Received Date : 27-May-2016

Revised Date : 21-Nov-2016

Accepted Date : 29-Nov-2016

Article type : Original Article

INVESTIGATING THE EXTENT OF PARALLELISM IN MORPHOLOGICAL AND GENOMIC DIVERGENCE AMONG LAKE TROUT ECOTYPES IN LAKE SUPERIOR

Alysse Perreault-Payette[1], Andrew M. Muir[2], Frederick Goetz[3], Charles Perrier[5], Eric Normandeau[1], Pascal Sirois[4], Louis Bernatchez[1]

[1] Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, 1030 avenue de la Médecine, Québec, Canada G1V 0A6

[2] Department of Fisheries and Wildlife, Michigan State University, 13 Natural Resources Building, East Lansing, Michigan 48824, USA; and Great Lakes Fishery Commission, 2100 Commonwealth Boulevard, Suite 100, Ann Arbor, Michigan 48105, USA[4]NOAA, Northwest Fisheries Science Center

[3] Northwest Fisheries Science Center, 98366, Port Orchard, Washington, USA

[4] Chaire de recherche sur les espèces aquatiques exploitées, Laboratoire des sciences aquatiques, Département des sciences fondamentales, Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada

[5] Centre d'Écologie Fonctionnelle et Évolutive (UMR CEFE CNRS 5175), Montpellier, France

Corresponding author: Alysse Perreault-Payette, Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, 1030 avenue de la Médecine, Québec, Canada, G1V 0A6, Phone : (1-418) 656-2131 ext. 8455, Email : alysse.perreault@gmail.com

Running title: Genomic divergence of Lake Tout ecotypes

Keywords: population genomics, RADseq, morphometrics, local adaptation, salmonid, ecological speciation

ABSTRACT

Understanding the emergence of species through the process of ecological speciation is a central question in evolutionary biology which also has implications for conservation and management. Lake Trout (*Salvelinus namaycush*) is renowned for the occurrence of different ecotypes linked to resource and habitat use throughout North America. We aimed to unravel the fine genetic structure of the four Lake Trout ecotypes in Lake Superior. A total of 486 individuals from four sites were genotyped at 6822 filtered SNPs using RADseq technology. Our results revealed different extent of morphological and genetic differentiation within the different sites. Overall, genetic differentiation was weak but significant and was on average three times higher between sites (Mean F_{ST} = 0.016) than between ecotypes within sites (Mean F_{ST} = 0.005) indicating higher level of gene flow or a more recent shared ancestor between ecotypes within each site than between populations of the same ecotype. Evidence of divergent selection was also found between ecotypes and/or in association with morphological variation. Outlier loci found in genes related to lipid metabolism and visual acuity were of particular interest in this context of ecotypic divergence. However, we did not find clear indication of parallelism at the genomic level, despite the presence of phenotypic parallelism among some ecotypes from different sampling sites. Overall, the occurrence of different levels of both genomic and phenotypic differentiation between ecotypes within each site with several differentiated loci linked to relevant biological functions support the presence of a continuum of divergence in Lake Trout.

INTRODUCTION

The study of diversification and ultimately speciation is central to evolution and relevant for conservation biology (Weissing et al. 2011). The most common and established mechanism of speciation is divergence in allopatry, where spatial and geographical barriers prevent gene flow, thus allowing genetic incompatibilities to accumulate, subsequently resulting in reproductive isolation following secondary contact (Coyne and Orr 2004; Tittes and Kane 2014). Some examples of allopatric isolation mechanisms

