Results

Figure 1 – Repetitive DNA of teleost fish. Stacked plots show transposable elements. Blue circles show relative amounts of simple repeats.

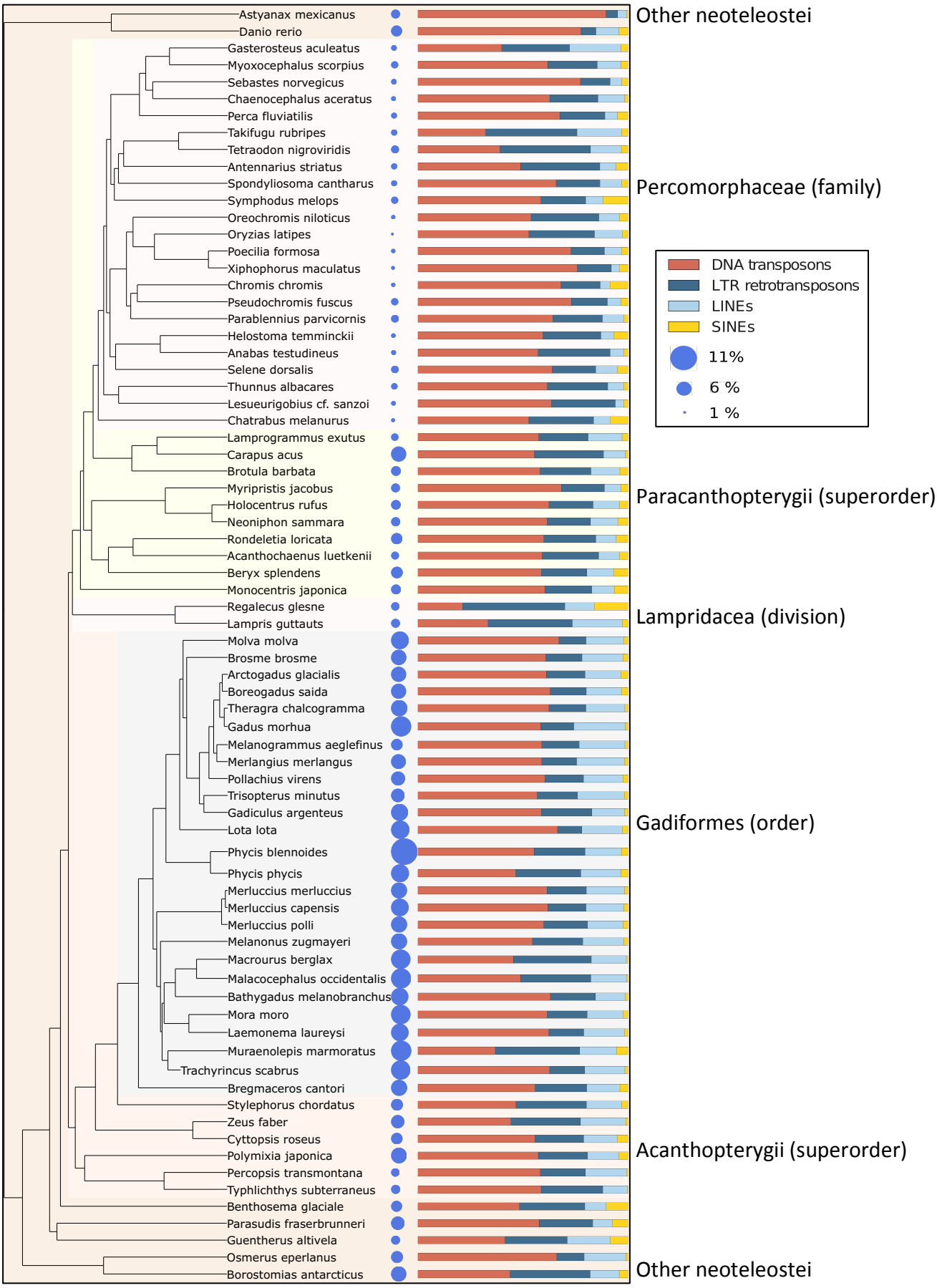

Table 1 – Reciprocal overlaps between published Zebrafish repeats and repeats detected in the custom pipeline. Of the 2.8 million repetitive elements in the downloaded Zebrafish data (repeatmasker.org), 2.6 million overlapped with at least one bp (94%), 2.3 million overlapped ≥ 80% and 1.4 million elements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

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

Figure X – 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).

