Growth genes are implicated in the evolutionary divergence of sympatric piscivorous and insectivorous rainbow trout (*Oncorhynchus mykiss*)

Jared A. Grummer\(^1\), Michael Whitlock\(^1\), Patricia M. Schulte\(^1\), and Eric B. Taylor\(^{1,2}\)

\(^1\)Department of Zoology and Biodiversity Research Centre, University of British Columbia, 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
\(^2\)Beaty Biodiversity Museum, University of British Columbia, 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada

Corresponding author: Jared A. Grummer; E-mail: grummer@zoology.ubc.ca
Abstract.—

Background: Identifying ecologically significant phenotypic traits and the genomic mechanisms that underly them are crucial steps in understanding the traits associated with population divergence. We used genome-wide data to identify genomic regions associated with a key trait that distinguishes two ecotypes of rainbow trout (Oncorhynchus mykiss) – insectivores and piscivores – that coexist in Kootenay Lake, southeastern British Columbia, for the non-breeding portion of the year. “Gerrards” are large-bodied (breeding maturity at >60cm) piscivores that spawn ∼50km north of Kootenay Lake in the Lardeau River, in contrast to the insectivorous populations that are on average smaller in body size, mainly forage on aquatic insects, and spawn in tributaries immediately surrounding Kootenay Lake. We used pool-seq data covering ∼60% of the genome to assess the level of genomic divergence between ecotypes, test for genotype-phenotype associations, and identify loci that may play functional or selective roles in their divergence.

Results: Analysis of nearly seven million SNPs provided a genome-wide mean $F_{ST}$ estimate of 0.18, indicating a high level of reproductive isolation between populations. The window-based $F_{ST}$ analysis did not reveal “islands” of genomic differentiation; however, the window with highest $F_{ST}$ estimate did include a gene associated with insulin secretion. Although we explored the use of the “Local score” approach to identify genomic outlier regions, this method was ultimately not used because simulations revealed a high false discovery rate (∼20%). Gene Ontology (GO) analysis identified several growth processes as enriched in genes occurring in the ∼200 most divergent genomic windows, indicating the importance of genetically-based growth and growth-related metabolic functions in the divergence of these ecotypes.

Conclusions: In spite of their sympatric coexistence, a high degree of genomic differentiation separates the populations of piscivores and insectivores, indicating little to no contemporary genetic exchange between ecotypes. Our results further indicate that the
large body piscivorous phenotype is likely not due to one or a few loci of large effect, rather it may be controlled by several loci of small effect, thus highlighting the power of whole-genome low-coverage sequencing in phenotypic association studies.

(Keywords: genome scan, small-effect loci, $F_{ST}$, GO analysis, rainbow trout, ecotype)
Understanding the divergence of populations and the mechanisms responsible for it are important for the fields of speciation, macroevolution, ecology, and conservation (Coyne and Orr 2004). In many cases, population-level divergences are associated with character displacement and phenotypic divergence of sympatric forms (e.g., Kitano et al. 2007). Recent advancements in sequencing technologies have enabled investigations into the genomics underlying traits associated with phenotypic divergences in non-model systems. In some cases, few highly differentiated loci of large effect can be responsible for population-level phenotypic or behavioural divergence in spite of the majority of genomic differentiation being very low between populations as the result of either recent divergence and/or ongoing gene flow (e.g., Heliconius butterflies; Supple et al. 2013, Morris et al. 2019).

In fishes, this is the case for the phenotypic divergence of two forms of brown trout (Salmo trutta) – piscivores and insectivores – where Jacobs et al. (2018) identified ~10 genomic outlier regions containing several genes involved in development, growth, and gene expression and regulation. In Atlantic cod (Gadus morhua), a ~17 MB (megabase) region containing two chromosomal inversions is associated with population divergence based on migration phenotype (Kirubakaran et al. 2016). Additionally, in Pacific salmon (genus Oncorhynchus), one small genomic region has been shown to control a vital phenotype for the natural history of two species – adult migration timing (Prince et al. 2017).

Theory predicts that the distribution of allelic effects on quantitative traits should be nearly exponential, where few loci have large effects and explain most of the variance in traits, but many loci with smaller effects explain the remaining variation in the trait (Robertson 1967; Mackay 2001). Across taxa, several phenotypes are explained by many loci of small effect. For instance, hundreds of genetic variants explain only ~10% of the phenotypic variation in human height (Allen et al. 2010). And a genome-wide analysis of 50,000 SNPs (single nucleotide polymorphism) in the collared flycatcher (Ficedula
albicollis) revealed that no SNPs were significantly associated with forehead patch size, a
sexually selected trait (Kardos et al. 2016). Furthermore, Kardos et al. (2016) report that
thousands of individuals and near-complete genome sequencing are necessary to reliably
detect large-effect loci.

Identifying the genomic regions underlying ecological forms can provide insights into
population divergence, especially with regards to the insectivorous–piscivorous divergence
that has been seen in many systems. Arctic charr (Salvelinus alpinus) have repeatedly
evolved sympatric forms of zooplanktivores and piscivores on multiple continents (Skúlason
1999; Jonsson and Jonsson 2001), including Iceland’s Lake Thingvallavatn (e.g., Skúlason
et al. 1993). Additionally, a large and piscivorous form of brown trout (“Ferox” trout;
Salmo trutta) exists in Loch Rannoch, Scotland (Thorne et al. 2016). Although individuals
usually breed with others of the same morph (assortative mating; Jonsson and Jonsson
2001) thereby leading to reproductive isolation between populations, evidence from Lake
Thingvallavatn suggests that the piscivorous morph may not be genetically distinguishable
from the other morphs (Guðbrandsson et al. 2019). It remains to be seen whether these
morphs a) have a genetic basis, b) share a genetic basis in independent waterbodies, and c)
are controlled by few loci of large-effect or many loci of small-effect.

We assessed whether particular genomic regions, and their putative functions, were
associated with the phenotypic differentiation of two sympatric populations of rainbow
tROUT in Kootenay Lake of southeastern British Columbia (Taylor et al. 2019). Phylogenetic
analyses suggest that O. mykiss diverged from other Oncorhynchus species (e.g., cutthroat
tROUT and Pacific salmon) in the early Miocene, ~20 million years ago (Lecaudey et al.
2018). Oncorhynchus mykiss spans a broad latitudinal range in western North America
from central Mexico to Alaska and likely originated in the south and expanded northward
over time (Hendrickson et al. 2002; Penaluna et al. 2016). Fish populations in most of
Canada were displaced by Pleistocene glaciations and current populations are therefore
assumed to be the results of postglacial range expansions in the past \(\sim 15,000\) years (Schluter 1996; Taylor 1999). A variety of divergent life-history patterns can be found in populations across the native range of \(O.\ mykiss\), including migration patterns (resident vs. anadromous), habitat type (lake vs. stream), and diet (piscivory vs. insectivory) (Keeley et al. 2007). While genomic regions have been found that are associated with migration and habitat type (Hess et al. 2016; Prince et al. 2017), little is known about the evolutionary mechanisms promoting divergence of piscivorous and insectivorous \(O.\ mykiss\) populations.

