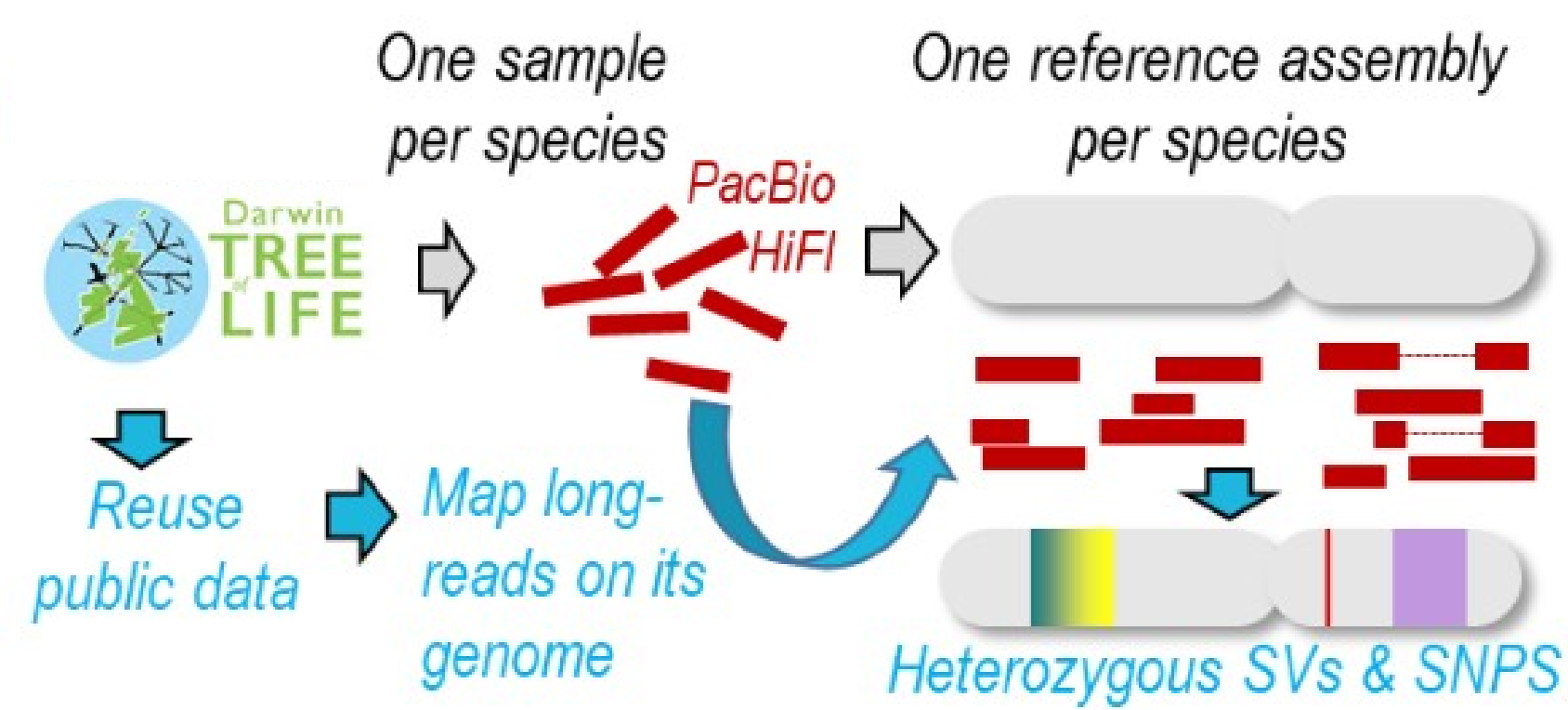


Structural genetic diversity across the Tree of Life: development of a long-read based pipeline for robust detection of SVs

Brazier, T.¹, Lemaitre, C.² & Mérot, C.¹ (contact: thomas.brazier@univ-rennes.fr)

¹ University of Rennes, CNRS, ECOBIO (Ecosystems, Biodiversity, Evolution) - UMR 6553, Rennes, France

² University of Rennes, Inria, CNRS, IRISA, Rennes, France



Shifting from Single Nucleotide Polymorphism (SNP)-based genetic diversity to **Structural Variation (SV) diversity**.