in fishes include the glacial cycles in North America responsible for the origin of many freshwater species (April et al. 2013), the rise and fall of Lake Tanganyika, and the barriers created by high water flow in large rivers such as the Amazon or Congo River (reviewed in Bernardi 2013). However, a geographic barrier is not always needed and speciation can emerge in sympatry, or in parapatry despite high gene flow, by divergent selection on ecologically important traits (Tittes and Kane 2014; Gavrilets et al. 2007). Divergent selection on ecological traits can be caused by biotic and abiotic influences where adaptations to different environments or ecological niches result in the emergence of reproductive incompatibilities (Bernardi 2013; Nosil et al. 2009). The latter may create a continuum of divergence from continuous variation within a single gene pool, to ecotype formation and finally to complete differentiation and reproductive isolation (Lu and Bernatchez 1999; Nosil et al. 2009; Hendry 2009; Gagnaire et al. 2013). Models and case studies have shown that sympatric speciation is possible under gene flow when few loci underlying the divergent trait undergo strong selection, whereas gene flow homogenises the rest of the genome (Gavrilets et al. 2007; Franchini et al. 2013).

Ecological speciation has been extensively documented in several geologically young fish species living in sympatry. For instance sympatric speciation has occurred in Midas cichlids (*Amphilophus* spp.) (Franchini et al. 2013), Lake Victoria cichlids (Wagner et al. 2013) but more commonly in several temperate freshwater fishes such as stickleback (*Gasterosteus* spp.), smelt (*Osmerus* spp.) and especially in salmonids such as whitefish (*Coregonus* spp.), trout (*Salmo* spp.), Pacific salmon (*Oncorhynchus* spp.) and charrs (*Salvelinus* spp.) (Taylor 1999; Jonsson and Jonsson 2001). Sympatric speciation is usually linked to trophic polymorphism in which intra-specific ecotypes use different habitats and resources (Blackie et al. 2003; Hansen et al. 2012; Smith and Skúlason 1996). Trophic polymorphism is common in post-glacial lakes where retreat of the ice sheet creates unoccupied niches and opportunities for intra-specific competition (Blackie et al. 2003; Zimmerman et al. 2009). These conditions are believed to be responsible for the extensive radiation in North American freshwater fishes where several species are adapted to different ecological niches (Schluter 2001). Parallel evolution of shared phenotypic traits linked to trophic resource use has been demonstrated in several postglacial systems. These shared morphological traits between populations can be accompanied by shared genetic architecture underlying the ecologically important traits or can arise from independent genetic processes (Ralph and Coop 2014). For example, the repeated divergence of marine and freshwater stickleback exhibiting similar phenotypic changes in body armour have been described and the repeated reduction in armour plates was found to be controlled by the same set of loci (Colosimo et al. 2005; Jones et al. 2012). On the other hand, convergent phenotypic traits may not always be controlled by similar developmental pathways as is the case for cavefish (*Astyanax* spp.), beach mice (*Peromyscus polionotus*), and fruit fly (*Drosophila* spp.) (reviewed in Arendt and

Reznick 2008; Bernatchez 2016). For instance, the evolution of parallel phenotypic divergence between benthic normal and limnetic dwarf whitefish (*Coregonus* spp.) in several North American lakes was found to be only partially associated with parallelism at the genome level (Gagnaire et al. 2013; Laporte et al. 2015).