In Kootenay Lake, a large piscivorous form (often \(> 60\) cm and \(5\) kg, but up to \(12\) kg), locally known as “Gerrards”, has an adult dietary specialization on Kokanee salmon – small, land-locked sockeye salmon (\(O.\ nerka\)) – and spawns near the abandoned town site Gerrard (thus the name “Gerrards”) on the Lardeau River that drains into the north arm of Kootenay Lake (Northcote 1973; Irvine 1978); the size of this population is believed to vary from several hundred to over 1,000 individuals, given annual census surveys. In contrast, the smaller form (typically \(\leq 50\) cm and \(\leq 2\) kg) is primarily insectivorous and made up of several populations that spawn in smaller tributaries that drain into Kootenay Lake (Northcote 1973). In spite of discrete spawning locations (Fig. 1), the piscivorous and insectivorous forms are sympatric in the lake for a majority of the year (e.g., non-breeding season) after spending their first 1–2 years separated as juveniles in their respective streams. Piscivorous individuals have significantly larger head and upper jaw sizes, likely due to their specialization on larger prey (M. Neufeld, BC Ministry of Forests, Lands, Natural Resource Operations and Rural Development, unpublished data). The parsimonious explanation for the origin of the piscivorous form is that it evolved in Kootenay Lake from the widespread insectivorous ancestor following deglaciation of the Cordilleran ice sheet \(\sim 15,000\) years ago (McPhail 2007), given its geographic restriction to this area, but this scenario has not been formally tested. Previous research has shown that populations of these ecotypes are genetically distinct at microsatellite DNA loci \(F_{ST} = \)
0.14; Taylor et al. 2019), that no interbreeding is occurring, and that phenotypic
differences between piscivorous and insectivorous forms show a high degree of heritability
(Keeley et al. 2007). Together, these studies provide evidence against phenotypic plasticity,
versus genetic mechanisms, as the source of this sympatric morphological polymorphism
(Skulason and Smith 1995).

We sought to answer three questions regarding the ecological and evolutionary
divergence of these two *O. mykiss* populations in Kootenay Lake. First, what is the level of
genomic divergence between piscivorous and insectivorous populations? Second, can we
identify genomic regions responsible for the phenotypic divergence of these two forms? And
third, is the phenotypic divergence between forms caused by few loci of large effect or many
loci of small effect?

We addressed these questions using a low-coverage whole-genome resequencing
dataset composed of millions of single nucleotide polymorphisms (SNPs) sampled from
across the genome, using the “version 1.0” *O. mykiss* reference genome (NCBI link). Based
on the natural history of these populations, we hypothesized that elevated regions of
genetic differentiation would harbour genes that fall into two classes: 1) those underlying
metabolic and/or physiological processes due to body size and growth rate differences
between the populations, and 2) genes associated with determining feeding behaviour
and/or faciocranial morphology in relation to the divergent diet between populations (e.g.,
Roberts et al. 2011; McGirr and Martin 2017).

**RESULTS**

**DNA Sequence Quality and Genomic Alignments**

Each of the three pools (including a reference insectivorous population; *n* = 80 per
population) produced ∼290 million raw read-pairs, with the majority (83-87%) passing
quality filters (Table 1). The percentage of the genome covered following quality control and coverage thresholds was 62 – 63% with mean depth of coverages at 23 – 24x (Supp. Fig. S1). Sequencing coverage varied across chromosomes, with the chromosomal ends often having higher sequencing coverage (Supp. Fig. S2). Levels of genome-wide nucleotide diversity (\(\pi\)) also showed a similar pattern with elevated levels on ends of chromosomes (Supp. Fig. S3), matching well to regions of known partial tetrasomy retained from the salmonid-specific whole-genome duplication event (Ss4rR; Berthelot et al. 2014, Campbell et al. 2019). In the dataset including all three populations, 8,140,802 SNPs met all filters and were included in analyses, whereas the Kootenay Lake piscivores-insectivores dataset had 6,896,554 SNPs. The density of SNPs varied across the genome, with areas of lower density occurring in centromeric regions (Supp. Fig. S4).

Population Differentiation

Weir and Cockerham \(F_{ST}\) estimates were highly congruent with the PoPoolation2 ("Karlsson") results, with 19 of the 20 most divergent windows shared and overall pairwise estimates very similar (results not shown); for consistency, only Weir and Cockerham \(F_{ST}\) estimates are presented. Our non-overlapping window \(F_{ST}\) analysis revealed high differentiation between all three populations (Fig. 2). The genome-wide divergence was higher between sympatric piscivores and insectivores (\(F_{ST} = 0.188\)) than between Kootenay Lake insectivores and Blackwater River insectivores (\(F_{ST} = 0.159\)), a population that is >500km (straight line) from Kootenay Lake. Although the modes of the Blackwater–Kootenay insectivore and Kootenay piscivore–Kootenay insectivore \(F_{ST}\) distributions are approximately the same (Fig. 2), the latter distribution is right-skewed, indicating a larger fraction of the genome has higher divergence in the sympatric population-pair. The Kootenay Lake piscivores and Blackwater River trout were the most divergent populations with an overall \(F_{ST} = 0.288\).
Although single-SNP-based estimates of $F_{ST}$ reached 1.0 in the empirical data (Supp. Fig. S5), the maximal window-based $F_{ST}$ estimate was 0.746 on chromosome 2 (Fig. 3). In general, the high overall level of genomic differentiation between piscivores and insectivores did not lead to obvious “islands of differentiation” surrounded by regions of low differentiation. The twenty most divergent windows (representing the most divergent 0.10% of all windows) all had $F_{ST}$ estimates > 0.609, and were spread across 13 chromosomes (Supp. Table S2).

Given a chromosome-wide error correction rate of 0.05 ($\alpha$) and thirty chromosomes in a genome (the rainbow trout “version 1” reference genome has 29), the local score test would be predicted to incorrectly identify significant regions (e.g., false positives) on 1-2 chromosomes per genome. In our simulations of only neutral loci, ~6 regions per genome on average were identified as significant outliers (5.58 for FLK and 5.50 for LK; Supp. Table S3), leading to a type I error rate of 19% for the FLK test and 18.3% for the LK test. Although FLK and LK are closely related statistics, approximately 7% of the outlier regions were significant based on the results of one test (either FLK or LK), but not the other.

According to the binomial test and false-positive rates calculated from the simulated data, we did not observe more significant local score results than expected by chance in either the FLK ($p = 0.0533$) or LK tests ($p = 0.8094$) in the empirical data, meaning that these numbers of significant genomic regions could be all false-positives. Due to the high false-positive rate from this test and weak power to differentiate neutral from “outlier” regions, we did not use the local score to identify outlier regions in our empirical data. We instead focus on divergent regions identified from the $F_{ST}$ analysis.

