

Abstract

The genome of vertebrates is made of a mosaic of long stretches of DNA, called isochores, which are compositionally uniform, and belong to a few families of GC-poor (L1 and L2) and GC-rich (H1, H2, and H3) components. Poikilotherms tend to have GC-poor genomes, while endotherms comprise both GC-poor and GC-rich isochores. The thermal theory claimed that temperature and natural selection played an active role in favoring GC-rich genomic regions, yet empirical evidence was difficult to obtain. Early work based on cesium chloride ultracentrifugation gradients showed that the Lake Magadi tilapia, a hot-water adapted fish species, displayed GC-rich regions that were absent from a close relative that lives in colder water. The goal of this study was to revisit the original study using full genome sequencing. We found that the original GC-rich regions are indeed present, that they are interspersed in the genome. Indeed, when comparing Lake Magadi tilapia with the temperate Nile tilapia, we found that 59.3 % of the genome of Lake Magadi tilapia had a base composition higher than 40 %GC, as opposed to 55.3 % of the genome of the Nile tilapia having a base composition higher than 40 % GC. We also found that their genomes comprised similar amounts of repetitive elements (20 % and 19.5 %, respectively) indicating that the shifts in base composition might not be due to repetitive elements. Further work on repetitive element analyses, protein coding genes and additional hot-water adapted fishes will provide clues as to the origin of GC-rich isochores in Lake Magadi tilapia.