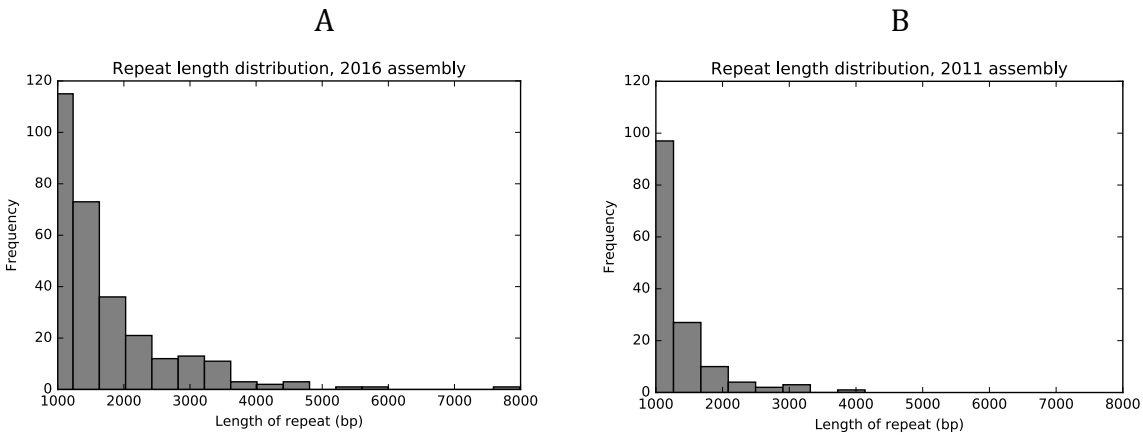

Figure 2 – Difference of masking genomes with using RepBase repeat libraries only (A) and species-specific custom *de novo* libraries (B). Selected orders are colored (Gadiformes: orange, Tetraodontiformes: yellow, Perciformes: purple, Cyprinidontiformes: red, Beryciformes: blue, Notiotheniodei: green).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

