

Results

Figure 1 – Repetitive DNA in teleost fish genomes. Stacked bars show relative fractions of transposable elements. Blue circles show relative amounts of simple repeats.

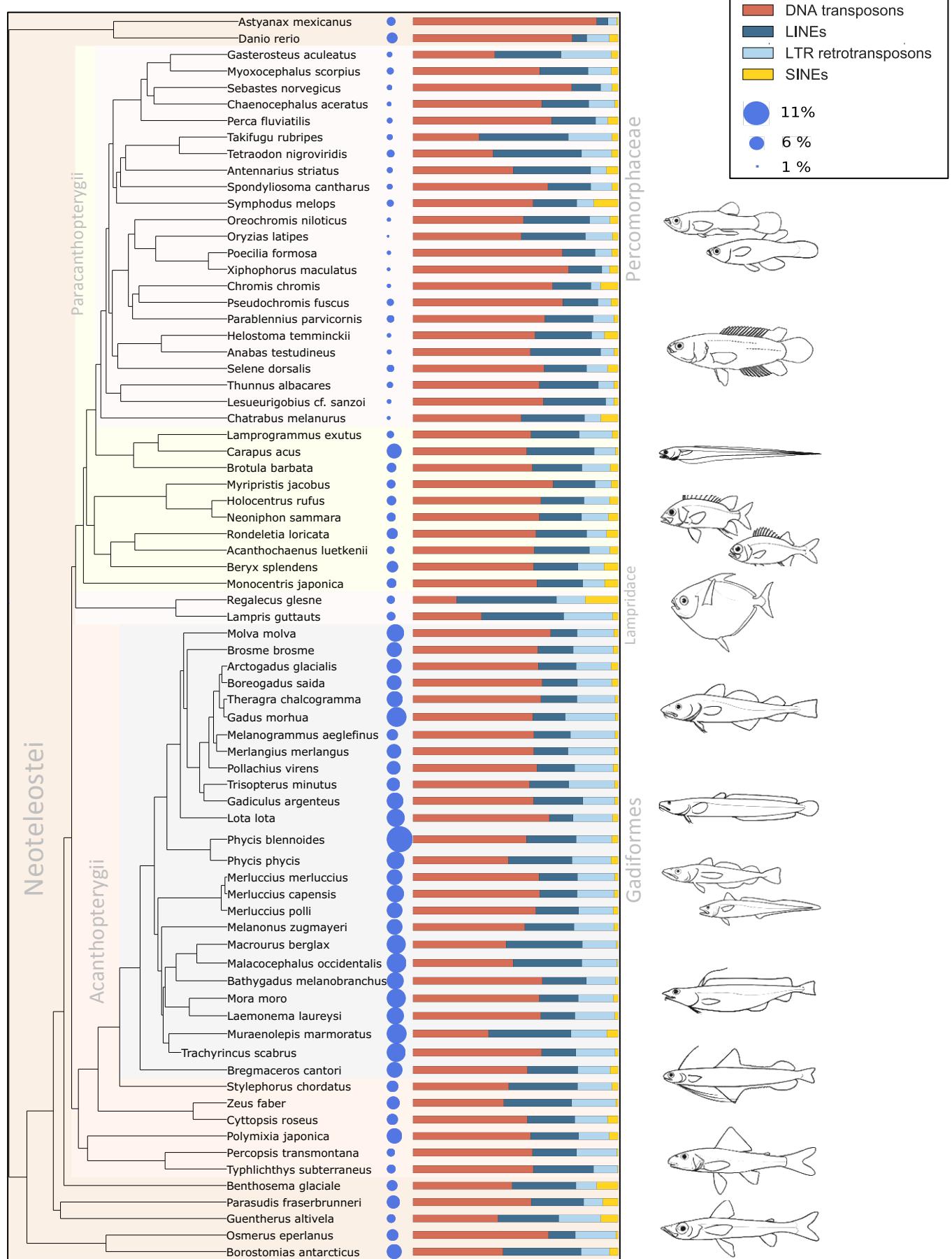


Table 1 – Reciprocal overlaps between published Zebrafish repeats and repeats detected in the custom pipeline. Of the 2.8 million repetitive elements in the published Zebrafish data (repeatmasker.org), 2.6 million overlapped with at least one bp (94%), 2.3 million overlapped $\geq 80\%$ and 1.4 million elements overlapped in full (51%). 15% of elements detected in the computational pipeline did not overlap with any repeats in the downloaded data. Assuming that repeats reciprocally overlapping $\geq 80\%$ are true positives, the pipeline has a repeat detection sensitivity of 80%. The repeat detection specificity is 85%.

