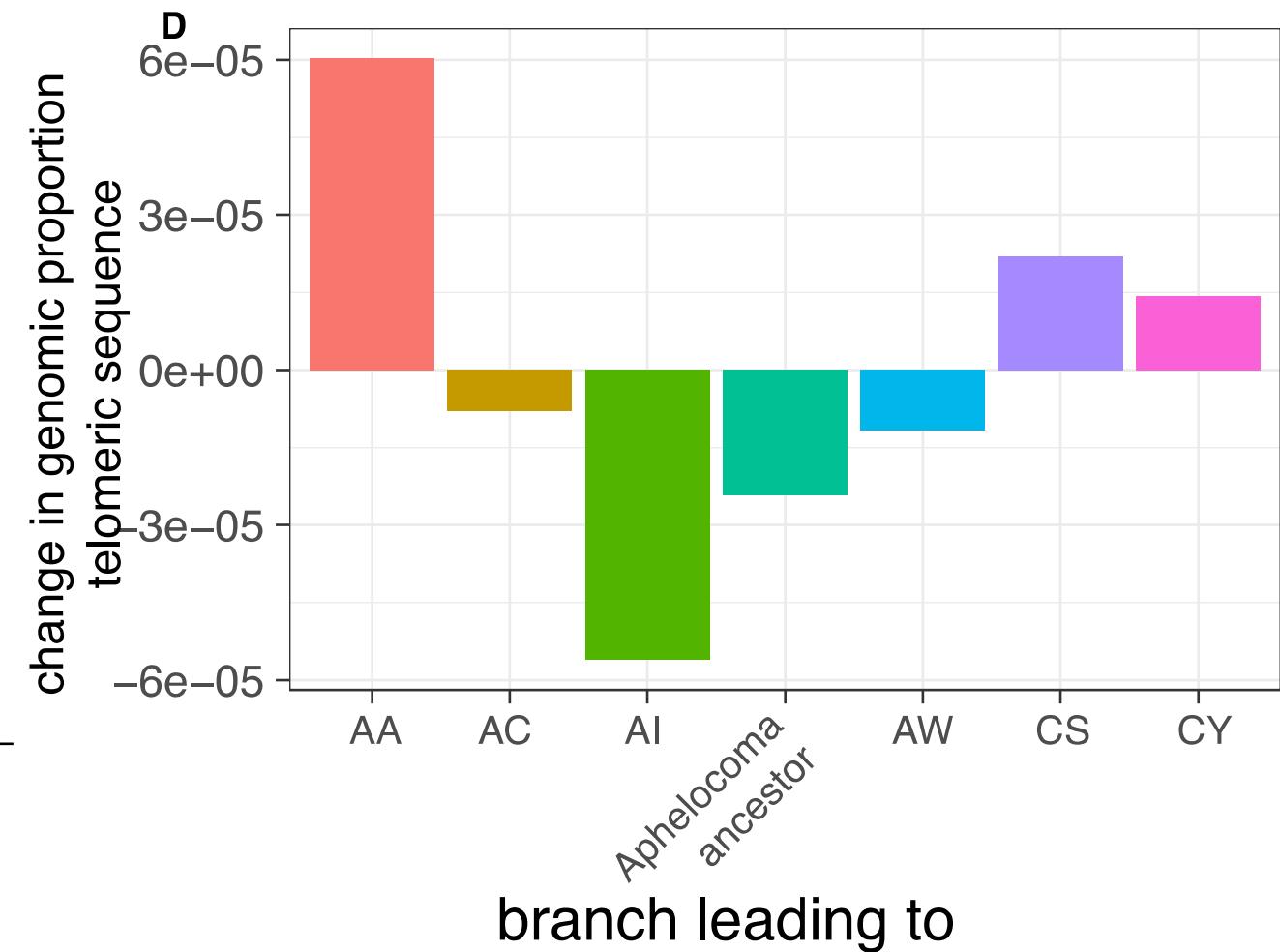
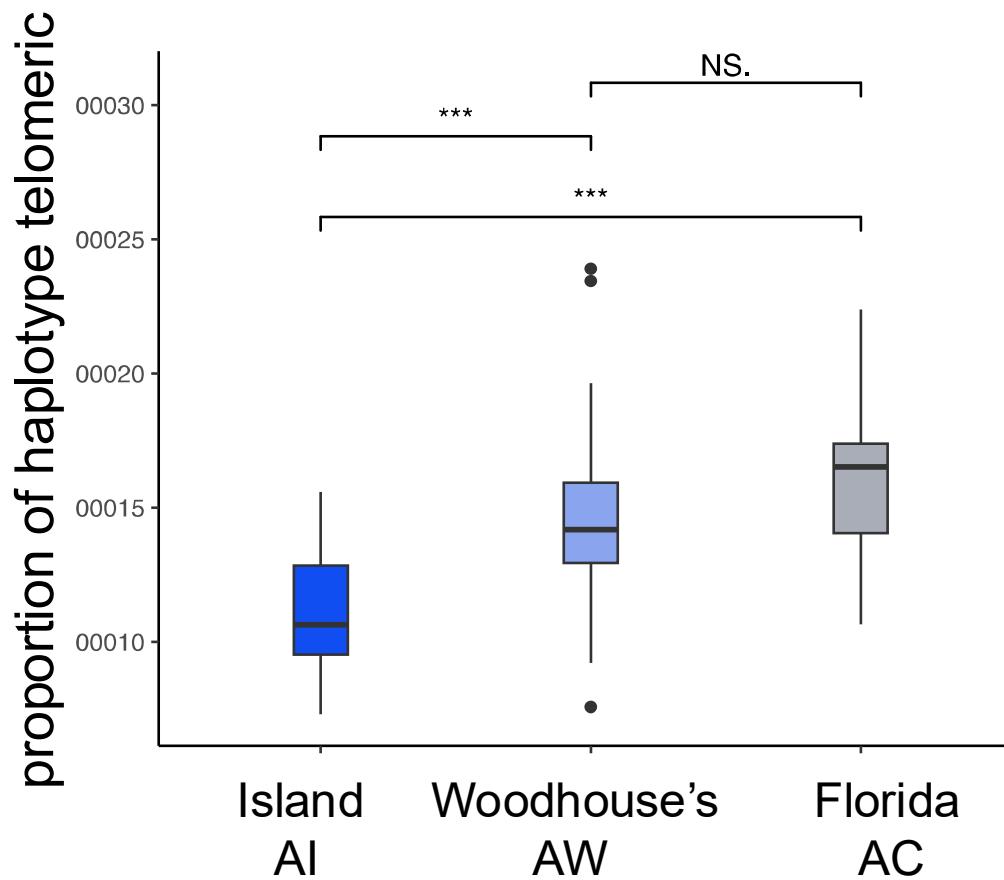
CHRONIC INFECTION

Hidden costs of infection: Chronic malaria accelerates telomere degradation and senescence in wild birds



Telomere abundances influenced by population size

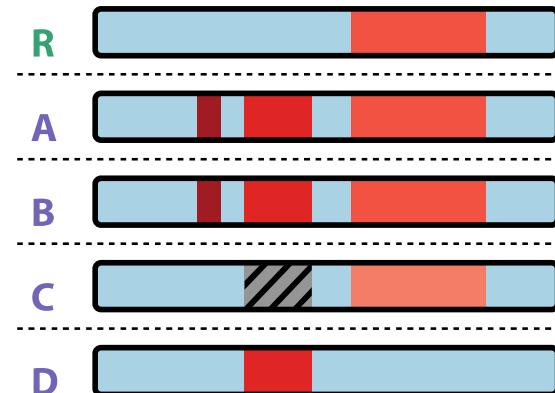
Species with smaller N_e are predicted to have shorter telomeres:
Brown et al. 2024. *Genome Biol. Evol.* evae111



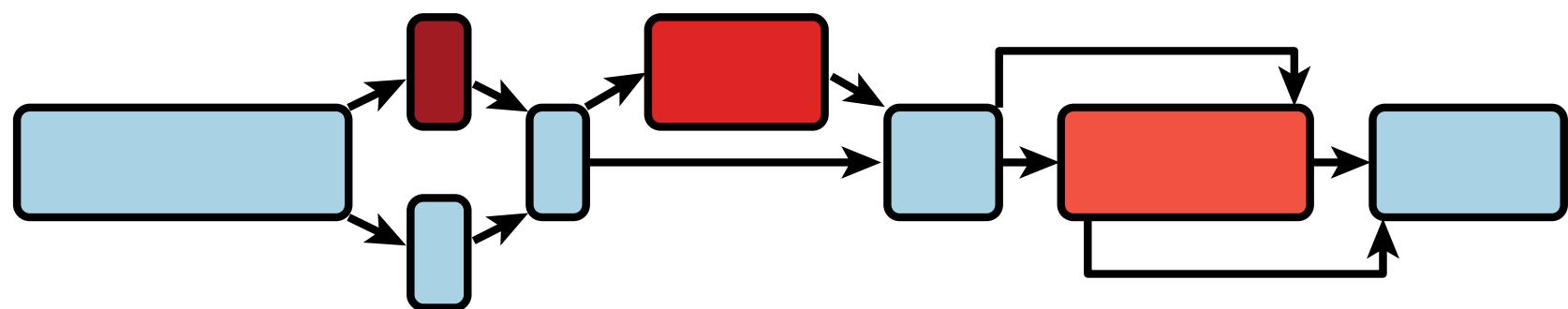
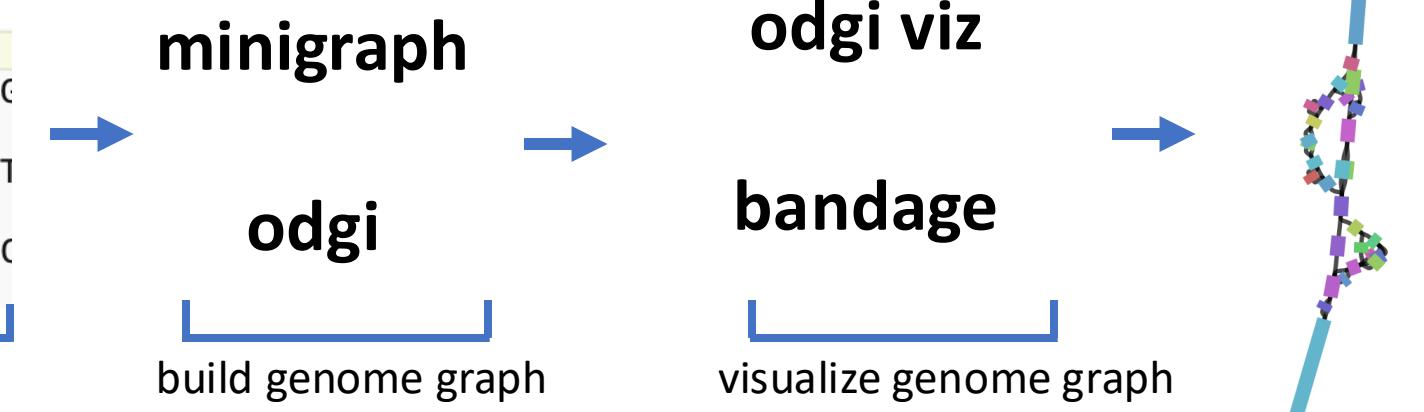
Pangenome graphs capture structural variation within species

```
>h1tg000104l  
GGCGGGGCCGGAGGGGCCGGGCGCTGAGGGCCGGTGCAGAGCC  
>h1tg000528l  
ATGGATACTTCAGTCAGAGCTTATAATAATTCCATAATTAAATATT  
>h1tg0007951  
ACTTTGGGGACACCTTGGGGACACCTCGGGGACACTTGGGCCACAAATCC
```

