Article: Discoveries

Evidence for epigenetically influenced mutagenesis and genetic assimilation in nascent species complexes (*Coregonus* sp.)

Clare J Venney[1], Claire Mérot[1,2], Eric Normandeau[1], Clément Rougeux[1], Martin Laporte[1,3], Louis Bernatchez[1]

1. Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada
2. UMR 6553 Ecobio, OSUR, CNRS, Université de Rennes, Rennes, France
3. Ministère des Forêts, de la Faune et des Parcs (MFFP), Québec, Québec

Corresponding author: Clare Venney (clarevenney@gmail.com)

Abstract

Phenotypic diversification and speciation are classically associated with genetic differentiation and gene expression variation. However, increasing evidence suggests that epigenetic mechanisms like DNA methylation may contribute to species divergence due to their effects on transcription and phenotype. Methylation can also increase mutagenesis and could lead to genetic assimilation of plastic phenotypes into genetic variants when the environment remains stable over many generations, though there has been minimal empirical research on assimilation. Whitefish are excellent systems to study speciation and genetic assimilation due to the repeated independent divergence of benthic-limnetic species pairs that have coexisted in lakes for ~12 000 years serving as natural replicates. Here we investigate whole genome genetic and epigenetic differentiation between benthic-limnetic species pairs in lake whitefish (*Coregonus clupeaformis*) and European whitefish (*C. lavaretus*) from four lakes (N=64). We found considerable, albeit variable, genetic and epigenetic differences between benthic and limnetic species in all four lakes. Polymorphism was enriched at CpG sites confirming the mutagenic nature of DNA methylation; all SNP types were enriched in all lakes, though C > T SNPs were most common. We also identified a surplus of potential sites undergoing genetic assimilation putatively associated with speciation, defined as sites harbouring both a differentially methylated site and a highly divergent SNP. Our results support the hypothesis that DNA methylation contributes to phenotypic divergence, influences mutagenesis, and drives genetic assimilation of phenotypes that are ultimately associated with ecological speciation in whitefish species complexes.

Introduction

Speciation has long been a focus of evolutionary biology, with recent research expanding into the role of genomics in reproductive isolation, phenotypic diversification, and species divergence (Seehausen et al. 2014; Marques et al. 2019). Speciation can occur rapidly despite the slow rate of genetic mutation, sometimes due to new combinations of standing genetic variation contributing to speciation (Marques et al. 2019). However, phenotypic plasticity can also promote speciation, particularly when allopatric populations acclimate to their respective environments, phenotypes are partially genetically controlled, and the capacity for plasticity is lost over time (Pfennig et al. 2010). Phenotypic changes can occur through altered gene expression (Whitehead and Crawford 2006) which can also lead to phenotypic divergence and speciation, as observed between Arctic charr ecotypes (*Salvelinus alpinus*; Gudbrandsson et al. 2018; Jacobs and Elmer 2021), between Chinook and Coho salmon and their hybrids (*Oncorhynchus tshawytscha* and *O. kisutch*; Mckenzie et al. 2021), and between sympatric lake whitefish species (*Coregonus* sp.; Derome et al. 2006; Rougeux, Gagnaire, Praebel, et al. 2019). Transcriptomic variation associated with lip size has been identified in the Midas cichlid species complex (*Amphilophus citrinellus;* Manousaki et al. 2013) and sequence variation in transcribed regions has been reported between *A. astorquii* and *A. zaliosus* (Elmer et al. 2010). Transcriptomic differences generally increase with taxonomic distance for closely related species (Whitehead and Crawford 2006) and drive phenotypic differences among species (Whitehead and Crawford 2006; Pavey et al. 2010). Therefore, the mechanisms underlying adaptive phenotypic differences during the course of speciation likely involve both phenotypic plasticity and stable genetic evolution. There has been recent interest in the role of DNA methylation as a plastic mechanism contributing to speciation (Vogt 2017; Ashe et al. 2021).

DNA methylation, the most studied epigenetic mechanism, is the addition of a methyl group to the cytosine of CpG sites (a cytosine followed by a guanine in the DNA sequence) in vertebrates, generally resulting in reduced transcription without a change in DNA sequence (Bird 2002). DNA methylation is highly sensitive to the environment, with differences in temperature (McCaw et al. 2020; Ryu et al. 2020; Beemelmanns et al. 2021; Venney et al. 2022), salinity (Artemov et al. 2017; Heckwolf et al. 2020; Hu et al. 2021), rearing environment (Le Luyer et al. 2017; Berbel-Filho et al. 2020; Venney et al. 2020; Leitwein et al. 2021; Wellband et al. 2021),

and other factors leading to epigenetic changes. DNA methylation can often alter phenotype (Anastasiadi et al. 2021; Vogt 2021), with evidence for phenotypically-linked epigenetic divergence associated with spawning life history tactics in capelin (*Mallotus villosus*; Venney et al. *in press*) and between four sympatric Arctic charr morphs (Matlosz et al. 2022). Due to the role of DNA methylation in plasticity and phenotypic diversification, it may contribute to speciation (Vogt 2017; Laporte et al. 2019; Ashe et al. 2021; Stajic and Jansen 2021). A recent simulation study showed that epigenetic plasticity can speed up speciation when it reduces the fitness of migrants and hybrid offspring but can prevent genetic adaptation and speciation if epigenetic adaptation occurs, precluding the need for genetic adaptation (Greenspoon et al. 2022). Empirical evidence for the role of DNA methylation in speciation is also emerging (Laporte et al. 2019). Methylation differences detected through methylation-sensitive amplified polymorphism, but not genetic differences, were predictive of the extent of behavioural isolation between species pairs in 16 species of darters (*Ulocentra*, *Nanostoma,* and *Etheostoma*; Smith et al. 2016). Another study showed considerable epigenetic differences between six phenotypically divergent species of Lake Malawi cichlids (Vernaz et al. 2022). DNA methylation can affect transcription and phenotype and could result in initial plastic phenotypic responses that lead to phenotypic diversification and speciation over generations.

DNA methylation is also mutagenic and can generate polymorphism at CpG sites. This is partially due to the spontaneous hydrolytic deamination of methylated cytosine to uracil which is rapidly converted to a thymine tautomer if not corrected by DNA repair enzymes (Gorelick 2003). Spontaneous deamination is ~3.5 times more likely to occur at methylated cytosines than unmethylated ones (Jones et al. 1992; Gorelick 2003) with methylated cytosines having mutation rates ~20 000 times higher than unmethylated cytosines after accounting for DNA repair efficiency (Gorelick 2003). It is estimated that ~8 deamination events occur per day in the ~6 billion bp diploid human genome (Jones et al. 1992) providing a consequential source of novel mutations, particularly in larger genomes. C > T transitions are the most common and explainable epigenetically induced mutation, though there is also evidence for increased C > A and C > G mutations at CpG sites due to mutagen exposure (Tomkova and Schuster-Böckler 2018). On the other hand, DNA methylation can also shield CpG sites from mutagenesis depending on the environmental stimulus or trigger, though the intricacies of when mutagenesis is favoured or prevented remain unclear (Tomkova and Schuster-Böckler 2018). A study in

humans (*Homo sapiens*) showed that cytosines with intermediate levels of methylation (20-60%) had the highest mutation rate, even compared to fully methylated sites (Xia et al. 2012). The shielding properties of DNA methylation could also be due to the inability to discern DNA methylation from other methylation-related marks through bisulfite sequencing. In particular, 5-methylcytosine can be converted to 5-hydroxymethylcytosine during demethylation (Li et al. 2019). The two cannot be differentiated through bisulfite sequencing (Li et al. 2019), though 5-hydroxymethylcytosine may protect CpG sites from mutation (Tomkova and Schuster-Böckler 2018). Therefore, DNA methylation can lead to transcriptional and phenotypic changes, influence mutation rates, and has thus been proposed as a mechanism for genetic assimilation of phenotypic changes (Nishikawa and Kinjo 2018; Danchin et al. 2019).

Genetic assimilation occurs when an environmentally induced phenotype becomes stable and genetically encoded, even in the absence of the original stimulus (Nishikawa and Kinjo 2018; Anastasiadi et al. 2021). A stable shift in environment results in altered DNA methylation which can be passed on to future generations through epigenetic inheritance (Anastasiadi et al. 2021; Vogt 2021). Over the course of many generations, methylation can influence the production of point mutations, either by directly inducing mutation or by shielding sites from mutation (Tomkova and Schuster-Böckler 2018). The latter could still lead to genetic assimilation if high methylation shields a CpG site from mutation in one cohort but lower methylation levels allow mutation at that site in another cohort. This can result in genetic assimilation of the phenotype when the environment remains predictable (Angers et al. 2010; Nishikawa and Kinjo 2018; Danchin et al. 2019). While most literature on epigenetically based genetic assimilation consists of review and perspective articles, there is evidence for genetic assimilation in *Saccharomyces cerevisiae*, with the speed of mutational assimilation dependent on the capacity to modify histone acetylation (Stajic et al. 2019). However, direct empirical evidence for methylated sites inducing point mutations at the same genomic positions and contributing to evolution remain sparse.

The two whitefish sister taxa, lake whitefish (*Coregonus clupeaformis*) in North America and the European whitefish (*C. lavaretus*) in Europe, are well-characterized and relevant systems to study phenotypic diversification and speciation. Lake and European whitefish have evolved separately since they became geographically isolated ~500 000 years ago (Bernatchez and

Dodson 1991; Bernatchez and Dodson 1994; Jacobsen et al. 2012). Both show repeated independent divergence of sympatric species complexes consisting of an ancestral benthic species and a derived limnetic species originating from different glacial refugia during the last glaciation period (Bernatchez and Dodson 1991; Pigeon et al. 1997; Østbye, Bernatchez, et al. 2005; Bernatchez et al. 2010; Rougeux et al. 2017) which came into secondary contact ~12 000 years ago when colonizing postglacial lakes (Rougeux et al. 2017; Rougeux, Gagnaire, Praebel, et al. 2019). In North America, the derived limnetic species evolved to colonize the limnetic zone, leading to differences in diet (Bernatchez et al. 1999), reduced size (Bernatchez et al. 1999), slower growth (Trudel et al. 2001), more slender body morphology (Laporte et al. 2015; Laporte et al. 2016), higher metabolic rate (Trudel et al. 2001; Dalziel et al. 2015), and greater swimming behaviour (Rogers et al. 2002). Benthic and limnetic species often coexist in Europe, though some Fenno-Scandinavian and alpine lakes contain more than two, and up to six sympatric whitefish species (Østbye, Næsje, et al. 2005; De-Kayne et al. 2022). In both North America and Europe, sympatric whitefish are reproductively isolated with variable amount of gene flow depending on the lake (Rogers and Bernatchez 2006; Rougeux et al. 2017) which translates into genetic (Østbye, Næsje, et al. 2005; Østbye et al. 2006; Rogers and Bernatchez 2007; Bernatchez et al. 2010; Siwertsson et al. 2013; Dion-Côté et al. 2017; Rougeux et al. 2017; Rougeux, Gagnaire, Praebel, et al. 2019; De-Kayne et al. 2022; Mérot et al. 2022) transposable element methylation (Laporte et al. 2019), and transcriptional differences (Derome et al. 2006; Jeukens et al. 2009; Rougeux, Gagnaire, Praebel, et al. 2019). Benthic-limnetic species pairs in different lakes thus present a naturally replicated system in which to study the molecular mechanisms associated with speciation.