	Number of elements (millions)	Percentage (%)
Any overlap	2.6	94
$\geq 80\%$ overlap	2.3	80
100% overlap	1.4	51
No overlap	0.47	15

Figure 2 – Repeat length distributions of repeats change with assembly quality. The genome assembly of Atlantic cod (Star et al. 2011) has fewer detected long (>1000 bp) genomic repeat sequences reported by RepeatModeler compared with the 2016 assembly (Tørresen et al. in print).

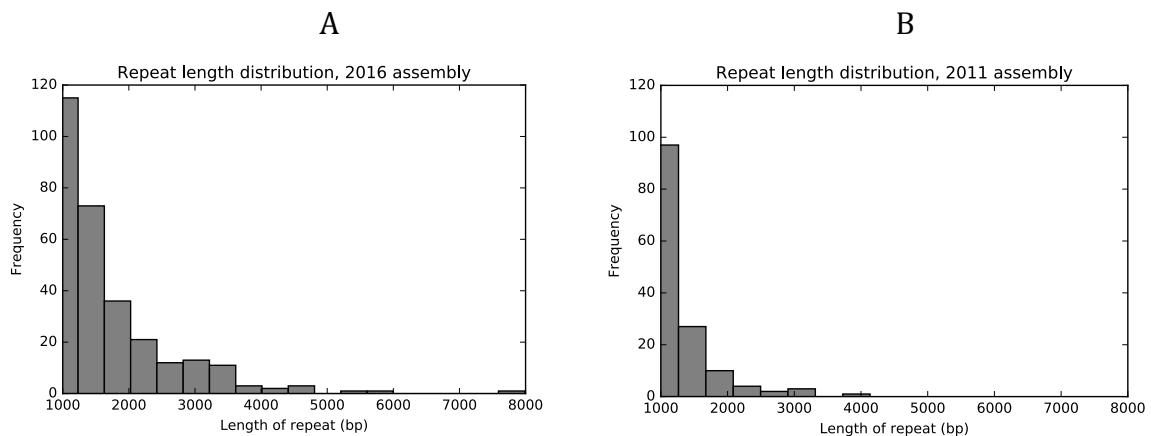


Figure 2 – Difference of masking genomes with using RepBase repeat libraries only (A) and species-specific custom *de novo* libraries (B). Selected orders are colored (Gadiformes: orange, Tetraodontiformes: yellow, Perciformes: purple, Cyprinodontiformes: red, Beryciformes: blue, Nototheniodei: green). Fish of other orders are colored gray.

