

Supplementary Material

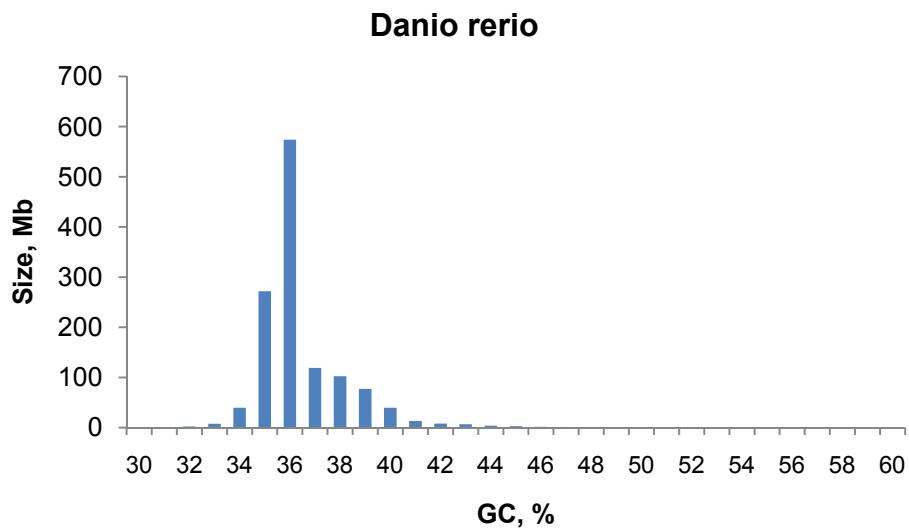
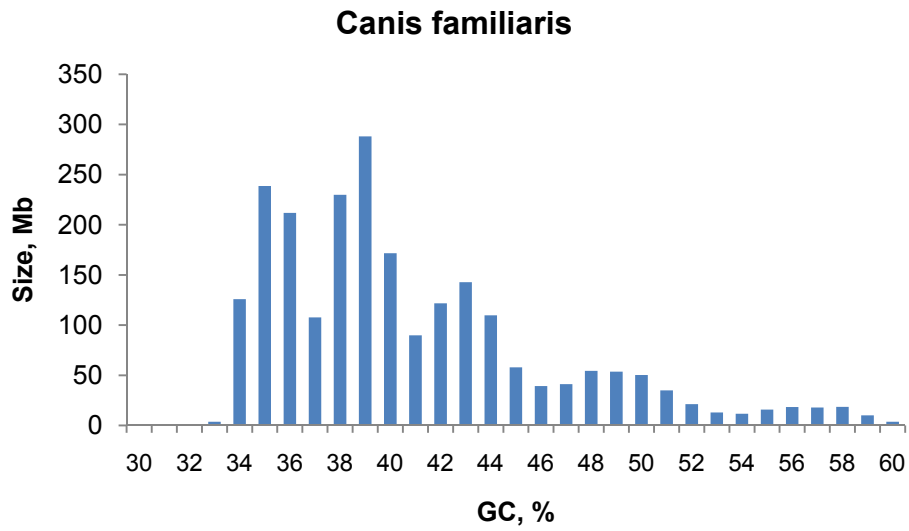
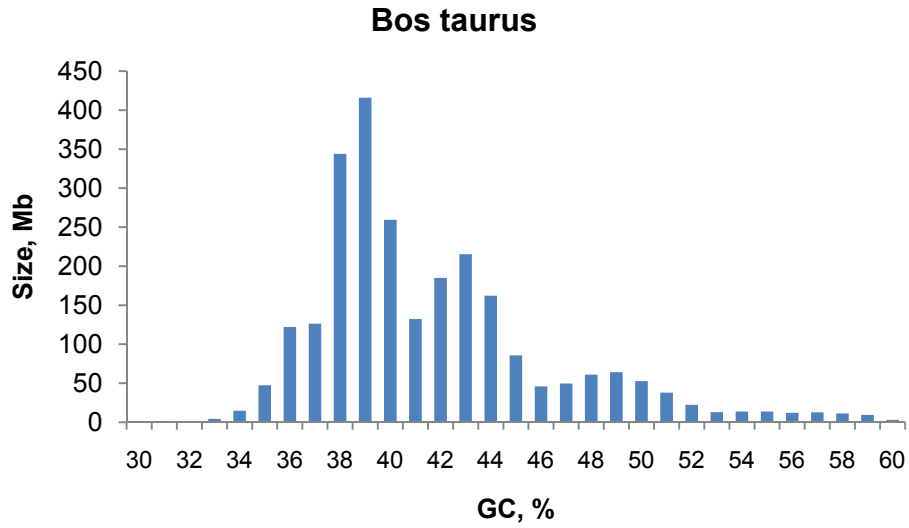
Content