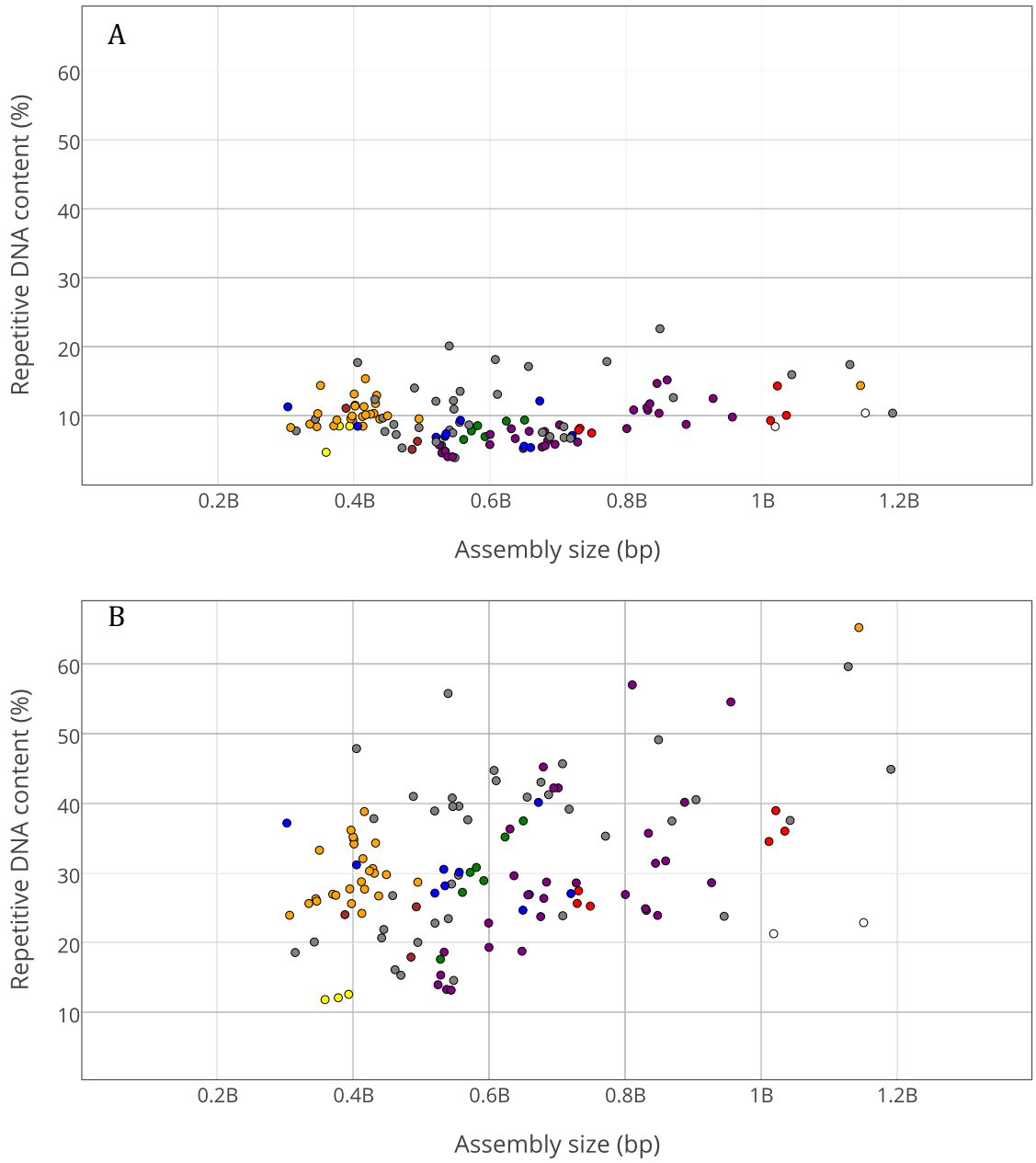

Figure 3 – Boxplots showing repetitive DNA variance in well-represented orders (n > 5). The codfishes (*Gadiformes*) have more tandem repeats than other fish (A), and transposable element content is more widely distributed (B), reflecting variance observed in genome assembly size (C)

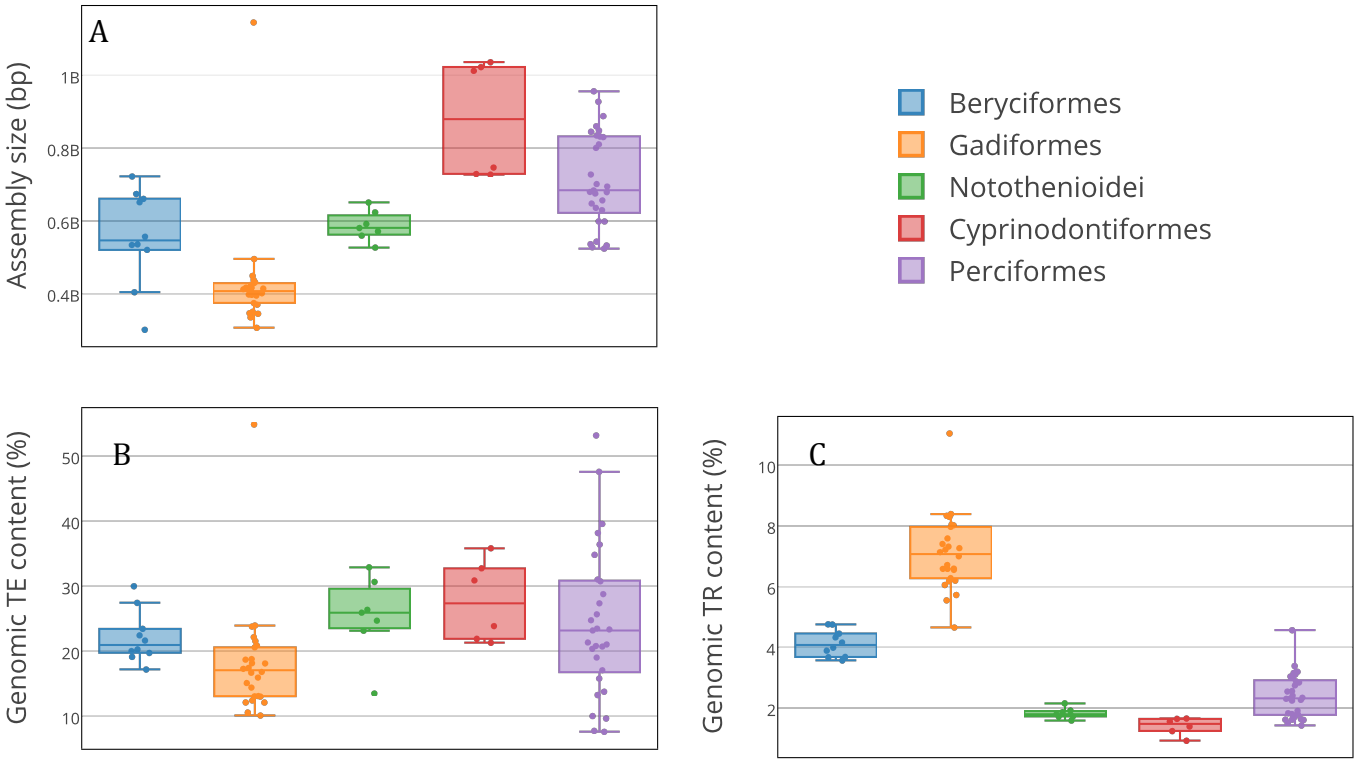

Figure 5 – Repetitive DNA and genome size.