Darwin Tree of Life project = **PacBio HiFi long reads** and high-quality genome assemblies.
Access to SV diversity across the tree of life.

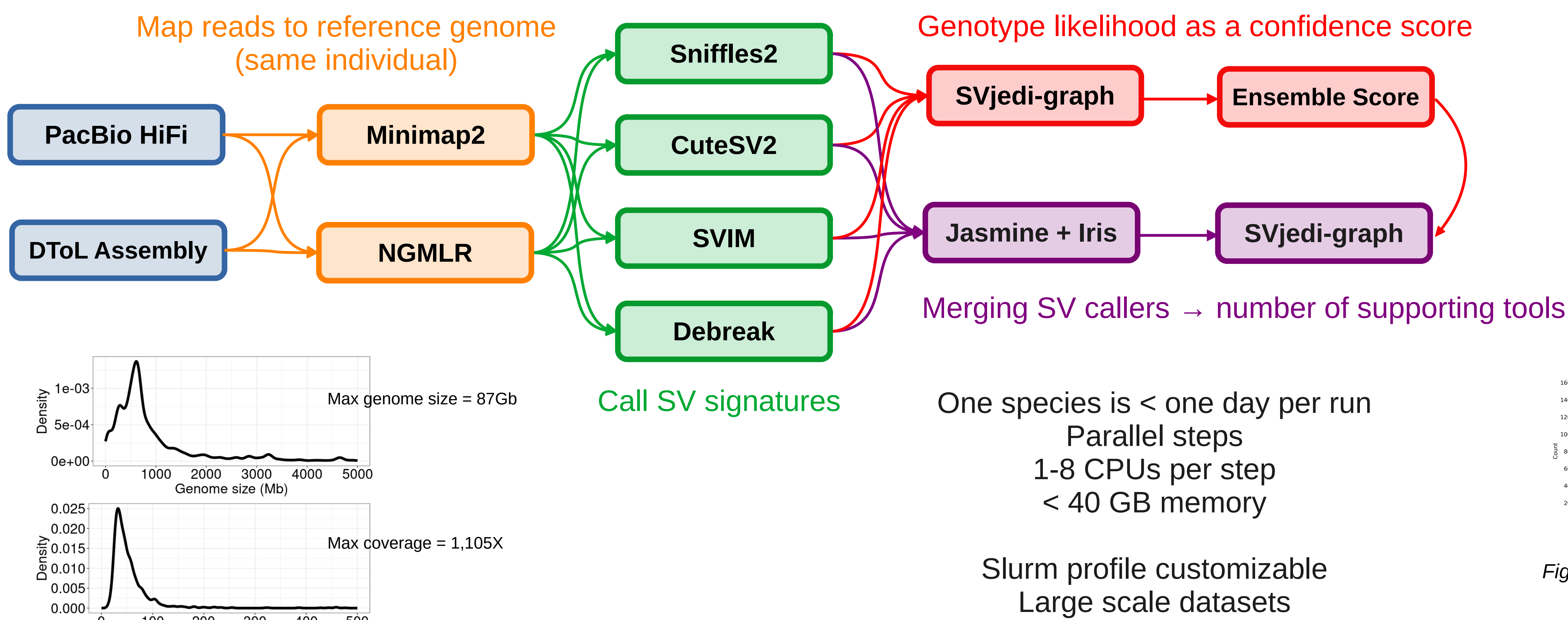
Need to develop a Snakemake pipeline for detecting SV.

Fig. 1. Analytical plan to address the variation of genetic diversity among taxa for different variants.

Understanding the variation of (structural) genetic diversity across species and examining the overall role of SVs in evolutionary processes.

