Phylogenomic analysis of the Lake Kronotskoe species flock of Dolly Varden charr reveals genetic signatures of sympatric radiation

Katherine C. Woronowicz[1], Evgeny V. Esin[2], Grigorii N. Markevich[3], Jacob M. Daane[4], Matthew P. Harris[1]*, Fedor N. Shkil[2,5]*

[1]Department of Orthopaedics, Boston Children's Hospital and Department of Genetics, Harvard Medical School, Boston, MA 02115
[2]AN Severtsov Institute of Ecology and Evolution, RAS; Leninskiy-33, 119071 Moscow, Russian Federation
[3]Kronotsky Nature Biosphere Reserve; Ryabikova-48, 68400 Yelizovo, Russian Federation
[4]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204
[5]NK Koltzov Institute of Developmental Biology, RAS; Vavilova-26, 119334 Moscow, Russian Federation

Corresponding authors:
*Matthew P. Harris
Email: harris@genetics.med.harvard.edu

*Fedor N. Shkil
Email: fedorshkil@yandex.ru

Author Contributions:

KCW, EVE, GNM, JMD, MPH, and FNS designed research. KCW, EVE, GNM, JMD, MPH, and FNS performed research. KCW, EVE, GNM, JMD, MPH, and FNS analyzed data. KCW, EVE, GNM, JMD, MPH, and FNS wrote the paper.

Competing Interest Statement: None
Classification: Biological Sciences, Evolution
Keywords: evolution, population genetics, radiation

Abstract

Recent adaptive radiations are powerful evolutionary case studies which can be parsed to reveal the relationship between genomic variation and the origin of distinct phenotypes. Sympatric radiations of the charr complex (genus *Salvelinus*), present a trove for comparative genomics of craniofacial evolution as charrs have repeatedly diversified into multiple sympatric morphs with distinct feeding specializations, with species flocks frequently manifesting as stereotypical assemblages of two to three resident lineages. In such repeated variation, commonalities in adaptive response can highlight potential constraints in biasing the type of phenotypes available for evolutionary change. A species flock of Dolly Varden charr (*Salvelinus malma*) in Lake Kronotskoe constitutes a unique case study of charrs as it contains at least seven true-breeding lineages, each with defining morphological and ecological traits. This is the most extensive radiation described for the genus. Here, we performed the first genome-wide analyses of variation within this species flock to provide a working foundation in which to address genetic foundations for adaptive change in character. Our data support distinct reproductively isolated lineages within the lake environment suggesting true breeding morphs and little hybridization. Analysis of changes within the lacustrine lineages, we find specific selection on thyroid and craniofacial genes as key shared genetic basis for the lake radiation. Further, we find a shift in thyroid signaling as a key modulator of subsequent lineage specification. Our data delineate a clear genetic basis for diversification of specialized lineages and provide a new model to understand commonalities in adaptive radiations through constraint on development.

Significance Statement

Salvelinus charrs have undergone repeated, independent adaptive radiations across Holarctic lakes, each involving large morphological transitions. Parsing the genetic and developmental mechanisms which facilitate such adaptive potential is important to understand adaptive radiations. A species flock of Dolly Varden charr (*Salvelinus malma*) from Lake Kronotskoe constitutes at least seven true breeding, morphologically distinct lineages; the broadest radiation described for the genus. We performed genome-wide analyses of variation within this flock and assessed shared changes within the lake morphs, as well as specific changes associated with adaptive morphologies. The data support reproductive isolation, albeit with shared genetic signatures that likely catalyzed the radiation. Genomic and physiological data suggest thyroid signaling modulation underlies the exceptional adaptations of this and other charr radiations.

Main Text

Introduction

The Salmonoid fishes of the genus *Salvelinus* represent an exceptional example of repeated morphological adaptations into new environments (1). Most charr lineages are represented by morphologically distinct populations implementing varied life histories, including complete and partial anadromy, as well as freshwater riverine and lacustrine residency (2-4). Frequently, freshwater resident populations establish species flocks of stereotypical morphologies, each associated with an ecological niche (5-12). Within a sympatric population, varying diets, depth preferences, and spawning intervals are commonly observed (1, 13). Within independent lacustrine *Salvelinus* radiations, specific morphological adaptations have repeatedly evolved. Such parallelism suggests that populations experience similar selective pressures, or, alternatively that there are genetic biases shaping the morphologies that can be embodied. The variability in charr radiations proffers important genetic case studies to uncover mechanisms of rapid morphological diversification. The genetic mechanisms that enable charrs to generate such distinct morphologies, however, are poorly understood. Unlike comparable models of repeated radiation and differentiation such as stickleback (14-16) or *Astanyax* cavefish (17), charrs have repeatedly diversified into multiple sympatric morphs with distinct feeding specializations rather analogous to those seen in the evolution and diversification of both African and South American cichlids (18-20), or African and Asian barbs (21-23). Shifts in developmental timing may underlie these common transitions to specialized morphologies among charrs (24). In this model,

development shapes particular radiations by imposing evolutionary constraint and imparting a common axis of change. Early and pleiotropic developmental shifts would permit parallel evolutionary events via few genetic modifications.

The Lake Kronotskoe radiation of Dolly Varden (*Salvelinus malma*) is unique amongst charrs, as it contains at least seven reproductively isolated phenotypes associating with different ecological niches (5). This radiation is currently the broadest observed within resident lacustrine populations of this genus and, indeed, among salmonids more broadly (5). Lake Kronotskoe is one of several volcanogenic lakes of the Kamchatka peninsula (**Fig. 1A**) originating approximately 12,000 years ago due to a lava flow damming the valley of the ancestral river (**Fig. 1B)** (25). The resulting landlocked Dolly Varden population diversified within the new lacustrine environment, leading to prominent changes in craniofacial form and new feeding strategies (**Fig. 1C, D**) (5, 12). Phenotypic plasticity among charrs is considered a driving-force behind their incipient speciation (26). A major axis of change is proportional jaw length, seen in piscivorous Longhead and deep-water benthivorous Bigmouth morphs, as well as the modulation of frontonasal proportions found in the Nosed morphs. The deep-water omnivorous Smallmouth morph demonstrates an increase in eye size in proportion with the cranium (**Fig. 1C, D**). Enlarged eye size is widespread among deep-dwelling resident charr lineages in other lakes and may suggest specialization for foraging in low light conditions (1). The varied Lake Kronotskoe charrs show strict natal homing with some morphs spawning in tributaries and other morphs spawning within the lake (27). The morphs also exhibit differential annual windows for the timing of reproduction, which reinforces reproductive isolation and facilitates sympatric radiation (28).

Here, we survey the genetics of the Dolly Varden radiation within Lake Kronotskoe to frame the genetic context of their diversification as well as the foundation for the morphological and physiological adaptations of these specialized lineages.

Results

Genome wide assessment of variability within and between Lake Kronotskoe Dolly Varden morphs

To investigate genome-wide variation across the charr genome, we performed genome-wide targeted capture, of coding and non-coding loci in pooled samples of Dolly Varden charr from Lake Kronotskoe to identify lineage-defining variation with the morphs. Targeted capture allows cross-clade sequence comparisons of species and facilitates analysis of genetic variance within a lineage. We have applied similar methodology to assess variation in Belonifomes (29), rockfishes (30), and notothenioids (31) allowing for clade-wide analysis of genomic variation.

We constructed a pan-Salmonidae capture-array consisting of 558,882 bait sequences designed against conserved Atlantic Salmon (*Salmo salar*) and Rainbow Trout (*Oncorhynchus mykiss*) genomic sequences and targeting 97Mb of total sequence **(Supp. Fig. 1**). Coding regions constitute 83.8% of targeted regions. Conserved non-coding, ultraconserved non-coding, and miRNA hairpins comprise the remaining 16.2% of targeted regions. With our capture methodology and analysis pipeline, bait sequences successfully hybridize with targeted elements harboring up to 15% deviation in sequence identity (32), allowing recovery of sequence variability within library pools of several individuals. Pairwise analyses between homologous loci across the genome were used to distinguish fixed and variable genetic loci within morphs. Genes associating with variation were identified by making strategic comparisons within a phylogeny that includes character diversification (33).

We pooled DNA from multiple individuals from each Lake Kronotskoe morph except the Nosed 2 morph, prior to performing targeted capture. For outgroup assessment, we sampled anadromous Dolly Varden charr from the local Kamchatka river basin, which is neighboring but isolated from that of Lake Kronotskoe (**Fig. 1B**). For each group, we recovered approximately 90% of the targeted elements sequenced to a mean depth of 25-52 reads (**Supp. Table 1**). We detected substantial variation within each lineage and the putative ancestral population (avg pi in Dolly Varden = 0.003) supporting further pairwise analysis of shared and unique variation within

each lineage. Underrepresented loci from our capture were limited and represent gene classes observed in other broad capture approaches (**Supp. Table 2**)(29-31).

Genetic differentiation and relationships among Dolly Varden morphs

An analysis of genetic variability between anadromous and lacustrine Dolly Varden morphs shows a clear differentiation of ecologically specialized lineages (**Fig. 2A**), having pairwise mean Fst consistent with distinct populations (**Fig. 2C**). The greatest pairwise differentiation is found between the riverine Dolly Varden charr and lacustrine deep-water Bigmouth morph (Fst = 0.127). Notably, each lacustrine lineage is more differentiated from the riverine Dolly Varden population that from any other lacustrine lineage. Furthermore, lacustrine lineages are equally differentiated from the putative ancestral Dolly Varden population. The least differentiated pairing is between the lacustrine Nosed 1 and Nosed 3 morphs (Fst = 0.047). Although informative of broad patterns of divergence, these values are likely underestimations of genetic differentiation due to the conservome-based dataset biasing analysis to regions having an inherent constraint on variation.

Based on changes in resource utilization, prior hypotheses of the charr radiation within Lake Kronotskoe argued for multiple step diversification accompanied by feeding specialization resulting in the extant morphs (5). We investigated relationships among Lake Kronotskoe charrs through construction of a phylogenetic tree using the genetic variation data obtained by pooled sequencing of the different morphs. We used IQ-TREE to derive a phylogeny from a dataset, which included 1,218,207 variant sites. Of these, 44,157 variants were parsimony informative. Our data support the existence of distinct morphs having high bootstrap support for each of the six sampled lineages (**Fig. 2A**). The genetic analysis clusters Longhead and White lineages as an outgroup to a clade consisting of the Bigmouth, Smallmouth, and Nosed morphs.