unaligned fasta files

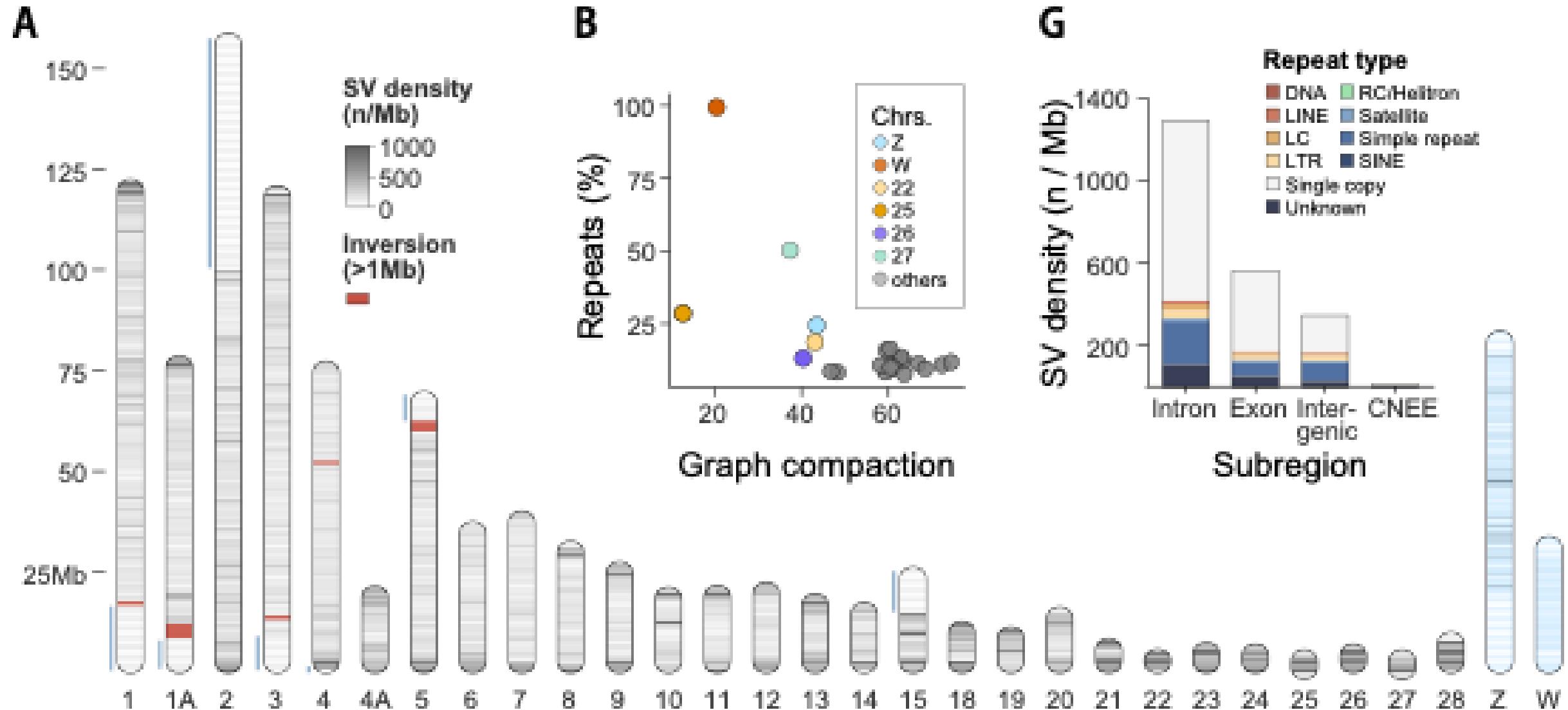


Multiple sequence alignment



Bidirected genome graph

Distribution of genome-wide structural variants

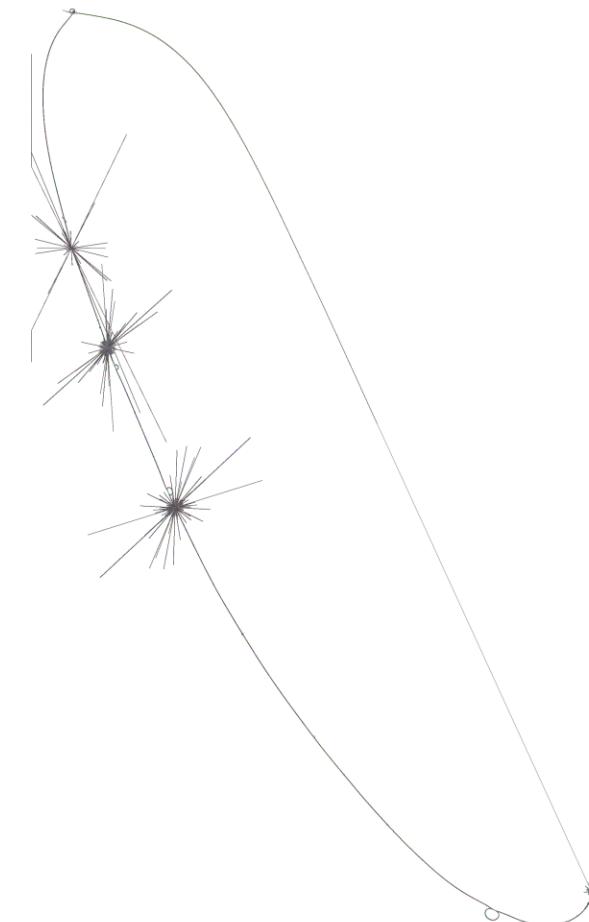


2D pangenome graph visualizations – PGGB/Odgi

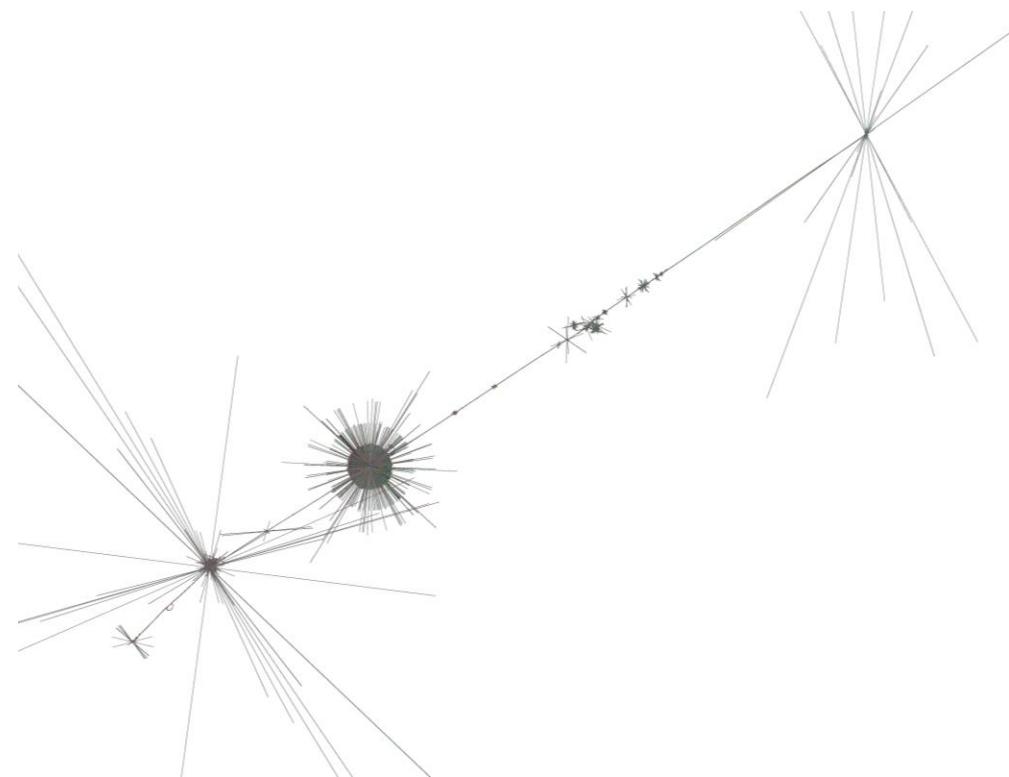
Chr 18 – 12 Mb



Chr 1 – 160 Mb
‘telomere kiss’