snakemake/ snakemake

SV calling and genotyping from PacBio HiFi long reads



QC

Fastqc
Sniffles2plot
Samplot
Rmarkdown

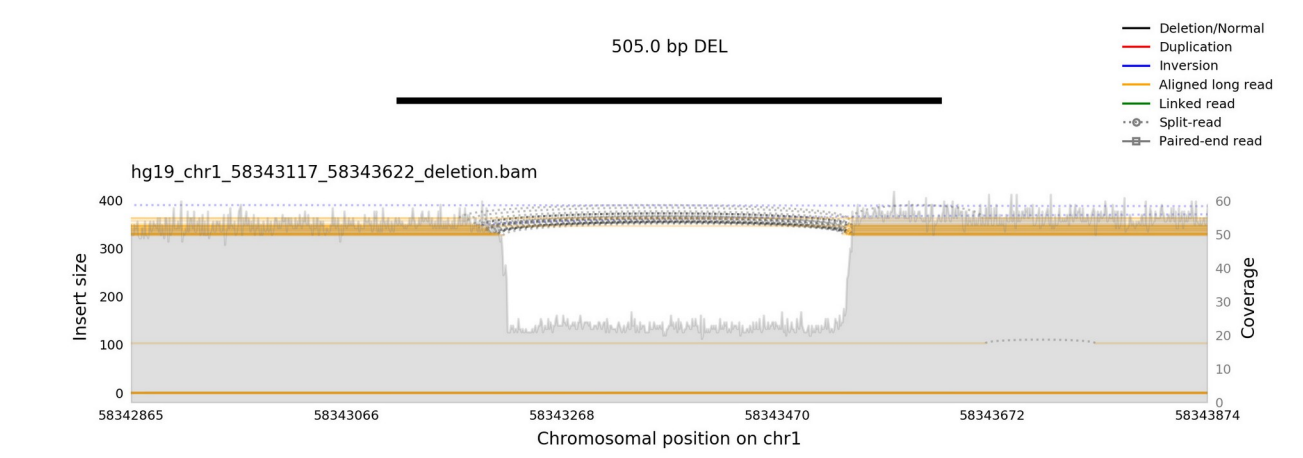


Fig. 3. QC figures generated automatically by the pipeline.

Fig. 2. Characteristics of the DToL dataset.

SV validation: Meta-learning to estimate calling uncertainty

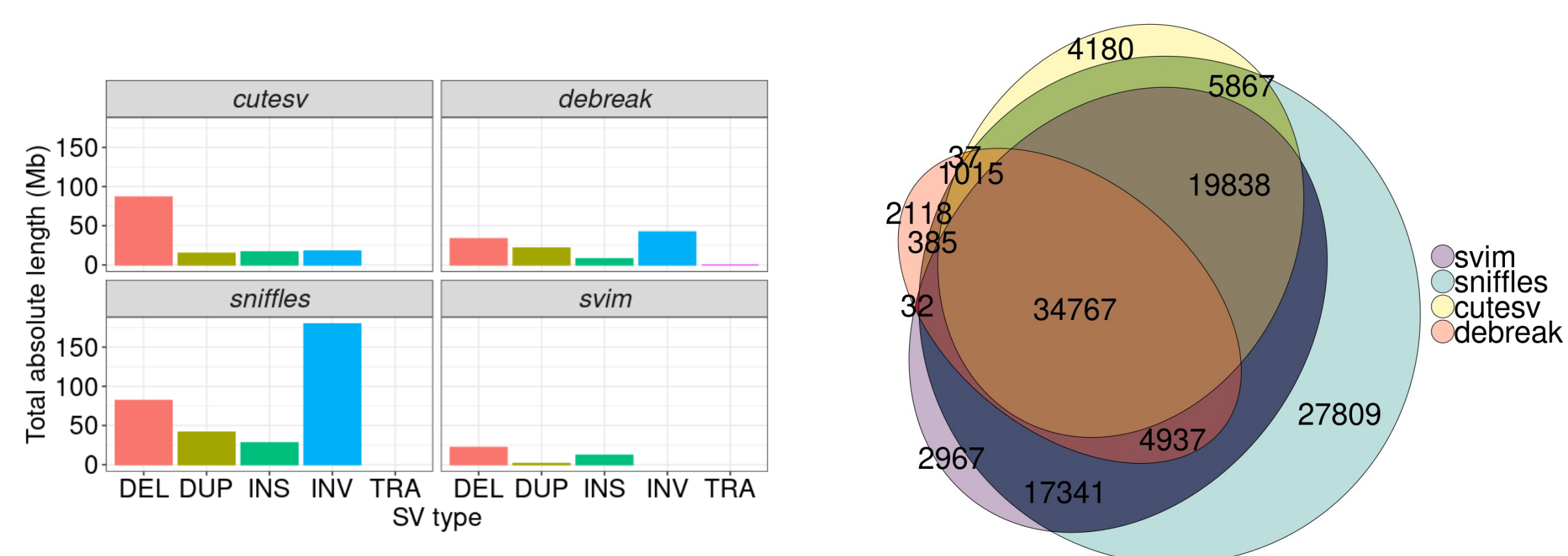


Fig. 4. Variability of SV detection among SV callers. (A) The proportion of genome covered by SV varies among SV callers. (B) Venn diagram of shared and unique SV calls among SV callers.