Introgression analysis further supports the existence of distinct lineages, albeit with the presence of some incomplete lineage sorting. Significant introgression was identified in 80% of the 20 possible trios (DSuite; Holm-Bonferoni FWER < 0.01) in patterns which deviated from the topology of the phylogeny (**Supp. Table 3**) (34). Despite observed significance, the introgression values are relatively small when compared to the recent timeline of the Lake Kronotskoe radiation. The Bigmouth and Nosed 1 morphs were ascribed the greatest excess of allele sharing (D_{tree} = 4.2%) (**Supp. Table 3**). We further calculated f_4-admixture ratios and used f-branch statistics to disentangle the interrelatedness of admixture signals among morphs, which share common internal phylogenetic branches (**Fig. 2B**). The greatest proportions of shared alleles are between the Bigmouth morph and the Nosed 1 morph (f-branch = 9.0%), between the Bigmouth morph and the Nosed 3 morph (f-branch = 8.3%), and between the White morph and the ancestor of the Nosed 1 and Nosed 3 morphs (f-branch = 5.9%). The f-branch statistic further supports that the Lake Kronotskoe lineages differentiated while maintaining relative genetic isolation.

To determine the extent to which Nosed 1 and Nosed 3 share common introgressed loci with Bigmouth, we calculated D in sliding windows (40 variant windows, 20 variants step-size) for Smallmouth, Bigmouth, Nosed 1 (mean = 0.05, SD = 0) and Smallmouth, Bigmouth, Nosed 3 (mean = 0.04, SD = 0.21 for S, B, N3). For all introgression analyses, riverine Dolly Varden served as the outgroup. Among sliding windows with D >= 0.8 (32 of 10,617 windows for S, B, N1; 28 of 10,790 windows for S, B, N3), we observed shared, isolated regions of the genome with signatures of admixture. The largest interval with evidence for shared introgression was an 11.8MB region of elevated D (mean D for the interval = 0.53 for S, B, N1; 0.54 for S, B, N3) with of more than one third of windows with D > 3 SDs from the mean (**Supp. Fig. 2**). However, most introgressed intervals are more spatially restricted. One locus of interest spanned an interval centered on *zyg11b* which has been implicated in craniofacial microsomia (**Fig. 2D**)(35). Another included multiple genes of interest including, *zswim*5, which is expressed in cranial neural crest cells (36), *rnf152*, which is involved in regulating neural crest formation (37), and *mc4r*, which is a crucial regulator of appetite and metabolism via leptin and thyroid signaling modulation (**Fig. 2E**)(38). This suite of craniofacial and metabolic genes may have functioned like a supergene,

which was spread via shared introgression among the Bigmouth and Nosed lineages and subsequently reinforced the differentiation of distinct lineages.

Patterns of differentiation between river and lake populations of Dolly Varden charr

To understand the shared variation that differentiates the Lake Kronotskoe residents from the riverine population of Dolly Varden, pairwise Fst was calculated per coding (25,373 genes) and conserved non-coding element (CNE)(22,575 CNEs) from our targeted capture (**Fig. 3A, B**). Fst values were computed by grouping all variation contained within the species flock and comparing against the variation contained in the riverine Dolly Varden charr population (lake versus river). There are 327 genes and 80 CNEs with Fst > 0.5. Each CNE was assigned flanking target genes by GREAT. Some intervals harbor multiple highly differentiating CNEs. For example, the close proximity of two CNEs led to *cholinesterase-like* and *dync2h1* to appear twice (**Fig. 3B**). Interestingly, craniofacial (orange cells in table) and thyroid (blue cells in table) related genes (5 craniofacial genes of top 20, 2 thyroid genes of top 20) and CNEs (12 craniofacial CNEs of top 20, 3 thyroid CNEs of top 20) are well represented among the most differentiating elements (**Fig. 3A, 3B**).

Craniofacial genes assigned to highly differentiating CNEs by GREAT included both those with predicted roles in neural crest (*tfap2a*) (39), as well as frontonasal (*meis2, pitx2*) (40, 41) and splanchnocraniun (*faf1*) development (42). *pitx2* is involved in release of thyroid stimulating hormone from the pituitary (43). *dync2h1* and *tfap2a* appear more than once as putative target genes, suggesting multiple alterations to gene expression domains, timing, or strength at these loci may contribute to the genetic landscape that distinguishes the riverine Dolly Varden from the lacustrine morphs. Some CNEs are flanked on both sides by genes implicated in craniofacial development (*ofcc1* and *tfap2a*, *meis2* and *spred1*, *fam172a* and *nr2f1a*). Genes connected to ion transport (*myh6, khnyn, or11a1, endod1, cacnb2,* and *hrh1*), and vesicular transport (*man2a1, shld2, gdpd2, onecut2,* and *clint1a*) according to KEGG and Reactome reconstructions are also well represented among the top differentiating genes and as targets of the top differentiating CNEs. The majority of nucleotide changes identified were single SNPs with unknown functional significance. However the high Fst is reflective of potential selection or bottleneck at the locus.

GO terms were assigned to all genes (2,825 GO terms) and CNEs (1,419 GO terms) that differentiate riverine Dolly Varden from Lake Kronotskoe residents (Fst > 0.5). Among GO terms appearing 6 or more times, themes emerged (**Fig. 3C**) (See **Supplemental File 5**). Highly differentiating genes are associated with roles in ion channel regulation and protein trafficking. Highly differentiating CNEs are associated with brain, kidney, and cartilage development. Other functions shared among highly differentiang coding and noncoding elements include functions such as DNA binding, transcription factors, and regulation of RNA polymerase II.

The gene *slc26a10,* encoding a sulfate transporter, functions as a part of thyroid hormone synthesis, and acts downstream of thyroid hormone receptor alpha (Thra) (44). We found that *slc26a10* has segregating nonsynonymous changes in conserved residues that are highly differentiated between riverine and lacustrine morphs (**Fig. 4A, 4B**). The function of this gene however is not well understood. Similarly, *catl1l,* encoding Cathepsin L1-like, is orthologous to cathepsin L1 and is predicted to contribute to normal function of the thyroid and have proteolytic functions which facilitate release of T4 from its prohormone thyroglobulin (45). Further, splicing factor 3b subunit 4 (*sf3b4)* is a key developmental regulator of frontonasal patterning and growth and is associated with craniofacial Nager syndrome in humans (46). Within this locus, we find synonymous variants at high Fst in all lacustrine morphs, and fixed within benthic groups (**Fig. 5A, 5B**). Elevated footprints of clustered, highly differentiating variants may indicate further modifications within poorly conserved non-coding regions which were not targeted by our conservome-derived bait sequences.

Thyroid hormone signaling activity is associated with shifts in craniofacial form

The pattern of fixation in thyroid associated genes and non-coding elements in lacustrine morphs compared with riverine Dolly Varden warranted further investigation, as modulation of thyroid

signaling has been implicated in the charrs phenotypic specialization in different environments (47). To address if changes in thyroid signaling were associated with evolution of different morphs in Lake Kronotskoe, we assessed levels of circulating thyroid hormone in adult riverine and lacustrine charrs. Intriguingly, we found a significant decrease in T_3 hormone in specific lacustrine populations (**Fig. 6A**). The pattern of reduced T_3 abundance across the species flock correlates with a clear change in craniofacial proportions: Nosed and Smallmouth morphs, with their subterminally positioned mouths, have significantly lower hormonal level compared to riverine Dolly Varden charr and the lacustrine White morph, which have comparatively 'wild-type' craniofacial form, as well as the piscivorous Longhead and deep-water benthivorous Bigmouth morphs, which are highly specialized (**Fig. 1D**).

We assessed specific mutations identified as fixed or nearly fixed (Fst > 0.9) within Nosed morphs that differentiate them from Longhead or White morphs. Our findings identify a disproportionate number of differentiated loci in Nosed lineages that are known to regulate thyroid signaling (**Fig. 6B**). A candidate locus that differentiates the low T_3 Nosed 1 and Nosed 3 morphs from the Longhead and White outgroups is a *leptin* homologue (**Fig. 6C**). All salmonids have two leptin A ohnologues derived from the salmonid shared whole genome duplication. LeptinA2 paralogues encode an N-terminal sequence extended by 66 amino acids compared to LeptinA1 or B orthologues within salmonids. The nonsynonymous change differentiating Nosed lineages lies within a conserved residue of the unique sequence of this paralogue. As Leptin hormone modulates thyroid axis activity and vice versa, both endocrine signaling pathways affect global metabolic activity in concert. Notably, *mc4r*, which shows evidence of allele sharing between the Bigmouth and Nosed lineages, serves as a relay through which leptin stimulates the thyroid axis, suggesting that these hormonal axes may be modulated at multiple levels by distinct patterns of modification in differentiated lineages (38). Exogenous leptin and thyroid hormone treatments are associated with enlarged craniofacial dimensions marking an intriguing relationship between these pathways and Nosed lineage diversification (48-51).

We also identified a highly differentiating variant within a conserved non-coding enhancer of *otx2b* (**Fig. 6D**). Nosed 1 and Nosed 3 morphs are fixed or nearly fixed for a variant allele which lies in a predicted *Pax1* transcription factor binding site. *otx2b* is involved in development of the skull as well as the anterior pituitary gland, which is a source of hormonal signals also involved in thyroid regulation (52, 53). Due to the significant overlapping domains of *pax1* and *otx2* in the pharyngeal arches and the oral endoderm from which the pituitary gland arises, the differentiating variant identified provides a plausible regulatory shift associated with evolution of craniofacial morphology (54, 55).

Discussion

Lake Kronotskoe harbors a unique radiation of Dolly Varden charr that provides a powerful new case study of the genetic and developmental foundations supporting vertebrate radiations. Our data redefine the evolutionary relationships among the lake morphs and supports differentiated true-breeding lineages. Previous models based on morphology, ecology and feeding behavior suggest two lacustrine clades: a deep-water clade consisting of Smallmouth and Bigmouth morphs, and a shallow-water clade composed of Longhead, White, and Nosed morphs (5). That our high confidence phylogeny upturned the previous expectations of evolutionary relationships highlights the necessity of genomic data for studying radiations.