Lake Trout (*Salvelinus namaycush*) are renowned for the occurrence of different ecotypes linked to resource and habitat use throughout North America. In small lakes, Lake Trout diverge mainly into a planktivorous and piscivorous ecotype (Vander Zander et al. 2000; Bernatchez et al. 2016), whereas several large lakes harbor at least four ecotypes associated with differential resource partitioning (Muir et al. 2015). For instance, four different ecotypes occur in Great Bear Lake and Lake Superior, three in Great Slave Lake and two in Lake Mistassini and Rush Lake (Muir et al. 2015). In Lake Superior, four distinct ecotypes have been reported that are recognized based on differences in morphology and coloration but also in life history traits, physiology and ecology (Muir et al. 2015). For instance, they differ in traits such as growth rate, asymptotic length and weight, size at sexual maturity, as well as developmental rate of fertilized eggs or fry. They also differ in physiology such as buoyancy and swim bladder retention (Muir et al. 2015; Hansen et al. 2016). The 'lean' ecotype has a slender, streamlined body with low body lipid content, and occupies shallow waters where it preys upon pelagic fishes (Bronte et al. 2003; Goetz et al. 2011; Burnham-Curtis and Smith 1994; Moore and Bronte 2001; Zimmerman et al. 2009). The 'humper' ecotype inhabits offshore, mid-water shoals, feeds on small prey and is sexually mature at relatively smaller sizes (< 500 mm) (Stafford et al. 2014; Burnham-Curtis and Smith 1994; Hansen et al. 2016). It also has a small head with moderately large eyes dorsally positioned and short-paired fins (Zimmerman et al. 2009; Moore and Bronte 2001; Bronte et al. 2003). The 'siscowet' ecotype is recognized by its sloping snout, moderately large eyes and high body fat content which may facilitate diel vertical migration to follow the migration of ciscoes (Ahrenstorff et al. 2011; Hrabik et al. 2014; Bronte et al. 2003; Bronte and Sitar 2008; Burnham-Curtis and Smith 1994; Hansen et al. 2012). Lastly, the 'redfin' ecotype has a robust body, a large head, a long deep peduncle and large fins (Muir et al. 2014). Several hypotheses have been proposed to explain the origin of these ecotypes (Wilson and Mandrak 2004; Eshenroder 2008). These could be the result of developmental plasticity in which a single genotype expresses different phenotypes matching selection optima or can be genetically based or a mix of both (Goetz et al. 2010). While this does not rule out a role for developmental plasticity, two lines of evidence suggest some genetic basis for the phenotypic differences observed between the ecotypes. First, progeny from wild lean and siscowet gametes have been raised in a common garden experiment and key phenotypic features that differentiate wild leans and siscowets such as condition factor, morphology and lipid content were maintained (Goetz et al. 2010). Furthermore, the same study uncovered transcriptional differences in lipid-related genes

between the two ecotypes (Goetz et al. 2010). Second, morphological differences in the operculum and supraethmoid bones have been documented between leans, siscowets and humpers. Cranial bones are of taxonomic significance in salmonids and are unlikely affected by environmental conditions and ontogenic shifts (Burnham-Curtis and Smith 1994).

Lake Trout (*Salvelinus namaycush*) were once the dominant predator in the Great Lakes. It historically supported one the most important freshwater commercial fisheries before being extirpated in the 1950s in all lakes except Lake Superior, where it is now considered restored and Lake Huron, where recruitment has been increasing (Riley et al. 2007), but it remains at relatively low abundance (Zimmerman and Krueger 2009, Bronte et al. 2003). The collapse of Lake Trout populations has been associated with anthropogenic factors, including habitat degradation, pollution and overfishing, as well as predation by invasive sea lamprey following the construction of navigation canals (Bronte and Sitar 2008; Page et al. 2003; Page et al. 2004). A review by Zimmerman and Krueger (2009) examined impediments to its recovery or restoration and provided guidelines to maintain, increase or reintroduce Lake Trout populations in the Laurentian Great Lakes. Here, understanding and evaluating genetic structure and diversity of remaining Lake Trout population was identified as a key research topic.

The general goal of this study was to gain insight into the nature and origin of the different Lake Trout ecotypes in Lake Superior. More specifically, we aimed to; 1) investigate the extent of both morphological and genome wide genetic differentiation and connectivity among the four Lake Trout ecotypes from different geographic locations within the lake; 2) identify possible adaptive genetic differentiation among ecotypes by means of genome scans and genotype-phenotype associations; and 3) examine the degree of parallelism at the phenotypic and genotypic level among ecotypes from the four sampling sites. To achieve this, we used RADseq to genotype Lake Trout from the four ecotypes and from four sites from Lake Superior. In parallel, geometric morphometric analyses were performed on head and body shape.