Ensemble methods

2 aligners x 4 SV callers

Average confidence score + confidence interval

Number of supporting calls

Single individual

Only heterozygote SV expected

Proportion of heterozygous SV = 0.93

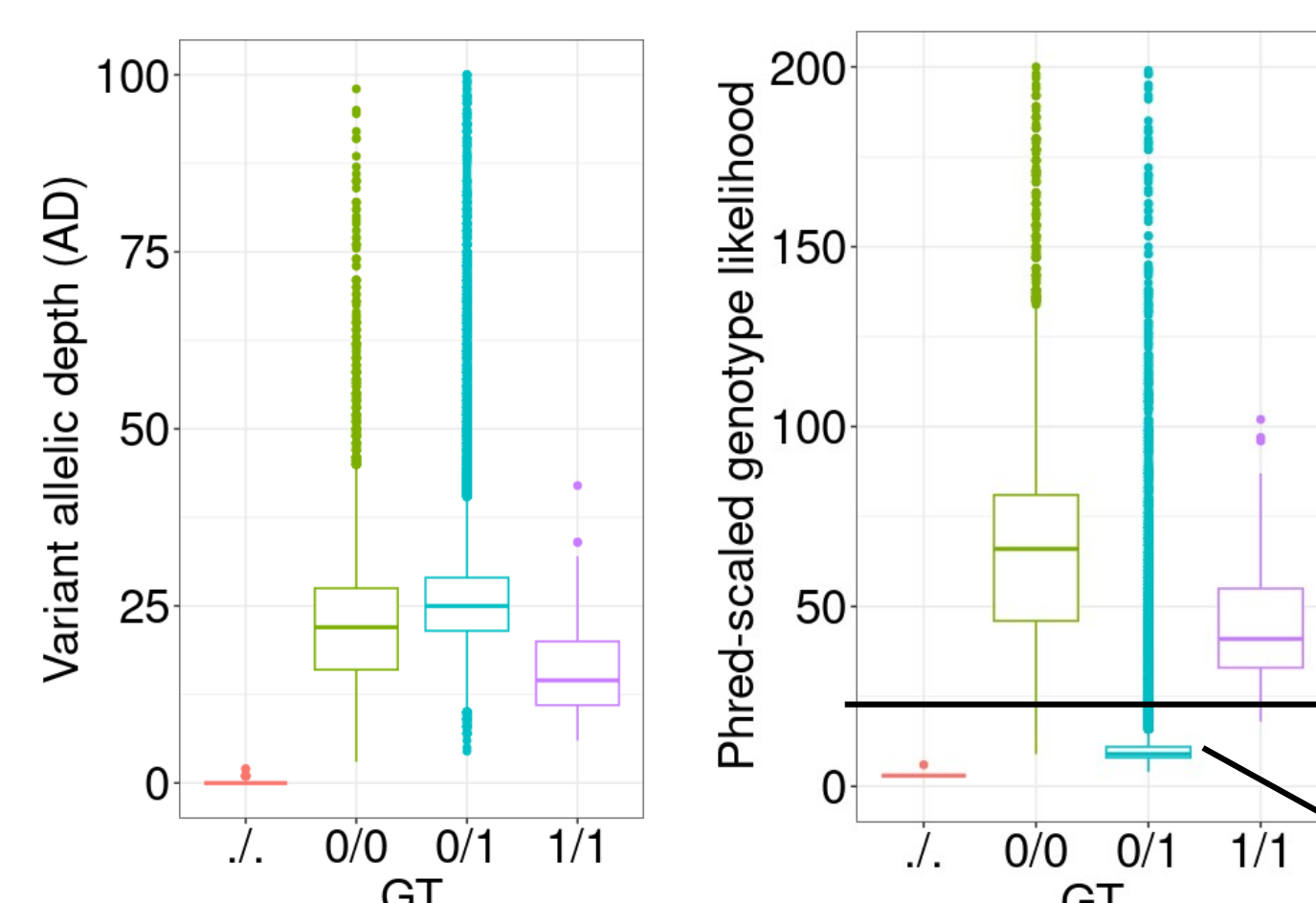


Fig. 5. SV validation by genotyping. Variant Allelic Depth and HET Genotype likelihood.