**Gene Functions in Highly Differentiated Genomic Regions**
The rainbow trout genome has annotations for 42,884 protein-coding genes and 71,223 mRNAs. After filtering, our dataset had sequence data from 90% of all annotated genes (38,532) and 93% of all mRNA transcripts. The gene ontology analysis of genes \((n = 47)\) in the 20 windows with the highest \(F_{ST}\) estimates did not produce any statistically significant results. GO analyses of the four bins containing \(F_{ST}\) estimates > 0.2, 0.3, 0.4, and 0.5 resulted in 95, 72, 22, and 19 GO terms/processes, respectively (Fig. 4; Supp. Figs. S6–8). Whether the target set of genes was included in the background set made no difference on which GO terms were identified (results not shown). Although GO terms for behaviour and growth are present in the bin of least divergent loci \((F_{ST} > 0.2);\) Supp. Fig. S6), they are among 90+ other biological processes. However, we observed behavioural processes in the \(F_{ST} > 0.4\) bin (among 21 other processes), and more noticeably, growth and developmental growth in the bin of most divergent loci as two of 19 processes (Fig. 4; Table 2). Furthermore, the GO term for retinal metabolism, which includes the regulation of genes involved in development and growth, was also enriched in genes in the most divergent genomic regions. Other significantly enriched processes in the most divergent genomic regions \((F_{ST} > 0.5)\) included blood vessel formation (“platelet-derived growth factor receptor signaling pathway”), cellular-level assembly of the cytoskeleton (“cytoskeleton organization”), and neuron cell-cell adhesion.

In searching annotations of the five most divergent \(F_{ST}\) windows, serine/threonine-protein kinase BRSK2-like was found on chromosome two in the window with the highest \(F_{ST}\) estimate (Supp. Table S4). This gene functions to regulate insulin secretion in fish, potentially playing a role in differential growth and development between ecotypes. Lastly, based on gene annotations, we observed no support for the functional importance of genes associated with morphological feeding adaptations between these two populations.
Population divergence resulting from divergent selection on distinct phenotypes has been documented in many cases (Schluter and McPhail 1992; McKinnon et al. 2004; Skúlason et al. 2019). The genetic mechanisms underlying ecologically relevant traits, however, are generally not well understood. Here, through a dataset consisting of millions of genome-wide SNPs, we identified divergent loci involved with cell growth and differentiation that are associated with the divergence of two ecotypes of rainbow trout – piscivores and insectivores. The genetic control of these ecotypes does not seem to be controlled by a few loci of large effect, but rather by many genomic regions that are highly diverged between the populations. Our results further indicate that these two populations are moderately to highly divergent from one another with a genome-wide $F_{ST}$ estimate of 0.18, suggesting little to no gene flow between these populations. These results are in agreement with previous research with microsatellite DNA loci that has shown that the two phenotypes are quite divergent from one another ($F_{ST} = 0.14$; Taylor et al. 2019), and that these ecotypes are largely genetically controlled (Keeley et al. 2007).

**Differential growth implicated in divergence of sympatric ecomorphs**

Rainbow trout, as well as salmon (both Atlantic and Pacific species), are economically important species in the aquaculture industry that produces meat for human consumption. The Food and Agriculture Organization of the United Nations estimated an average annual production of $\sim$800,000 tonnes and >$3 billion from world aquaculture production of *O. mykiss* alone (Fisheries 2017). Accordingly, much research has been done to determine the genetic mechanisms determining growth and muscle mass (e.g., “fillet weight”) in these species (e.g., Salem et al. 2012; Yoshida et al. 2017). Genetic association studies indicate that body weight in *O. mykiss* is a polygenic trait, with the most important genomic
regions typically explaining <5% of the genetic variance (e.g., Gonzalez-Pena et al. 2016; Neto et al. 2019). Two previous studies on *O. mykiss* have identified loci on chromosomes 21 (Neto et al. 2019) and 27 (Gonzalez-Pena et al. 2016) that explain 2.5 and 1.7%, respectively, of the genetic variance in body weight. One of the 20 most divergent $F_{ST}$ windows we identified on chromosome 21 is less 150kb from the region identified by Neto et al. (2019).

We hypothesized that genes either involved with growth and related metabolic processes or faciocranial morphology and feeding behaviours would underly the divergence of sympatric rainbow trout ecomorphs in Kootenay Lake. Identifying single SNPs or genes in the most divergent genomic regions as causal variants related to these traits is difficult given the high genome-wide divergence we estimated between insectivorous and piscivorous populations. Nonetheless, a gene that plays a role in the regulation of insulin secretion – serine/threonine-protein kinase BRSK2-like (100% similar to 987 bases of mRNA sequence) – was found in the most divergent $F_{ST}$ window (Chen et al. 2012). In teleost fishes, insulin is involved in somatic growth, reproduction, and development (Wood et al. 2005; Garcia de la Serrana and Macqueen 2018), indicating that it could play a key role in differential growth between these two populations. This was the only gene of potentially large effect that we found in support of our hypothesis.

In addition to inspecting individual genes for their potential roles in population-level divergence, we also looked at sets of genes and their roles in emergent biological functions through gene ontology (GO) analyses. We hypothesized that biological functions and processes likely playing a role in the divergence of these ecomorphs would make up a greater proportion of the more strongly differentiated loci. In the least divergent bin of genomic regions that we analyzed ($F_{ST} \geq 0.2$), 238 GO terms were found spanning a variety of biological functions and processes (Supp. Fig. S6). In contrast, only 19 significantly enriched GO terms occurred in the most divergent genomic regions ($F_{ST} \geq 0.5$), and growth
functions accounted for more than half (10/19) of these GO terms. Interestingly, under the same diet, piscivores’ growth and standing metabolic rates during ontogenetic development significantly increase relative to those of insectivores (Monnet et al. 2020), corroborating our genetic results here. The gene ontology results indicate that the piscivorous phenotype is controlled by many loci of smaller effect and provide support to our hypothesis that growth plays a key role in leading to ecological divergence between populations.

Beyond the “growth” and “developmental growth” GO terms, gene regions associated with “retinal metabolism” were also significantly enriched in the most divergent $F_{ST}$ bin. Retinoids (including retinal and other vitamin A derivatives) have two physiological functions – visual pigments in the eye and hormonal retinoids that regulate the expression of target genes involved in embryonic development, differentiation, and postnatal growth (Wolf 1984; Mangelsdorf 1994). Regulated amounts of retinoic acid and retinol are needed, particularly within mammalian embryonic tissues, for normal cell differentiation, proliferation, and morphogenesis (Morriss-Kay 1997). Though less is known about the roles of these molecules in fishes, it is conceivable that they play analogous physiological roles and therefore could help explain some of the morphological differences between insectivores and piscivores.

Although not part of our search strategy, we identified via BLAST a combination of four transforming growth factor-β family genes that were present in multiple, un-linked chromosomal regions in the top 20 $F_{ST}$ windows. We were not able to confirm, however, that these sequences indeed are from these genes (versus an unidentified repeating element in the rainbow trout genome) and therefore code for these growth proteins. Interestingly, a recent study examining the evolutionary origin of piscivory in brown trout (Salmo trutta) also identified transforming growth factor β genes associated with the evolution of piscivory (Jacobs et al. 2018). It is possible that the unidentified fragments in our BLAST searches represent paralogs or isoforms of proteins in the TGF-β superfamily. The common
ancestor of all salmonids underwent a whole genome duplication event approximately 125 million years ago (“salmonid-specific fourth vertebrate whole-genome duplication, Ss4R”; Macqueen and Johnston 2014), in addition to a teleost-specific whole genome duplication event approximately 300–350 million years ago (Vandepoele et al. 2004; Near et al. 2012). Therefore, the current rainbow trout genome assembly may be lacking annotations of growth factor paralog isoforms that we have potentially identified in outlier regions. Genes controlling morphological traits are not well annotated as well, potentially explaining why we did not find any such genes significantly diverged between populations. Furthermore, this could be because our approach for detecting outliers is not optimal for detecting loci of small effect.