We assessed the role of genomic and epigenomic variation in whitefish speciation by performing low-coverage whole genome sequencing (WGS) and whole genome bisulfite sequencing (WGBS) for four benthic-limnetic species pairs (N=64 individuals sequenced in both datasets): lake whitefish from Cliff and Indian Lake, Maine, USA, and European whitefish from Langfjordvatn Lake, Norway and Zurich Lake, Switzerland (Figure 1). The objectives of this study were to (i) determine the extent of whole genome genetic divergence between benthic-limnetic whitefish species pairs, (ii) measure the rate of polymorphism at CpG sites relative to the rest of the genome, (iii) characterize whole genome differences in DNA methylation among limnetic-benthic whitefish species pairs, and (iv) determine whether there was an enrichment of

outlier SNPs at differentially methylated loci (DMLs), which would support the hypothesis that epigenetically influenced mutagenesis may be driving genetic variation and be involved in speciation. As such, our results provide novel support for the role of DNA methylation in interspecies variation, and in driving mutagenesis and genetic evolution between species.

Results

Genetic divergence between benthic-limnetic species pairs

The extent of genetic differentiation between the benthic and the limnetic species varied among lakes. F_{ST} estimates for Cliff Lake and Indian Lake in North America were previously published in Mérot et al. (2022). In North America, moderate to high divergence between lakes both within and between species was observed (F_{ST} between lakes: 0.084 between Cliff Lake and Indian Lake limnetic species to 0.182 between Cliff Lake and Indian Lake benthic species), with greater F_{ST} between species in Cliff Lake than Indian Lake (0.098 for Indian Lake and 0.175 for Cliff Lake; Table 1A). In Europe, F_{ST} estimates for Langfjordvatn Lake and Zurich Lake showed greater differentiation between lakes than between species (F_{ST} between lakes: 0.237 between Zurich Lake benthic and Langfjordvatn Lake limnetic to 0.242 between Zurich Lake limnetic and Langfjordvatn Lake benthic; F_{ST} between species within lake: 0.034 in Langfjordvatn Lake to 0.043 in Zurich Lake; Table 1B). In all lakes, genetic differentiation is widespread along the genome (Figure S1).

Elevated rate of polymorphism in CpG sites

We analyzed between 11 861 765 and 30 919 358 SNPs per lake (Table 2) to assess whether SNPs were enriched in CpG sites relative to the rest of the genome to investigate the prediction that CpG sites are mutagenic. We also tested whether the enrichment was specific to (i) C/T and G/A SNPs (to account for the G position of CpG sites), (ii) C/G and G/C SNPs, and (iii) C/A and G/T SNPs. Permutation tests showed that SNPs were significantly enriched in CpG sites for all four lakes ($p < 0.00001$; Table 2), with none of the 10 000 permutations per test reaching or exceeding the observed rate of polymorphism at CpG sites. Between 10.5 and 12.3% of SNPs occurred in CpG sites for each lake in contrast with an average of 3.8% of polymorphic sites across the genome. This means that 2.8 to 3.2 times more SNPs occur in CpG sites than in the rest of the genome. There was significant enrichment for all SNP types according to

Pearson's chi-squared tests ($p < 0.001$; Table 2). C/T and G/A SNPs were most common in CpG sites, with 3.0 to 3.4 times more C/T SNPs occurring in CpG sites than in all C and G sites across the genome. The other SNP types were slightly less common, with an enrichment of 1.1 to 1.2 times more C/G and G/C SNPs and 1.2 to 1.3 times more C/A and G/T SNPs in CpG sites relative to the rest of the C and G sites in the genome (Table 2).

Epigenetic differentiation between benthic and limnetic whitefish

We assessed the level of epigenetic differentiation between benthic-limnetic species pairs by performing differential methylation analysis for each lake. After all quality trimming, we analyzed between 13 236 657 and 19 605 766 CpG sites per lake with approximately 7.5 to 8.1X coverage per sample (see Table 3 for detailed information). We ran generalized linear models in DSS (Park and Wu 2016) to identify differentially methylated loci (DMLs; i.e., CpG sites) and differentially methylated regions (DMRs; regions of the genome showing differences in methylation levels between experimental groups) between species. We identified variable but significant epigenetic differentiation between benthic-limnetic species pairs in all lakes: 41 116 DMLs and 3 019 DMRs in Cliff Lake, 26 583 DMLs and 2 383 DMRs in Indian Lake, 4 030 DMLs and 402 DMRs in Langfjordvatn Lake, and 8 004 DMLs and 754 DMRs in Zurich Lake (Figure 2, Table 3, Tables S1-S8). The DMRs covered between 63 597 and 1 005 874 CpG sites in each lake (Table 3). Overall epigenetic differentiation between species was greater in North America than in Europe, consistent with higher interspecies genetic differentiation as measured by F_{ST} in North American relative to Europe.

Enrichment for overlaps between outlier SNPs and DMLs

For each lake, we assessed whether the most differentiated SNPs (5% highest F_{ST}) in each lake overlap more often than expected by chance with CpG sites identified as DMLs between species. Such overlaps in genetic and epigenetic differentiation represent potential sites of ongoing genetic assimilation where divergent methylation patterns could be undergoing epigenetically influenced mutagenesis into stable genetic variants between species. We considered only polymorphic CpG sites in this analysis to account for the elevated mutation rate at CpG sites. Pearson's chi-squared tests showed that the rate of outlier SNP-DML overlaps was much greater than expected, with an 11.4 to 36.9-fold enrichment of overlaps. The enrichment

was greater in North America (36.9-fold in Cliff Lake and 22.9-fold in Indian Lake) than in Europe (11.4-fold in Langfjordvatn Lake and 17.4-fold in Zurich Lake).

We also considered the relationship between DML percent methylation and the allelic frequency of cytosine at outlier SNP-DML overlaps; high percent methylation and low frequency of the cytosine allele would suggest that methylation has induced mutation at these sites, leading to genetic differentiation between species based on differences in methylation. The inverse relationship (positive correlation between percent methylation and frequency of cytosine) would suggest that high methylation levels shield the site from mutation whereas less methylated sites are more prone to mutate. Due to the use of low coverage WGS data, there is some uncertainty around allele frequency calls when the minor allele is at low frequency (i.e., some cytosine alleles with low frequency were missed and the frequency of cytosine at these CpG sites was determined to be 0). Despite this, we found that DML percent methylation was clearly positively associated with the frequency of cytosine alleles at the same sites for the potential sites undergoing genetic assimilation (Figure 3). There were strong correlations between percent methylation and the allele frequency of cytosine at the potential sites undergoing genetic assimilation (correlation coefficients between 0.884 and 0.946), especially when we compared the relationship in other polymorphic CpGs (not limited to outliers). Permutation tests with 100 000 permutations showed that the observed correlation in the potential genetic assimilation sites was far greater than the same correlations with random subsets of data, with subsets never approaching the observed correlation (Figure S2). GO enrichment analysis for the outlier SNP-DML overlaps showed enrichment for 24 terms involved in behaviour and immune function in Cliff Lake, two terms involving the Golgi apparatus and other membrane structures in Indian Lake, four terms associated with catalytic activity in Langfjordvatn Lake, and 32 terms in Zurich Lake associated with behaviour, immune function, metabolism, and muscle development (Table 4).

Discussion

Recent speciation research has expanded in scope to include plastic mechanisms such as DNA methylation (Pál and Miklós 1999; Richards et al. 2010; Ashe et al. 2021; Greenspoon et al. 2022) which could contribute to phenotypic diversification. Here we provide evidence for both genetic and epigenetic differentiation between sympatric benthic-limnetic whitefish species

pairs from two continents. We show that there is an elevated rate of genetic polymorphism at CpG sites, including high rates of outlier SNPs at CpG sites showing differential methylation between species which may represent sites of ongoing genetic assimilation. Together, our results provide empirical support for the proposed role of DNA methylation in driving phenotypic diversification (Anastasiadi et al. 2021; Vogt 2021), mutagenesis (Tomkova and Schuster-Böckler 2018), and speciation (Pál and Miklós 1999; Richards et al. 2010; Ashe et al. 2021; Greenspoon et al. 2022).

Considerable but variable genetic divergence between species and lakes

We observed genetic differentiation between species in all four lakes, consistent with previous studies showing genetic differences between whitefish benthic-limnetic species (Østbye, Næsje, et al. 2005; Østbye et al. 2006; Rogers and Bernatchez 2007; Bernatchez et al. 2010; Siwertsson et al. 2013; Dion-Côté et al. 2017; Rougeux et al. 2017; Feulner and Seehausen 2019; Rougeux, Gagnaire, Praebel, et al. 2019; De-Kayne et al. 2022; Mérot et al. 2022). There was relatively high genetic differentiation between species and lakes in North America, though differentiation between lakes was greater in Europe than in North America. There was greater differentiation between species in Cliff Lake than Indian Lake, as previously reported in our sister study (Mérot et al. 2022). Genetic differentiation between lakes was greater than between species in Europe (Table 1B), likely due to the Zurich Lake and Langfjordvatn Lake populations having occupied two different glacial refugia during the last glaciation period (Østbye, Bernatchez, et al. 2005). However, our within-lake F_{ST} estimates are slightly low compared to other estimates between European whitefish species pairs (0.042 to 0.096 based on 16 microsatellite loci (Siwertsson et al. 2013) and 0.01 to 0.075 based on six microsatellite loci (Østbye et al. 2006)), likely due to the use of whole genome vs. targeted methods and the low MAF filter including many low frequency SNPs in this analysis. Variable interspecies F_{ST} between lakes is consistent with varying degrees of reproductive isolation/gene flow between species in different systems (Lu and Bernatchez 1999; Østbye et al. 2006; Gagnaire et al. 2013; Rougeux, Gagnaire, and Bernatchez 2019; De-Kayne et al. 2022). The variation in interspecies genetic divergence between lakes could be due to the extent of differences between the species' trophic niches, with species occupying vastly different trophic niches also showing greater morphological differences (Lu and Bernatchez 1999). Overall, genetic differentiation between

species was widespread along the genome consistent with previous work (Feulner and Seehausen 2019; De-Kayne et al. 2022; Mérot et al. 2022), though large-effect loci (De-Kayne et al. 2022) and genomic islands of divergence (Gagnaire et al. 2013) have also been reported. Thus, we observed genome-wide genetic differentiation between species which may be associated with phenotypic divergence and reproductive isolation between species.