Lake Kronotskoe arose from a volcanogenic event and, presently drains via a waterfall which is considered impassable by charr (56). Pairwise Fst analyses found that each lake lineage is more differentiated from riverine Dolly Varden than from any other lineage. Furthermore, each lake lineage is differentiated from riverine Dolly Varden charr at roughly equivalent levels. Equal genetic differentiation from the putative ancestral population suggests that this sympatric radiation occurred in the absence of secondary contact. We also determined that the Lake Kronotskoe lineages maintain reproductive isolation through low hybridization. By contrast, the much older Lake Malawi cichlid radiation, has f-branch values commonly exceeding 5% and up to 14.2% (19) and *Coregonus* salmonids also have extensive introgression (57). These new data suggest that the species flock in Lake Kronotskoe was established by a single founding population having

shared genetic signatures that quickly established and maintained thoroughly reproductively isolated populations.

We find specific genetic differentiation between lacustrine resident lineages and anadromous Dolly Varden populations with selective signatures in genes regulating craniofacial development and thyroid function. While the relationship to morphological radiation is less clear, the abundance of ion transport and protein trafficking genes among the highly differentiating genes, suggests that these processes may also be important drivers of charr evolution. Crucially, ion homeostasis is central to effective osmoregulation during freshwater adaptation (58). Whether and how vesicular function contributes to adaptations in Lake Kronotskoe is less obvious. Differentiating CNEs are associated with roles in brain, kidney, and cartilage development. The kidneys are indispensable for osmoregulation and obligate freshwater populations of salmonids place different demands on their kidneys their anadromous counterparts (59). Cartilage templates lay foundations for most craniofacial structures. The data suggest that CNEs are particularly influential in the evolution of kidney function and cartilage morphology within the Lake Kronotskoe radiation. Additionally, genes that function in thyroid hormone regulation are differentially selected, suggesting modulation of the hormone may underlie stereotypical phenotypic shifts in lacustrine charr morphology. As many specialized morphologies are hypothesized to arise from heterochronic shifts in development (7), the thyroid axis may prove to be one of the common mechanisms underlying the adaptive potential permitting incipient radiations and may contribute to the remarkable similarities exhibited by lacustrine charr species flocks (47, 60). Indeed, the pairs of morphs in each of the Lake Kronotskoe clades exhibit alternative heterochronic tendencies. For example, the Smallmouth morph with proportionally large eyes and blunt, rounded rostra shows hallmarks of paedomorphosis, while the sister Bigmouth morph possesses peramorphic traits such as overdeveloped mandibula. The enlarged jaws and frontonasal protrusion of the Longhead morph, as well as the drastically modulated frontonasal proportion of the Nosed 3 morph are peramorphic features in comparison to their sister lineages. Similar craniofacial adaptive changes arose during the adaptive radiation of phylogenetically distant fishes, the large African barbs (g. *Labeobarbus*; Cypriniformes; Teleostei), inhabiting Lake Tana (21). The age of the Lake Tana species flock of barbs is comparable with the Lake Kronotskoe species flock, yet genetic differences between Lake Tana morphs are comparatively subtle (61, 62), and ecomorphological differentiation is a result of heterochronic shifts presumably induced by thyroid axis alterations (63). Such similarities suggest that genetic modification of thyroid signaling may be a widespread mechanism facilitating rapid freshwater teleost adaptive radiations and provide a genetic foundation from which more specialized morphologies may be further elaborated.

In anadromous salmonids, smoltification in preparing for migration out to sea requires orchestration of hormonal and physiological switches. A hypothesis stemming from lacustrine populations is that selective pressure on the ancestral riverine charr population was relaxed upon colonization of the lake, as the newly resident population of charr became obligate freshwater residents. This initial shift in developmental programs may fall within a common program of lacustrine-locked charr that biases the direction of adaptation to generate similar forms among independent lineages and lays the foundation upon which more trophically specialized morphologies may arise. In this context, exclusion of the smoltification stage from the charr life cycle may permit lacustrine adaptive diversification. Endocrine signaling pathways, including the growth, hypothalamic-pituitary-interrenal, and thyroid axes drive physiological, morphological, and behavioral changes during smoltification (64). In landlocked salmonid populations, the growth and hypothalamic-pituitary-interrenal axes, which ordinarily induce physiological changes for a transition to seawater, are upregulated, while the thyroid axis likely maintains its developmental, physiological, and adaptive significance (65). In support of this model, thyroid hormone signaling is selectively modified during freshwater colonization and subsequent adaptive radiations of the threespine stickleback, *Gasterosteus aculeatus* (66).

The lineage-specific pattern of highly differentiating loci identified in Nosed morphs, suggests an initial developmental state of extensive modifications to thyroid signaling and craniofacial development shared among all lineages, was further refined. Lineage-specific

secondary elaborations of an initial suite of genetic variants may have catalyzed the exceptional species flock diversification in Lake Kronotskoe. The fixation of variants *in leptina2* and *slc26a10* found in Nosed morphs over other lake groups suggests further modulation to the thyroid axis in these lineages. The data are supported by findings of altered T_3 levels within these lineages as adults and the presence of a *mc4r* within an interval of excess allele sharing with the Bigmouth morph. These may have facilitated a "two-step" differentiation enabling the sympatric radiation in Lake Kronotskoe to expand beyond the two or three morph radiation paradigm.

Such repeated and parallel derivations of morphotypes across the *Salvelinus* complex suggest that there is an underlying genetic framework biasing the radiations of resident lacustrine populations. Our data suggest that shared modulation of the thyroid signaling axis in tandem with craniofacial regulators may enforce such biases. The patterns of variation identified in the Lake Kronotskoe radiation point to a fundamental genetic groundwork for craniofacial evolution and a common axis for morphological change. Our genomic data and description of this unique radiation opens the door for systematic in-depth studies of the genetic roots of morphological evolution.

Materials and Methods

Field material collection

Charrs skipping the spawning season, adults without spawning changes in color and head shape, and reproductive state, were sampled in Lake Kronotskoe, and adult riverine Dolly Varden charrs were collected in the nearest watercourse draining the opposing slope of Valaginskiy range. Blood, pectoral fin tissue and pictures were collected. Blood for thyroid hormone test was carefully collected from the caudal vessel with a Vacuette serum tube. The distal part of the right pectoral fin (finclip, 0.2-0.3 cm^2) was taken with scissors and fixed in pure alcohol for DNA analysis. Fish were photographed, treated with the antibacterial and antifungal solution (Melafix and Pimafix, API) for 30 min, and released if the fish did not display any signs of injury and/or infection in 48 hrs. All catches were carried out in accordance with the Russian Federal Act on Specially Protected Natural Areas (OOPT) N33-Φ3 14/03/1995, article 10, and Plan of the research activities of Kronotsky Nature Reserve. The procedures with fish were approved by the Bioethics Commission of the AN Severtsov Institute of Ecology and Evolution, Russian Academy of Science.

Phylochip targeted sequence enrichment design

We aimed to create a pan-Salmoniformes targeted sequence capture design that can enrich sequencing libraries for conserved genetic regions across a broad diversity of salmon genomes. This design targets protein-coding exons as well as a set of conserved non-protein coding elements (CNEs), miRNA hairpins, and ultraconservative non-coding elements (UCNEs). The majority of capture baits were derived from the Atlantic salmon genome (*Salmo salar*, ICSASG_v2)(67), with inclusion of regions from the genome of rainbow trout (*Oncorhynchus mykiss*, AUL_PRJEB4421_v1)(68) that were either not represented in the Atlantic salmon genome or were <85% identity to a capture target within the rainbow trout genome. As these fish bracket both sides of the salmon phylogeny (**Supp. Fig. 1A**), the 'Phylochip' design strategy enables DNA from the majority of salmonids target regions to be efficiently enriched using this one capture design.

As the Atlantic salmon genome was not annotated at the time of capture design, annotated coding sequences were isolated from the rainbow trout and northern pike (*Esox lucius*, GCF_000721915.2_ASM72191v2)(69) genomes. These were then identified within the Atlantic salmon genome via BLASTN (ncbi-blast-2.2.30+; parameters '-max_target_seqs 1 -outfmt 6'), and these hits used in the capture design. Genes from rainbow trout that were not identified in Atlantic salmon or that had <85% identity to the best BLAST hit within the Atlantic salmon genome were also retained in the capture design. CNEs were defined from the constrained regions in the Ensembl compara 11-way teleost alignment (Ensembl release-84)(70). To reserve space in the capture design, only CNEs ≥75bp in length were included in the capture baits. These

CNEs were extracted from the Japanese medaka (*Oryzias latipes*, MEDAKA1), three-spined stickleback (*Gasterosteus aculeatus*, BRAOD S1), and zebrafish (*Danio rerio*, GRCz10.84) genomes using Bedtools (v2.23.0) intersectBed (71). miRNA hairpins were extracted from miRbase and ultraconservative elements (UCNEs) from UCNEbase (72, 73). As with protein coding exons, these elements were identified within each reference genome using BLASTN (ncbi-blast-2.2.30+; parameters '-max_target_seqs 1 -outfmt 6'). miRNA hairpins were padded to be at least 100 bp to improve capture specificity.

From these targeted regions, the specific SeqCap EZ Developer (cat #06471684001) oligonucleotide capture baits were made in collaboration with the Nimblegen design team. Capture baits are strategically designed to standardize oligo annealing temperature, remove low complexity DNA regions and to reduce the oligo sequence redundancy. The capture design targeted sequence from 558,882 genomic regions (97,049,118 total bp) across the two salmonid genomes. This included including 460,210 protein coding exons, 93,973 CNEs, 1,082 miRNAs and 3,617 UCNEs (**Supp. Fig. 1**).

DNA extraction and preparation of sequencing libraries

Tissue from finclips was digested and genomic DNA was column purified using QIAGEN DNeasy Blood & Tissue Kit (QIAGEN 69506). Genomic DNA was extracted from finclips of 3 riverine Dolly Varden charr, 8 Bigmouth morphs, 10 Longhead morphs, 5 Nosed 1 morphs, 5 Nosed 3 morphs, 6 Smallmouth morphs, and 6 White morphs. Pools of genomic DNA were produced for each lineage such that genomic DNA from every individual in a lineage pool was equally represented. The pooled samples were sheared to a target size of 200bp in Tris-HCl EDTA shearing buffer (10mM Tris-HCl, 0.1mM EDTA, pH 8.0). Shearing was performed on a Covaris E220 ultrasonicator (duty cycle, 10%; intensity, 5; cycles/burst, 200; time, 400 seconds; temperature, 8°C) using Covaris microTUBE Snap-Cap tubes (Covaris 520045). Sequencing libraries were produced by KAPA HyperPrep Kit (Roche 07137923001) using 500ng of starting material for each library. Library preparation was conducted by following the SeqCap EZ HyperCap Workflow Version 1.2. A pre-capture amplification of 4 PCR cycles was performed. Samples were hybridized with the targeted capture library at 45°C for 24 hours. A post-capture amplification of 11 PCR cycles was conducted. Fragment size and DNA concentration were quantified using Agilent 2100 BioAnalyzer and High Sensitivity DNA Chips (Agilent 5067-4626). The BioAnalyzer also confirmed absence of daisy chaining, which is an indication of PCR overamplification. Paired-end, 150bp Illumina HiSeq sequencing was performed on a pool consisting of multiple barcoded libraries.