Evolutionary divergence of sympatric O. mykiss forms

Populations of piscivores and insectivores have been sympatric in Kootenay Lake for at most ~15,000 years, since the end of the last glacial maximum. It is not known whether these populations diverged in sympatry or allopatry and subsequently established sympatry, but it is assumed that they evolved in Kootenay Lake and did not originate there via human introduction (some human-mediated introductions of Gerrard trout have occurred to other areas for angling purposes). The high divergence we estimated between these populations ($F_{ST} = 0.18$) implies a long period of isolation, potentially pre-dating postglacial colonization. However, population divergence estimates between trout populations are often high, even for geographically proximate populations (e.g., strong IBD or IBE [isolation-by-environment]; Tamkee et al. 2010). Conversely, high genome-wide $F_{ST}$ estimates can result from population bottlenecks/founder effects that cause reductions in genetic diversity in either or both populations. Although Kootenay Lake is large, the piscivorous population is small with an estimated census size of ~300–1,000 spawning individuals recorded annually since 1961 (Ltd 2016). Indeed, we observed on average lower
levels of nucleotide diversity in piscivores than in insectivores across most of the genome (Supp. Fig. S3). However, the higher levels of nucleotide diversity we observed in insectivores may be because this ecotype was sampled from multiple populations around Kootenay Lake, which likely increased estimates of $\pi$.

With a high overall genome-wide divergence between ecotypes, we did not find clear candidate “genomic islands” harbouring genes potentially involved in causing population divergence. In some cases, a suite of genes involved in local adaptation or divergence of phenotypes may be found linked together in a non-recombining block, for instance in an inverted chromosomal region (e.g., Kirubakaran et al. 2016; Pearse et al. 2019). However, we did not find any such candidate genomic regions. While trying to identify genomic outliers, using only $F_{ST}$ estimates could be misleading given correlations between divergence and other features such as genetic diversity (Cruickshank and Hahn 2014) and recombination (Booker et al. 2020). Without a recombination map for wild rainbow trout, we are unable to account for this correlation in our analyses. Nevertheless, the excess enrichment of genes associated with growth in the most highly differentiated genomic regions argues against recombination being the main factor driving the overall pattern of genomic divergence in this dataset.

The “pace-of-life syndrome” has often been invoked to explain the divergence of closely related populations and species (Ricklefs and Wikelski 2002; Wikelski et al. 2003; Réale et al. 2010). The hypothesis is that closely related populations or species should differ in physiological traits that have co-evolved with the particular life histories of each population/species. Thus, ecology can have a direct impact on the evolution of particular life-history strategies (Roff 1993). Growth was one of the functions enriched in the most divergent loci ($F_{ST}> 0.5$) in our gene ontology (GO) analyses. Furthermore, Monnet et al. (2020) discovered that piscivores have higher growth and standing metabolic rates than insectivores, and were behaviourally more bold (e.g., took less time to explore novel
habitat). Results from both our study and Monnet et al. highlight the importance of growth traits in differentiating these ecotypes, thus producing a hypothesis regarding the functions of these genes that can be tested in an integrated physiological-genetic study.

**CONCLUSIONS**

Through the analysis of pool-seq data, we demonstrated that two feeding ecotypes of rainbow trout that are sympatric for much of their life spans in Kootenay Lake are highly genetically differentiated from each other. The ecological and evolutionary distinctiveness of these ecotypes – including contrasting diets, non-overlapping breeding sites, distinct morphologies, and divergent genomes – highlights the critical role of managing them separately to promote their persistence into the future. Our results demonstrate an association between genes controlling growth and highly differentiated genomic regions, indicating a potential functional role of these loci in the divergence of these two rainbow trout feeding ecotypes. Our results inform future studies that should perform transcriptomic analyses of these growth genes to evaluate gene expression during piscivorous *O. mykiss* development and the ontogenetic shift that accompanies a changing diet from primarily invertebrates to fish (*Northcote 1973*). Along with previous research, our study indicates the potential role of growth genes in causing phenotypic and ecological divergence of natural populations.
Acknowledgments

We thank S. Narum and the Columbia River Inter-Tribal Fish Commission for DNA library preparations, S. Northrup and the Freshwater Fisheries Society of BS for access to tissues, and S. Micheletti for his troubleshooting assistance with PoolParty. We would also like to give a big thanks T. Booker for his assistance with the Local Score simulations in SLiM.

Ethics Statement

All sampling was conducted under scientific permit and was approved by the University of British Columbia Animal Care and Use committee (UBC-AC) under protocol number A16-0329. Permits issued under the UBC-AC are reviewed by expert committees to ensure ethical treatment of animals and were performed under the Canadian Committee on Animal Care (CCAC) Good Animal Practice certificate. This certificate demonstrates full adherence to CCAC’s internationally recognized standards of animal ethics and care in research. This included collection of non-lethal samples and precautions to reduce stress and ensure adequate recovery after release.

Consent for Publication

Not applicable

Data Availability

- Raw, un-edited DNA Sequences are available in NCBI’s Short Read Archive, BioProject ID PRJNA695019.
- Sampling locality information is available in the Supplementary Table S1.
- Scripts for analysis are available in the Dryad repository (Dryad link; https://datadryad.org/stash/share/jYk11SUP6azAZKMNbbdwt3JRZLe-fVdtmLeTby0vbTQ).

Competing Interests

The authors declare that they have no competing interests.
Funding
The study was funded by a Genome Canada Large-Scale Applied Research Project in Natural Resources and the Environment (Project code 242RTE), NSERC grant no. 81696 to EBT, and NSERC grant no. RGPIN-2016-03779 to MCW.

Authors’ Contributions
JAG designed the study, developed analysis scripts, performed analyses, and wrote the manuscript. MCW, PMS, and EBT designed the study and edited the manuscript. All authors read and approved the final manuscript.
METHODS

Sampling and DNA Extraction

Fish were sampled from three separate populations: 80 piscivores from the Lardeau River (spawning site ~50km north of Kootenay Lake, BC), eight insectivores each from 10 tributaries immediately surrounding Kootenay Lake (all 80 individuals combined together as a single “insectivore” population), and 80 insectivorous rainbow trout from the Blackwater River of central British Columbia (Fig. 1; Supp. Table S1). The Blackwater population was used as an outgroup for identifying significantly differentiated genomic regions (see below). Spawning adult piscivores were collected by angling in the spring, with some piscivore juveniles collected in the late summer/fall in the Lardeau River. Kootenay Lake insectivores were collected via electrofishing in the late summer/early fall. Fish were collected during the spawning season to facilitate population identification. Whole genomic DNA was extracted from fin tissues, or in a few cases, dried fish scales, with a Qiagen DNeasy extraction kit (Qiagen, Valencia, CA) and quantified with a Qubit fluorometer (Life Technologies, Carlsbad, CA). DNA from 80 individuals per population was combined into a single pool (for a total of three pools) in equimolar ratios for preparation of pool-seq libraries (Futschik and Schlötterer 2010), then libraries were sequenced with 125 bp paired-end reads on an Illumina HiSeq 2500 at the Génome Québec Innovation Centre of McGill University.