Epigenetic divergence between whitefish species pairs

We found considerable epigenetic divergence between species, though the extent of divergence varied between lakes and was greater in North American than in Europe (Figure 2, Table 3). Our results highlight the role of DNA methylation as a mechanism contributing to phenotypic diversification and speciation, similar to results in North American whitefish showing differential methylation of transposable elements between species (Laporte et al. 2019). These methylation changes may have arisen due to environmental differences between benthic and limnetic habitats leading to character displacement between species since the environment can have profound effects on the methylome. Previous studies have shown that methylation is affected by temperature (Anastasiadi et al. 2017; Metzger and Schulte 2017; Ryu et al. 2018; McCaw et al. 2020; Venney et al. 2022), salinity (Artemov et al. 2017; Heckwolf et al. 2020), rearing environment (Le Luyer et al. 2017; Gavery et al. 2018; Leitwein et al. 2021; Wellband et al. 2021), and other factors. Differences in habitat use can influence DNA methylation, as observed between capelin (*Mallotus villosus*) utilizing beach- and demersal-spawning life history tactics (Venney et al. *in press*), among sympatric Arctic charr morphs occupying different habitats and dietary niches (Matlosz et al. 2022), between phenotypically divergent cichlid species (Vernaz et al. 2022), and in freshwater snails (*Potamopyrgus antipodarum*) which exhibit differences in shell shape associated with water current speed (Thorson et al. 2017). DNA methylation can also be influenced by genetic variation (Richards 2006; Lallias et al. 2021), therefore epigenetic differences may in part be driven by genetic divergence between species. Given the greater epigenetic divergence and F_{ST} between species pairs in North American than in Europe, it is possible that the methylation differences are a function of genetic divergence between species. Parallel transcriptional differences were previously reported between benthic-limnetic whitefish species pairs in both captive and natural conditions, indicating that some transcriptional differences between species are under genetic control and may have been subject

to selection (St-Cyr et al. 2008). Transcriptional divergence between benthic-limnetic species pairs was also explored in 48 of 64 samples used in this study; differences were often parallel across lakes in both lake whitefish and European whitefish (Rougeux, Gagnaire, Praebel, et al. 2019). We generally observed similar ratios of interspecies DMRs among lakes from our study and differentially expressed gene (DEG) counts in the earlier transcriptome study by Rougeux et al. (3 019 DMRs and 3 175 DEGs in Cliff Lake, 2 383 DMRs and 238 DEGs in Indian Lake, 402 DMRs and 276 DEGs in Langfjordvatn Lake, and 754 DMRs and 1 392 DEGs in Zurich Lake). This indicates that Cliff Lake and Zurich Lake have greater epigenomic and transcriptomic differences between species relative to the other lakes, though there are considerably more DMRs than DEGs in Indian Lake, possibly due to greater genetic differentiation between species. Thus, DNA methylation divergence between species may be due to differences in habitat and genetic background and could provide a mechanistic basis for phenotypic diversification between these nascent species pairs due to its effects on transcription (Bird 2002) and phenotype (Anastasiadi et al. 2021; Vogt 2021).

Mutational enrichment at CpG sites suggests epigenetically influenced mutagenesis

Our results suggest that epigenetically influenced mutagenesis has occurred at CpG sites, generating genetic variation between species pairs. DNA methylation influences mutagenesis (Flores et al. 2013; Tomkova and Schuster-Böckler 2018; Ashe et al. 2021) either through promoting or inhibiting mutation (Tomkova and Schuster-Böckler 2018) and has been suggested to play a role in genetic assimilation of phenotypic traits (Nishikawa and Kinjo 2018; Danchin et al. 2019). The heritable differences in behaviour (e.g., Rogers et al. 2002) and minimal plasticity in morphological traits differentiating benthic and limnetic lake whitefish (Laporte et al. 2016) suggest that these putatively adaptive traits might be genetically encoded. Transcriptomic differences between species are also retained when rearing lake whitefish in a controlled environment (St-Cyr et al. 2008) and are related to parallel genetic differences between species in both lake and European whitefish (Rougeux, Gagnaire, Praebel, et al. 2019). Therefore, the occurrence of genetic assimilation of phenotypic differences between benthic and limnetic whitefish seems possible and may lead to transcriptional differences between species becoming genetically encoded. We show that CpG sites, the main sites where DNA methylation occurs in vertebrate genomes, have higher rates of polymorphism compared to the rest of the genome,

suggesting that DNA methylation may be inducing point mutations that could accumulate between species over generations (Table 2). We observed a 2.8- to 3.2-fold polymorphism enrichment at CpG sites which is slightly lower than the previously reported 3.5-fold increased mutation in methylated relative to unmethylated cytosines (Jones et al. 1992; Gorelick 2003), possibly due to (i) not all CpG cytosines being methylated in the genome, (ii) the background mutation rate in non-CpG and unmethylated cytosines, and (iii) selection against epigenetically induced mutations. While epigenetically induced C > T transitions are most common due to spontaneous deamination of cytosines to uracil (Tomkova and Schuster-Böckler 2018), we show that there is significant enrichment of all types of point mutations at CpG sites (Table 2). The exact mechanisms behind spontaneous C > A and C > G mutations without the involvement of mutagens are less clear in the context of epigenetically influenced mutagenesis (Tomkova and Schuster-Böckler 2018), though we provide evidence that all types of polymorphism are enriched at CpG sites.

When environments change and then remain stable (e.g., in range expansions, colonization of novel habitats, or shifts to novel habitat use), DNA methylation may serve as an initial plastic response to the new environment, slowly leading to mutagenesis over many generations. Selection could then act upon these mutations, which would be removed from the population if they are unfavourable; it is inferred that most of these mutations will be deleterious based on mismatches between experimentally observed epigenetically induced mutation rate and the observed rates of these mutations in populations (Danchin et al. 2019). However, the mutations could be retained through random genetic drift if neutral or selected for if they prove beneficial, leading to increased frequency of the novel mutation over time. If these mutations affect transcription of proximal genes, they could explain the stable transcriptional differences observed between whitefish species (St-Cyr et al. 2008; Rougeux, Gagnaire, Praebel, et al. 2019). We therefore propose that DNA methylation provides a mechanism to increase genetic differentiation between populations, potentially leading to phenotypic and genetic divergence between nascent species pairs (Vogt 2017; Danchin et al. 2019).

Novel evidence for epigenetically influenced genetic assimilation in speciation

Our results also show that SNPs with high F_{ST} between species are enriched at DMLs, which is suggestive of potential sites of ongoing genetic assimilation contributing to genetic

evolution between benthic-limnetic species pairs. Several theoretical articles have discussed the role of DNA methylation in speciation (Pál and Miklós 1999; Richards et al. 2010; Danchin et al. 2019; Ashe et al. 2021; Greenspoon et al. 2022) and the possibility of genetic assimilation of epigenetic marks (Nishikawa and Kinjo 2018; Danchin et al. 2019). A new, stable methylation state can arise when the environment changes then remains relatively consistent (Anastasiadi et al. 2021) which may be acted upon by natural selection (Ashe et al. 2021). If the encoded phenotype approaches the hypothetical "optimal" phenotype under those conditions, canalization of the phenotype would be supported (Pál and Miklós 1999). Over time, epigenetically influenced mutagenesis can lead to methylation changes becoming genetically encoded (Danchin et al. 2019; Ashe et al. 2021). Modeling studies have provided support for epigenetic plasticity becoming replaced with genetic changes (Nishikawa and Kinjo 2018; Danchin et al. 2019). Our study is one of the very few empirical studies attempting to empirically address these topics. Stajic et al. (2019) showed that mutational assimilation was dependent on the capacity of *S. cerevisiae* to modify histone acetylation, showing the involvement of epigenetic mechanisms in genetic assimilation. Another study in humans found that mutation rates in embryonic stem cells were higher at methylated cytosines, though cytosines with intermediate levels of methylation (20-60% methylation) showed the highest mutation rate (Xia et al. 2012). Interestingly, a study in great apes (*Homo*, *Pan*, *Gorilla*, and *Pongo* sp.) found that mutational signatures across the genome differed between species and were associated with chromatin state and cytosine hydroxymethylation (Goldberg and Harris 2022). The role of epigenetic processes in influencing mutagenesis is becoming clearer, though we provide some of the first direct evidence of DNA methylation influencing mutagenesis in the context of speciation.

The sites we identified as undergoing genetic assimilation associated with species differentiation showed both differential methylation between species (i.e., when CpG sites are still present in both species) and high genetic differentiation between species (i.e., when one species has partially assumed the "assimilated" state and a cytosine has mutated to another nucleotide). There was a greater enrichment for outlier SNP-DML overlaps in North American than Europe consistent with greater genetic differentiation, possibly reflecting more pronounced reproductive isolation between species pairs in North America, though this could also reflect differences in mutational signatures between species (e.g., Goldberg and Harris 2022). Gene ontology analysis of the overlaps showed that they occurred in genes associated with behaviour,

immune function, muscle development, metabolism, membrane function, and catalytic activity depending on the lake (Table 4), consistent with the previous transcriptomic study showing enrichment of immune and metabolic functions in DEGs (Rougeux, Gagnaire, Praebel, et al. 2019). Furthermore, we show that percent methylation is positively correlated with the frequency of cytosine alleles at these sites (Figure 3), providing further support for our hypothesis that differential methylation between species can influence mutation and genetic divergence. The correlation is polarized in Cliff Lake, with higher methylation and higher frequency of cytosine in the benthic species and the opposite observation in the limnetic species (Figure 3A). This could be due to the reference genome representing a benthic fish from Cliff Lake, or due to reduced interspecies gene flow in Cliff Lake relative to the other lakes leading to less mixture of epigenetic signatures between species, furthering the epigenetic divergence in this lake. Overall, the positive relationship between methylation rate and the frequency of the cytosine allele indicates that, at these sites, higher methylation in one species likely shields the CpG from mutation, whereas lower methylation in the other species does not inhibit mutation. This is consistent with a previous study in humans which showed that cytosines with intermediate levels of methylation (20-60%) had the highest mutation rates (Xia et al. 2012), suggesting that higher levels of methylation may impede mutagenesis. This may also be due to the inability to discern 5-methylcytosine from 5-hydroxymethylcytosine in bisulfite sequencing data and other methylation analysis methods (Li et al. 2019); the latter may play a role in protecting CpG sites from mutations (Tomkova and Schuster-Böckler 2018). Further study of the independent effects of 5-methylcytosine and 5-hydroxymethylcytosine in ecological and speciation studies is warranted to understand their relative effects on mutagenesis and contributions to phenotypic diversification and speciation. To our knowledge, ours is the first study to relate epigenetic divergence and epigenetically influenced mutagenesis to speciation, providing support for previous theories on the role of DNA methylation in genetic assimilation.