Trimming Adapters and Read Mapping

Illumina adapter sequences were removed from reads using Trimmomatic v0.36 (74). Trimmed and masked reads were aligned to the *Salvelinus alpinus* reference genome using NextGenMap v0.5.5 (75).

Coverage of Targeted elements

Coverage statistics were derived using the BEDTools v2.21.1 (71) coverage function. For each targeted element, the average depth of coverage was quantified.

Deriving Phylogeny

The Dolly Varden phylogeny was derived using IQ-TREE v1.6.12 (76). The input consisted of 6,833,254 nucleotide sites, including 44,157 parsimony informative variants. The ModelFinder function (77) determined the base substitution model of best fit to be a transversion model with empirical base frequencies and a proportion of invariable sites (TVM + F + I). 1,000 ultrafast bootstrap replicates (78) quantified support for the phylogeny.

Introgression Analysis

Introgression was quantified for all trios in the phylogeny using Dsuite v0.4 (34) Dtrios. Riverine Dolly Varden was specified as the outgroup. The f-branch statistic was depicted as a matrix by

taking the output from Dsuite Fbranch and running the Dsuite dtools.py script to generate a plot. Dinvestigate was used to generate sliding windows of 40 variants per window and 50% overlap (See **Supplemental File 2**).

Calculating pairwise Fst and Tajima's Pi

To quantify genetic differentiation, pairwise Fst was calculated using PoPoolation2 v1201 (79). The software package allows Fst to be quantified in sliding windows or in a genewise manner. Sequencing alignment data were converted into the mpileup format using SAMtools v1.13 (80). The PoPoolation2 program mpileup2sync (--min-qual 20) generated the sync file used as input to calculate Fst in sliding windows. Popoolation2 calculated Fst based on allele frequency (81). The PoPoolation2 program fst-sliding (--min-coverage 20 --min-count 3 --max-coverage 200) was used to calculate Fst in non-overlapping sliding windows (See **Supplemental File 1**). To assess which genes were most differentiating between populations, the function create-genewise-sync was utilized to intersect the sync file with a gtf containing all targeted regions in the *S. alpinus* genome, and filtered according to the same depth and minor allele count criteria as sliding windows analyses (See **Supplemental File 3**). Prior to filtering for depth, Fst was calculated for 59,478 genes and 22,590 CNEs. Noncoding elements were associated with putative regulatory targets by following the GREAT workflow to establish basal regulatory windows. A BED file of CNE loci was intersected with Intervals spanning 5kb upstream of and 1kb downstream from transcriptional start sites, with up to a 1 Mb extension (See **Supplemental File 4**) (82). To quantify nucleotide diversity, Tajima's Pi was calculated using PoPoolation v1.2.2 (83). The software also enables quantification of Tajima's Pi in sliding windows. The same depth criteria that were used for Fst sliding window quantifications were used to calculate Tajima's Pi in sliding windows for individual lineages.

Elisa for Thyroid hormone in blood

Serum samples were transferred to 2 ml specimen tubes and centrifuged at 12 000 g for 10 min with Eppendorf MiniSpin. Serum was then collected into Eppendorf 1.5 ml tubes and placed in in freezer at 24-26C. The total triiodothyronine (T3, bioactive form of thyroid hormone) concentration in plasma was evaluated by enzyme-linked immunosorbent assay (Monobind Total Triiodothyronine (tT3) test system, Monobind Inc, USA) and measured the hormone in accordance with the manufacture protocol using StatFax 303 Plus strip reader (Awareness Technology Inc, USA).

Acknowledgments

We thank Dr. Stephen Treaster for thoughtful conversations and technical support. This work was funded by F32 5F32DE029362 (NIDCR) to KCW, and by the Harvard Medical School Fund for Genetics of Climate Change to JMD.

References

1. A. Klemetsen, The Charr Problem Revisited: Exceptional Phenotypic Plasticity Promotes Ecological Speciation in Postglacial Lakes. *Freshwater Reviews* **3**, 49-74 (2010).
2. E. B. Taylor, The Arctic char (Salvelinus alpinus) “complex” in North America revisited. *Hydrobiologia* **783**, 283-293 (2016).
3. L. A. Lecaudey *et al.*, Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution* **124**, 82-99 (2018).
4. A. G. Osinov, A. A. Volkov, N. S. Mugue, Charrs of the genus Salvelinus (Salmonidae): hybridization, phylogeny and evolution. *Hydrobiologia* **848**, 705-726 (2021).
5. G. Markevich, E. Esin, L. Anisimova, Basic description and some notes on the evolution of seven sympatric morphs of Dolly Varden Salvelinus malma from the Lake Kronotskoe Basin. *Ecol Evol* **8**, 2554-2567 (2018).
6. A. Jacobs *et al.*, Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. *PLoS Genet* **16**, e1008658 (2020).
7. M. K. Simonsen *et al.*, Allometric trajectories of body and head morphology in three sympatric Arctic charr (Salvelinus alpinus (L.)) morphs. *Ecology and Evolution* **7**, 7277-7289 (2017).
8. O. T. Sandlund *et al.*, The Arctic Charr Salvelinus alpinus in Thingvallavatn. *Oikos* **64**, 305-351 (1992).
9. H. Nordeng, Solution to the "Char Problem" based on Arctic Char (Salvelinus alpinus) in Norway. *Canadian Journal of Fisheries and Aquatic Sciences* **40**, 1372-1387 (1983).
10. P. S. Maitland, I. J. Winfield, I. D. McCarthy, F. Igoe, The status of Arctic charr Salvelinus alpinus in Britain and Ireland. *Ecology of Freshwater Fish* **16**, 6-19 (2007).
11. A. F. Walker, R. B. Greer, A. S. Gardner, Two ecologically distinct forms of arctic charr Salvelinus alpinus (L.) in Loch Rannoch, Scotland. *Biological Conservation* **43**, 43-61 (1988).
12. E. V. Esin, E. S. Bocharova, E. A. Borisova, G. N. Markevich, Interaction among morphological, trophic and genetic groups in the rapidly radiating Salvelinus fishes from Lake Kronotskoe. *Evolutionary Ecology* **34**, 611-632 (2020).
13. B. Jonsson, N. Jonsson, Polymorphism and speciation in Arctic charr. *Journal of Fish Biology* **58**, 605-638 (2001).
14. P. F. Colosimo *et al.*, Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science* **307**, 1928-1933 (2005).
15. M. D. Shapiro, M. A. Bell, D. M. Kingsley, Parallel genetic origins of pelvic reduction in vertebrates. *Proceedings of the National Academy of Sciences* **103**, 13753-13758 (2006).
16. C. T. Miller *et al.*, cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. *Cell* **131**, 1179-1189 (2007).
17. R. W. Mitchell, W. H. Russell, W. Elliott, Mexican eyeless characin fishes, genus Astyanax: Environment, distribution, and evolution. *Spec. Publ. Mus. Texas Tech. Univ.* **12**, 1-89 (1977).
18. M. Malinsky *et al.*, Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350**, 1493-1498 (2015).
19. M. Malinsky *et al.*, Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol* **2**, 1940-1955 (2018).
20. M. Barluenga, K. N. Stölting, W. Salzburger, M. Muschick, A. Meyer, Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* **439**, 719-723 (2006).
21. L. Nagelkerke, F. A. Sibbing, The large barbs (Barbus SPP., cyprinidae, teleostei) of Lake Tana (Ethiopia), with a description of a new species, Barbus osseensis. *Netherlands Journal of Zoology 50 (2000) 2* **50** (2000).
22. G. S. Myers, The Endemic Fish Fauna of Lake Lanao, and the Evolution of Higher Taxonomic Categories. *Evolution* **14**, 323-333 (1960).