Precision/recall for each caller

GT HET for this SV caller

GT HOM for this SV caller

TP FP

TN FN

GT ./. for this SV caller but HET for another (one vs all)

Lower is better

How much genetic diversity across the Tree of Life?

Diversity and determinants of Structural Variation across hundreds of species, e.g. Lewontin Paradox

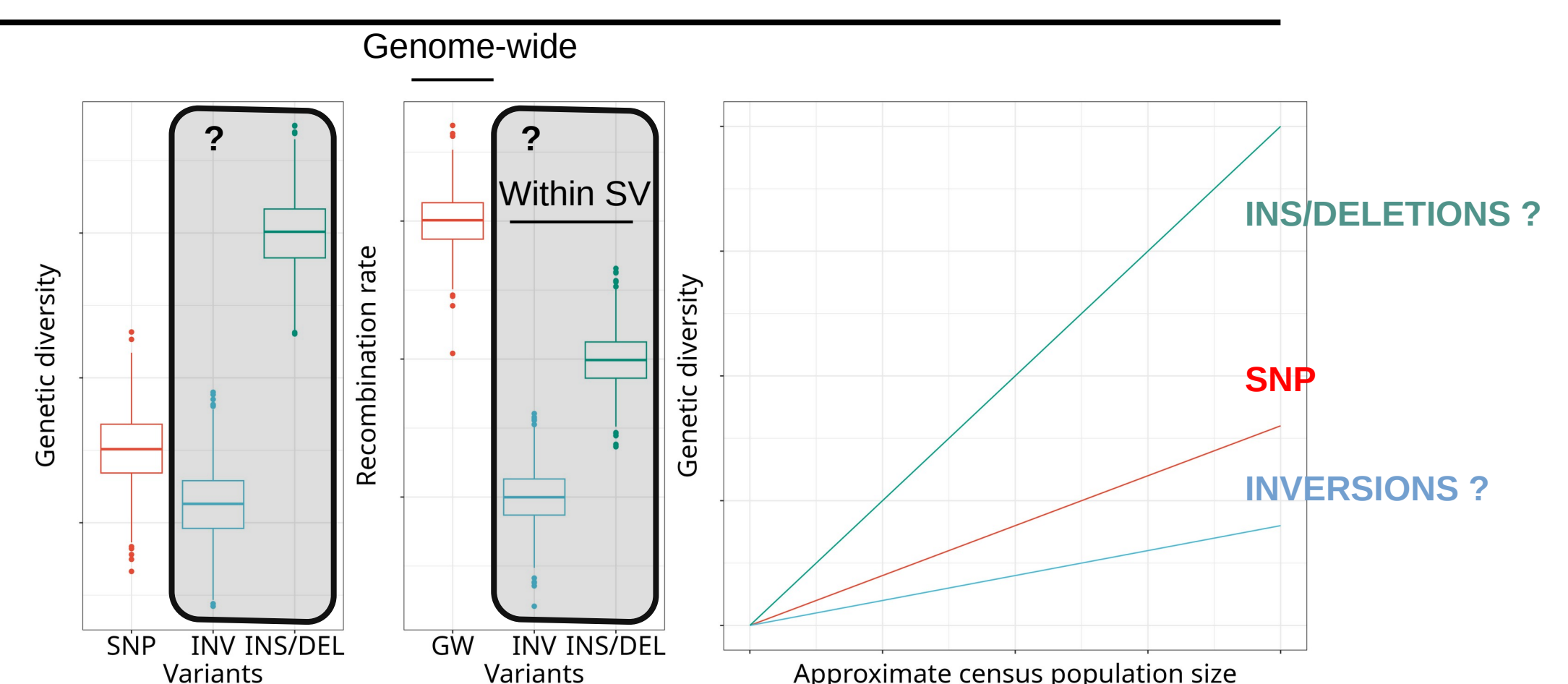


Fig. 6. Unknown diversity and conceptual predictions to address.

Acknowledgements. Joana Meier, Mark Blaxter, Matthieu Muffato, Darwin Tree of Life project members, L. Angevin, K. Lohse, S. Martin