Given the rapid speciation rate (~3-4K years) between benthic and limnetic species in all lakes, we propose that methylation differences arose quickly, contributing to the multi-trait rapid evolution that occurred between species. These methylation differences between species could lead to stable genetic variants (both by directly inducing mutations, and by protecting sites from mutation in one species but not the other), with divergent selection acting on both epigenetic and genetic marks. Over generations, this could lead to reduced phenotypic (Laporte et al. 2016) and

transcriptional plasticity (St-Cyr et al. 2008), resulting in genetic assimilation of phenotypic differences between nascent species pairs. An alternative hypothesis is that both DNA methylation and gene expression-related polymorphism at the same CpG site could have similar effects on transcription and thus provide two different mechanisms for controlling transcription at the same site, though this does not rule out epigenetically influenced mutagenesis at that CpG site resulting in the observed polymorphism. Overall, we provide some of the first experimental evidence directly associating variation in DNA methylation with increased polymorphism and genetic assimilation, potentially contributing to early diversification and speciation.

Conclusions

We found substantial genetic and epigenetic divergence between independently derived benthic-limnetic species pairs in lake and European whitefish. We provide evidence that DNA methylation may generate genetic polymorphism due to its mutagenic nature. We characterized potential sites of ongoing genetic assimilation wherein differential methylation levels between species may be influencing the rate of mutagenesis and resulting in divergent genetic variation between benthic-limnetic species pairs. As such, our results shed light on the diverse ways DNA methylation can contribute to plasticity and evolution, from plastic responses to environmental changes to the induction of mutagenesis and genetic assimilation of plastic phenotypes.

Materials and Methods

Sample preparation and sequencing

We sampled limnetic-benthic whitefish species pairs from two lakes in North America and two lakes in Europe: lake whitefish (*C. clupeaformis*) from Cliff Lake and Indian Lake, Maine, USA, and European whitefish (*C. lavaretus*) from Langfjordvatn Lake, Norway and Zurich Lake, Switzerland (Fig. 1). Animal care was performed humanely under Université Laval animal care permit 126316. Fish were caught with gillnets, humanely euthanized, and immediately dissected to obtain fresh tissue samples. Liver tissue was sampled from 64 fish, including 48 previously used in Rougeux et al. (2019): eight individuals per species per lake except for Indian Lake where we sampled seven benthic and nine limnetic fish. Samples were stored either at -80°C or in RNAlater; all samples from Europe were stored in RNAlater. Liver tissue was chosen due to its homogeneous tissue characteristics and involvement in growth and

metabolism (Trefts et al. 2017).

Genomic DNA was isolated from all 64 fish using a modified version of a salt extraction protocol (Aljanabi and Martinez 1997). DNA integrity was checked on a 1% agarose gel and quantified on a NanoDrop spectrophotometer. Individual whole genome sequencing (WGS) and whole genome bisulfite sequencing (WGBS) libraries were built at the McGill University and Genome Quebec Innovation Centre (Montreal, Canada). WGS was performed using paired end 150 bp sequencing on an Illumina HiSeq4000 with estimated 5X coverage. The North American samples were previously used and sequenced in Mérot et al. (2022). WGBS was performed using paired end 150 bp sequencing on the Illumina HiSeqXwith samples randomly distributed across 16 lanes (four individuals pooled per lane) with estimated 10X coverage.

Whole genome sequencing analysis

Reads were trimmed and filtered for quality with fastp (Chen et al. 2018). The North American samples were aligned to the genome of *Coregonus clupeaformis* (Mérot et al. (2022) and the European samples were aligned to the genome of the European whitefish, *Coregonus* sp. Balchen, (De-Kayne et al. 2020) using BWA-MEM (Li 2013). Aligned reads were filtered to keep mapping quality over 10 with Samtools v1.8 (Li et al. 2009). Duplicate reads were removed with MarkDuplicates (PicardTools v1.119., http://broadinstitute.github.io/picard). We realigned around indels with GATK IndelRealigner (McKenna et al. 2010) and soft clipped overlapping read ends using clipOverlap in bamUtil v1.0.14 (Breese and Liu 2013). The pipeline is available at https://github.com/enormandeau/wgs_sample_preparation.

Bam alignments were analysed with the program ANGSD v0.931 (Korneliussen et al. 2014) which accounts for genotype uncertainty and is appropriate for low and medium- coverage whole genome sequencing (Lou et al. 2020). Input reads were filtered to remove low-quality reads and to keep mapping quality above 30 and base quality above 20. We ran ANGSD on each lake separately to detect polymorphic positions (SNPs), estimate the spectrum of allele frequency, minor allele frequency (MAF), and genotype likelihoods. Genotype likelihoods were estimated with the GATK method (-GL 2). The major allele was the most frequent allele (-doMajorMinor 1). We filtered to keep positions covered by at least one read in at least 75% of individuals, with a total coverage below 400 (25 times the number of individuals) to avoid

including repeated regions in the analysis. After such filtering, we exported a list of covered positions for each lake for further analysis, including 1 998 994 058 positions in Cliff Lake, 1 981 247 030 in Indian Lake, 1 658 793 591 in Langfjordvatn, and 1 702 852 520 in Zurich Lake.

From this list of variant and invariant positions, we extracted a list of SNPs as the variable positions with an MAF above 5% and subsequently used this list with their respective major and minor alleles for most analyses (i.e., F_{ST}, overlap with CpG sites). Differentiation between benthic and limnetic species in each lake was measured with F_{ST} statistics, using ANGSD to estimate joint allele frequency spectrum, realSFS functions to compute F_{ST} in sliding windows of 100 KB with a step of 25 KB. Positions were restricted to the polymorphic SNPs (>5% MAF) previously polarized as major or minor allele (options –sites and –doMajorMinor 3), and which were covered in at least 75% of the samples in each species. The pipeline used to call and analyse genetic variation with ANGSD is available at https://github.com/clairemerot/angsd_pipeline.

Frequency of SNPs in CpG sites

We determined whether CpG sites were more polymorphic than the rest of the genome, which would be indicative of CpG methylation-induced mutagenesis (Jones et al. 1992; Gorelick 2003). We started by using fastaRegexFinder.py (https://github.com/dariober/bioinformatics-cafe/blob/master/fastaRegexFinder/) to find all CpG sites in each reference genome. Next, we restricted the CpG site lists to include only sites with sufficient coverage in the WGS data using bedtools *intersect* (Quinlan and Hall 2010) with the list of covered positions for each lake exported earlier (>75% of samples covered). We used regioneR (Gel et al. 2016) to determine how many SNPs fell within CpG sites using the *numOverlaps* command. We assumed a uniform distribution of SNPs across the genome as a null hypothesis for both tests, thus a higher proportion of SNPs in CpGs relative to average polymorphism in the genome would indicate an enrichment of point mutations in CpGs.

$$\frac{n(SNPs\ in\ CpGs)}{n(CpG\ nucleotides)\ covered\ by\ WGS\ data}\ v.s.\ \frac{n(SNPs)}{length\ of\ genome\ covered\ by\ WGS\ data}$$

We also used a permutation test to determine if the enrichment of SNPs in CpG sites would occur by chance over 10 000 iterations. We determined the proportion of the genome made up of CpG sites by dividing the number of CpG nucleotides by the total length of the genome with sufficient WGS coverage. We assumed a uniform distribution of SNPs across the genome as a null hypothesis for both tests, thus a higher proportion of SNPs in CpG sites relative to the proportion of the genome made up of CpG sites would indicate an enrichment of point mutations in CpGs.

$$\frac{n(SNPs\ in\ CpGs)}{n(SNPs)}\ v.s.\ \frac{n(CpG\ nucleotides)\ covered\ by\ WGS\ data}{length\ of\ genome\ covered\ by\ WGS\ data}$$

We also tested whether there was an enrichment of specific SNP substitution types as C/T polymorphisms are expected to be more common due to deamination of methylated cytosines (Jones et al. 1992; Gorelick 2003; Tomkova and Schuster-Böckler 2018). For this analysis, we only considered C and G sites in the genome due to the increased mutagenicity of these nucleotides (Kiktev et al. 2018). We split the list of SNPs into three separate files for each possible mutation: (i) all C/T and G/A SNPs (where G/A SNPs would indicate a C/T mutation at the G position in the reverse complement of the DNA), (ii) all C/A and T/G SNPs, and (iii) all C/G SNPs. We also generated a list of all C and G sites in the genome with sufficient WGS coverage. We compared the rate of finding each SNP type in a CpG site to the rate of finding that SNP type in C and G sites, then used Pearson's chi-squared test to determine if the proportions were significantly different. All scripts are available at https://github.com/cvenney/ga_permutation.

$$\frac{n(specific\ SNPs\ in\ CpGs)}{n(CpG\ nucleotides)}\ v.s.\ \frac{n(specific\ SNPs\ in\ genome)}{n(C\ and\ G\ sites\ in\ the\ genome\ covered\ by\ WGS\ data)}$$

Whole genome DNA methylation tabulation

Raw methylation data were trimmed using fastp (Chen et al. 2018) to remove sequences with phred quality less than 25, length less than 100 bp, and to remove the first and last nucleotides which have high sequencing error rate. After trimming, lake and European whitefish methylation data were analysed separately. Trimmed sequences were aligned to the same reference genomes for each continental lineage as described in the WGS methods using bwa-

meth (https://github.com/brentp/bwa-meth). Alignments with mapping quality greater than 10 were outputted to a BAM file using samtools (Li et al. 2009) and duplicate sequences were removed using Picard tools. Methyldackel's *mbias* function was used to inform trimming of biased methylation calls at the beginning and end of reads, then CpG-specific methylation calling was performed on bias-trimmed data using methyldackel's *extract* function (https://github.com/dpryan79/MethylDackel). The pipeline is available at https://github.com/enormandeau/bwa-meth_pipeline.

Individual-level whitelisting of methylation data using SNP data

C/T SNPs cannot be discerned from true methylation reads because bisulfite conversion leaves methylated cytosine unchanged but converts unmethylated cytosine to uracil which is sequenced as a thymine. Therefore, CpG sites which overlap with C/T or G/A SNPs must be removed from the analysis. More specifically, we can use an individual-level approach where we filter CpG sites separately for each sample based on that sample's genotype. Relying on our WGS data, we thus created a whitelist of CpG sites for each sample, keeping the sites homozygous for C and G and masking individuals heterozygous or homozygous for the T or A allele at that site. To do so, we kept CpG sites that were covered in at least 12 individuals in a given lake (75% of samples for each lake). Then we filtered the whitelists, requiring sites to be either (i) non-variant positions, (ii) not a C/T SNP for the C position of CpGs in the lake, (iii) not a G/A SNPs for the G position of CpGs in the lake, or (iv) a C/T SNP or a G/A SNP in other individuals within the lake, but where this individual has a likelihood greater than 0.7 to be a C/C and G/G homozygote. This resulted in individual-level SNP masking for each sample where CpG sites with C/T and G/A SNPs were removed in heterozygous individuals, but not blindly across all samples. BedGraph files were filtered to include only CpG sites covered in the whitelist (i.e., where both the C and G positions were whitelisted). The pipeline for SNP masking and all subsequent methylation analysis is available at https://github.com/cvenney/ga_methyl.