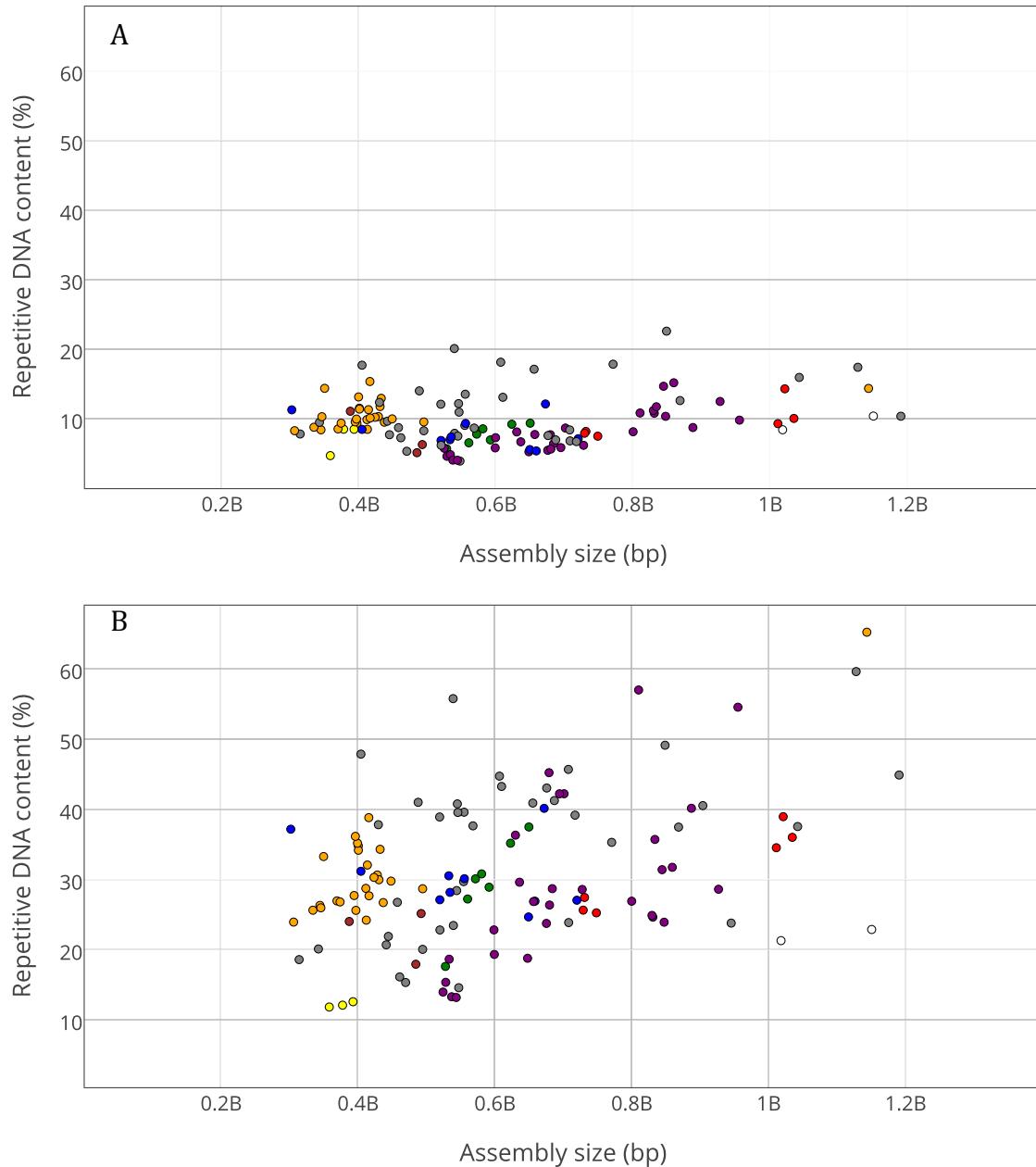


Figure 3 – Boxplots showing repetitive DNA variance in well-represented orders ($n > 5$). The codfishes (*Gadiformes*) have more tandem repeats than other fish (A), and transposable element content is more widely distributed (B), reflecting variance observed in genome assembly size (C)