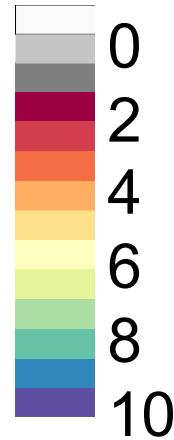


Chr 23 – 8 Mb

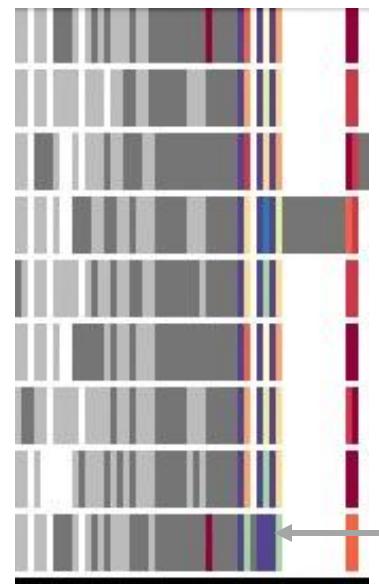


Variation in depth of a pangenome graph

depth of
MHC
graph (x)



low node-depth
regions – SNPs
or small
indels

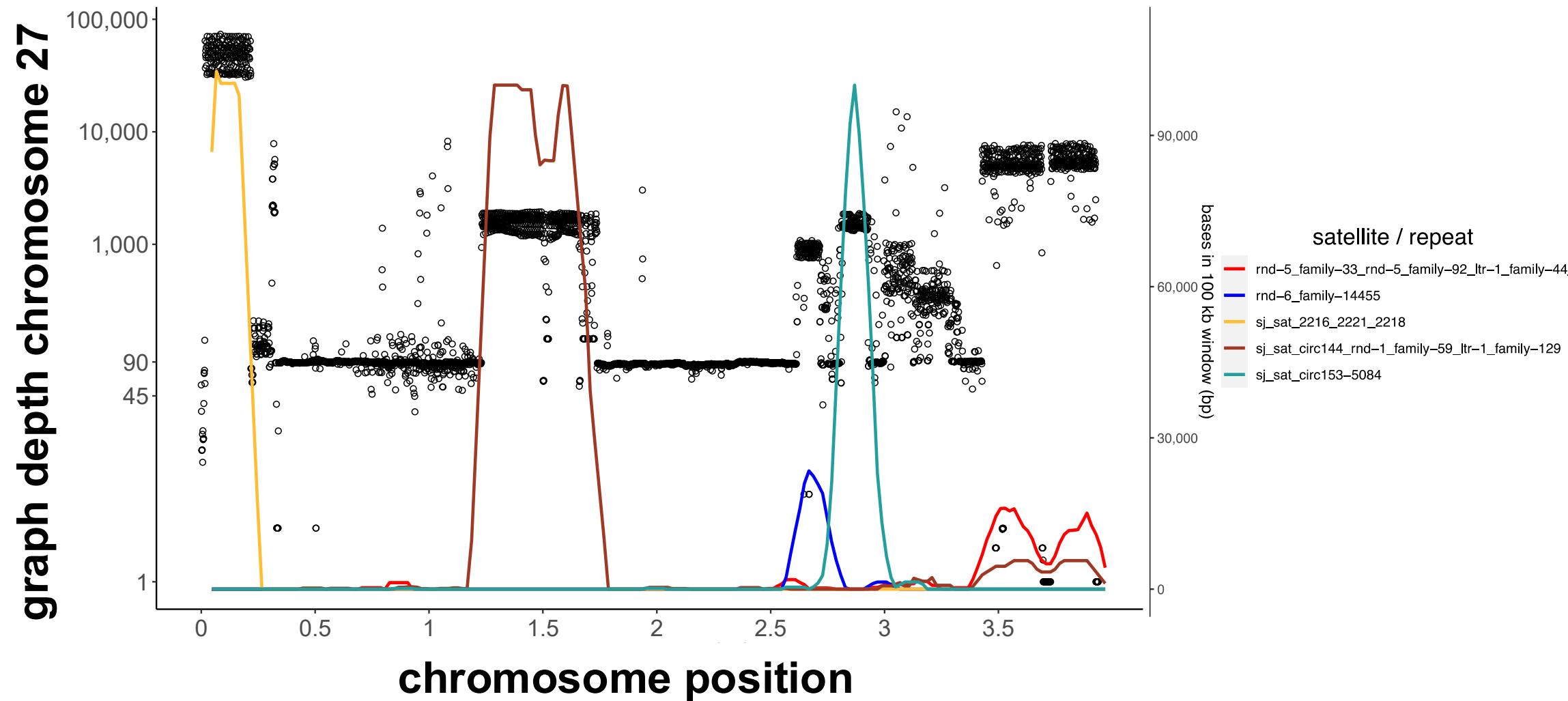


medium
(self-)node-depth regions
depth = ~ n haplotypes

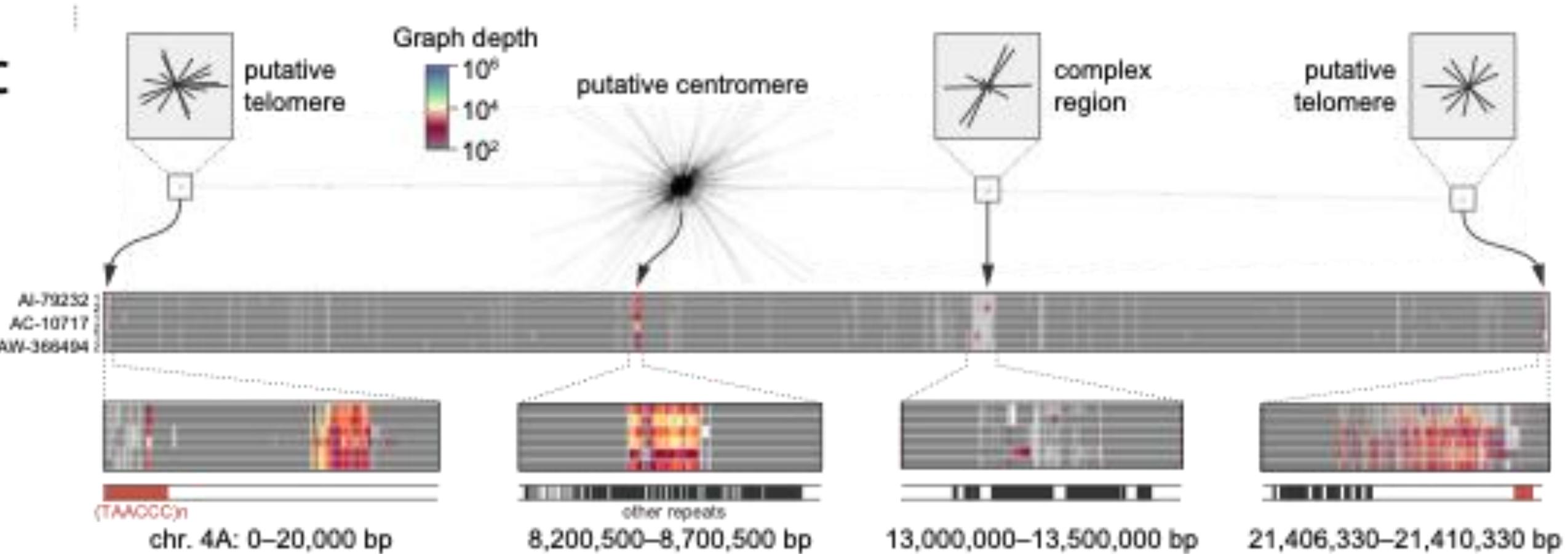
TGGGGATGATTTGGGAGGTTGGGTTGATTAAAGACATTTGGGATATT
TGAGGGGATTTGGGACATTTGAGGGGTTGGGACACTTGGGACAATT
TTTGGGAGGTTGGGCTGATTTGAGAGGTTGGATCCAACTTGGAGCAT
TTTGAGGGGTTGGGCCACTTTGGGCACTTTGGGTGGCCTCAGTCCCC
TTGGGGACACCTGGAGGGCTGGGGACAACCTGGGTGCTGGGAACACTT
TAAGGATATTTGGGACATTTGAGGGTTGGGACACTTTGGGACAATT