METHODS

Sampling

Fish from the four Lake Trout ecotypes were sampled in 2013-2014 from four sites in Lake Superior; Big Reef (2014), Stannard Rock (2013-2014), Superior Shoals (2013) and Isle Royale (2013) (Figure 1, Table 1). For the first three sites, a nylon gill net, 183 m long by 1.8 m high, was used with 30.5 m long panels of different mesh sizes (50.8-114.3 mm). Nets were deployed for 24 hours at different depth ranges (0-50 m, 50-100 m and >100 m) approximately representing preferred depths of the different ecotypes. A picture of each fish was taken following the protocol in Muir et al. (2012) and a biopsy of either the adipose or

pectoral fin was collected and preserved in 95% ethanol. The fourth site, Isle Royale, was sampled in 2013 using overnight sets of 274-823 m long gill nets with nine panels (91.4 m long by 1.83 m high) of single mesh size (5.1 cm, 6.4 cm, 7.6 cm, 8.9 cm, 10.2 cm, 11.4 cm, 12.7 cm, 14.0 cm, 15.2 cm). Pictures were taken using the same protocol (Muir et al. 2012) and liver or gonads were conserved in RNA Later. Samples without pictures or genetic material were removed from subsequent analyses. Information about total length (mm), wet weight (g) and sex, and depth of capture were recorded for each sampled individual.

Ecotype assignment

Consensus of both morphometric analyses (body and head) and visual identification as visual interpretation of fish pictures by Lake Trout experts (see Acknowledgements) was used to assign an ecotype to each fish per Muir et al. (2014). Fish less than 430 mm long with the exception of humper-like fish, that are < 430 mm when sexually mature, were excluded to remove the confounding effect of ontogenetic shifts in morphology (Zimmerman et al. 2009). Body and head were analysed separately to distinguish locomotion (body) from feeding habit (head). In addition, morphometric analyses were conducted separately for each site to investigate morphological variation among sites. Landmarks and semi-landmarks were digitized and analysed with the Thin Plate Spline suite (TPS; State University of New York at Stony Brook; http://life.bio.sunysb.edu/morp). First, for each fish a rectangular grid was overlaid to identify belly curvature corresponding to 20-30-40-50% of body length using the program REVIT (Autodesk) (Figure S1a). The body grid was anchored at the tip of the snout and the midpoint of the hypural plate. Second, 16 homologous landmarks and four semi-landmarks were digitized on each fish with the program TpsDig2 and semi-landmarks were slid using TpsUtil (Figure S1a). Semi-landmarks were used to represent belly curvature which is known to be distinctive between the two major ecotypes, leans and siscowets (Muir et al. 2014). Similarly, a squared grid was overlaid on each fish head dividing it into 10 equally spaced sections using the program REVIT (Figure S1b). The head grid was anchored at the tip of the snout and the posterior edge of the opercula. Eight homologous landmarks and 20 semi-landmarks were digitized on each fish head with the program TpsDig2 and semi-landmarks were slid using TpsUtil (Figure S1b). Distortions from rotation and size were removed by the program TpsRelw producing partial warps scores which are size-free variables. A principal component analysis (PCA) was performed to reduce the number of morphometric variables or scores and extract divergent morphometric patterns. Subsequently, relevant axes were supplied to a Bayesian clustering analysis implemented in the R package Mclust v.4. Mclust is a normal mixture modeling for model-based cluster analysis, classification and density estimation that include the Bayesian Information Criterion (BIC) for model selection and that do not require priori information about groups such as discriminant function analysis