23. B. A. Levin, E. Simonov, Y. Y. Dgebuadze, M. Levina, A. S. Golubtsov, In the rivers: Multiple adaptive radiations of cyprinid fishes (Labeobarbus) in Ethiopian Highlands. *Scientific Reports* **10**, 7192 (2020).
24. E. V. Esin, G. N. Markevich, M. Y. Pichugin, Juvenile divergence in adaptive traits among seven sympatric fish ecomorphs arises before moving to different lacustrine habitats. *Journal of Evolutionary Biology* **31**, 1018-1034 (2018).
25. O. A. Braitseva, I. V. Melekestsev, V. V. Ponomareva, L. D. Sulerzhitsky, Ages of calderas, large explosive craters and active volcanoes in the Kuril-Kamchatka region, Russia. *Bulletin of Volcanology* **57**, 383-402 (1995).
26. K. J. Parsons, H. D. Sheets, S. SkÚLason, M. M. Ferguson, Phenotypic plasticity, heterochrony and ontogenetic repatterning during juvenile development of divergent Arctic charr (Salvelinus alpinus). *Journal of Evolutionary Biology* **24**, 1640-1652 (2011).
27. G. N. Markevich *et al.*, Natural Barriers and Internal Sources for the Reproductive Isolation in Sympatric Salmonids from the Lake–River System. *Evolutionary Biology* **48**, 407-421 (2021).
28. E. V. Esin, G. N. Markevich, N. O. Melnik, D. V. Zlenko, F. N. Shkil, Ambient temperature as a factor contributing to the developmental divergence in sympatric salmonids. *PLOS ONE* **16**, e0258536 (2021).
29. J. M. Daane *et al.*, Modulation of bioelectric cues in the evolution of flying fishes. *Curr Biol* **31**, 5052-5061 e5058 (2021).
30. S. Treaster *et al.*, *Nature Portfolio* 10.21203/rs.3.rs-1126287/v1 (2022).
31. J. M. Daane *et al.*, Historical contingency shapes adaptive radiation in Antarctic fishes. *Nature Ecology & Evolution* **3**, 1102-1109 (2019).
32. V. C. Mason, G. Li, K. M. Helgen, W. J. Murphy, Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome research* **21**, 1695-1704 (2011).
33. J. M. Daane, N. Rohner, P. Konstantinidis, S. Djuranovic, M. P. Harris, Parallelism and Epistasis in Skeletal Evolution Identified through Use of Phylogenomic Mapping Strategies. *Molecular Biology and Evolution* **33**, 162-173 (2015).
34. M. Malinsky, M. Matschiner, H. Svardal, Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour* **21**, 584-595 (2021).
35. A. Tingaud-Sequeira *et al.*, Functional and genetic analyses of ZYG11B provide evidences for its involvement in OAVS. *Molecular Genetics & Genomic Medicine* **8**, e1375 (2020).
36. T. C. B. Wong *et al.*, Genes regulated by potassium channel tetramerization domain containing 15 (Kctd15) in the developing neural crest. *The International journal of developmental biology* **60**, 159-166 (2016).
37. G. H. Yoon, K. Kim, D. S. Park, S. C. Choi, RNF152 negatively regulates Wnt/β-catenin signaling in Xenopus embryos. *BMB Rep* (2022).
38. S. Decherf *et al.*, Thyroid hormone exerts negative feedback on hypothalamic type 4 melanocortin receptor expression. *Proc Natl Acad Sci U S A* **107**, 4471-4476 (2010).
39. M. Rothstein, M. Simoes-Costa, Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Research* **30**, 35-48 (2020).
40. J. Fabik, K. Kovacova, Z. Kozmik, O. Machon, Neural crest cells require Meis2 for patterning the mandibular arch via the Sonic hedgehog pathway. *Biology Open* **9** (2020).
41. A. L. Evans, P. J. Gage, Expression of the homeobox gene Pitx2 in neural crest is required for optic stalk and ocular anterior segment development. *Human Molecular Genetics* **14**, 3347-3359 (2005).
42. Z. Ma *et al.*, A novel inducible mutagenesis screen enables to isolate and clone both embryonic and adult zebrafish mutants. *Sci Rep* **7**, 10381 (2017).
43. F. Castinetti *et al.*, PITX2 and PITX1 Regulate Thyrotroph Function and Response to Hypothyroidism. *Molecular Endocrinology* **25**, 1950-1960 (2011).

44. S. Richard *et al.*, A Pivotal Genetic Program Controlled by Thyroid Hormone during the Maturation of GABAergic Neurons. *iScience* **23**, 100899-100899 (2020).
45. B. Friedrichs *et al.*, Thyroid functions of mouse cathepsins B, K, and L. *The Journal of Clinical Investigation* **111**, 1733-1745 (2003).
46. F. P. Bernier *et al.*, Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome. *American journal of human genetics* **90**, 925-933 (2012).
47. E. V. Esin, G. N. Markevich, N. O. Melnik, D. V. Kapitanova, F. N. Shkil, Natural toxic impact and thyroid signalling interplay orchestrates riverine adaptive divergence of salmonid fish. *J Anim Ecol* **90**, 1004-1019 (2021).
48. Y. Yagasaki *et al.*, The role of craniofacial growth in leptin deficient (ob/ob) mice. *Orthodontics & Craniofacial Research* **6**, 233-241 (2003).
49. T. Zimmermann-Belsing, G. Brabant, J. Holst, U. Feldt-Rasmussen, Circulating leptin and thyroid dysfunction. *European Journal of Endocrinology Eur J Endocrinol* **149**, 257-271 (2003).
50. D. L. Copeland, R. J. Duff, Q. Liu, J. Prokop, R. L. Londraville, Leptin in teleost fishes: an argument for comparative study. *Front Physiol* **2**, 26-26 (2011).
51. F. N. Shkil, D. V. Kapitanova, V. B. Borisov, B. Abdissa, S. V. Smirnov, Thyroid hormone in skeletal development of cyprinids: effects and morphological consequences. *Journal of Applied Ichthyology* **28**, 398-405 (2012).
52. D. Diaczok, C. Romero, J. Zunich, I. Marshall, S. Radovick, A Novel Dominant Negative Mutation of OTX2 Associated with Combined Pituitary Hormone Deficiency. *The Journal of Clinical Endocrinology & Metabolism* **93**, 4351-4359 (2008).
53. H. Bando *et al.*, Otx2b mutant zebrafish have pituitary, eye and mandible defects that model mammalian disease. *Human Molecular Genetics* **29**, 1648-1657 (2020).
54. Y.-H. Liu, T.-C. Lin, S.-P. L. Hwang, Zebrafish Pax1a and Pax1b are required for pharyngeal pouch morphogenesis and ceratobranchial cartilage development. *Mechanisms of Development* **161**, 103598 (2020).
55. X. Liu, H. Wang, G. Li, H.-Z. Huang, Y.-Q. Wang, The function of *DrPax1b* gene in the embryonic development of zebrafish. *Genes & Genetic Systems* **88**, 261-269 (2013).
56. R. M. Viktorovskii, *Mekhanizm vidoobrazovaniya u gol'tsov Kronotskogo Ozera* (Изд-во Наука, 1978).
57. R. De-Kayne *et al.*, Hybridization and a mixture of small and large-effect loci facilitate adaptive radiation. *bioRxiv* 10.1101/2022.02.18.481029, 2022.2002.2018.481029 (2022).
58. S. D. McCormick, A. M. Regish, W. R. Ardren, B. T. Bjornsson, N. J. Bernier, The evolutionary consequences for seawater performance and its hormonal control when anadromous Atlantic salmon become landlocked. *Sci Rep* **9**, 968 (2019).
59. C. K. Tipsmark, K. J. Sørensen, S. S. Madsen, Aquaporin expression dynamics in osmoregulatory tissues of Atlantic salmon during smoltification and seawater acclimation. *Journal of Experimental Biology* **213**, 368-379 (2010).
60. E. V. Esin, G. N. Markevich, D. V. Zlenko, F. N. Shkil, Thyroid-Mediated Metabolic Differences Underlie Ecological Specialization of Extremophile Salmonids in the Arctic Lake El'gygytgyn. *Frontiers in Ecology and Evolution* **9** (2021).
61. L. A. Nagelkerke *et al.*, Shallow genetic divergence and species delineations in the endemic Labeobarbus species flock of Lake Tana, Ethiopia. *J Fish Biol* **87**, 1191-1208 (2015).
62. M. de Graaf, H. J. Megens, J. Samallo, F. Sibbing, Preliminary insight into the age and origin of the Labeobarbus fish species flock from Lake Tana (Ethiopia) using the mtDNA cytochrome b gene. *Mol Phylogenet Evol* **54**, 336-343 (2010).
63. F. N. Shkil *et al.*, Ontogenetic mechanisms of explosive morphological divergence in the Lake Tana (Ethiopia) species flock of large African barbs (Labeobarbus; Cyprinidae; Teleostei). *Russian Journal of Developmental Biology* **46**, 294-306 (2015).

64. S. D. McCormick, "5 - Smolt Physiology and Endocrinology" in Fish Physiology, S. D. McCormick, A. P. Farrell, C. J. Brauner, Eds. (Academic Press, 2012), vol. 32, pp. 199-251.
65. S. D. McCormick, A. M. Regish, W. R. Ardren, B. T. Björnsson, N. J. Bernier, The evolutionary consequences for seawater performance and its hormonal control when anadromous Atlantic salmon become landlocked. *Scientific Reports* **9**, 968 (2019).
66. J. Kitano *et al.*, Adaptive divergence in the thyroid hormone signaling pathway in the stickleback radiation. *Curr Biol* **20**, 2124-2130 (2010).
67. W. S. Davidson *et al.*, Sequencing the genome of the Atlantic salmon (Salmo salar). *Genome Biology* **11**, 403 (2010).
68. C. Berthelot *et al.*, The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* **5**, 3657 (2014).
69. E. B. Rondeau *et al.*, The Genome and Linkage Map of the Northern Pike (Esox lucius): Conserved Synteny Revealed between the Salmonid Sister Group and the Neoteleostei. *PLOS ONE* **9**, e102089 (2014).
70. J. Herrero *et al.*, Ensembl comparative genomics resources. *Database (Oxford)* **2016**, bav096 (2016).
71. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
72. S. Dimitrieva, P. Bucher, UCNEbase--a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic acids research* **41**, D101-D109 (2013).
73. A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* **39**, D152-D157 (2010).
74. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
75. F. J. Sedlazeck, P. Rescheneder, A. von Haeseler, NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791 (2013).
76. B. Q. Minh *et al.*, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530-1534 (2020).
77. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587-589 (2017).
78. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518-522 (2017).
79. R. Kofler, R. V. Pandey, C. Schlötterer, PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics (Oxford, England)* **27**, 3435-3436 (2011).
80. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
81. D. L. Hartl, A. G. Clark, A. G. Clark, *Principles of population genetics* (Sinauer associates Sunderland, MA, 1997), vol. 116.
82. C. Y. McLean *et al.*, GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).
83. R. Kofler *et al.*, PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLOS ONE* **6**, e15925 (2011).