high node-depth regions –
large structural variants, satellites

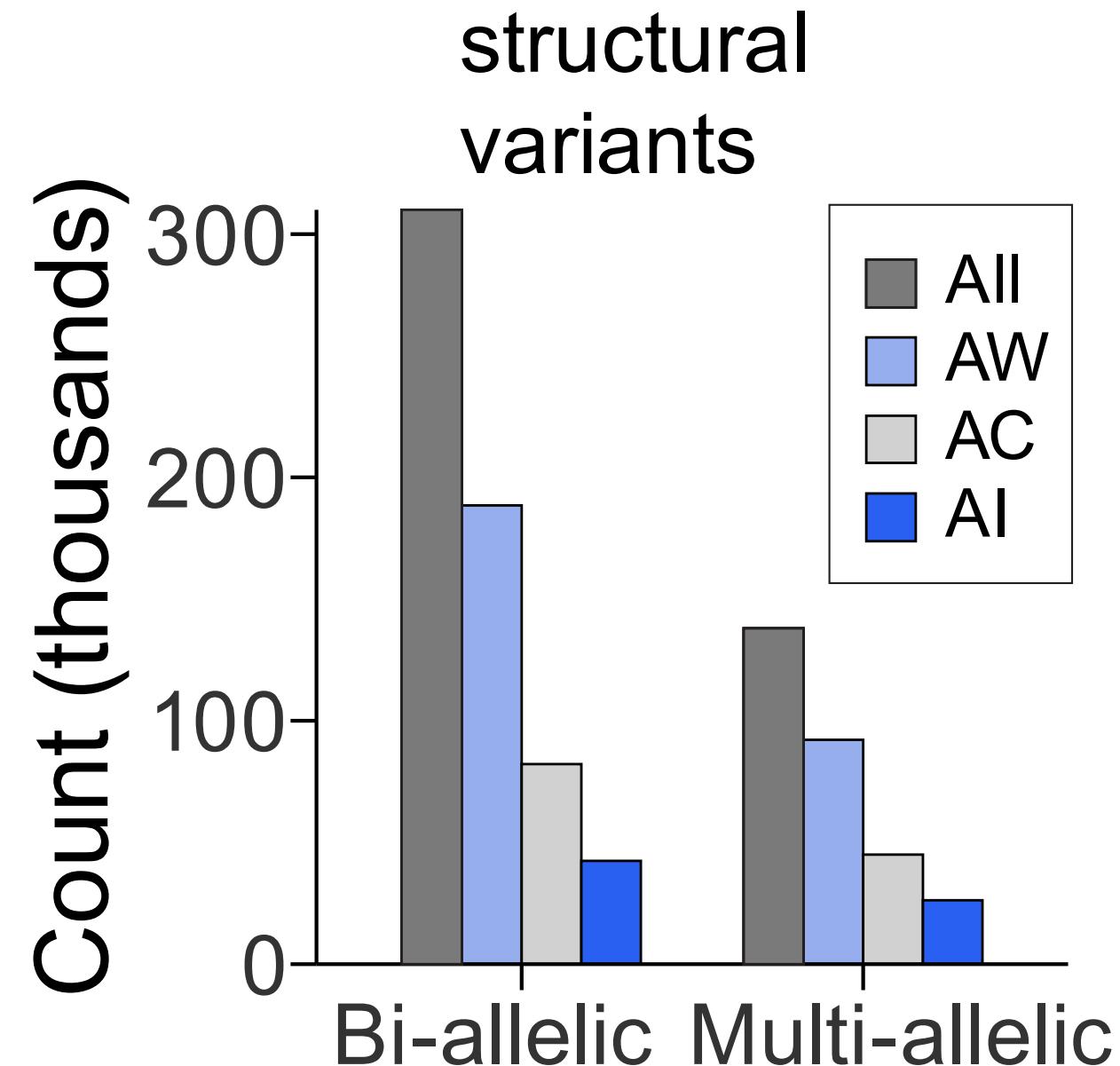
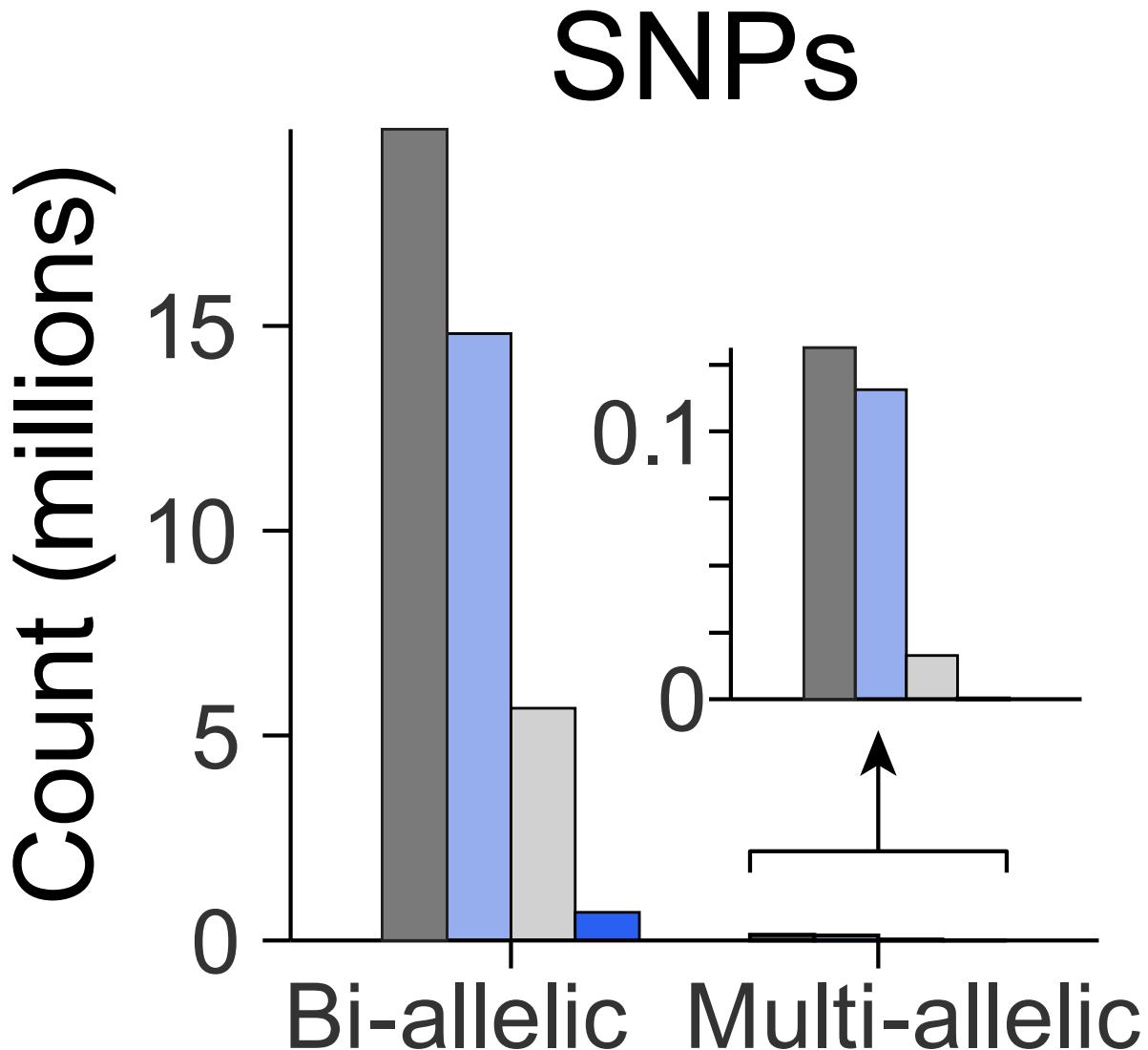
Graph depth of microchromosome 27 correlates with LTRs and satellites



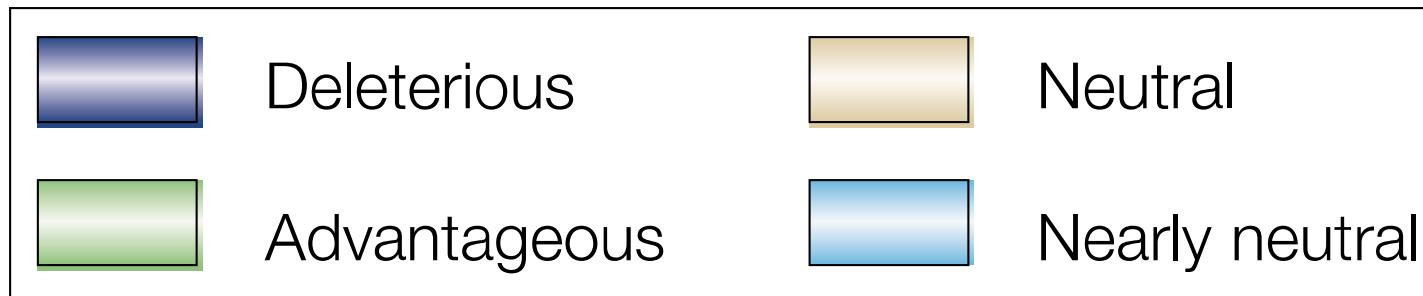
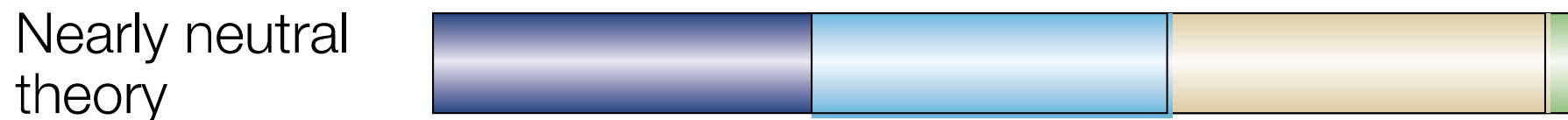
Pangenome depth reveals chromosomal features