(Fraley & Raftery 2012). Components accounting for more than 65% of the variance were supplied to the Mclust algorithm. The best model (with highest BIC) was the one able to separate leans from siscowets since they are the most morphologically differentiated ecotypes (Fraley and Raftery 2012; Muir et al. 2014). Group classification resulting from the chosen model was retrieved for each individual. The visual identification of each collected fish from Big Reef, Stannard Rock and Superior Shoals was conducted by visual consensus of three trained biologists. Visual identification of Isle Royale fish was provided by an experienced biologist. An ecotype was assigned to each fish based on the consensus from body shape, head shape and visual identification. Two out of three similar ecotype assignments were needed to assign to each fish a particular ecotype. In the case of different head, body and visual assignments; the fish were assigned « no consensus» and removed from subsequent analyses. Fish for subsequent genetic analyses were chosen as follows: (1) fish with 100% consensus having the lowest group uncertainty and (2) fish with 2/3 consensus having the lowest group uncertainty. However, if no individual of a given ecotype was identified based on morphometric analysis, the visual identification only was used and taken into account in subsequent analyses since ecotypes differ in several life history traits (eg. size and age at maturity, color) that are not taken into account in morphometric analyses but that are still used commonly by local expert fishery biologists to distinguish ecotypes.

Morphometric analysis

Two multivariate analyses were used to test for morphological differences between the four ecotypes at the four sites and to investigate among site differences for the same ecotype. First, a principal component analysis (PCA) was performed to reduce variable dimensionality, and components explaining most of the variance were selected based on the broken stick method. Then a multivariate analysis of variance (MANOVA) was conducted in R (package “stats”) on the selected components. Since partial scores derived from a configuration that included semi-landmarks do not have the same number of free variables as degrees of freedom, a requisite of MANOVA, a between-group analysis (groupPCA) implemented in the R package “morpho” was conducted on partial warps (Webster and Sheets 2010). This analysis takes into account uneven group size and does not require normality or homogeneity of variance (Mitteroecker and Bookstein 2011). The Euclidean distance between group mean was tested using 10 000 permutations. For both analyses, the effects of the sampling site, sex and ecotype were tested.

Sample DNA extraction and sequencing

Genomic DNA was extracted from individuals representing each ecotype at the four sites using a salt-extraction protocol adapted from Aljanabi and Martinez (1997). Sample quality and concentration were checked on 1% agarose gels and using the NanoDrop 2000 spectrophotometer (Thermo Scientific). Each individual’s genomic DNA was normalized to 20 ng/µl of 10 µl (200ng total) using PicoGreen

(Fluoroskan Ascent FL, Thermo Labsystems) in 96 well plates. The ddRAD libraries were constructed and sequenced on the Ion Torrent Proton platform (IBIS, Laval University) following the protocol in Mascher et al. (2013). Briefly, restriction digest buffer (NEB4) and two restriction enzymes (*PstI* and *MspI*) were added to each sample. Digestion was completed by incubation at 37°C for two hours and enzymes were inactivated by incubation at 65°C for 20 minutes. Two adaptors (one unique to each sample and the second common) were added to each sample and ligation was performed using a ligation master mix followed by the addition of T4 ligase. The ligation reaction was completed at 22°C for 2 hours followed by 65°C for 20 minutes to inactivate the enzymes. Samples were pooled in 48-plex and cleaned-up using QIAquick PCR purification kits. The library was then amplified by PCR and sequenced on the Ion Torrent Proton P1v2 chip. The detailed methods for SNP identification, SNP filtering and genotyping using STACKS v.1.32 (Catchen et al. 2011) are presented in Supplementary materials. Resulting VCF was converted to various formats necessary for other programs using PGDSpider 2.0.7.2 (Lischer and Excoffier 2012) and VCFtools (Danecek et al. 2011).