Bioinformatic Dataset Assemblies

Raw sequence data (fastq format) were processed using the PoolParty pipeline (Micheletti and Narum 2018) as follows. We first used FastQC (Andrews 2010) to assess raw read
sequence quality. Raw paired-end reads were trimmed using the trim-fastq.pl script (part of PoPoolation v1; Kofler et al. 2011a) based on a Phred quality cutoff score of 20 and trimmed sequence length >50 bp. We then used BWA (Li and Durbin 2009) to align reads to a reference O. mykiss genome (NCBI link), then Samblaster (Faust and Hall 2014) and SamTools (Li et al. 2009) were used to remove PCR duplicates and ambiguously aligned reads (e.g., reads with a low mapping quality), respectively. Reads were then sorted by coordinate and removed if unpaired in Picard Tools (v2.4.1; http://broadinstitute.github.io/picard/) and Samtools. Lastly, indel regions were then identified using the PoPoolation2 (Kofler et al. 2011b) identify-genomic-indel-regions.pl script, and SNPs were discarded if within 5 bp of the indel regions using the filter-sync-by-gtf.pl script in PoPoolation2. SNPs were retained that had 1) two alleles, 2) a minor allele frequency (MAF) cutoff >0.05 globally (e.g., across all populations), 3) at least two copies of the minor allele at a locus (to mitigate the effect of sequencing errors), and 4) a sequencing depth of $15 \leq x \leq 50$ (data filtering script available in the Dryad repository). The minimum bound of 15x increases our confidence in SNP calls, while the upper bound of 50x was established to eliminate paralogs while retaining true homologs.

**Population Differentiation**

We used the fixation index ($F_{ST}$) to estimate population differentiation with a window-based approach. We calculated $F_{ST}$ using two different approaches: the method of Weir and Cockerham (1984) (R script available in the Dryad repository), and the “–karlsson” option within PoPoolation2. Both $F_{ST}$ methods were used on two datasets, one containing only Kootenay Lake piscivores and insectivores for identifying outlier regions, and a second dataset containing the Kootenay Lake samples along with Blackwater River samples, to estimate pairwise genome-wide differentiation between the three populations. $F_{ST}$ was calculated for 100,000 bp windows (along the reference genome) with a step-size of
100,000 bp (e.g., non-overlapping windows), giving ~19,500 $F_{ST}$ estimates across the genome. Results by genomic position were visualized as a Manhattan plot in the R package *qqman* (Turner 2017).

**Identifying Genomic Regions of High Differentiation**

Two approaches were used to identify genomic regions that were highly differentiated between piscivores and insectivores and therefore potentially responsible in underlying their phenotypic divergence. The first method was with a window-based $F_{ST}$ approach as described above (Weir & Cockerham 1984 calculation, 100,000 bp non-overlapping windows). This is a reliable indicator of locus-specific estimates of population divergence, however, determining genomic outliers based solely on $F_{ST}$ estimates can result in false-positives that may arise in neutral loci due to other processes such as recombination rate variation (Booker et al. 2020). We focused on windows with the highest 0.1% of $F_{ST}$ estimates ($n = 20$) as a first step in identifying the genomic regions that may be associated with the phenotypic differentiation of these two ecotypes. Non-overlapping windows were used because a sliding-window approach leads to non-independent tests from overlapping windows with no clear statistical adjustment for multiple comparisons.

As a complementary approach to outlier identification, we also used the “local score” approach developed by Fariello et al. (2017). The local score method identifies genomic regions of high differentiation between populations by using a Lindley process to identify stretches of adjacent/linked SNPs that show significant differentiation as determined by single-locus analyses. This method has the benefit over a fixed window approach in that the outlier region bounds are not user-specified and are instead determined during analysis. As in the $F_{ST}$ window-based analysis, this approach takes advantage of linkage by using divergence information from adjacent sites and is therefore superior to single-locus outlier detection methods. The local score method implements a chromosome-wide error rate
correction, meaning $\alpha$ fraction (e.g., 0.05) of the chromosomes in the genome will produce false positives. We used the significance of LK (Lewontin-Krakauer) and FLK (an extension of the LK test using a population kinship matrix “$F$”; Bonhomme et al. 2010) single-locus test results expressed as $p$-values that were calculated in R scripts provided by Bonhomme et al. (2010), as the input for the local score analysis. We considered loci as outliers at the 0.05 significance level.

One potential downside to the local score approach is that the true type I error rate (e.g., false-positive rate) has not been previously reported. We therefore estimated it through simulations mimicking our empirical dataset as follows. A single ancestral population of 1,000 individuals evolved for 1,000 generations and then split into three populations, each consisting of 1,000 individuals. These three populations evolved independently (e.g., no migration) for 2,500 more generations (generating an $F_{ST}$ estimate of $\sim0.18$ between all three populations) at which point an average of 23 chromosomes (from a truncated distribution of 15–50 and $\sigma = 5$) were sampled from each population, based on the average sequencing depth per population from the empirical data (see Results). Each chromosome was 75 MB in length and simulated in SLiM v3 (Haller and Messer 2019; Haller et al. 2019) with a recombination rate of $1.5 \times 10^{-8}$ (chromosome was $\sim1$ Morgan in length) and mutation rate of $2 \times 10^{-7}$ to generate $\sim15$ million SNPs per chromosome after implementing a MAF filter (> 0.05); only neutral mutations were introduced. Thirty chromosomes were then combined into a “pseudo-genome” replicate to mimic the empirical dataset (29 chromosomes), from which allelic frequencies were calculated that were then used in FLK and LK analyses using the R scripts provided in Bonhomme et al. (2010). Local score tests were then performed on the FLK and LK results of each genomic replicate using the R scripts provided by Fariello et al. (2017) and results were examined at the 0.05 significance level, per genome. One hundred pseudo-genome replicates were simulated and analyzed (e.g., 3,000 chromosomal replicates). To determine if significantly
more genomic regions were identified in our empirical data than expected given the false discovery rate estimated in our simulations (see Results below), we ran an exact binomial test in R with significance determined at the $p \leq 0.05$ level. The null hypothesis is that we expect as many false positives in the empirical data as are found in the simulations.

Functions of Divergent Loci

We performed gene ontology (GO) enrichment analyses to identify putative functions of genes in the most divergent genomic regions between piscivores and insectivores. The GO database is designed to be species-neutral, and therefore annotations were transferred from *Homo sapiens* to *O. mykiss* (Primmer et al. 2013). We looked at gene function in two sets of divergent loci. The first set contained the genes in the 20 most divergent $F_{ST}$ windows ($n$ genes = 45). And secondly, by separating our data into four bins based on estimates from our window-based $F_{ST}$ analyses: windows with $F_{ST}$ estimates > 0.2, 0.3, 0.4, and 0.5. These bins represent the 37% ($n$ windows = 7,219), 14% (2,737), 4.4% (854), and 0.9% (187) most divergent regions in the genome, respectively. We organized our data into bins based on divergence estimates because we expected the loci associated with biological processes most likely involved in the divergence of these populations to emerge in the more highly differentiated genomic regions (e.g., the genomic bins with higher $F_{ST}$ estimates) as compared to the less differentiated genomic regions (genomic bins with lower $F_{ST}$ estimates). According to our hypothesis, the most divergent genomic regions would contain genes enriched for processes related to growth and metabolism.