Number of structural variants scales with population size

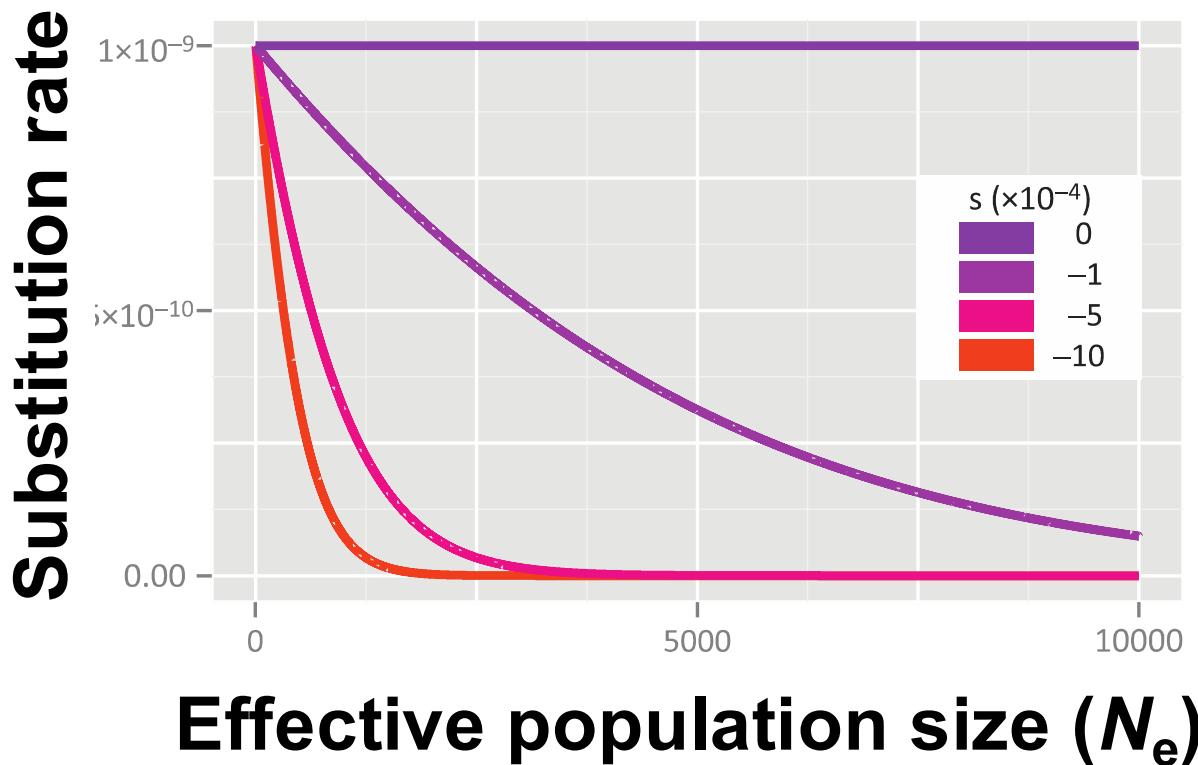


Nearly neutral molecular evolution

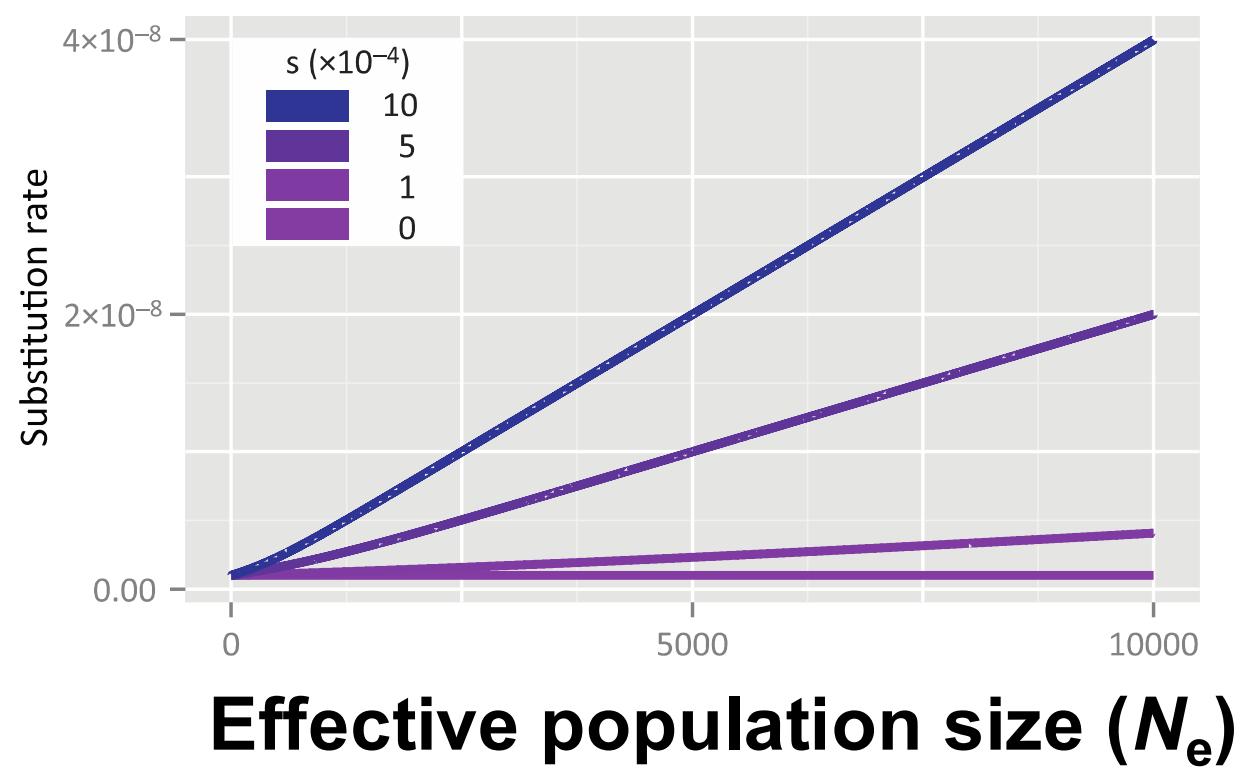


Population size and rate of molecular evolution

deleterious mutations



advantageous mutations

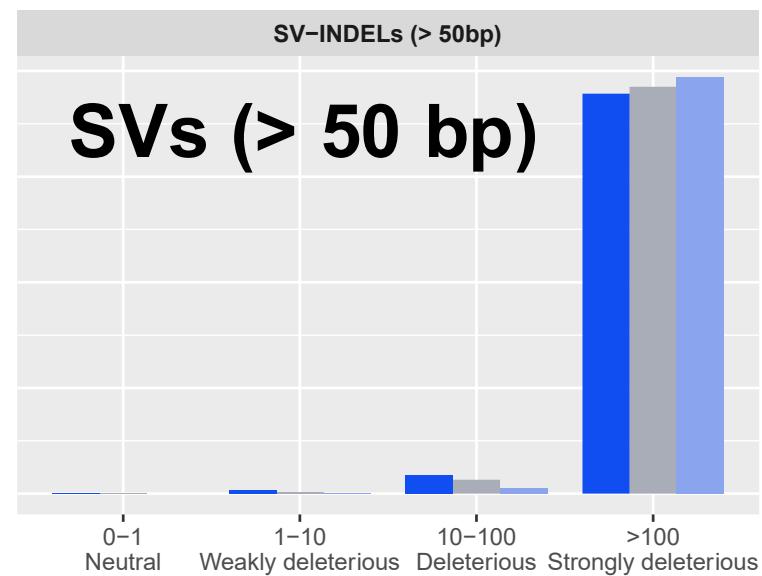
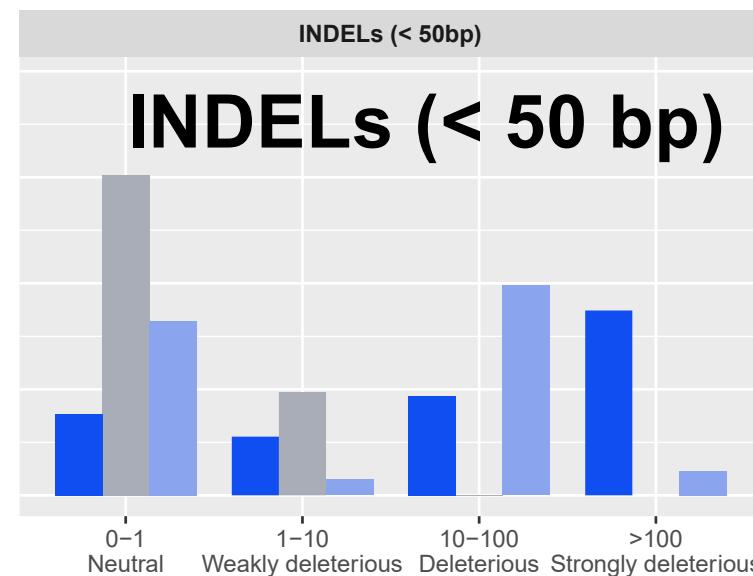
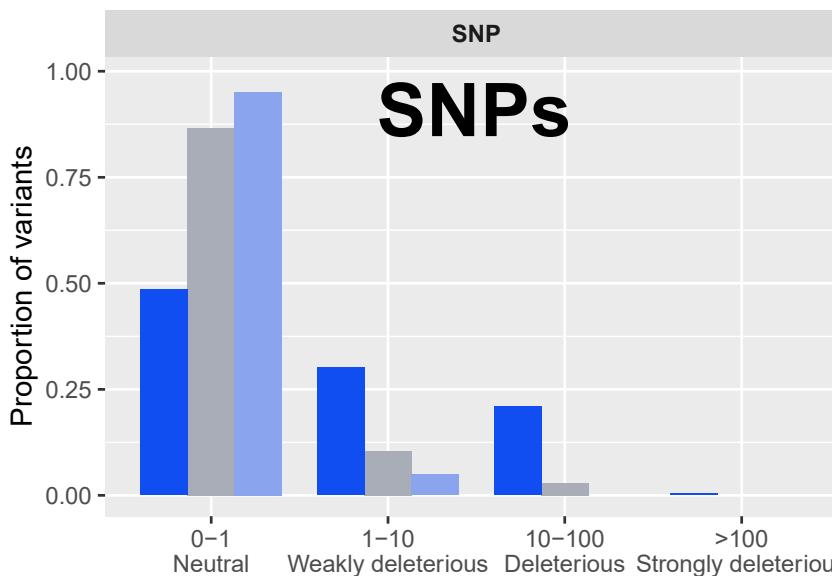