Coverage filtration and differential methylation analysis

Whitelisted bedGraph files for each sample were filtered to exclude CpG sites with less than five and more than 100 reads. The files were imported into R (R Core Team 2022) for further coverage filtration. To test for differential methylation between species, we kept only

CpG sites with sufficient coverage in at least four limnetic and four benthic samples in the lake. BedGraph files were reformatted for further analysis with DSS (Park and Wu 2016). Methylation data were smoothed over 500 bp regions using the built-in moving average algorithm in DSS to control for spatial correlation of methylation levels among proximal CpGs. GLMs were used to identify DMLs which were considered significant if the false discovery rate corrected p-value was less than 0.05. DMRs were identified in DSS as regions with many statistically significant DMLs with p-values less than 0.05.

Overlap between DMLs and outlier SNPs

We assessed whether there was an enrichment of outlier SNP-DML overlaps compared to the number of overlaps expected by chance given the frequency of polymorphic DMLs and the frequency of outlier SNPS in CpG sites. Such an enrichment could indicate epigenetically influenced mutagenesis and genetic assimilation of methylation changes into stable genetic variants. We generated a list of highly differentiated SNPs for each lake, hereafter called "outlier SNPs", retaining only the top 5% of SNPs with the greatest F_{ST} between species. DML test files were converted to bed format for input into bedtools, then *intersect* was used to determine the number of overlaps between outlier SNPs and DMLs. We used only CpG sites that were (i) covered by WGS data, (ii) covered by WGBS data, and (iii) polymorphic to account for the observed elevated mutation rate in CpG sites compared to the rest of the genome. We then used Pearson's chi-squared test to determine if there was an enrichment of observed DML-outlier SNP overlaps in polymorphic CpG sites compared to the expected rate of finding DML-outlier SNP overlaps in CpG sites.

$$\frac{n(DMLs\ in\ SNPs)}{n(polymorphic\ CpG\ sites)}\ x\ \frac{n(outlier\ SNPs\ in\ covered\ CpG\ sites)}{n(polymorphic\ CpG\ sites)}\ \ vs.\ \ \frac{n(overlaps\ between\ DMLs\ and\ outlier\ SNPs)}{n(polymorphic\ CpG\ sites)}$$

We determined if percent methylation was associated with the frequency of cytosine alleles at the outlier SNP-DML overlaps, providing further evidence for methylation influencing the conversion of cytosine to alternative alleles. We compared the frequency of the cytosine allele for each species in each population to the average percent methylation of the DML at the same genomic position. We also performed permutation tests to estimate the correlation between percent methylation and the frequency of cytosine alleles using subsets of polymorphic CpG

sites to determine if the same correlation holds true across the entire genome, or whether the sites of potential genetic assimilation show a stronger correlation, implying that mutations at these sites might be favoured or retained. Subsets of CpG sites were randomly selected based on the total number of outlier SNP-DML overlaps for each lake then 100 000 permutations were performed to estimate the possible range of correlations using genome-wide CpG sites.

We performed GO enrichment analysis for the outlier SNP-DML overlaps for each lake. The whitefish genomes were first annotated with the GAWN pipeline (https://github.com/enormandeau/gawn) using the *Salvelinus namaycush* (GCF_018398675.1) transcriptome from GenBank. We then retrieved the genes that overlapped the SNP-DML overlap positions for each lake and proceeded to GO enrichment tests using the go_enrichment pipeline (https://github.com/enormandeau/go_enrichment). Results were filtered to remove broad terms (depths 0 and 1) and to require a Benjamini-Hochberg false discovery rate corrected p-value of 0.1 or less.

Acknowledgements

We would like to thank Anne-Laure Ferchaud and Kyle Wellband for helpful discussions on the project. We thank Kim Praebel, Shripathi Bhat and the Freshwater Ecology group at UiT (Tromso) for providing samples from Norway, and Ole Seehausen for providing samples from Switzerland. This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to L.B (CG053918). WGS data for *Coregonus clupeaformis* and *C. lavaretus* are available online at SRA BioProject PRJNA820751 and [currently being uploaded], respectively. WGBS data are available online at BioProject PRJNA559821.

References

Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res. 25:4692–4693.

Anastasiadi D, Díaz N, Piferrer F. 2017. Small ocean temperature increases elicit stage-dependent changes in DNA methylation and gene expression in a fish, the European sea bass. Sci. Rep. 7:12401.

Anastasiadi D, Venney CJ, Bernatchez L, Wellenreuther M. 2021. Epigenetic inheritance and reproductive mode in plants and animals. Trends Ecol. Evol. 36:1124–1140.

Angers B, Castonguay E, Massicotte R. 2010. Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. Mol. Ecol. 19:1283–1295.

Artemov A V., Mugue NS, Rastorguev SM, Zhenilo S, Mazur AM, Tsygankova S V., Boulygina ES, Kaplun D, Nedoluzhko A V., Medvedeva YA, et al. 2017. Genome-wide DNA methylation profiling reveals epigenetic adaptation of stickleback to marine and freshwater conditions. Mol. Biol. Evol. 34:2203–2213.

Ashe A, Colot V, Oldroyd BP. 2021. How does epigenetics influence the course of evolution? Philos. Trans. R. Soc. B Biol. Sci. 376:20200111.

Beemelmanns A, Ribas L, Anastasiadi D, Moraleda-Prados J, Zanzuzzo FS, Rise ML, Gamperl AK, Zanuzzo FS, Rise ML, Gamperl AK. 2021. DNA methylation dynamics in Atlantic salmon (*Salmo salar*) challenged with high temperature and moderate hypoxia. Front. Mar. Sci. 7:3.

Berbel-Filho WM, Berry N, Rodríguez-Barreto D, Rodrigues Teixeira S, Garcia de Leaniz C, Consuegra S. 2020. Environmental enrichment induces intergenerational behavioural and epigenetic effects on fish. Mol. Ecol. 29:2288–2299.

Bernatchez L, Chouinard A, Lu G. 1999. Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, *Coregonus* sp., as a case study. Biol. J. Linn. Soc. 68:173–194.

Bernatchez L, Dodson JJ. 1991. Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. Evolution (N. Y). 45:1016–1035.

Bernatchez L, Dodson JJ. 1994. Phylogenetic relationships among Palearctic and Nearctic whitefish (*Coregonus* sp.) populations as revealed by mitochondrial DNA variation. Can. J. Fish. Aquat. Sci. 51:240–251.

Bernatchez L, Renaut S, Whiteley AR, Derome N, Jeukens J, Landry L, Lu G, Nolte AW, Østbye K, Rogers SM, et al. 2010. On the origin of species: insights from the ecological genomics of lake whitefish. Philos. Trans. R. Soc. B Biol. Sci. 365:1783–1800.

Bird A. 2002. DNA methylation patterns and epigenetic memory. Genes Dev. 16:6–21.

Breese MR, Liu Y. 2013. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics 29:494–496.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.

Dalziel AC, Martin N, Laporte M, Guderley H, Bernatchez L. 2015. Adaptation and acclimation of aerobic exercise physiology in lake whitefish ecotypes (*Coregonus clupeaformis*). Evolution (N. Y). 69:2167–2186.

Danchin E, Pocheville A, Rey O, Pujol B, Blanchet S. 2019. Epigenetically facilitated mutational assimilation: epigenetics as a hub within the inclusive evolutionary synthesis. Biol. Rev. 94:259–282.

De-Kayne R, Selz OM, Marques DA, Seehausen O, Feulner PGD, Frei D. 2022. Genomic architecture of adaptive radiation and hybridization in Alpine white fish. Nat. Commun. 13:1–13.

De-Kayne R, Zoller S, Feulner PGD. 2020. A de novo chromosome-level genome assembly of *Coregonus* sp. “Balchen”: one representative of the Swiss Alpine whitefish radiation. Mol. Ecol. Resour. 20:1093–1109.

Derome N, Duchesne P, Bernatchez L. 2006. Parallelism in gene transcription among sympatric

lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. Mol. Ecol. 15:1239–1249.

Dion-Côté AM, Symonová R, Lamaze FC, Pelikánová Š, Ráb P, Bernatchez L. 2017. Standing chromosomal variation in lake whitefish species pairs: the role of historical contingency and relevance for speciation. Mol. Ecol. 26:178–192.

Elmer KR, Fan S, Gunter HM, Jones JC, Boekhoff S, Kuraku S, Meyer A. 2010. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. Mol. Ecol. 19:197–211.

Feulner PGD, Seehausen O. 2019. Genomic insights into the vulnerability of sympatric whitefish species flocks. Mol. Ecol. 28:615-629.

Flores KB, Wolschin F, Amdam G V. 2013. The role of methylation of DNA in environmental adaptation. Integr. Comp. Biol. 53:359–372.

Gagnaire PA, Pavey SA, Normandeau E, Bernatchez L. 2013. The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution (N. Y). 67:2483–2497.

Gavery MR, Nichols KM, Goetz GW, Middleton MA, Swanson P. 2018. Characterization of genetic and epigenetic variation in sperm and red blood cells from adult hatchery and natural-origin steelhead, *Oncorhynchus mykiss*. G3 Genes, Genomes, Genet. 8:3723–3736.

Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016. RegioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 32:289–291.

Goldberg ME, Harris K. 2022. Mutational signatures of replication timing and epigenetic modification persist through the global divergence of mutation spectra across the great ape phylogeny. Genome Biol. Evol. 14:1–18.

Gorelick R. 2003. Evolution of dioecy and sex chromosomes via methylation driving Muller's ratchet. Biol. J. Linn. Soc. 80:353–368.

Greenspoon PB, Spencer HG, M'Gonigle LK. 2022. Epigenetic induction may speed up or slow down speciation with gene flow. Evolution (N. Y). 76:1170–1182.

Gudbrandsson J, Franzdóttir SR, Kristjánsson BK, Ahi EP, Maier VH, Kapralova KH, Snorrason SS, Jónsson ZO, Pálsson A. 2018. Differential gene expression during early development in recently evolved and sympatric Arctic charr morphs. PeerJ 2018:1–35.

Heckwolf MJ, Meyer BS, Häsler R, Höppner MP, Eizaguirre C, Reusch TBH. 2020. Two different epigenetic information channels in wild three-spined sticklebacks are involved in salinity adaptation. Sci. Adv. 6.

Hu J, Wuitchik SJSS, Barry TN, Jamniczky HA, Rogers SM, Barrett RDH. 2021. Heritability of DNA methylation in threespine stickleback (*Gasterosteus aculeatus*). Genetics 217.

Jacobs A, Elmer KR. 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. Mol. Ecol. 30:4955-4969.

