

****ABSTRACT NOT FOR CITATION WITHOUT AUTHOR**

PERMISSION. The title, authors, and abstract for this completion report are provided below. For a copy of the full completion report, please contact the author via e-mail at Louis.Bernatchez@bio.ulaval.ca . Questions? Contact the GLFC via email at frp@glfc.org or via telephone at 734-669-3020.**

Lake Trout Whole Genome Sequencing

Louis Bernatchez², Andrew Muir³

² Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Pavillon C-H Marchand, Québec QC G1V 0A6

³ Great Lakes Fishery Commission, 2200 Commonwealth Blvd., Ann Arbor, Michigan 48105 USA

April 2021

ABSTRACT:

Here we present an annotated, chromosome-anchored, genome assembly for Lake Trout (*Salvelinus namaycush*) – a highly diverse salmonid species of notable conservation concern and an excellent model for research on adaptation and speciation. We leveraged Pacific Biosciences long-read sequencing, paired-end Illumina sequencing, proximity ligation (Hi-C), and a previously published linkage map to produce a highly contiguous assembly composed of 7,378 contigs (contig N50 = 1.8 mb) assigned to 4,120 scaffolds (scaffold N50 = 44.975 mb). 84.7% of the genome was assigned to 42 chromosome-sized scaffolds and 93.2% of Benchmarking Universal Single Copy Orthologs were recovered, putting this assembly on par with the best currently available salmonid genomes. Estimates of genome size based on k-mer frequency analysis were highly similar to the total size of the finished genome, suggesting that the entirety of the genome was recovered. A mitome assembly was also produced. Self-vs-self synteny analysis allowed us to identify homeologs resulting from the Salmonid specific autotetraploid event (Ss4R) and alignment with three other salmonid species allowed us to identify homologous chromosomes in other species. We also generated multiple resources useful for future genomic research on Lake Trout including a repeat library and a sex averaged recombination map. A novel RNA sequencing dataset was also used to produce a publicly available set of gene annotations using the National Center for Biotechnology Information Eukaryotic Genome Annotation Pipeline. Potential applications of these resources to population genetics and the conservation of native populations are discussed.