SVs are on average more deleterious than SNPs



Tomoko Ohta

- Florida
- Woodhouse's
- Island

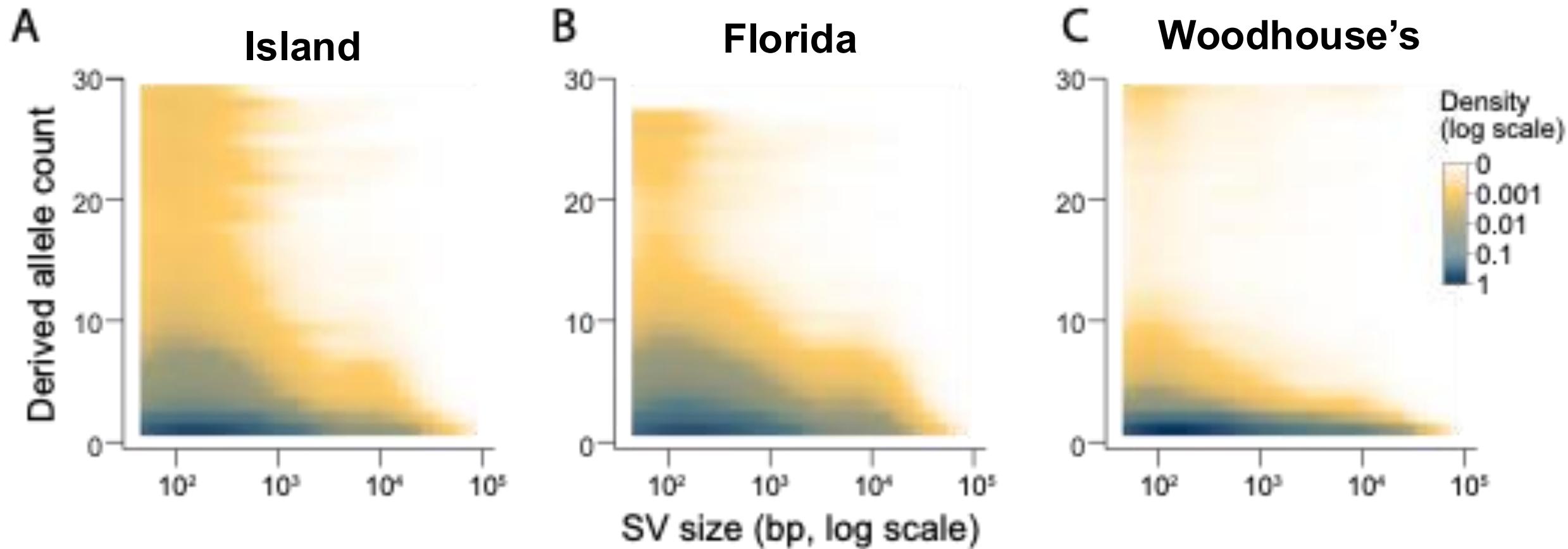


$-N_e S$

$-N_e S$

$-N_e S$

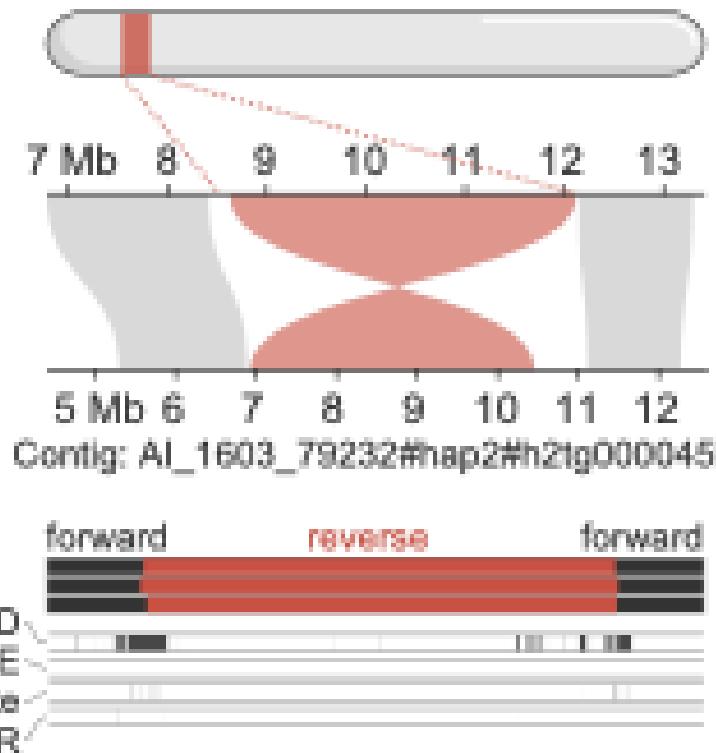
Longer SVs rise to higher frequencies in island population



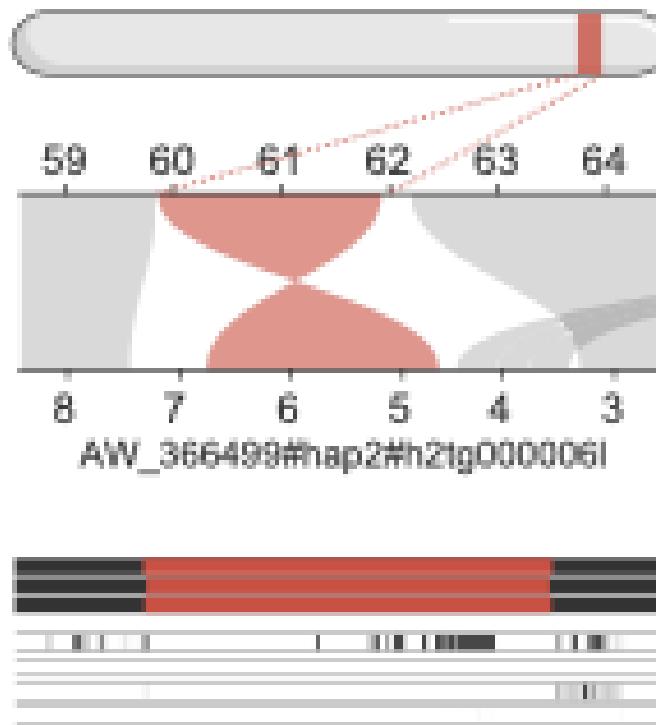
Density of ~485,000 structural variants extracted from PGGB pangenome graph