1. Analysis of isochore families in all genomes
2. Analysis of methods' differences

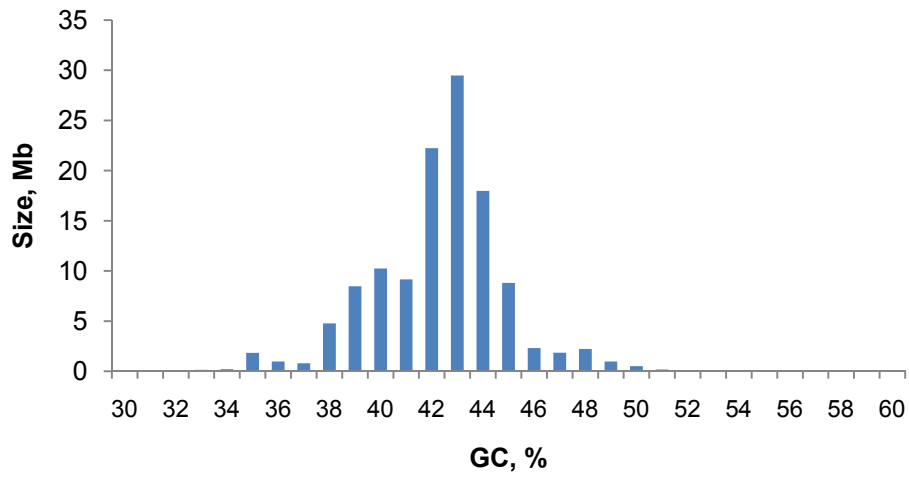
1. Analysis of isochore families in all genomes

In Figure S 3 we present histograms of the amount of genomic DNA located in segments with a given GC level for the consensus isochore segmentation for all genomes. In the case of warm-blooded vertebrates isochore families (as previously experimentally described) clearly stand out. The observed local maxima obtained from the peaks in the Figures S 3 are in fair agreement with the isochore families reported and used in the field over the last three decades [45-47] throughout all genomes. Possible evolutionary reasons for this conservation are discussed in [28]. Hypotheses about isochore evolution such as the biased gene conversion (BGC), mutation bias or selection, are reviewed in [24].