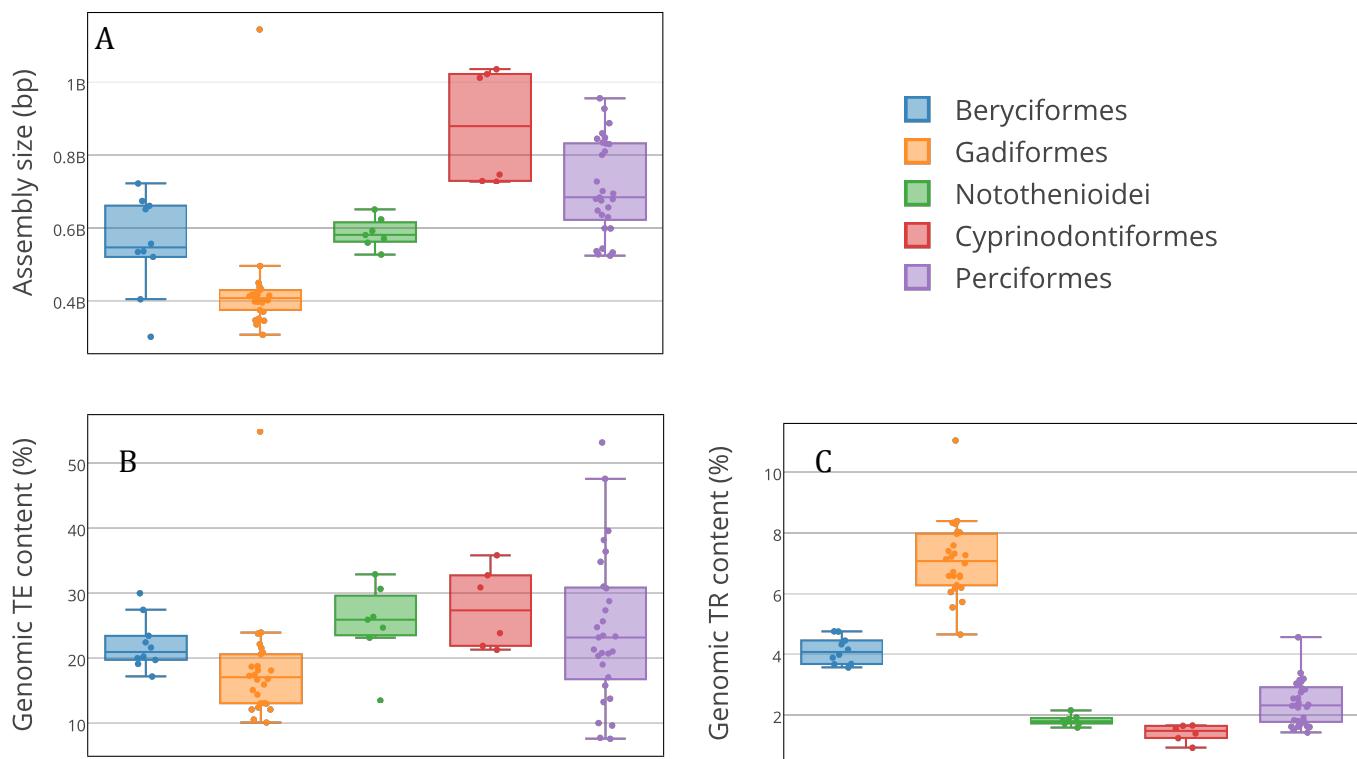


Figure 5 – Transposable elements, tandem repeats and genome size.