Inversions are common and track population size

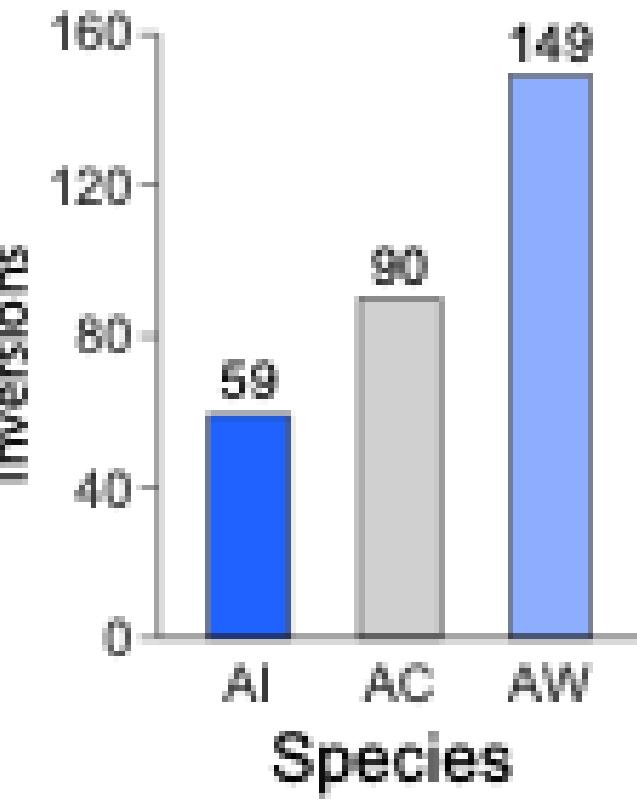
Chr. 1A



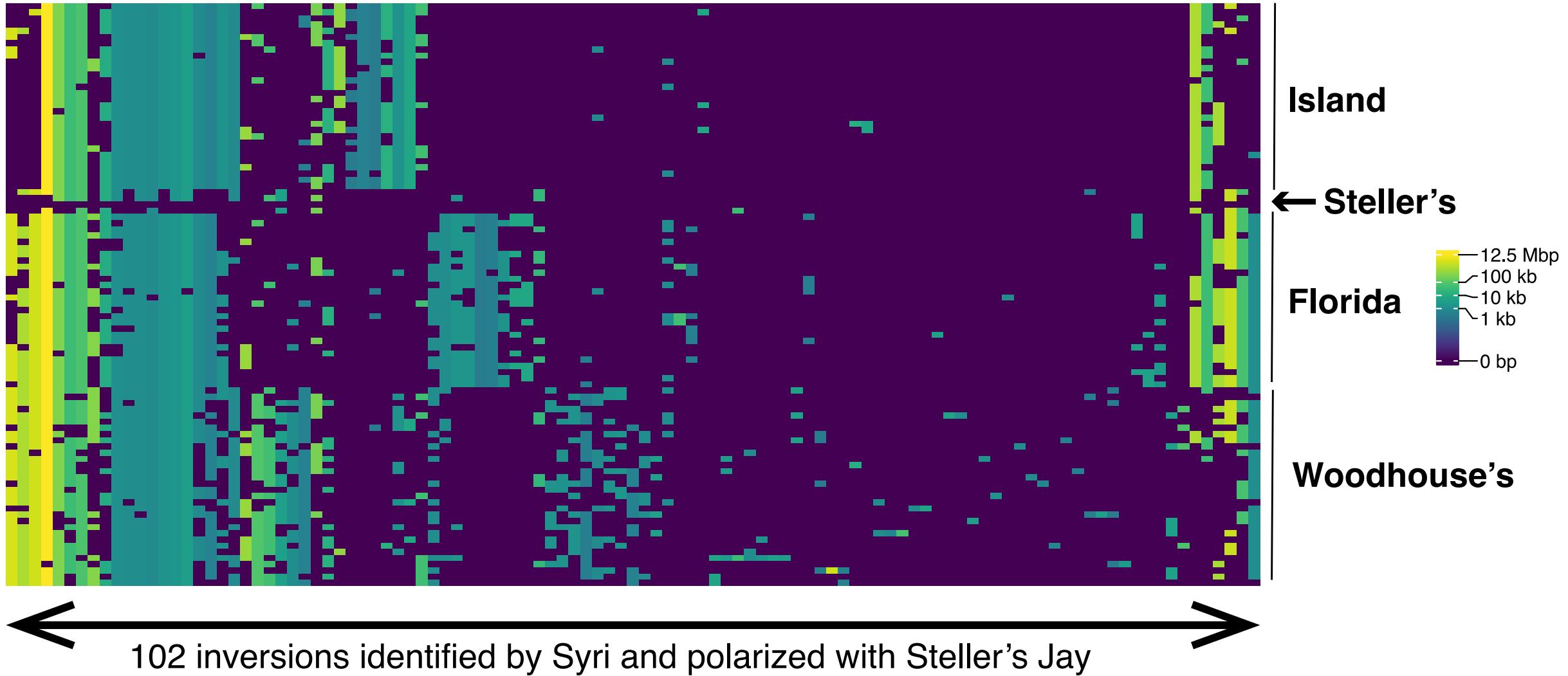
Chr. 5



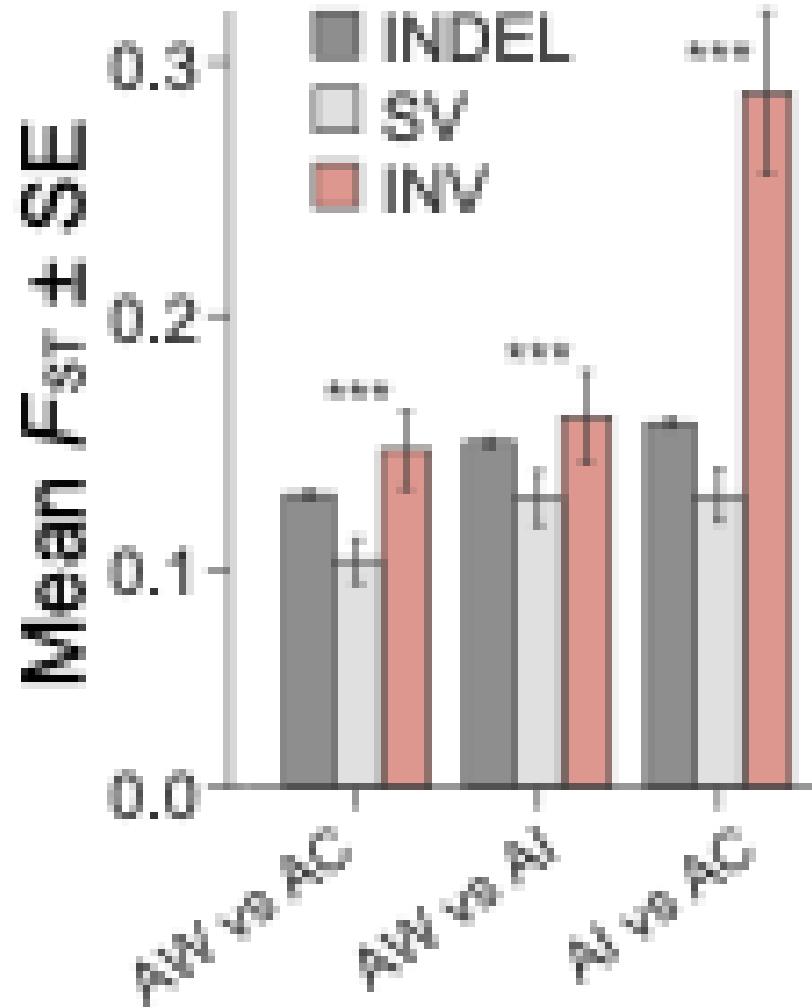
Number of segregating inversions



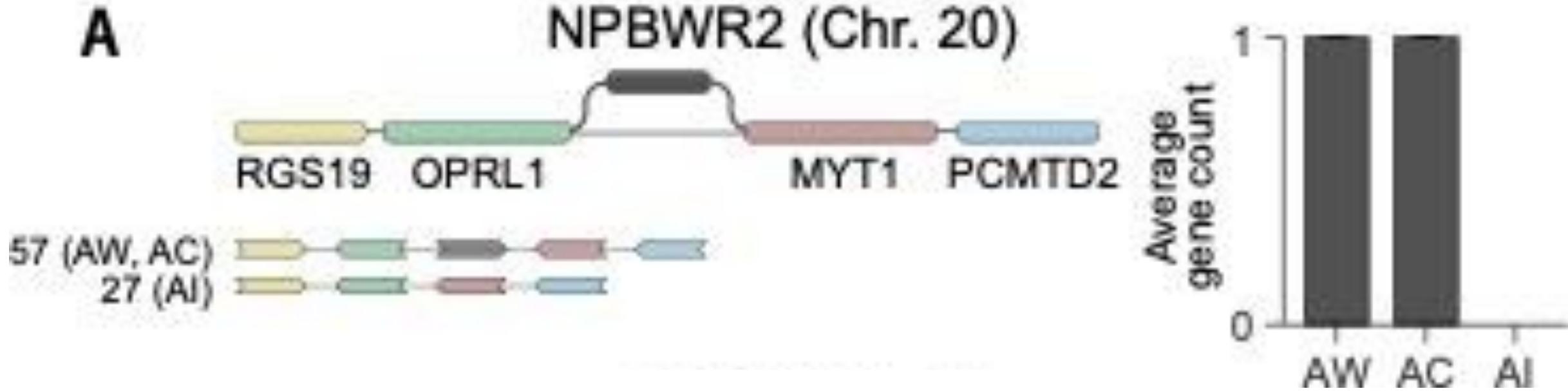
Numerous inversions identified by pangenome and reference-based methods



Inversions show enhanced differentiation between species

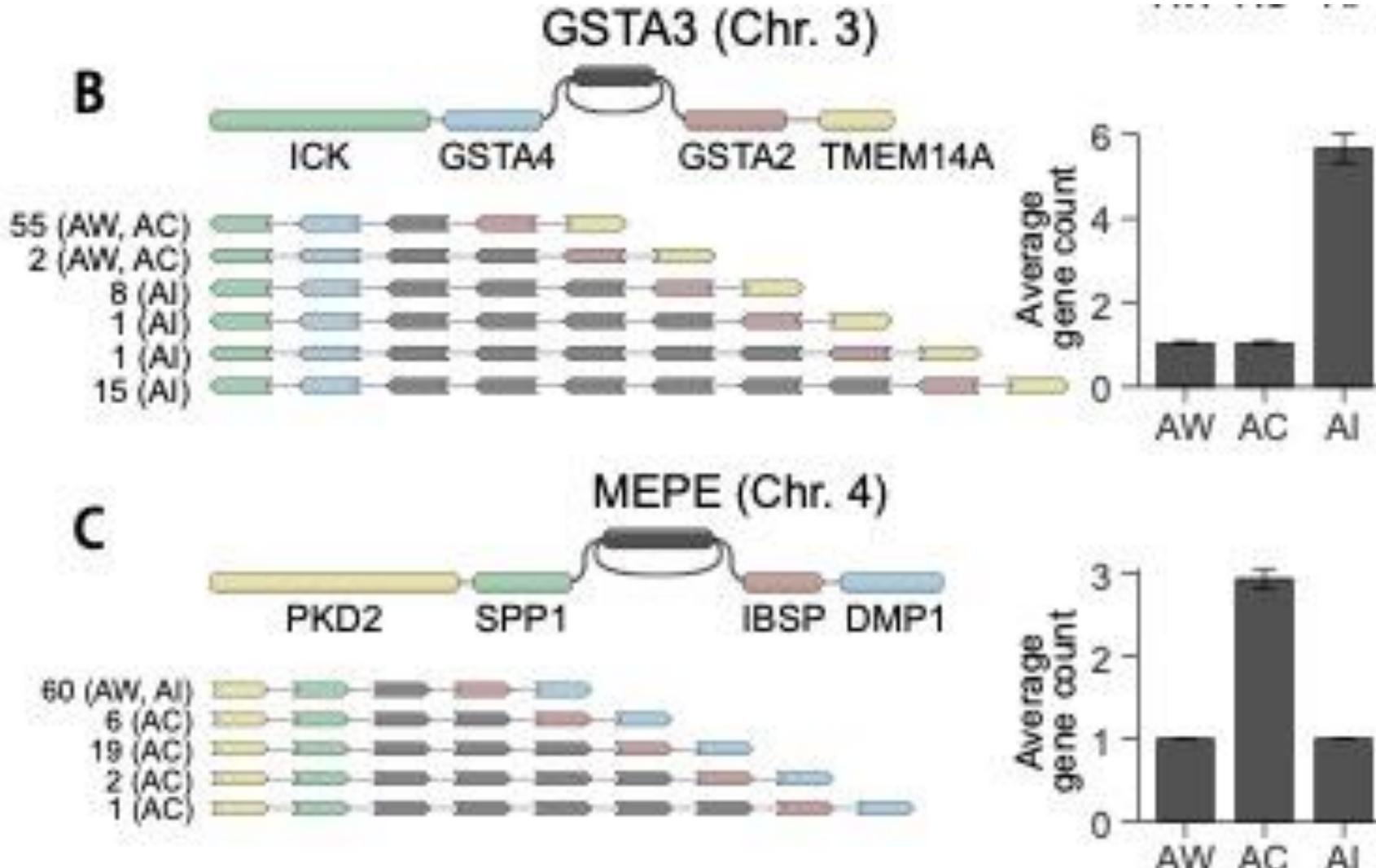


Abundant copy-number variants and gene deletions

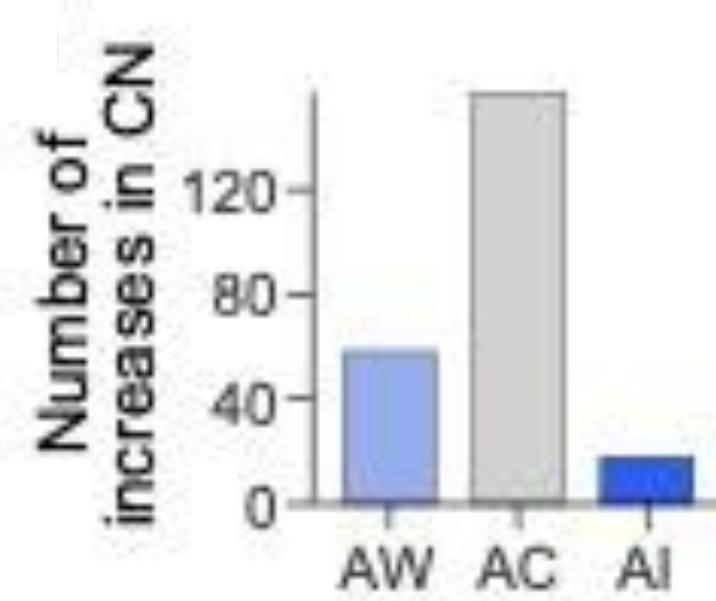
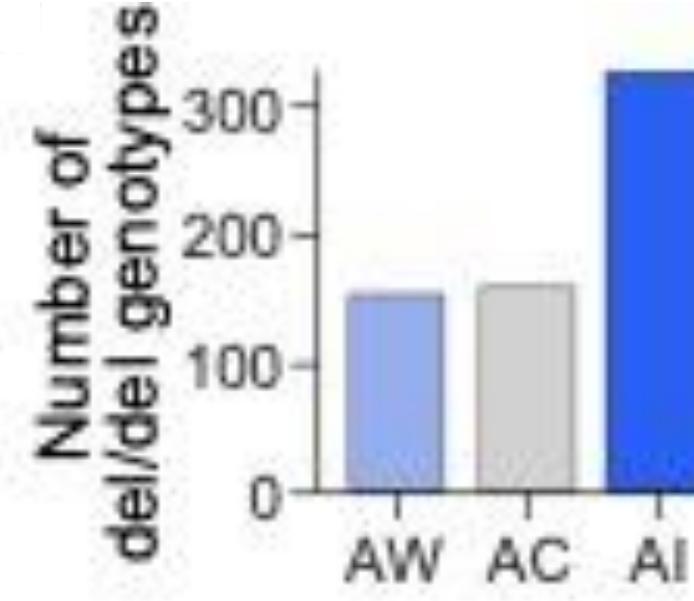
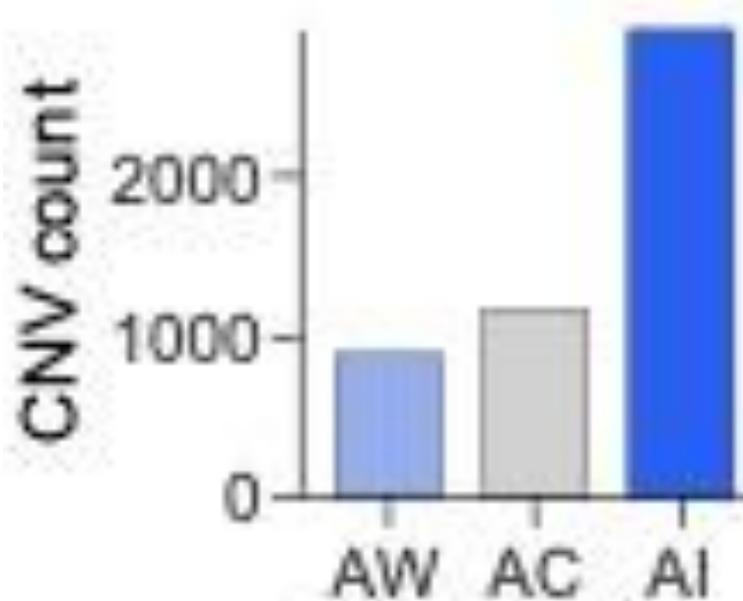