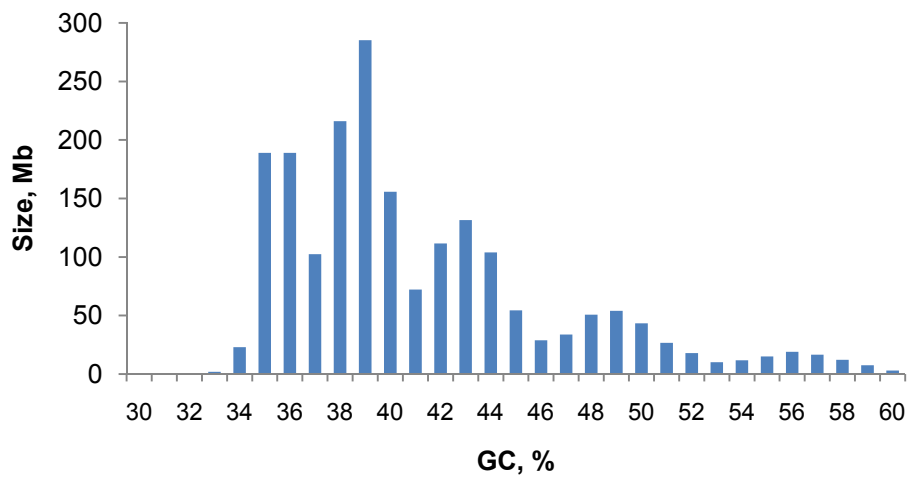
In the case of fishes (here *D. rerio*, *O. latipes*, *G. Aculeatus*, *T. nigroviridis*), we confirm that cold-blooded vertebrates show a remarkably less heterogeneous composition than warm-blooded ones [46, 48]. In each fish genome analyzed at most two isochore families can be observed. This is in agreement with large-scale experimental analysis of Bucciarelli et al [49]. The genomic differences and hence the preference of different isochore families are assumed to be related to the ecological niche where the respective fishes are living. The fishes analyzed in this study live at distinct water temperatures, the ones that are comfortable at higher temperatures show a clear bias towards GC richer isochore families.