Jacobsen MW, Hansen MM, Orlando L, Bekkevold D, Bernatchez L, Willerslev E, Gilbert MTP. 2012. Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). Mol. Ecol. 21:2727–2742.

Jeukens J, Bittner D, Knudsen R, Bernatchez L. 2009. Candidate genes and adaptive radiation: insights from transcriptional adaptation to the limnetic niche among coregonine fishes (*Coregonus* spp., *Salmonidae*). Mol. Biol. Evol. 26:155–166.

Jones PA, Rideout WM, Shen J-C, Spruck CH, Tsai YC. 1992. Methylation, mutation and cancer. BioEssays 14:33–36.

Kiktev DA, Sheng Z, Lobachev KS, Petes TD. 2018. GC content elevates mutation and recombination rates in the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U. S. A. 115:E7109–E7118.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15:1–13.

Lallias D, Bernard M, Ciobotaru C, Dechamp N, Labbé L, Goardon L, Le Calvez JM, Bideau M, Fricot A, Prézelin A, et al. 2021. Sources of variation of DNA methylation in rainbow trout: combined effects of temperature and genetic background. Epigenetics 16:1031–1052.

Laporte M, Dalziel AC, Martin N, Bernatchez L. 2016. Adaptation and acclimation of traits associated with swimming capacity in lake whitefish (*Coregonus clupeaformis*) ecotypes. BMC Evol. Biol. 16:1–13.

Laporte M, Le Luyer J, Rougeux C, Dion-Côté AM, Krick M, Bernatchez L. 2019. DNA methylation reprogramming, TE derepression, and postzygotic isolation of nascent animal species. Sci. Adv. 5:1–9.

Laporte M, Rogers SM, Dion-Côté AM, Normandeau E, Gagnaire PA, Dalziel AC, Chebib J, Bernatchez L. 2015. RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in lake whitefish (*Coregonus clupeaformis*) species pairs. G3 Genes, Genomes, Genet. 5:1481–1491.

Leitwein M, Laporte M, Le Luyer J, Mohns K, Normandeau E, Withler R, Bernatchez L. 2021. Epigenomic modifications induced by hatchery rearing persist in germ line cells of adult salmon after their oceanic migration. Evol. Appl. 14:2402–2413.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv 1303.3997v.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Li L, Gao Y, Wu Q, Cheng ASL, Yip KY. 2019. New guidelines for DNA methylome studies regarding 5-hydroxymethylcytosine for understanding transcriptional regulation. Genome Res. 29:543–553.

Lou RN, Jacobs A, Wilder A, Therkildsen NO. 2020. A beginner's guide to low-coverage whole genome sequencing for population genomics. Mol. Ecol. 30:5966-5993.

Lu G, Bernatchez L. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. Evolution (N. Y). 53:1491–1505.

Le Luyer J, Laporte M, Beacham TD, Kaukinen KH, Withler RE, Leong JS, Rondeau EB, Koop BF, Bernatchez L. 2017. Parallel epigenetic modifications induced by hatchery rearing in a

Pacific salmon. Proc. Natl. Acad. Sci. 114:12964–12969.

Manousaki T, Hull PM, Kusche H, MacHado-Schiaffino G, Franchini P, Harrod C, Elmer KR, Meyer A. 2013. Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. Mol. Ecol. 22:650–669.

Marques DA, Meier JI, Seehausen O. 2019. A combinatorial view on speciation and adaptive radiation. Trends Ecol. Evol. 34:531–544.

Matlosz S, Sigurgeirsson B, Rut S, Pálsson A, Jónsson ZO, Franzdóttir SR, Pálsson A, Jónsson ZO. 2022. DNA methylation differences during development distinguish sympatric morphs of Arctic charr (*Salvelinus alpinus*). Mol. Ecol. 31:4739–4761.

McCaw BA, Stevenson TJ, Lancaster LT. 2020. Epigenetic responses to temperature and climate. Integr. Comp. Biol. 60:1469–1480.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Mckenzie JL, Araújo HA, Smith JL, Schluter D, Devlin RH, Mckenzie JL. 2021. Incomplete reproductive isolation and strong transcriptomic response to hybridization between sympatric sister species of salmon. Proc. R. Soc. B Biol. Sci. 288:20203020.

Mérot C, Stenløkk KSR, Venney C, Laporte M, Moser M, Normandeau E, Árnyasi M, Kent M, Rougeux C, Flynn JM, et al. 2022. Genome assembly, structural variants, and genetic differentiation between lake whitefish young species pairs (*Coregonus* sp.) with long and short reads. Mol. Ecol:https://doi.org/10.1111/mec.16468.

Metzger DCH, Schulte PM. 2017. Persistent and plastic effects of temperature on DNA methylation across the genome of threespine stickleback (*Gasterosteus aculeatus*). Proc. R. Soc. B Biol. Sci. 284:20171667.

Nishikawa K, Kinjo AR. 2018. Mechanism of evolution by genetic assimilation. Biophys. Rev. 10:667–676.

Østbye K, Amundsen PA, Bernatchez L, Klemetsen A, Knudsen R, Kristoffersen R, Næsje TF, Hindar K. 2006. Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. Mol. Ecol. 15:3983–4001.

Østbye K, Bernatchez L, Næsje TF, Himberg KJM, Hindar K. 2005. Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. Mol. Ecol. 14:4371–4387.

Østbye K, Næsje TF, Bernatchez L, Sandlund OT, Hindar K. 2005. Morphological divergence and origin of sympatric populations of European whitefish (*Coregonus lavaretus* L.) in Lake Femund, Norway. J. Evol. Biol. 18:683–702.

Pál C, Miklós I. 1999. Epigenetic inheritance, genetic assimilation and speciation. J. Theor. Biol. 200:19–37.

Park Y, Wu H. 2016. Differential methylation analysis for BS-seq data under general experimental design. Bioinformatics 32:1446–1453.

Pavey SA, Collin H, Nosil P, Rogers SM. 2010. The role of gene expression in ecological speciation. Ann. N. Y. Acad. Sci. 1206:110–129.

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP. 2010. Phenotypic plasticity's impacts on diversification and speciation. Trends Ecol. Evol. 25:459–467.

Pigeon D, Chouinard A, Bernatchez L. 1997. Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). Evolution (N. Y). 51:196–205.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

R Core Team. 2022. R: A language and environment for statistical computing. Available from: https://www.r-project.org

Richards CL, Bossdorf O, Pigliucci M. 2010. What role does heritable epigenetic variation play

in phenotypic evolution? Bioscience 60:232–237.

Richards EJ. 2006. Inherited epigenetic variation - revisiting soft inheritance. Nat. Rev. Genet. 7:395–402.

Rogers SM, Bernatchez L. 2006. The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). J. Evol. Biol. 19:1979–1994.

Rogers SM, Bernatchez L. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs. Mol. Biol. Evol. 24:1423–1438.

Rogers SM, Gagnon V, Bernatchez L. 2002. Genetically based phenotype-environment association for swimming behavior in lake whitefish ecotypes (*Coregonus clupeaformis* Mitchill). Evolution (N. Y). 56:2322–2329.

Rougeux C, Bernatchez L, Gagnaire PA. 2017. Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). Genome Biol. Evol. 9:2057–2074.

Rougeux C, Gagnaire PA, Bernatchez L. 2019. Model-based demographic inference of introgression history in European whitefish species pairs'. J. Evol. Biol. 32:806–817.

Rougeux C, Gagnaire PA, Praebel K, Seehausen O, Bernatchez L. 2019. Polygenic selection drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic sister species complex. Mol. Ecol. 28:4388–4403.

Ryu T, Veilleux HD, Donelson JM, Munday PL, Ravasi T. 2018. The epigenetic landscape of transgenerational acclimation to ocean warming. Nat. Clim. Chang. 8:504–509.

Ryu T, Veilleux HD, Munday PL, Jung I, Donelson JM, Ravasi T. 2020. An epigenetic signature for within-generational plasticity of a reef fish to ocean warming. Front. Mar. Sci. 7:1–15.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre GP, Bank C, Brännström Å, et al. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176–192.

Siwertsson A, Knudsen R, Præbel K, Adams CE, Newton J, Amundsen PA. 2013. Discrete foraging niches promote ecological, phenotypic, and genetic divergence in sympatric whitefish (*Coregonus lavaretus*). Evol. Ecol. 27:547–564.

Smith TA, Martin MD, Nguyen M, Mendelson TC. 2016. Epigenetic divergence as a potential first step in darter speciation. Mol. Ecol. 25:1883–1894.

St-Cyr J, Derome N, Bernatchez L. 2008. The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). Mol. Ecol. 17:1850–1870.

Stajic D, Jansen LET. 2021. Empirical evidence for epigenetic inheritance driving evolutionary adaptation. Philos. Trans. R. Soc. B Biol. Sci. 376:20200121.

Stajic D, Perfeito L, Jansen LET. 2019. Epigenetic gene silencing alters the mechanisms and rate of evolutionary adaptation. Nat. Ecol. Evol. 3:491–498.

Thorson JLM, Smithson M, Beck D, Sadler-Riggleman I, Nilsson E, Dybdahl M, Skinner MK. 2017. Epigenetics and adaptive phenotypic variation between habitats in an asexual snail. Sci. Rep. 7:1–11.

Tomkova M, Schuster-Böckler B. 2018. DNA modifications: naturally more error prone? Trends Genet. 34:627–638.

Trefts E, Gannon M, Wasserman DH. 2017. The liver. Curr. Biol. 27:R1147–R1151.

Trudel M, Tremblay A, Schetagne R, Rasmussen JB. 2001. Why are dwarf fish so small? An energetic analysis of polymorphism in lake whitefish (*Coregonus clupeaformis*). Can. J. Fish. Aquat. Sci. 58:394–405.

Venney CJ, Cayuela H, Rougeux C, Laporte M, Mérot C, Normandeau E, Leitwein M, Dorant Y, Præbel K, Kenchington EL, et al. (forthcoming) Genome-wide DNA methylation predicts environmentally-driven life history variation in a marine fish. Evolution (N. Y).

Venney CJ, Wellband KW, Heath DD. 2020. Rearing environment affects the genetic architecture and plasticity of DNA methylation in Chinook salmon. Heredity (Edinb).:38–49.

Venney CJ, Wellband KW, Normandeau E, Houle C, Garant D, Audet C, Bernatchez L. 2022.

Thermal regime during parental sexual maturation, but not during offspring rearing, modulates DNA methylation in brook charr (*Salvelinus fontinalis*). Proc. R. Soc. B Biol. Sci. 289:20220670.

Vernaz G, Hudson AG, Santos ME, Fischer B, Carruthers M, Shechonge AH, Gabagambi NP, Tyers AM, Ngatunga BP, Malinsky M, et al. 2022. Epigenetic divergence during early stages of speciation in an African crater lake cichlid fish. Nat. Ecol. Evol.:1–12.