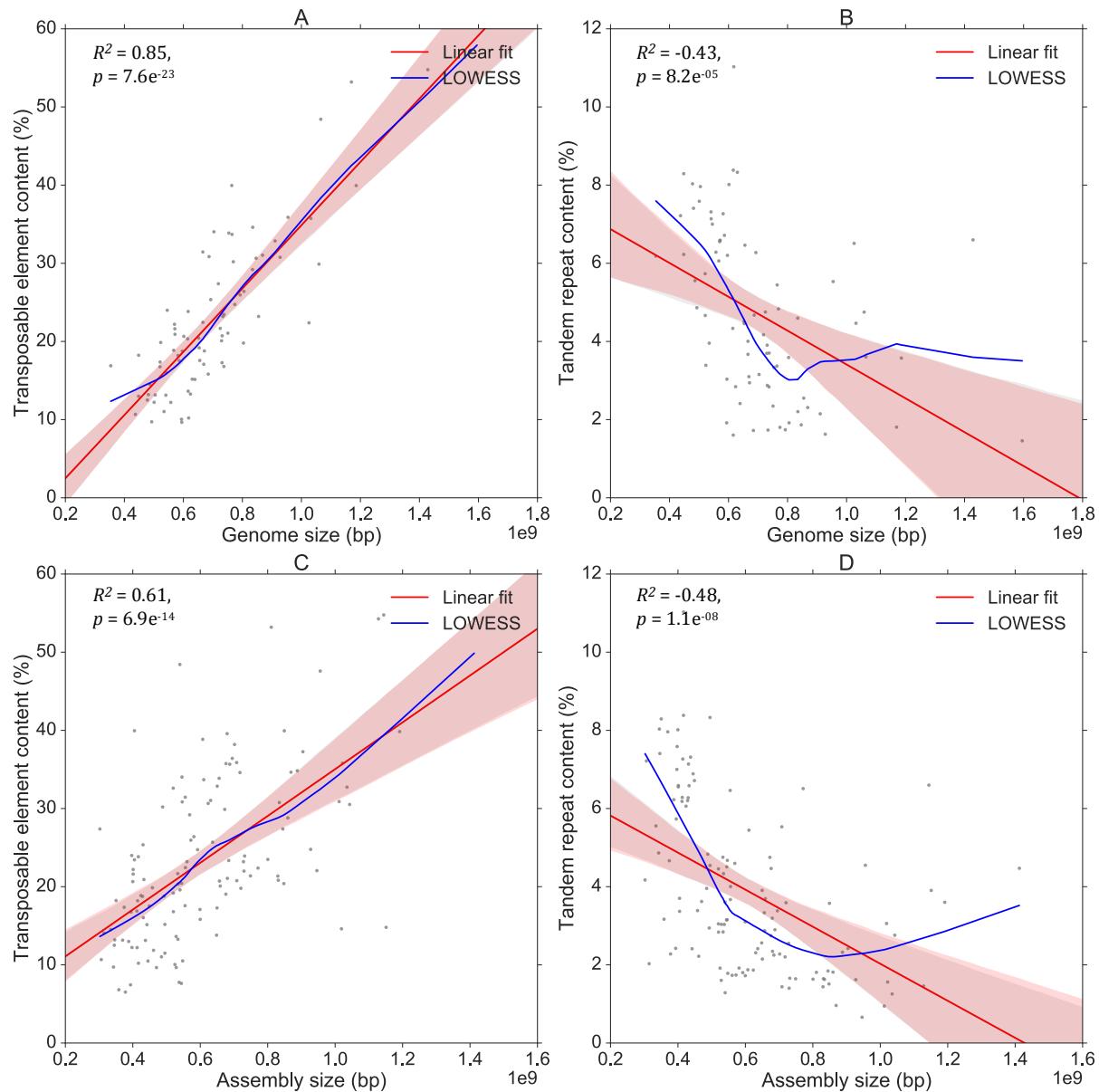


Table 2 – Teleost fish in which DNA transposons are not the most frequent transposon class.

Species	Common name	Most frequent TE class
<i>Tetraodon nigroviridis</i>	Green spotted puffer	LINE
<i>Lampris guttauts</i>	Opah	LINE
<i>Regalecus glesne</i>	King of herring	LINE
<i>Cetomimus sp</i>	Cetomimus sp	LINE
<i>Muraenolepis marmoratus</i>	Marbled moray cod	LINE
<i>Pungitius pungitius</i>	Ninespine stickleback	LTR
<i>Takifugu rubripes</i>	Japanese puffer	LINE
<i>Takifugu flavidus</i>	Yellowbelly pufferfish	LINE

Figure 4 – Transposon activity over time are similar in recently diverged species.