Figures and Tables

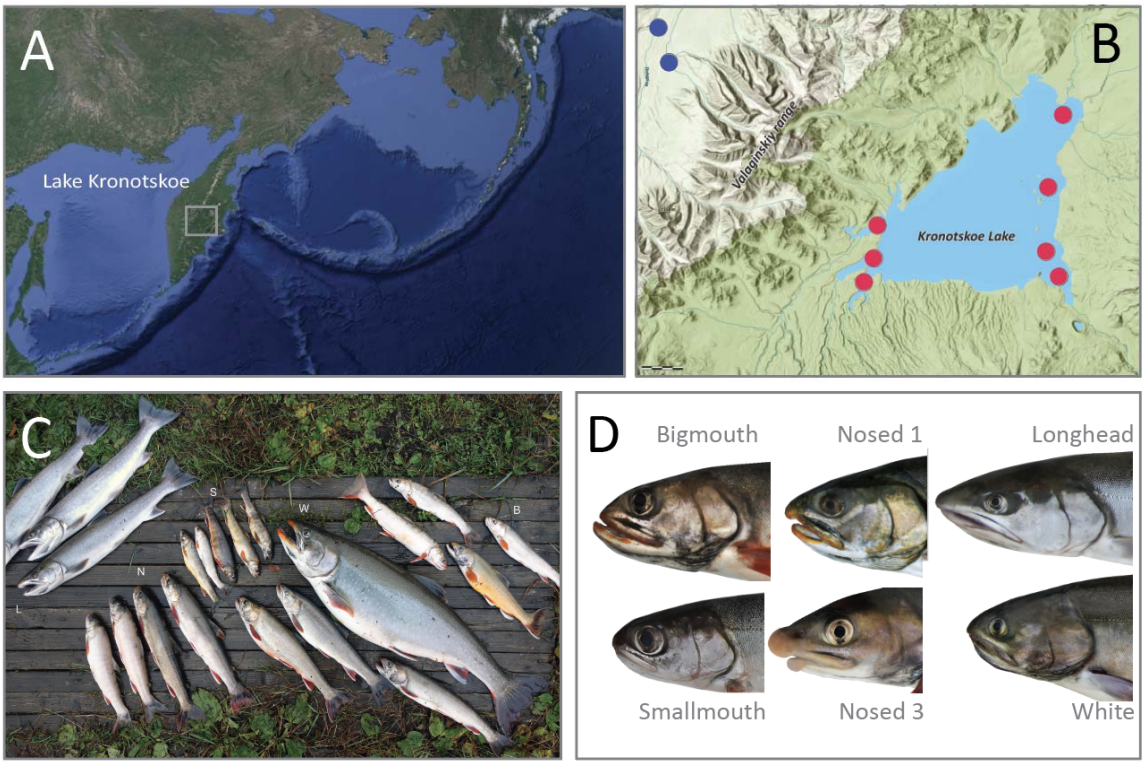


Figure 1. Lake Kronotskoe radiation of Dolly Varden Charr. (A) A map of the Kamchatka Peninsula. Lake Kronotskoe is centered within the highlighted box. **(B)** A map of Lake Kronotskoe geography and locations from which anadromous (blue dots) and resident lacustrine (red dots) Dolly Varden morphs were collected. **(C)** The Lake Kronotskoe morphs from left to right: Longhead (L), Nosed lineages (N), Smallmouth (S), White (W), Bigmouth (B). **(D)** Detailed images of representative adult heads for the six lineages sequenced.

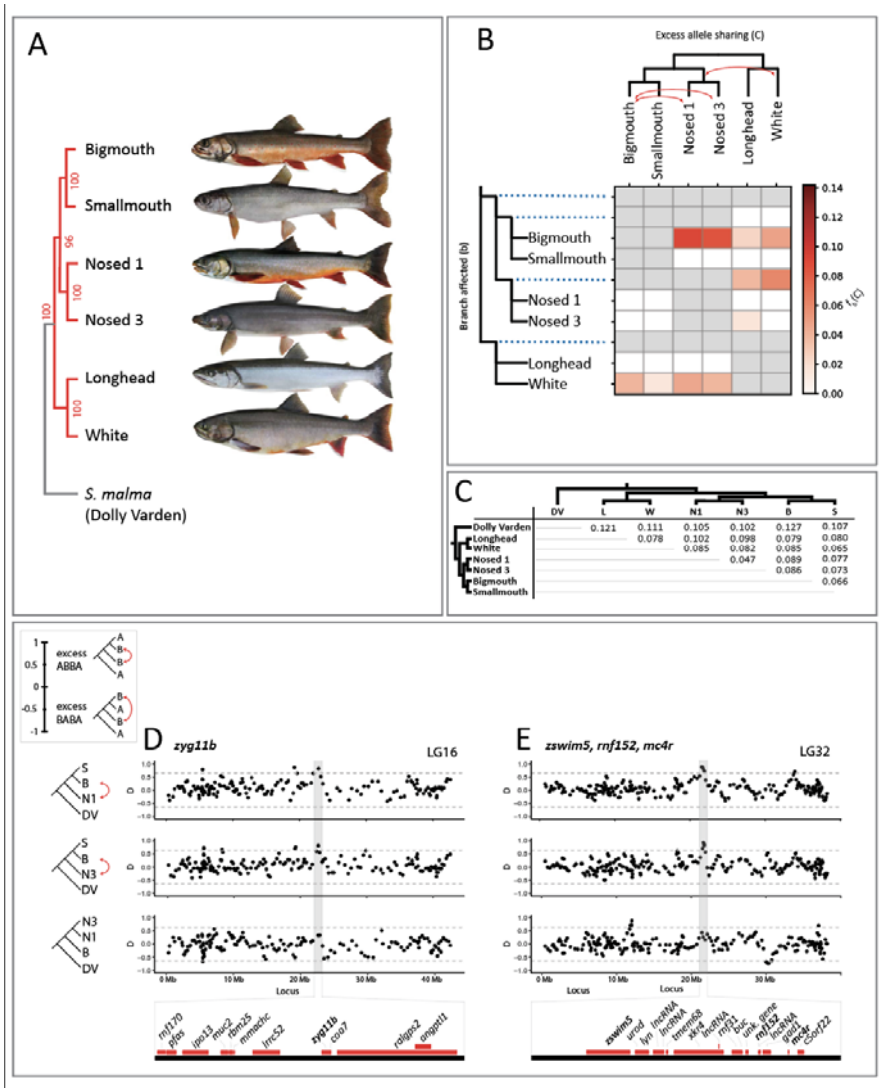


Figure 2. Phylogenetics and population differentiation among the sequenced lineages. (A) Phylogenic relationship of Lake Kronotskoe Dolly Varden lineages studied; anadromous *S. malma* from the Kamchatka River serve as the outgroup. **(B)** The branch-specific statistic f_b shows evidence for elevated gene flow between the Bigmouth and Nosed 1 morphs, the Bigmouth and Nosed 3 morphs, and the ancestor of the Nosed lineages and the White morph. The levels of admixture quantified among Lake Kronotskoe charr is much lower than quantifications of admixture among radiations which are orders of magnitude older. The color bar maximum, set to 14% reflects this differential. White cells represent combinations for which p-values are not significant. Gray cells represent arrangements which are not topologically feasible for calculating f-branch scores. **(C)** Distribution of genome-wide, pairwise Fst values calculated for non-overlapping 10kb sliding windows. See **Supplemental File 1**. **(D)** Sliding window plots of D for trios consisting of Smallmouth (S), Bigmouth (B), and Nosed 1 (N1), of S, B, and Nosed 3 (N3), and of N3, N1, and B. Positive D values indicate an excess of the ABBA pattern (red arrows), while negative values indicate an excess of the BABA pattern. The three plots show a common pattern of excess of allele sharing overlapping with *zyg11b* between B and N1 and B and N3, while there is no excess of allele sharing between B and N1 over B and N3. Horizontal lines signify 3SDs from the mean. Genes contained by the highlighted peak represented by red bars below plots. See **Supplemental File 2**. (E) three genes of interest, *zswim5*, *rnf152*, and

mc4r, possibly constitute a supergene contained by a shared peak between B and N1 and B and N3. See **Supplemental File 2**.

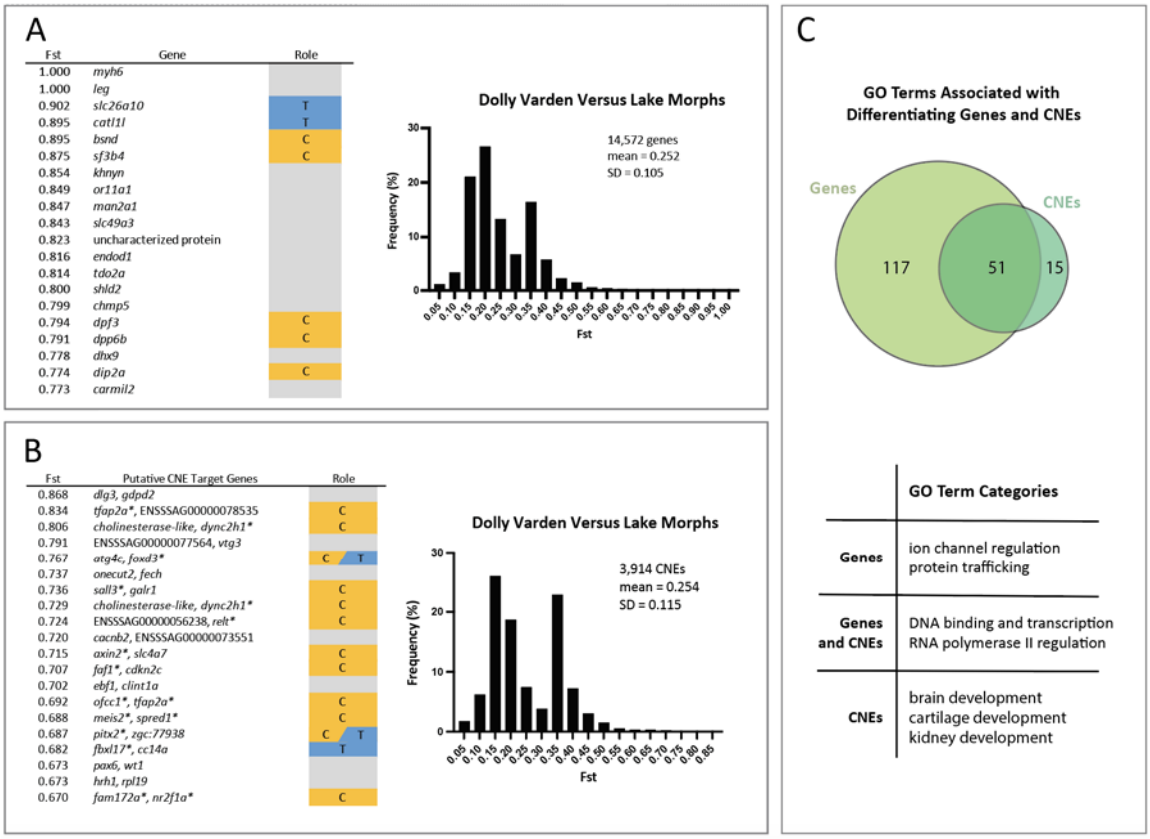


Figure 3. Coding and non-coding conserved elements differentiating anadromous Dolly Varden charr from the Lake Kronotskoe species flock. A) Table showing the top 20 genes differentiating riverine Dolly Varden from Lake Kronotskoe inhabitants along with distribution of Ft values of genes. See **Supplemental File 3**. B) Table of the top 20 differentiating CNEs and the distribution of Fst values. See **Supplemental File 4**. C) Overview of shared and unique GO terms associated with genes and CNEs with Fst > 0.5. Venn diagram values represent GO terms which occur six or more times. Table summarizes broad categories of most frequent GO term associations. See **Supplemental File 5**. The distributions and the numbers of elements represent all non-zero Fst values; orange, genes with known roles in modulating craniofacial morphology (C); blue, genes with known roles in thyroid function (T). Asterisk denotes gene classified as (C), (T), or both.