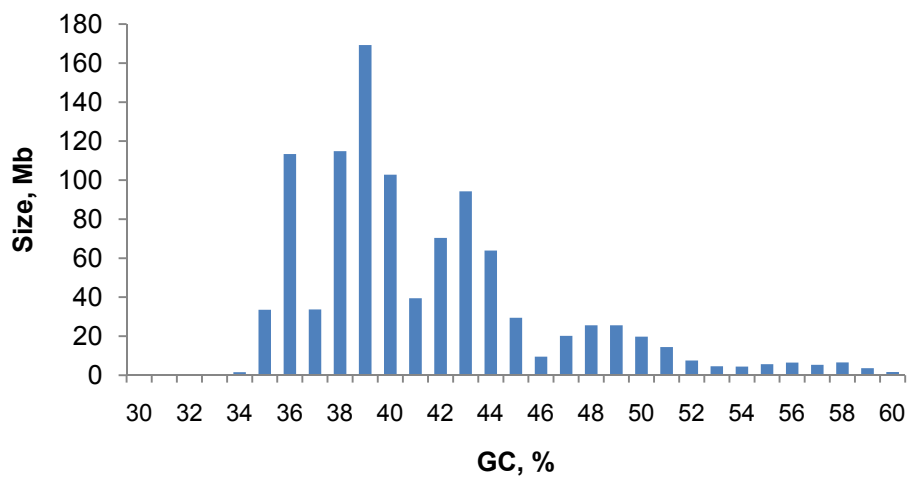
Drosophila melanogaster



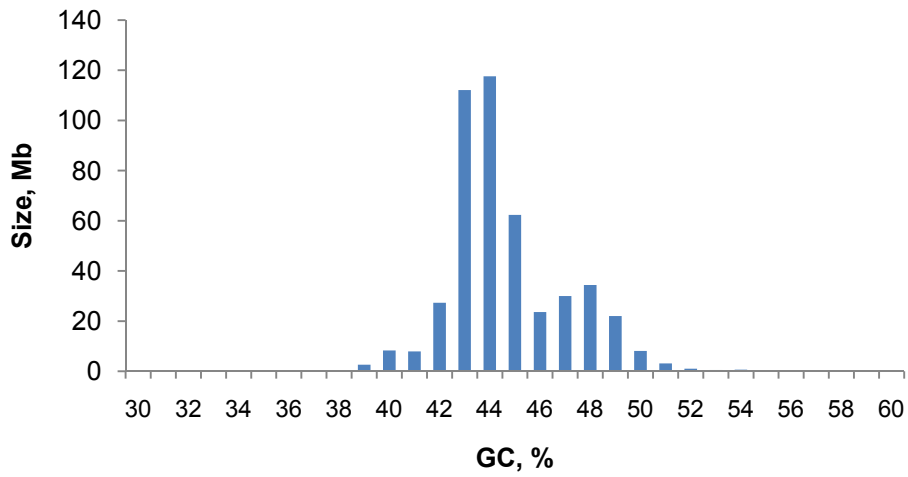
Equus ferus caballus



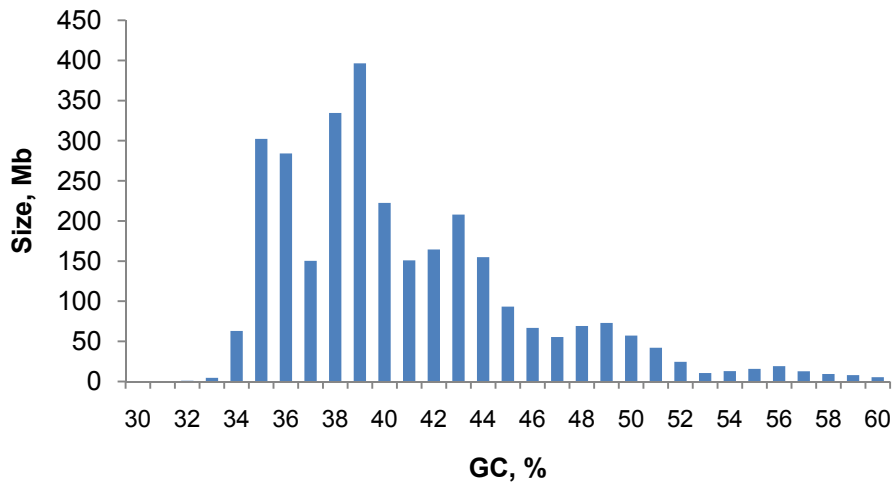
Gallus gallus



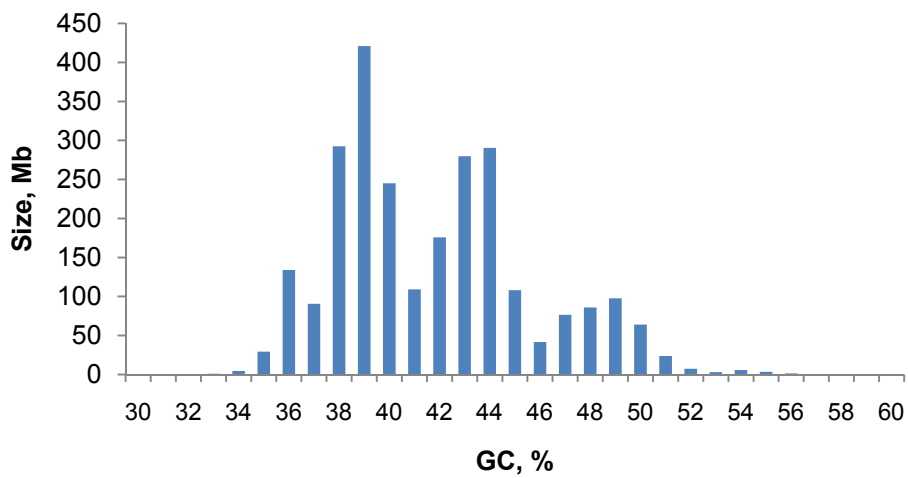
Gasterosteus aculeatus



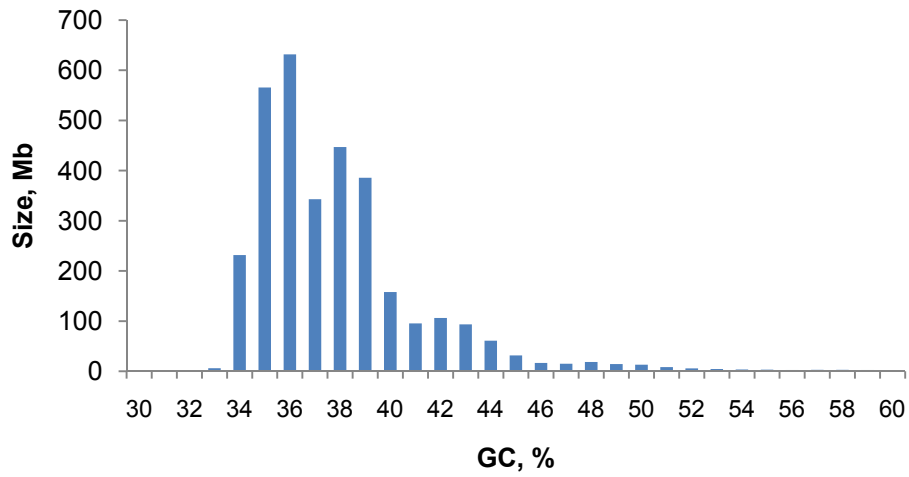
Homo sapiens



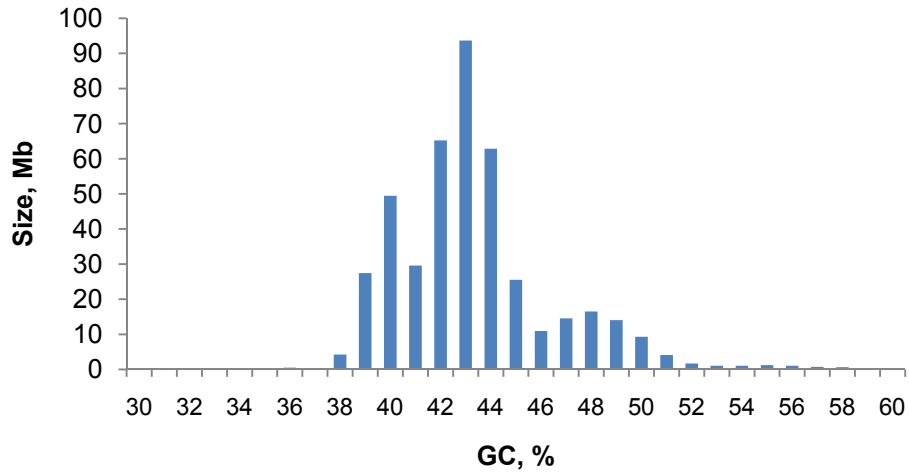
Mus musculus



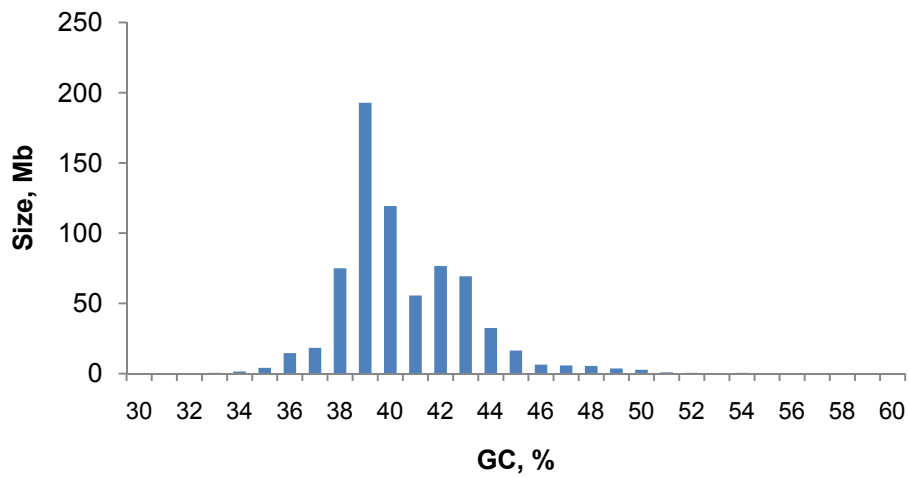
Monodelphis domestica



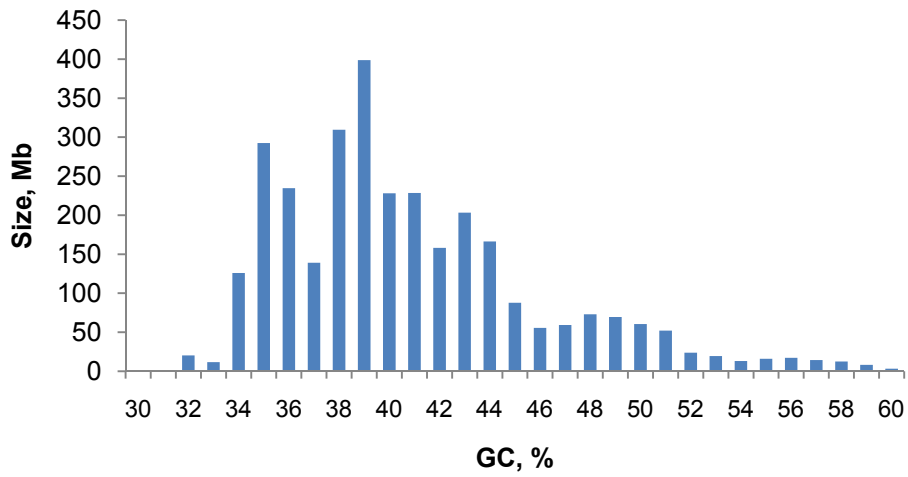
Ornithorhynchus anatinus



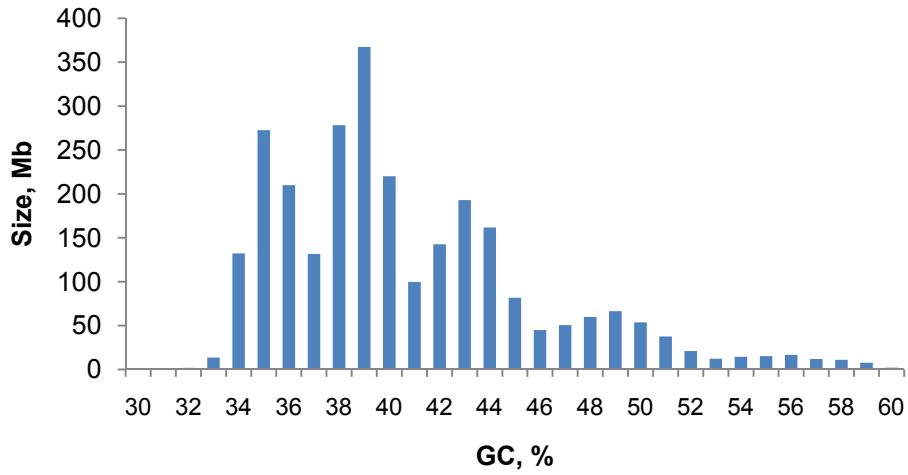
Oryzias latipes



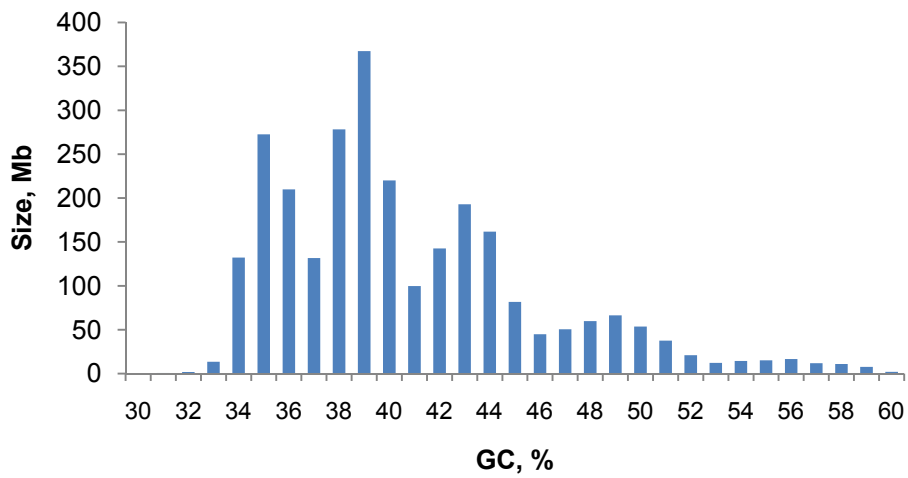
Pan troglodytes



Macaca mulatta



Rattus norvegicus



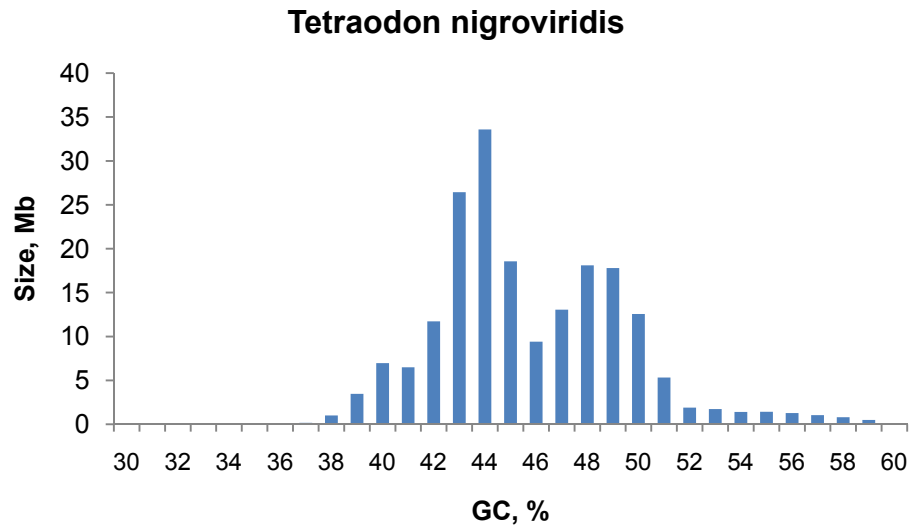