We used bedtools (Quinlan and Hall 2010) to extract gene and protein annotations from the available rainbow trout reference genome. We used the GOrilla online platform (Eden et al. 2007, 2009) to identify significantly enriched biological processes in differentiated genomic regions with two slightly different methods: 1) including the target gene set (all loci in an $F_{ST}$ bin) in the background set (all genes in the rainbow trout
genome), vs. 2) excluding the target gene set from the background set. GOrilla uses a minimum hypergeometric (mHG) score to assign the significance of a term occurring in the target set vs. background set of genes. The mHG score reflects the surprise of seeing a particular GO term in the target set compared to its probability of occurrence in the background set, under the null assumption that all GO terms in the background set occur with equal probability. An exact \( p \)-value of this score corrected for multiple testing is then calculated. All GO analyses were performed with a significance threshold of \( p \leq 0.001 \) (as recommended by the authors). Gene ontology analysis results were visualized with Revigo (Supek et al. 2011) as “treemaps”, where the size of each functional category was scaled by its \(-\log_{10} p\)-value (e.g., a smaller \( p \)-value equals a larger tile size). Beyond the GO analyses, we also used annotations of genes in the five most divergent \( F_{ST} \) windows to identify plausible candidates underlying the phenotypic divergence of insectivorous and piscivorous forms.

* References

Allen, H. L., K. Estrada, G. Lettre, S. I. Berndt, M. N. Weedon, F. Rivadeneira, C. J. Willer, A. U. Jackson, S. Vedantam, S. Raychaudhuri, et al., 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832–838.

Andrews, S., 2010. Fastqc: a quality control tool for high throughput sequence data.

Berthelot, C., F. Brunet, D. Chalopin, A. Juanchich, M. Bernard, B. Noël, P. Bento, C. Da Silva, K. Labadie, A. Alberti, et al., 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nature communications 5:1–10.
Bonhomme, M., C. Chevalet, B. Servin, S. Boitard, J. Abdallah, S. Blott, and M. SanCristobal, 2010. Detecting selection in population trees: the lewontin and krakauer test extended. Genetics 186:241–262.

Booker, T. R., S. Yeaman, and M. Whitlock, 2020. Variation in recombination rate affects detection of outliers in genome scans under neutrality. Molecular ecology 29:4274–4279.

Campbell, M. A., M. C. Hale, G. J. McKinney, K. M. Nichols, and D. E. Pearse, 2019. Long-term conservation of ohnologs through partial tetrasomy following whole-genome duplication in salmonidae. G3: Genes, Genomes, Genetics 9:2017–2028.

Chen, X.-Y., X.-T. Gu, H. Saiyin, B. Wan, Y.-J. Zhang, J. Li, Y.-L. Wang, R. Gao, Y.-F. Wang, W.-P. Dong, et al., 2012. Brain-selective kinase 2 (brsk2) phosphorylation on pctaire1 negatively regulates glucose-stimulated insulin secretion in pancreatic β-cells. Journal of Biological Chemistry 287:30368–30375.

Coyne, J. and H. Orr, 2004. Speciation. Sunderland, MA P. 281.

Cruickshank, T. E. and M. W. Hahn, 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular ecology 23:3133–3157.

Eden, E., D. Lipson, S. Yoge, and Z. Yakhini, 2007. Discovering motifs in ranked lists of dna sequences. PLoS computational biology 3:e39.

Eden, E., R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, 2009. Gorilla: a tool for discovery and visualization of enriched go terms in ranked gene lists. BMC bioinformatics 10:48.

Fariello, M. I., S. Boitard, S. Mercier, D. Robelin, T. Faraut, C. Arnould, J. Recoquillay, O. Bouchez, G. Salin, P. Dehais, et al., 2017. Accounting for linkage disequilibrium in
genome scans for selection without individual genotypes: the local score approach. Molecular ecology 26:3700–3714.

Faust, G. G. and I. M. Hall, 2014. Samblaster: fast duplicate marking and structural variant read extraction. Bioinformatics 30:2503–2505. URL +http://dx.doi.org/10.1093/bioinformatics/btu314.

Fisheries, F., 2017. Fishery and aquaculture statistics 2017/fao annuaire. statistiques des pêches et de l’aquaculture 2017/fao anuario. estadísticas de pesca y acuicultura 2017. Global aquaculture production statistics for the year .

Futschik, A. and C. Schlötterer, 2010. The next generation of molecular markers from massively parallel sequencing of pooled dna samples. Genetics 186:207–218.

Gonzalez-Pena, D., G. Gao, M. Baranski, T. Moen, B. M. Cleveland, P. B. Kenney, R. L. Vallejo, Y. Palti, and T. D. Leeds, 2016. Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (oncorhynchus mykiss). Frontiers in genetics 7:203.

Guðbrandsson, J., K. H. Kapralova, S. R. Franzdóttir, Þ. M. Bergsveinsdóttir, V. Hafstað, Z. O. Jónsson, S. S. Snorronson, and A. Pálsson, 2019. Extensive genetic differentiation between recently evolved sympatric arctic char morphs. Ecology and evolution 9:10964–10983.

Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph, 2019. Tree-sequence recording in slim opens new horizons for forward-time simulation of whole genomes. Molecular ecology resources 19:552–566.

Haller, B. C. and P. W. Messer, 2019. Slim 3: Forward genetic simulations beyond the wright–fisher model. Molecular biology and evolution 36:632–637.
Hendrickson, D. A., H. E. Pérez, L. T. Findley, W. Forbes, J. R. Tomelleri, R. L. Mayden, J. L. Nielsen, B. Jensen, G. R. Campos, A. V. Romero, et al., 2002. Mexican native trouts: a review of their history and current systematic and conservation status. Reviews in Fish Biology and Fisheries 12:273–316.

Hess, J. E., J. S. Zendt, A. R. Matala, and S. R. Narum, 2016. Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. Proceedings of the Royal Society B: Biological Sciences 283:20153064.

Irvine, J. R., 1978. The gerrard rainbow trout of kootenay lake, british columbia: A discussion of their life history with management, research and enhancement recommendations. BC Hydro Power Authority Fish Management Report 72:1–62.

Jacobs, A., M. R. Hughes, P. C. Robinson, C. E. Adams, and K. R. Elmer, 2018. The genetic architecture underlying the evolution of a rare piscivorous life history form in brown trout after secondary contact and strong introgression. Genes 9:280.

Jonsson, B. and N. Jonsson, 2001. Polymorphism and speciation in arctic charr. Journal of Fish Biology 58:605–638.

Kardos, M., A. Husby, S. E. McFarlane, A. Qvarnström, and H. Ellegren, 2016. Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. Molecular Ecology Resources 16:727–741.

Keeley, E., E. Parkinson, and E. Taylor, 2007. The origins of ecotypic variation of rainbow trout: a test of environmental vs. genetically based differences in morphology. Journal of Evolutionary Biology 20:725–736.