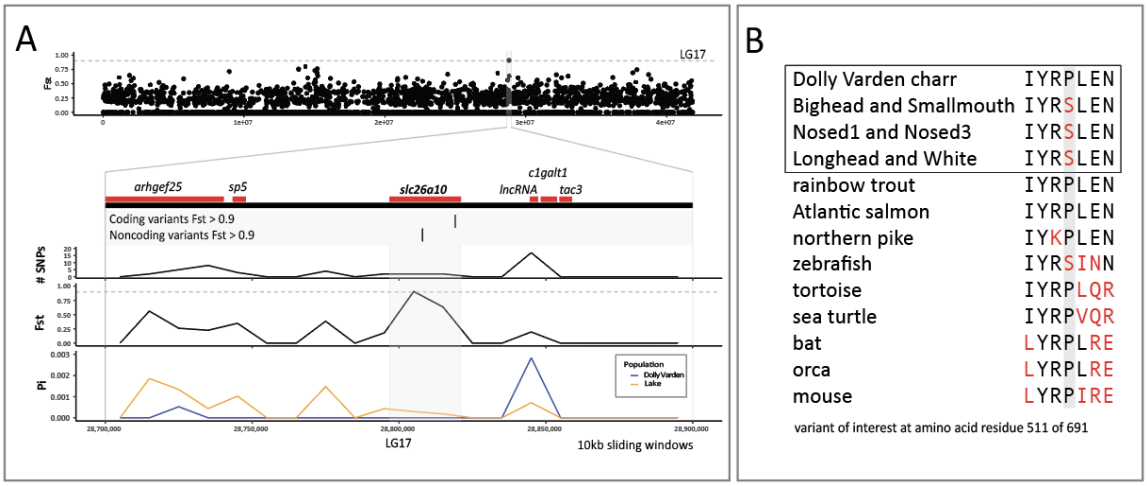


Figure 4. Differentiation of slc26a10 in pairwise comparisons between anadromous Dolly Varden charr and the Lake Kronotskoe species flock. **(A)** The slc26a10 locus shows high differentiation (Fst). The gene locus contains one highly differentiating non-coding and one coding variant. Dolly Varden slc26a10 is homologous to human pendrin (SLC26A4), a known thyroid regulator. The broader locus has low nucleotide diversity as illustrated by sliding window plots of Tajima's Pi. Horizontal dotted gray line represents Fst = 0.9. Plots are included for 10kb sliding windows along LG17, the coding elements within the broader locus, the number of variants per sliding window, Fst, and Tajima's Pi per sliding window in pairwise comparisons between riverine Dolly Varden charr and the Lake Kronotskoe species flock. **(B)** Slc26a10 contains a fixed amino acid substitution in a conserved proline that differentiates Dolly Varden charr from each of the major clades of the lacustrine morphs.

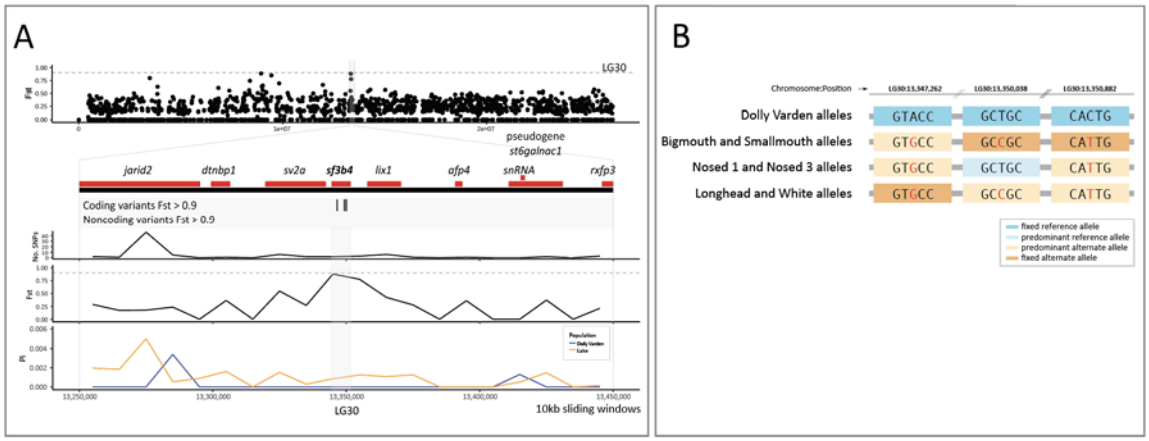


Figure 5. Fixation of variation in s*f3b4* between anadromous Dolly Varden charr and the Lake Kronotskoe species flock. **(A)** Three highly differentiating synonymous variants lie within *sf3b4*. Fst is plotted in non-overlapping 10kb sliding windows along the length of the chromosome LG30. *Sf3b4* is contained within a 200kb interval of low nucleotide diversity as illustrated by sliding window plots of Tajima's Pi. Horizontal dotted gray line represents Fst = 0.9. The broad locus encompassing *sf3b4* is shown in detail including plots of the number of variants, Fst, and Tajima's Pi in 10kb non-overlapping sliding windows. **(B)** 3 synonymous variants in *sf3b4* are fixed in lacustrine Dolly Varden charr; *dark orange,* non-reference allele fixed*; light orange,* non-reference allele predominant (alt. allele freq. > 50%) in sequence pool; *Light blue*, a lineage for which the reference allele is predominant (ref. allele freq. > 50%) at the locus.

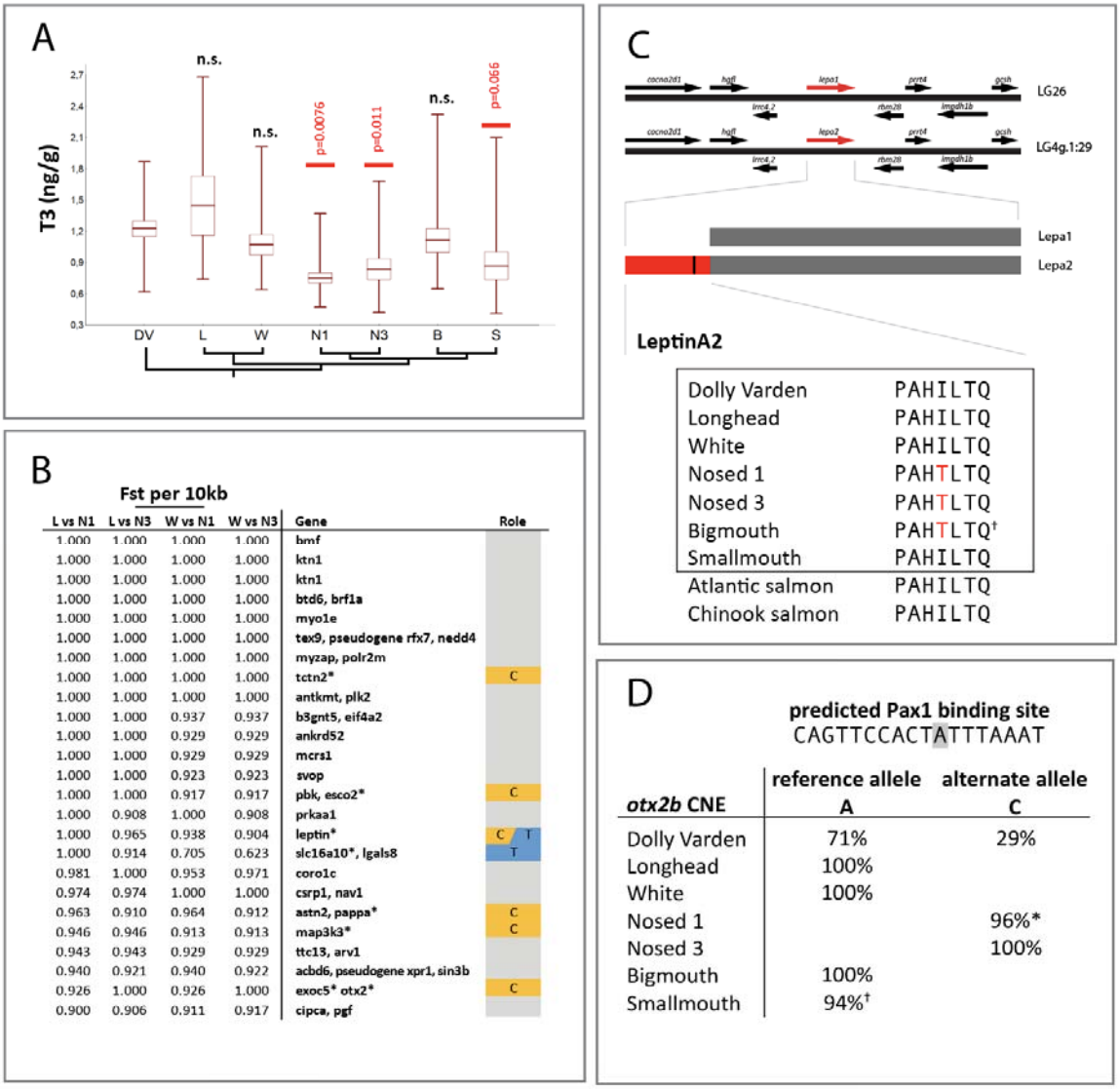
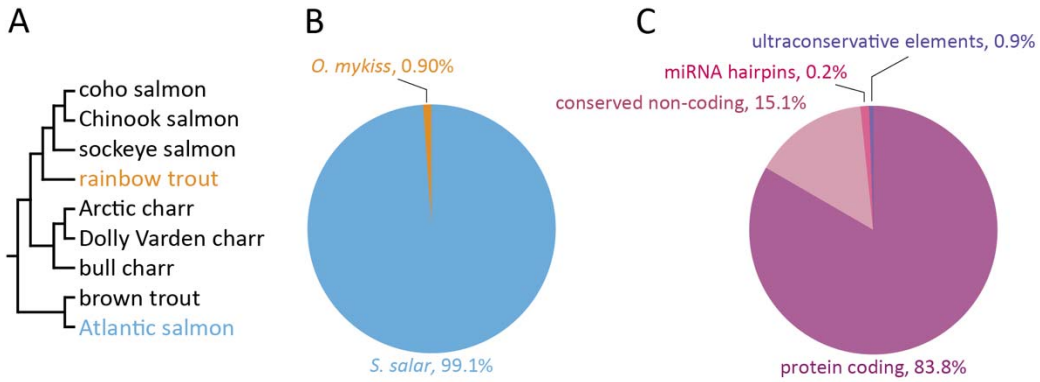
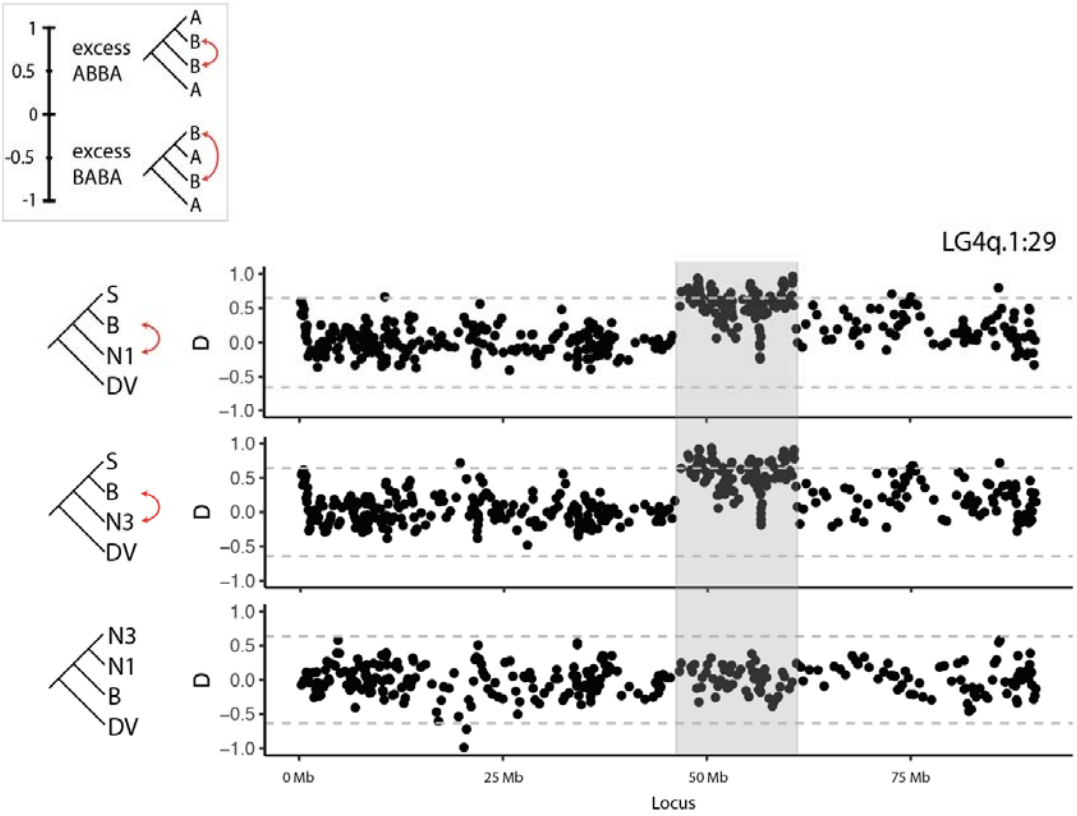