Vogt G. 2017. Evolution of epigenetic mechanisms in animals and their role in speciation. In: Handbook of Epigenetics: The New Molecular and Medical Genetics. Second Edi. Elsevier Inc. p. 409–426.

Vogt G. 2021. Epigenetic variation in animal populations: sources, extent, phenotypic implications, and ecological and evolutionary relevance. J. Biosci. 46.

Wellband K, Roth D, Linnansaari T, Curry RA, Bernatchez L. 2021. Environment-driven reprogramming of gamete DNA methylation occurs during maturation and is transmitted intergenerationally in Atlantic salmon. G3 Genes, Genomes, Genet. 11:jkab353.

Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: raw material for evolution. Mol. Ecol. 15:1197–1211.

Xia J, Han L, Zhao Z. 2012. Investigating the relationship of DNA methylation with mutation rate and allele frequency in the human genome. BMC Genomics 13 Suppl 8:S7.

Tables

Table 1: Pairwise F_{ST} matrices for (A) *C. clupeaformis* and (B) *C. lavaretus* reveal variable but significant genetic divergence between lakes and species.

A)

Cliff benthic	0.175		
Indian limnetic	0.084	0.162	
Indian benthic	0.146	0.182	0.098
	Cliff limnetic	**Cliff benthic**	**Indian limnetic**

B)

Langfjordvatn benthic	0.034		
Zurich limnetic	0.240	0.242	
Zurich benthic	0.237	0.239	0.043
	Langfjordvatn limnetic	**Langfjordvatn benthic**	**Zurich limnetic**

Table 2: Pearson's chi-squared test results for the rate of polymorphism in CpG sites (i) for all SNP types relative to the rate of polymorphism across the entire genome, and (ii) for specific SNP types relative to all C and G positions in the genome.

SNP Type	Value	Cliff	Indian	Langfjordvatn	Zurich
All	SNPs	11 861 765	12 727 697	30 919 358	24 221 888
	SNPs in CpGs	1 250 313	1 349 717	3 795 855	2 879 669
	Polymorphic CpGs (%)	10.5	10.6	12.3	11.9
	Polymorphic sites (%)	3.8	3.8	3.8	3.8
	Fold change	2.8	2.8	3.2	3.1
	p-value (Chi-squared test)	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***
	p-value (permutation test)	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***
C/T and G/A	SNPs	2 804 547	3 080 961	8 321 890	6 251 941
	SNPs in CpGs	750 311	819 314	2 487 283	1 834 278
	SNPs in CpGs (%)	0.98	1.09	3.94	2.81
	SNPs in all Cs and Gs (%)	0.32	0.36	1.17	0.85
	Fold change	3.0	3.0	3.4	3.3
	p-value (Chi-squared test)	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***
C/G and G/C	SNPs	956 737	1 012 702	2 253 168	1 810 523
	SNPs in CpGs	95 721	100 779	239 633	198 157
	SNPs in CpGs (%)	0.13	0.13	0.38	0.30
	SNPs in all Cs and Gs (%)	0.11	0.12	0.32	0.25
	Fold change	1.1	1.1	1.2	1.2
	p-value (Chi-squared test)	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***
C/A and G/T	SNPs	2 124 638	2 314 058	5 422 615	4 126 260
	SNPs in CpGs	234 886	248 499	590 917	476 256
	SNPs in CpGs (%)	0.31	0.33	0.94	0.73
	SNPs in all Cs and Gs (%)	0.25	0.27	0.76	0.56
	Fold change	1.3	1.2	1.2	1.3
	p-value (Chi-squared test)	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***

Table 3: Statistics from differential methylation analysis, including coverage depth and number of CpGs analysed after all coverage trimming, number of DMLs and DMRs identified using a significance level of $p < 0.05$, and information on outlier SNP-DML overlaps and Pearson chi-squared test results.

Statistic	**Cliff**	**Indian**	**Langfjordvatn**	**Zurich**
Average coverage depth (WGBS)	7.5	7.6	8.1	8.1
Total CpGs analysed	17 280 816	19 605 766	13 236 657	18 685 215
DMLs ($p < 0.05$)	41 116	26 583	4 030	8 004
DMRs	3 019	2 383	402	754
CpGs in DMRs	1 005 874	667 887	63 597	122 700
Outlier SNPs (5% highest F_{ST})	593 088	636 384	1 545 963	1 211 091
Outlier SNP-DML overlaps	577	483	151	388
Expected outlier SNP-DML overlaps	15.6	21.1	13.2	22.3
Fold enrichment	36.9	22.9	11.4	17.4
p-value	***< 0.001***	***< 0.001***	***< 0.001***	***< 0.001***

Table 4: Gene ontology results for outlier SNP-DML overlaps in lake and European whitefish. There were 24 enriched terms from Cliff Lake, four from Indian Lake, two from Langfjordvatn Lake, and 32 from Zurich Lake. Significant results had a Benjamini-Hochberg false discovery rate corrected p-value less than 0.1 and were filtered for GO term depth greater than 1 to remove broad terms.

GO term	Domain	GO term name	Depth	p-value (FDR)
		Cliff Lake		
GO:0050912	BP	detection of chemical stimulus involved in sensory perception of taste	5	0.050
GO:0002414	BP	immunoglobulin transcytosis in epithelial cells	6	0.006
GO:0001580	BP	detection of chemical stimulus involved in sensory perception of bitter taste	6	0.036
GO:0002415	BP	immunoglobulin transcytosis in epithelial cells mediated by polymeric immunoglobulin receptor	7	0.006
GO:0071746	CC	IgA immunoglobulin complex, circulating	4	0.001
GO:0042571	CC	immunoglobulin complex, circulating	3	0.001
GO:0009986	CC	cell surface	2	0.018
GO:0043235	CC	receptor complex	2	0.033
GO:0019814	CC	immunoglobulin complex	2	0.040
GO:0016020	CC	membrane	2	0.061
GO:0071745	CC	IgA immunoglobulin complex	3	0.001
GO:0005886	CC	plasma membrane	3	0.011
GO:0045202	CC	synapse	3	0.040
GO:0031226	CC	intrinsic component of plasma membrane	3	0.095
GO:0098978	CC	glutamatergic synapse	4	0.029
GO:0005887	CC	integral component of plasma membrane	4	0.033
GO:0071749	CC	polymeric IgA immunoglobulin complex	5	0.001
GO:0030425	CC	dendrite	5	0.029
GO:0030424	CC	axon	5	0.093
GO:0071751	CC	secretory IgA immunoglobulin complex	6	0.001
GO:0043167	MF	ion binding	2	0.063
GO:0019763	MF	immunoglobulin receptor activity	4	0.013
GO:0001792	MF	polymeric immunoglobulin receptor activity	5	0.004
GO:0005154	MF	epidermal growth factor receptor binding	5	0.004
		Indian Lake		
GO:0031224	CC	intrinsic component of membrane	2	0.082
GO:0031226	CC	intrinsic component of plasma membrane	3	0.089
GO:0000139	CC	Golgi membrane	5	0.094
GO:0042588	CC	zymogen granule	9	0.028
		Langfjordvatn Lake		
GO:0004598	MF	peptidylamidoglycolate lyase activity	5	0.027
GO:0004504	MF	peptidylglycine monooxygenase activity	5	0.027
		Zurich Lake		

GO:0050789	BP	regulation of biological process	2	0.033
GO:0097485	BP	neuron projection guidance	3	0.033
GO:0050794	BP	regulation of cellular process	3	0.036
GO:0061061	BP	muscle structure development	3	0.081
GO:0002326	BP	B cell lineage commitment	4	0.018
GO:0007411	BP	axon guidance	4	0.033
GO:0007517	BP	muscle organ development	4	0.045
GO:0007400	BP	neuroblast fate determination	4	0.065
GO:0002336	BP	B-1 B cell lineage commitment	5	0.010
GO:0050920	BP	regulation of chemotaxis	5	0.018
GO:0035290	BP	trunk segmentation	5	0.033
GO:0033860	BP	regulation of NAD(P)H oxidase activity	5	0.033
GO:0060268	BP	negative regulation of respiratory burst	5	0.035
GO:0035206	BP	regulation of hemocyte proliferation	5	0.040
GO:0048749	BP	compound eye development	6	0.033
GO:0050922	BP	negative regulation of chemotaxis	6	0.035
GO:0060264	BP	regulation of respiratory burst involved in inflammatory response	7	0.033
GO:0061332	BP	Malpighian tubule bud morphogenesis	7	0.033
GO:0002232	BP	leukocyte chemotaxis involved in inflammatory response	7	0.045
GO:0010726	BP	positive regulation of hydrogen peroxide metabolic process	7	0.093
GO:0060266	BP	negative regulation of respiratory burst involved in inflammatory response	8	0.033
GO:0010729	BP	positive regulation of hydrogen peroxide biosynthetic process	8	0.065
GO:1902623	BP	negative regulation of neutrophil migration	9	0.033
GO:2000508	BP	regulation of dendritic cell chemotaxis	9	0.033
GO:0010758	BP	regulation of macrophage chemotaxis	9	0.052
GO:1905522	BP	negative regulation of macrophage migration	9	0.066
GO:0090025	BP	regulation of monocyte chemotaxis	9	0.086
GO:0002689	BP	negative regulation of leukocyte chemotaxis	9	0.090
GO:2000509	BP	negative regulation of dendritic cell chemotaxis	10	0.010
GO:0090027	BP	negative regulation of monocyte chemotaxis	10	0.033
GO:0010760	BP	negative regulation of macrophage chemotaxis	10	0.033
GO:0090383	BP	phagosome acidification	12	0.074

Figures

Figure 1: Sympatric benthic-limnetic species pairs were sampled from four lakes. Lake whitefish (*C. clupeaformis*) were sampled from two lakes in North America (Cliff Lake and Indian Lake in Maine, USA) and European whitefish (*C. lavaretus*) were sampled from two lakes in Europe (Langfjordvatn Lake, Norway and Zurich Lake, Switzerland).