Kirubakaran, T. G., H. Grove, M. P. Kent, S. R. Sandve, M. Baranski, T. Nome, M. C.
De Rosa, B. Righino, T. Johansen, H. Otterà, et al., 2016. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of atlantic cod. Molecular ecology 25:2130–2143.

Kitano, J., S. Mori, and C. L. Peichel, 2007. Phenotypic divergence and reproductive isolation between sympatric forms of japanese threespine sticklebacks. Biological Journal of the Linnean Society 91:671–685.

Kofler, R., P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte, A. Futschik, C. Kosiol, and C. Schlötterer, 2011a. Popoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PloS one 6:e15925.

Kofler, R., R. V. Pandey, and C. Schlötterer, 2011b. Popoolation2: identifying differentiation between populations using sequencing of pooled dna samples (pool-seq). Bioinformatics 27:3435–3436.

Lecaudey, L. A., U. K. Schliewen, A. G. Osinov, E. B. Taylor, L. Bernatchez, and S. J. Weiss, 2018. Inferring phylogenetic structure, hybridization and divergence times within salmoninae (teleostei: Salmonidae) using rad-sequencing. Molecular phylogenetics and evolution 124:82–99.

Li, H. and R. Durbin, 2009. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, et al., 2009. The sequence alignment/map format and samtools. Bioinformatics 25:2078–2079.

Ltd, R. C., 2016. Kootenay lake action plan-2016. The Ministry of Forests, Lands and Natural Resource Operations, Nelson, BC.
Mackay, T. F., 2001. The genetic architecture of quantitative traits. Annual review of genetics 35:303–339.

Macqueen, D. J. and I. A. Johnston, 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proceedings of the Royal Society B: Biological Sciences 281:20132881.

Mangelsdorf, D. J., 1994. Vitamin a receptors. Nutrition reviews 52:S32.

McGirr, J. A. and C. H. Martin, 2017. Novel candidate genes underlying extreme trophic specialization in caribbean pupfishes. Molecular biology and evolution 34:873–888.

McKinnon, J. S., S. Mori, B. K. Blackman, L. David, D. M. Kingsley, L. Jamieson, J. Chou, and D. Schluter, 2004. Evidence for ecology’s role in speciation. Nature 429:294–298.

McPhail, J. D., 2007. The freshwater fishes of British Columbia. University of Alberta.

Micheletti, S. J. and S. R. Narum, 2018. Utility of pooled sequencing for association mapping in nonmodel organisms. Molecular ecology resources 18:825–837.

Monnet, G., J. S. Rosenfeld, and J. G. Richards, 2020. Adaptive differentiation of growth, energetics and behaviour between piscivore and insectivore juvenile rainbow trout along the pace-of-life continuum. Journal of Animal Ecology 89:2717–2732.

Morris, J., N. Navarro, P. Rastas, L. D. Rawlins, J. Sammy, J. Mallet, and K. K. Dasmahapatra, 2019. The genetic architecture of adaptation: convergence and pleiotropy in heliconius wing pattern evolution. Heredity 123:138–152.

Morriss-Kay, G. M., 1997. Retinoids in mammalian embryonic development. Pp. 79–92, in Advances in Organ Biology, vol. 3. Elsevier.
Near, T. J., R. I. Eytan, A. Dornburg, K. L. Kuhn, J. A. Moore, M. P. Davis, P. C. Wainwright, M. Friedman, and W. L. Smith, 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proceedings of the National Academy of Sciences 109:13698–13703.

Neto, R. V. R., G. M. Yoshida, J. P. Lhorente, and J. M. Yáñez, 2019. Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (oncorhynchus mykiss). Molecular Genetics and Genomics 294:563–571.

Northcote, T., 1973. Some impacts of man on kootenay lake and its salmonoids. Great Lakes Fishery Commission 25:1–46.

Pearse, D. E., N. J. Barson, T. Nome, G. Gao, M. A. Campbell, A. Abadía-Cardoso, E. C. Anderson, D. E. Rundio, T. H. Williams, K. A. Naish, et al., 2019. Sex-dependent dominance maintains migration supergene in rainbow trout. Nature ecology & evolution 3:1731–1742.

Penaluna, B. E., A. Abadía-Cardoso, J. B. Dunham, F. J. García-Dé León, R. E. Gresswell, A. R. Luna, E. B. Taylor, B. B. Shepard, R. Al-Chokhachy, C. C. Muhlfeld, et al., 2016. Conservation of native pacific trout diversity in western north america. Fisheries 41:286–300.

Primmer, C., S. Papakostas, E. Leder, M. Davis, and M. Ragan, 2013. Annotated genes and nonannotated genomes: cross-species use of gene ontology in ecology and evolution research. Molecular ecology 22:3216–3241.

Prince, D. J., S. M. O’Rourke, T. Q. Thompson, O. A. Ali, H. S. Lyman, I. K. Saglam, T. J. Hotaling, A. P. Spidle, and M. R. Miller, 2017. The evolutionary basis of
premature migration in pacific salmon highlights the utility of genomics for informing conservation. Science advances 3:e1603198.

Quinlan, A. R. and I. M. Hall, 2010. Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Réale, D., D. Garant, M. M. Humphries, P. Bergeron, V. Careau, and P.-O. Montiglio, 2010. Personality and the emergence of the pace-of-life syndrome concept at the population level. Philosophical Transactions of the Royal Society B: Biological Sciences 365:4051–4063.

Ricklefs, R. E. and M. Wikelski, 2002. The physiology/life-history nexus. Trends in Ecology & Evolution 17:462–468.

Roberts, R. B., Y. Hu, R. C. Albertson, and T. D. Kocher, 2011. Craniofacial divergence and ongoing adaptation via the hedgehog pathway. Proceedings of the National Academy of Sciences 108:13194–13199.

Robertson, A., 1967. The nature of quantitative genetic variation. Heritance from Mendel Pp. 265–280.

Roff, D., 1993. Evolution of life histories: theory and analysis. Springer Science & Business Media.

Salem, M., R. L. Vallejo, T. D. Leeds, Y. Palti, S. Liu, A. Sabbagh, C. E. Rexroad III, and J. Yao, 2012. Rna-seq identifies snp markers for growth traits in rainbow trout. PLoS One 7.

Schluter, D., 1996. Ecological speciation in postglacial fishes. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 351:807–814.
Schluter, D. and J. D. McPhail, 1992. Ecological character displacement and speciation in sticklebacks. The American Naturalist 140:85–108.

Garcia de la Serrana, D. and D. J. Macqueen, 2018. Insulin-like growth factor-binding proteins of teleost fishes. Frontiers in endocrinology 9:80.

Skúlason, S., 1999. Sympatric morphs, populations and speciation in freshwater fish with emphasis on arctic charr. Evolution of biological diversity Pp. 71–92.

Skúlason, S., K. J. Parsons, R. Svanbäck, K. Räsänen, M. M. Ferguson, C. E. Adams, P.-A. Amundsen, P. Bartels, C. W. Bean, J. W. Boughman, et al., 2019. A way forward with eco evo devo: an extended theory of resource polymorphism with postglacial fishes as model systems. Biological Reviews 94:1786–1808.

Skulason, S. and T. B. Smith, 1995. Resource polymorphisms in vertebrates. Trends in ecology & evolution 10:366–370.