Figure 6. Thyroid hormone (T3) levels and associating highly differentiating candidates in lacustrine Dolly Varden morphs related to thyroid function and craniofacial morphology. **(A)** Serum T3 levels were significantly less in Nosed 1, Nosed 3, and Smallmouth lineages relative to anadromous Dolly Varden; mean +/- 1 SD, with boundary values indicated (min/max). **(B)** Top 20 candidate regions differentiating Nosed lineages from Longhead and White lineages. Sites uniquely differentiating Longhead morphs from Nosed 1 and Nosed 3 morphs, and White morphs from Nosed 1 and Nosed 3 morphs identify loci associated with thyroid function (blue, T) or craniofacial developmental (orange, C) modulating elements. **(C)** Schematic of Leptin A paralogues in salmonids detailing previously unknown gene *leptin a2* having a unique 66 amino acid N-terminus extension compared to its paralogues. Nosed lineages contain a fixed, non-synonymous SNP in this conserved N-terminal sequence. (**D**) An *otx2b* CNE contains fixed or highly differentiating variants within a predicted Pax1 binding domain that associates with lake morphs exhibiting significantly different thyroid signaling activities; percentage of reference and alternate allele reads indicated per morph. Low level detection of variant or reference alleles noted (1 reference allele (asterisk), 2 reads with non-reference reads (dagger).



Supplemental Figure 1. Design of pan-Salmoniformes targeted capture array. (A) Conserved bait sequences were derived from Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) reference genomes. **(B)** Relative contributions of *S. salar* and *O. mykiss* derived bait sequences to the capture array. **(C)** Classifications and relative abundances of conserved elements targeted for capture.



Supplemental Figure 2. Sliding window plots of D showing a roughly 11.8MB interval of excess allele sharing. Sliding window plots for trios consisting of Smallmouth (S), Bigmouth (B), and Nosed 1 (N1), of S, B, and Nosed 3 (N3), and of N3, N1, and B. Positive D values indicate an excess of the ABBA pattern (red arrows), while negative values indicate an excess of the BABA pattern. The three plots show a common pattern of excess of allele sharing overlapping with between B and N1 and B and N3, while there is no excess of allele sharing between B and N1 over B and N3. Horizontal lines signify 3SDs from the mean.

Supplemental Table 1. Summary of reads aligned and targeted element coverage.

Morphotype	No. Individuals Pooled	No. Reads	No. Reads Mapped	Median Depth	Mean Depth	"Conservome" Coverage 2x	"Conservome" Coverage 5x
Bigmouth	8	106,550,218	57,565,377	62 reads	48 reads	90.52%	88.83%
Dolly Varden	3	106,421,268	59,900,530	64 reads	49 reads	90.69%	89.06%
Longhead	10	70,812,872	38,817,454	41 reads	31 reads	90.09%	87.77%
Nosed 1	5	112,191,596	62,639,040	67 reads	52 reads	90.64%	89.04%
Nosed 3	5	113,476,184	62,633,235	66 reads	51 reads	90.69%	89.09%
Smallmouth	6	65,114,828	35,612,592	37 reads	28 reads	89.64%	87.09%
White	6	53,788,232	30,134,300	32 reads	25 reads	89.53%	86.62%

Supplemental Table 2. Table of significantly underrepresented GO terms. This table contains the set of significantly underrepresented GO terms among all sample libraries (Benjamini-Hochberg FDR < 0.05).

GO name	GO accession	bin size
activation of MAPK activity	GO:0000187	15
anaphase-promoting complex	GO:0005680	51
anaphase-promoting complex-dependent catabolic process	GO:0031145	41
ATP-dependent chromatin remodeling	GO:0043044	133
ATPase activity	GO:0016887	262
calcium ion binding	GO:0005509	1185
calcium, potassium:sodium antiporter activity	GO:0008273	52
carbohydrate binding	GO:0030246	111
carbohydrate metabolic process	GO:0005975	204
cell adhesion	GO:0007155	541
chemokine activity	GO:0008009	72
cullin-RING ubiquitin ligase complex	GO:0031461	23
DNA binding	GO:0003677	2972
DNA integration	GO:0015074	1571
DNA-templated transcription, initiation	GO:0006352	36
ferric iron binding	GO:0008199	24
fructose-bisphosphate aldolase activity	GO:0004332	27
hexose metabolic process	GO:0019318	34
homophilic cell adhesion via plasma membrane adhesion molecules	GO:0007156	362
iron ion transport	GO:0006826	23
isomerase activity	GO:0016853	45
MAP kinase activity	GO:0004707	52
nucleic acid binding	GO:0003676	2731
oligopeptide transport	GO:0006857	20
phosphatidylinositol metabolic process	GO:0046488	49
phosphatidylinositol phosphate kinase activity	GO:0016307	47
receptor tyrosine kinase binding	GO:0030971	28
regulation of mitotic metaphase/anaphase transition	GO:0030071	44
sensory perception of sound	GO:0007605	69
skeletal muscle fiber development	GO:0048741	44
SWI/SNF complex	GO:0016514	147
transcription coactivator activity	GO:0003713	164
transcription factor TFIID complex	GO:0005669	38
transmembrane transport	GO:0055085	1148
transmembrane transporter activity	GO:0022857	306
transporter activity	GO:0005215	218
transposition, DNA-mediated	GO:0006313	1551
ubiquitin protein ligase binding	GO:0031625	28
vesicle-mediated transport	GO:0016192	250

Supplemental Table 3. Table of D_{tree} scores, f_4-admixture ratios, and Z-scores for each of the 20 trios contained within the Lake Kronotskoe species flock. 16 trios were found to have a significant, though minimal, contribution of introgressed alleles (asterisks)(Holm-Bonferoni, FWER < 0.01). B, Bigmouth; L, Longhead; N1, Nosed 1; N3, Nosed 3; S, Smallmouth; W, White.

Pop 1	Pop 2	Pop 3	D_{tree} (%)	f_4-ratio (%)	p-value	Z-score
S	N3	W	3.9	8.6	0	13.0*
S	N1	W	3.7	8.2	0	11.4*
S	N3	L	2.8	6.3	0	11.4*
S	N1	L	2.2	4.9	1.18E-16	8.3*
S	B	N1	4.2	9.0	1.56E-15	8.0*
S	B	W	2.2	4.7	1.26E-14	7.7*
B	N3	W	1.8	4.1	2.44E-14	7.6*
L	W	N1	2.2	4.5	2.73E-13	7.3*
L	W	B	1.7	3.9	2.76E-13	7.3*
B	N1	W	1.7	3.7	2.11E-12	7.0*
S	B	N3	3.6	8.3	2.31E-11	6.7*
L	W	N3	1.8	3.8	4.65E-09	5.9*
B	N3	L	1.8	4.0	6.52E-07	5.0*
S	B	L	1.1	2.4	1.01E-04	3.9*
B	N1	L	1.2	2.6	3.59E-04	3.6*
N1	N3	L	0.7	1.5	1.13E-03	3.3*
L	W	S	0.5	1.4	2.86E-03	3.0
N1	N3	S	0.4	0.9	5.72E-02	1.9
N1	N3	W	0.2	0.5	2.79E-01	1.1
N3	N1	B	0.3	0.7	2.86E-01	1.1