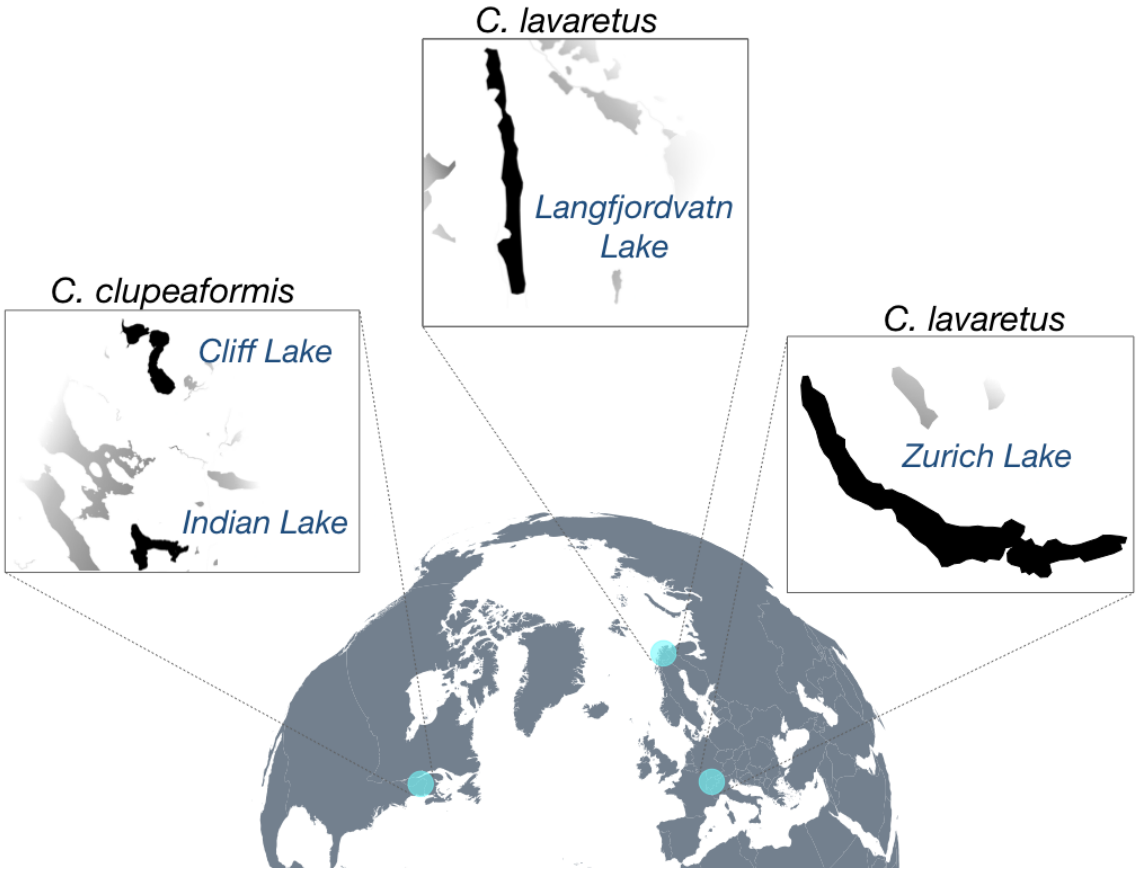

Figure 2: Differential methylation analysis comparing limnetic-benthic species pairs identified (A) 3 019 DMRs in Cliff Lake, (B) 2 383 DMRs in Indian Lake, (C) 402 DMRs in Langfjordvatn Lake, and (D) 754 DMRs in Zurich Lake. Limnetic (navy) and benthic (gold) species are generally neatly differentiated in the dendrograms shown on the x-axes. Percent methylation is displayed for each DMR from 0% (yellow) to 100% (indigo) with clear DNA methylation differences between species.

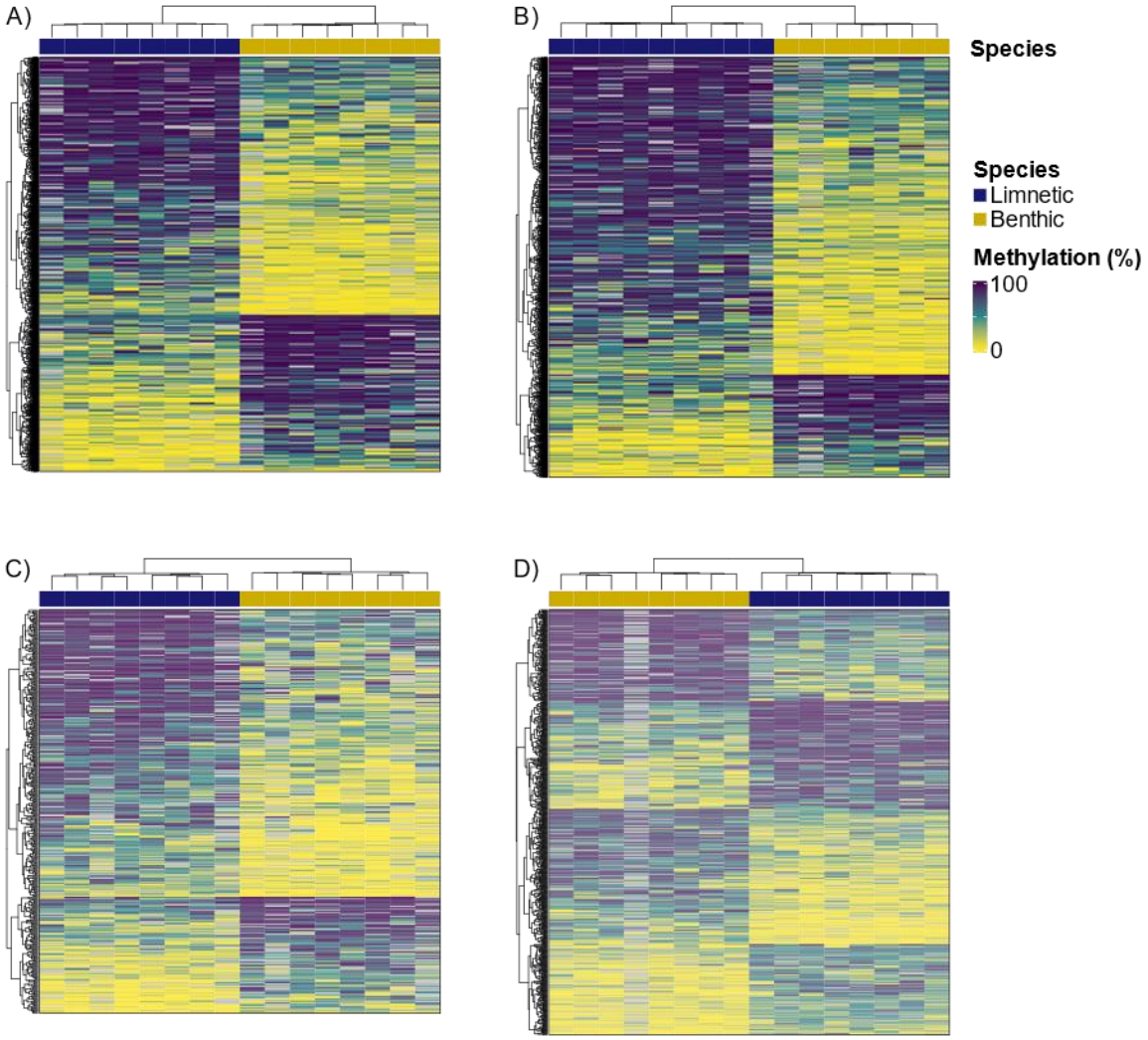
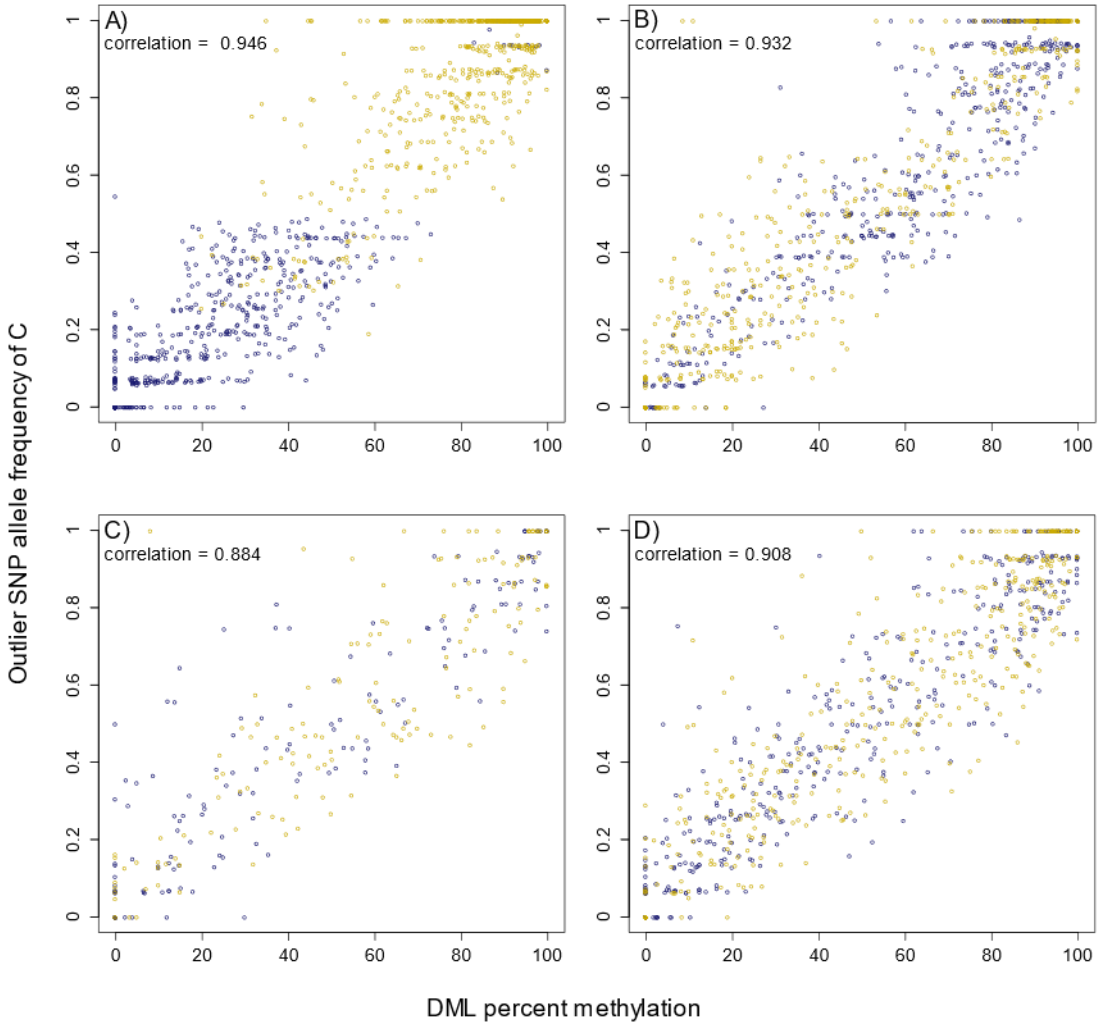
Figure 3: Relationship between DNA methylation and the allele frequency of cytosine alleles for sites undergoing potential genetic assimilation (overlaps between outlier SNPs and DMLs). Each datapoint shows the average percent methylation and frequency of the cytosine allele for that position in the respective species. Limnetic (blue) and benthic (yellow) species show a positive relationship between cytosine allele frequency and percent methylation in (A) Cliff Lake, (B) Indian Lake, (C) Langfjordvatn Lake, and (D) Zurich Lake, suggesting that higher DML methylation may inhibit mutation and maintain the presence of the cytosine allele at that locus.