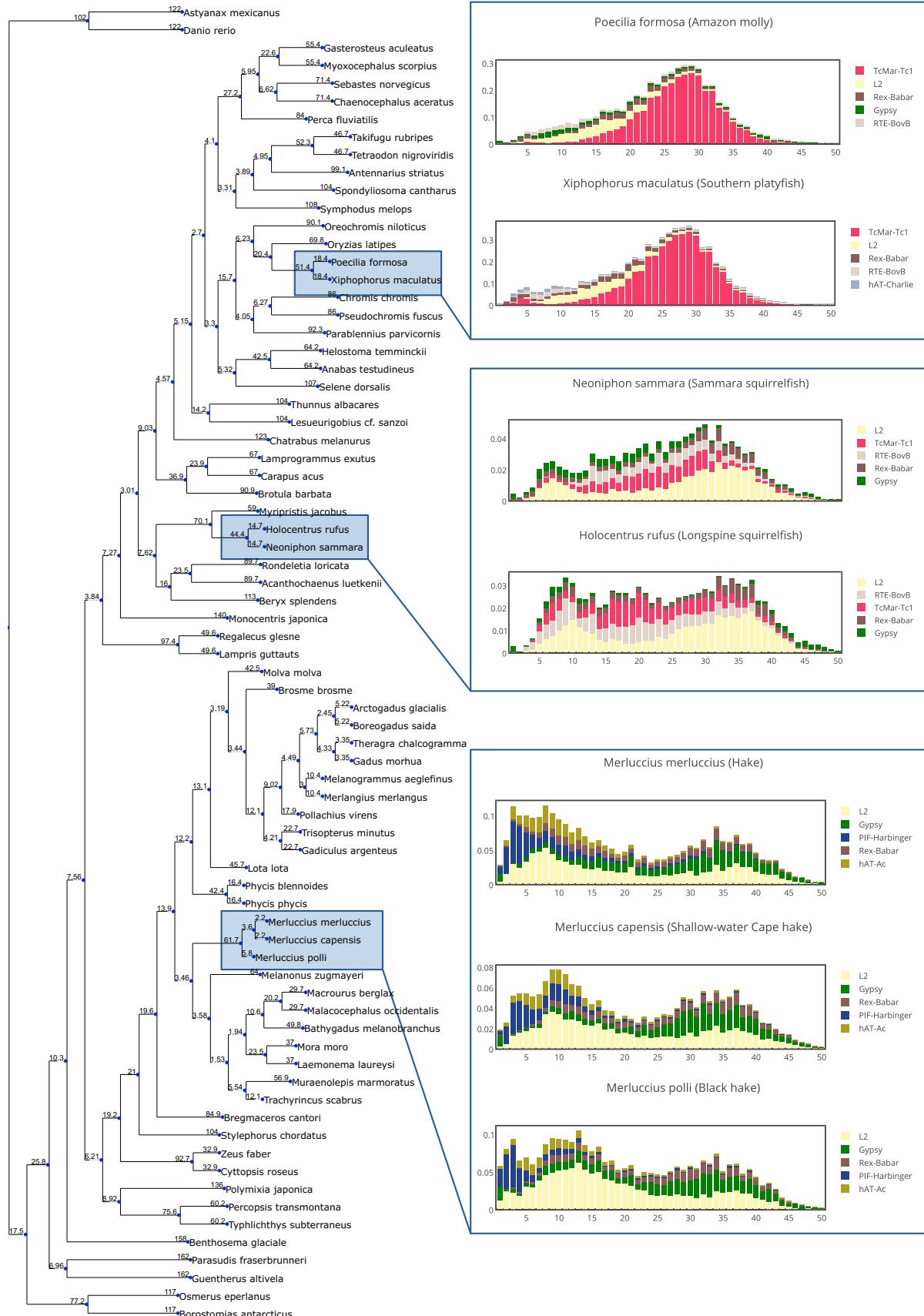


Figure X – Assembly size, fraction of repetitive DNA and TE family diversity.