Skúlason, S., S. S. Snorrason, D. Ota, and D. L. Noakes, 1993. Genetically based differences in foraging behaviour among sympatric morphs of arctic charr (pisces: Salmonidae). Animal Behaviour 45:1179–1192.

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011. Revigo summarizes and visualizes long lists of gene ontology terms. PloS one 6.

Supple, M. A., H. M. Hines, K. K. Dasmahapatra, J. J. Lewis, D. M. Nielsen, C. Lavoie, D. A. Ray, C. Salazar, W. O. McMillan, and B. A. Counterman, 2013. Genomic architecture of adaptive color pattern divergence and convergence in heliconius butterflies. Genome research 23:1248–1257.

Tamke, P., E. Parkinson, and E. Taylor, 2010. The influence of wisconsinan glaciation and
contemporary stream hydrology on microsatellite dna variation in rainbow trout (oncorhynchus mykiss). Canadian Journal of Fisheries and Aquatic Sciences 67:919–935.

Taylor, E. B., 1999. Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation. Reviews in Fish Biology and Fisheries 9:299–324.

Taylor, E. B., C. Foley, and M. Neufeld, 2019. Genetic mixture analyses in support of restoration of a high value recreational fishery for rainbow trout (oncorhynchus mykiss) from a large lake in interior british columbia. Conservation Genetics 20:891–902.

Thorne, A., A. I. MacDonald, and J. L. Thorley, 2016. The abundance of large, piscivorous ferox trout (salmo trutta) in loch rannoch, scotland. PeerJ 4:e2646.

Turner, S., 2017. qqman: Q-Q and Manhattan Plots for GWAS Data. URL https://CRAN.R-project.org/package=qqman. R package version 0.1.4.

Vandepoele, K., W. De Vos, J. S. Taylor, A. Meyer, and Y. Van de Peer, 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proceedings of the National Academy of Sciences 101:1638–1643.

Weir, B. S. and C. C. Cockerham, 1984. Estimating f-statistics for the analysis of population structure. Evolution Pp. 1358–1370.

Wikelski, M., L. Spinney, W. Schelsky, A. Scheuerlein, and E. Gwinner, 2003. Slow pace of life in tropical sedentary birds: a common-garden experiment on four stonechat populations from different latitudes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:2383–2388.

Wolf, G., 1984. Multiple functions of vitamin a. Physiological reviews 64:873–937.
Wood, A. W., G. Duan, and H. A. Bern, 2005. Insulin-like growth factor signaling in fish. International review of cytology 243:215–285.

Yoshida, G. M., J. P. Lhorente, R. Carvalheiro, and J. Yáñez, 2017. Bayesian genome-wide association analysis for body weight in farmed atlantic salmon (salmo salar l.). Animal genetics 48:698–703.
Figure 1: Study region of Kootenay Lake in southeastern British Columbia showing sampling locations, including the Blackwater River (shown in inset). See Supplemental Table S1 for more detailed sampling location information. Note that all individuals are sympatric in the main portion of Kootenay lake during the non-breeding portion of the year.
Figure 2: Distributions of 100kbp non-overlapping window Weir and Cockerham $F_{ST}$ estimates, 19,512 in total, between all pairs of rainbow trout groups. Orange represents Blackwater insectivores – Kootenay Lake piscivores, blue is Blackwater insectivores – Kootenay Lake insectivores, and red is Kootenay Lake insectivores – Kootenay Lake piscivores. The table shows mean values of each comparison, which are drawn with dashed lines in the graph.
Figure 3: Manhattan plot of 100kbp non-overlapping window-based $F_{ST}$ estimates between piscivorous and insectivorous rainbow trout from Kootenay Lake. The “outlier” cutoff line is shown in red, along with the windows that were above this value (0.609).
Figure 4: Results from a gene ontology (GO) enrichment analysis of loci from the most divergent 100kbp windows ($F_{ST} > 0.5$, $n = 187$) between piscivorous and insectivorous rainbow trout from Kootenay Lake visualized as a “treemap” when the target gene set is not included in the background set. Processes are grouped by functional class (colour), and box size of each process is scaled by the significance ($-\log_{10} p$-value; see Table 2) of enrichment of that process in relation to all processes in the genome. E.g., a larger tile size indicates a more significant result.
Table 1: Sequence read data for the three groups of rainbow trout in this study.

|                      | Piscivores   | Insectivores | Blackwater   |
|----------------------|--------------|--------------|--------------|
| No. Reads            | 592,736,824  | 574,066,240  | 590,881,476  |
| No. Reads Discarded  | 62,411,176   | 51,402,968   | 66,731,872   |
| No. Duplicate Reads Removed | 28,161,256  | 23,892,440   | 33,282,020   |
| No. Reads passing QC | 502,164,392  | 498,770,832  | 490,867,584  |
| No. Reads Mapped     | 365,737,358  | 364,938,474  | 363,369,176  |
| Proportion of genome covered | 0.628       | 0.633        | 0.635        |
| Mean coverage after filters | 23.16 ± 5.42 | 23.06 ± 5.33 | 24.14 ± 5.86 |
Table 2: Results from GOrilla gene ontology (GO) analysis of genes in the windows with $F_{ST}$ estimates $\geq 0.5$, arranged by $p$-value (smallest to largest). Terms match those in Figure 4.

| GO Term     | Description                                           | $p$-value   |
|-------------|-------------------------------------------------------|-------------|
| GO:0042574  | Retinal metabolic process                            | 9.88e$^{-05}$ |
| GO:0048008  | Platelet-derived growth factor receptor signaling pathway | 1.49e$^{-04}$ |
| GO:0050731  | Positive regulation of peptidyl-tyrosine phosphorylation | 2.18e$^{-04}$ |
| GO:0030029  | Actin filament-based process                          | 3.28e$^{-04}$ |
| GO:0007158  | Neuron cell-cell adhesion                             | 3.78e$^{-04}$ |
| GO:0048589  | Developmental growth                                  | 3.92e$^{-04}$ |
| GO:0051279  | Regulation of release of sequestered calcium ion into cytosol | 4.38e$^{-04}$ |
| GO:0060020  | Bergmann glial cell differentiation                   | 4.63e$^{-04}$ |
| GO:0040007  | Growth                                                | 5.02e$^{-04}$ |
| GO:0070895  | Negative regulation of transposon integration         | 5.89e$^{-04}$ |
| GO:0070894  | Regulation of transposon integration                  | 5.89e$^{-04}$ |
| GO:0048697  | Positive regulation of collateral sprouting in absence of injury | 5.89e$^{-04}$ |
| GO:0048696  | Regulation of collateral sprouting in absence of injury | 5.89e$^{-04}$ |
| GO:0035264  | Multicellular organism growth                         | 6.51e$^{-04}$ |
| GO:0030036  | Actin cytoskeleton organization                       | 6.57e$^{-04}$ |
| GO:0007010  | Cytoskeleton organization                             | 7.08e$^{-04}$ |
| GO:0090148  | Membrane fission                                      | 7.28e$^{-04}$ |
| GO:0016079  | Synaptic vesicle exocytosis                           | 9.66e$^{-04}$ |
| GO:0019730  | Antimicrobial humoral response                        | 9.66e$^{-04}$ |