Figure S 3 Distribution of genomic segments according to GC levels in the consensus maps. Genome sizes are calculated from the sum of the segments within the pooled GC bins.

2. Analysis of methods' differences

We evaluated the contribution and the overlap of each method in all genomes to the resulting consensus map. As evaluation measure we used the conditional entropies as proposed by Haiminen et al [44] to compare segmentations. The conditional entropy values indicate the amount of information that a segmentation gives about another one. To be precise, “the conditional entropy of segmentation P given segmentation Q is the expected amount of information we need to identify the segment of P a point belongs to, given that we know the segment of this point in Q”[44]. Figure S 4 shows the conditional entropy values of all methods and genomes. It can be immediately seen, that all four methods are clearly distinct and make a complementary contribution to the consensus map throughout all genomes.

We further evaluated whether the differences of the contributing methods are significantly distinguishable or whether the differences are merely by chance. Indeed, a SimpleLogistics [50] classifier (using the WEKA implementation version 3.5.7 [51] with default parameters) is able to correctly distinguish all four methods with high accuracy of more than 80%. As input features the conditional entropy values were used. The dataset is balanced as all four methods were applied in all genomes. The evaluation was done by ten-fold cross-validation. This shows (with respect to entropy information criteria), that all methods are predominately disjunctive.

Additionally, Table S 1 shows a break-down of the classifier performance for each of the four isochore assignment approaches. It can be seen, that all four methods i.e. IsoFinder, GC-Profile, Least-Squares and BASIO are correctly distinguishable with high precision (78% to 88%). This demonstrates the significant differences and the complementary contribution of the used datasets in the consensus approach.