Using miniprot and pangene
Li et al. 2024. *Bioinformatics*
Li, H. 2023. *Bioinformatics*

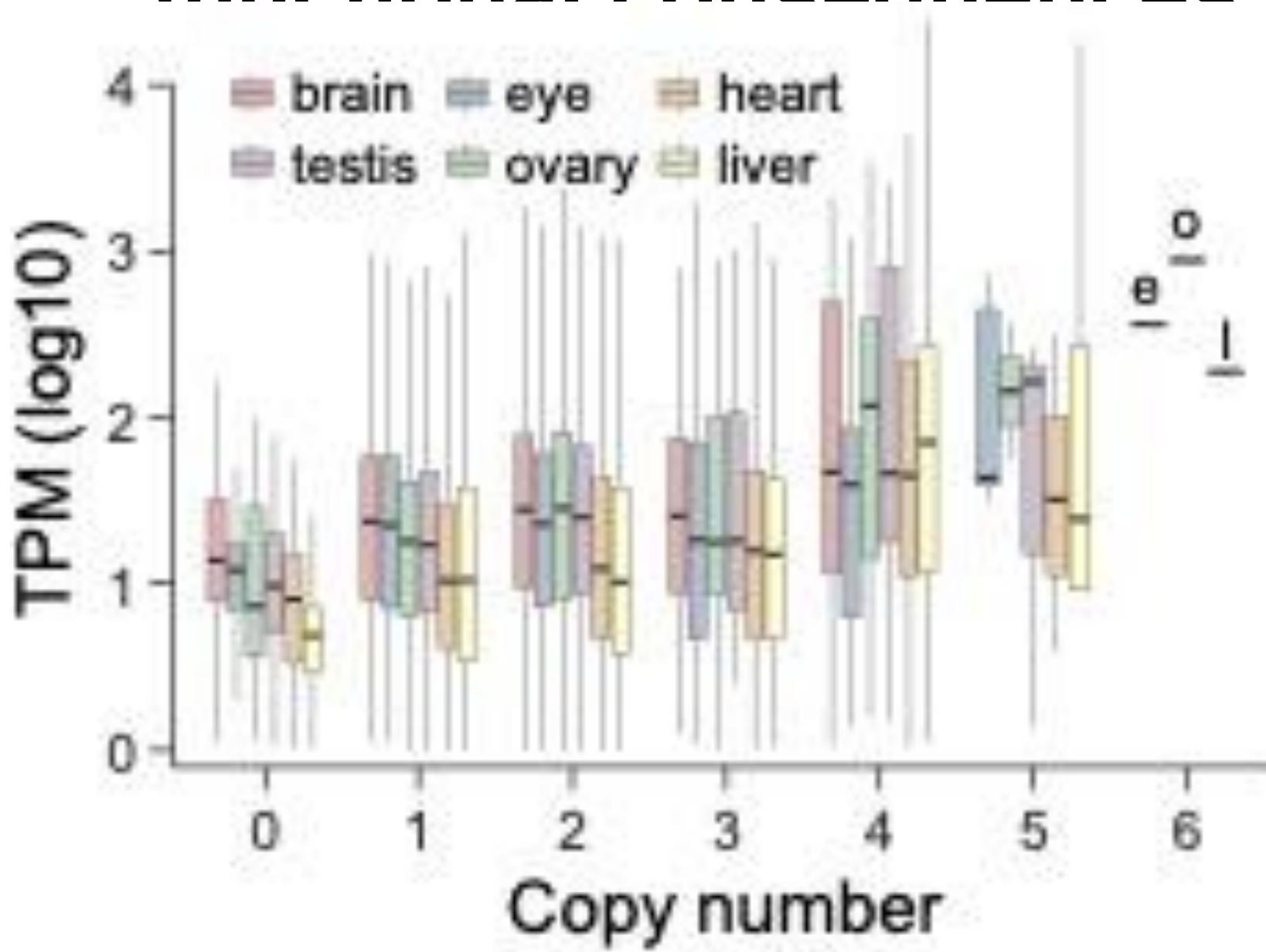
Abundant copy-number variants and gene deletions



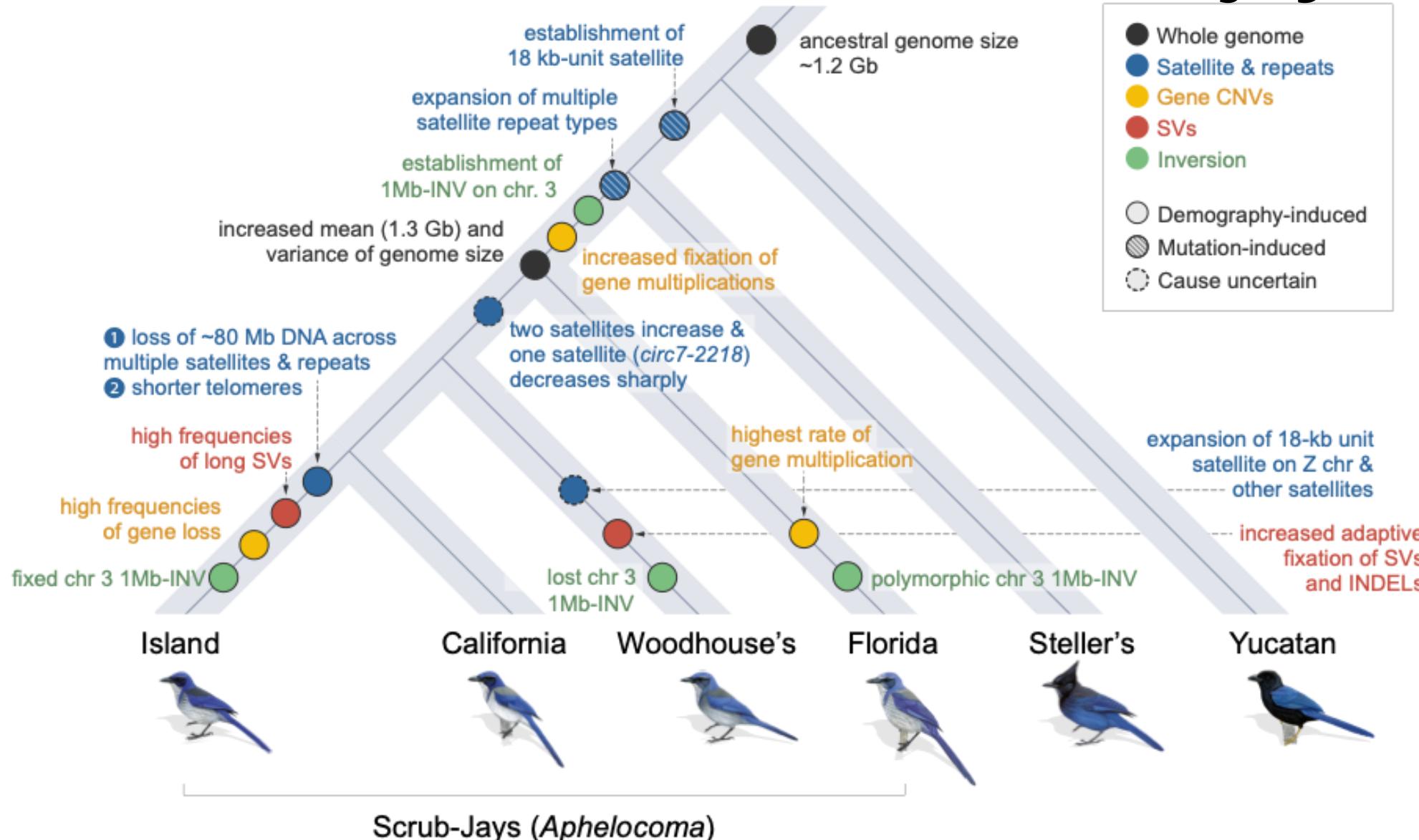
Gene copy number variants exhibit a surprising pattern



Copy number variants have functional consequences



Population size influences diversity of structural variants in scrub-jays



Conclusions



- Scrub-jay genomes are repeat-rich
- Structural variants generally track population size
- Structural variants are estimated to be slightly deleterious
- Gene copy number variants look strongly deleterious
- Pangenome analysis will likely become the common standard

Acknowledgements

Colorado team - Island Scrub Jay

Chris Funk
Rebecca Cheek
Paul Hohenlohe
Cameron Ghalambor



Fieldwork
Greg and Donna

Florida team - Florida Scrub Jay

Nancy Chen
Reed Bowman
John Fitzpatrick

Pangenome informatics
Erik Garrison
Andrea Guaracino

Harvard team - Woodhouse's Scrub Jay and Informatics

Tim Sackton
Danielle Khost
Heng Li
Bohao Fang
George Kolyfetis