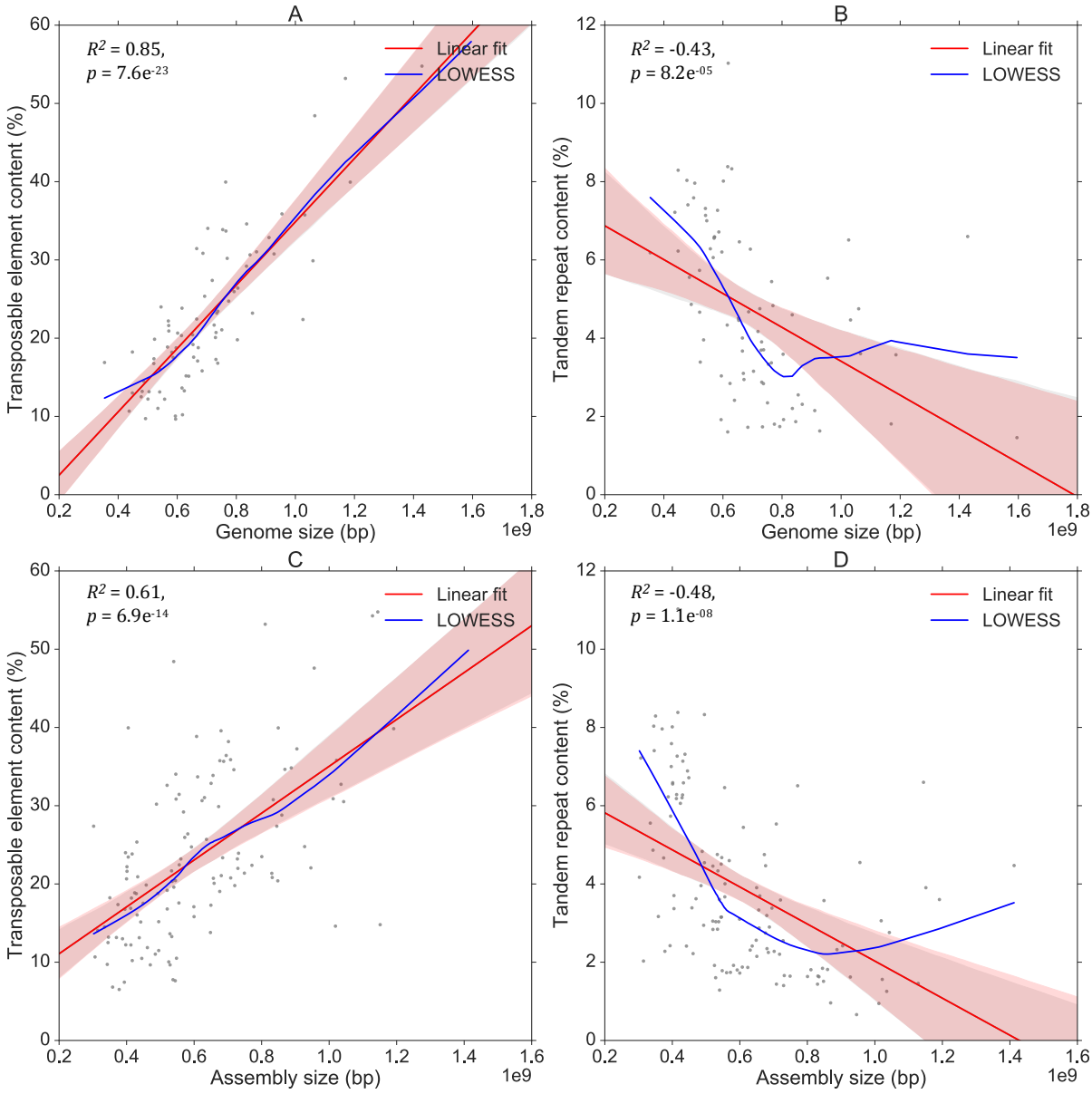

Figure 4 – Transposon activity over time are similar in recently diverged species.