Genetic diversity and differentiation

We first estimated pairwise population differentiation using Weir's and Cockerham's estimator of pairwise F_{ST} (Weir and Cockerman 1984) in GenoDive 2.0b23 (Meirmans and Van Tienderen 2004) with 10 000 permutations. Similarly, measures of observed (H_o) and expected heterozygosity (H_e) and inbreeding (F_{IS}) were estimated using GenoDive 2.0b23b. Effective population size (N_e) and number of polymorphic loci (N) for each sampling site was estimated using the program NeEstimator v.2.01 (Do et al. 2014). Briefly, the program was run with the linkage disequilibrium model, the random mating system and a critical value of 0.05 (Pcrit) to exclude alleles that occur in only a single copy in the sample. Genome-wide diversity (π) and the increase in individual homozygosity relative to mean Hardy-Weinberg expected homozygosity (F_h) were estimated for each site with the dataset prior to filtration using the R package *stackr* (https://github.com/thierrygosselin/stackr). Lastly, an analysis of molecular variance (AMOVA) was conducted to quantify the proportion of genetic variance explained by sites relative to that explained by variation among the four ecotypes (Meirmans and Van Tienderen 2004). The AMOVA was run with two different levels of hierarchical subdivision; first with sites nested within ecotypes and then ecotypes nested within sites. A total of 10 000 permutations were used to access significance and an infinite allele model was chosen. Because AMOVA does not allow missing data, missing values were replaced by randomly selecting alleles proportional to total allele frequency in Genodive 2.0b23. A Mantel test between genetic divergence (F_{ST} matrix) and phenotypic divergence (head and body Euclidean distances matrices) was conducted using the R package "vegan" (Oksanen et al. 2016) to assess the extent of association/parallelism in morphology and genetic among ecotypes and sampling sites.

Population clustering

Population clustering and connectivity was estimated with the program ADMIXTURE 1.23 (Alexander et al. 2009). This program estimates ancestry in a model-based manner where individuals are considered unrelated and allows choosing the best number of possible genetic groups present in the data based on a cross-validation procedure. The program was run with values of K ranging from 1 to 20. A population tree was built using the program TreeFit (Kalinowski 2009) and visualized with the program FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). Genetic distances were calculated using θ (Weir and Cockerham 1984) between each pair of population and the neighbor-joining (NJ) distance-based method was used for tree construction. Support for each branch was assessed by bootstrapping using 1000 permutations (Kalinowski 2009).

Population assignment

Population assignment was conducted to investigate the power to classify an unknown individual to either a sampling site or an ecotype. This analysis was run using Genodive 2.0b23 with the home likelihood criteria (the likelihood that an individual comes from the population where it was sampled), which is more appropriate when only part of all possible source population have been sampled (Meirmans and Van Tienderen 2004). A significance threshold of 0.05 was applied and zero frequencies were replaced by 0.005 as suggested by Meirmans and Van Tienderen (2004). To avoid bias due to the calculation of allele frequencies from the same individuals which are subsequently assigned, the program uses the leave one out (LOO) validation procedure in which a targeted individual is removed from its source population before calculation of the allele frequency. For this analysis, missing values were replaced by randomly picking alleles from the global allele pool. All loci were used for this analysis such that no correction was necessary to avoid high grading bias associated with using a subset of markers based on their ranking of level of differentiation (Anderson 2010).

Outlier detection and phenotype-genotype associations

We used two different types of approaches to detect outlier SNPs potentially under divergent selection between ecotypes and sites: 1) genome scans performed among the different ecotypes and/or sites; and 2) association tests between genotypes and continuous phenotypic values.

For the first approach, two different methods were used to detect outlier SNPs potentially under divergent selection (1) among the four sites (individuals from the different ecotypes were pooled), (2) among the four ecotypes (individuals from the different sites were pooled) and (3) among the four ecotypes within each site, independently. First, the program Bayescan v1.2 was used to detect outliers based on locus-specific F_{ST} with a prior odd of 10 000 and a false discovery rate (FDR) of 0.05. Bayescan was run with 5000 iterations and a burn-in length of 100 000 as recommended by Foll and Gaggiotti

(2008). Second, the program LFMM (Latent Factor Mixed Models) from the R package LEA was used to detect outliers based on allele frequencies exhibiting significant statistical association with selected phenotypes. Categorical variables were coded as orthogonal matrices on which a principal component analysis was applied and resulting scores were supplied to the LFMM analysis. LFMM was run with five repetitions, 10 000 cycles and 5000 burn-in as recommended by Frichot and François (2015). P-values were adjusted from their distribution and possible associations corrected for population structure detected from the admixture analysis as suggested by Frichot and François (2015).

For the second approach, phenotype-genotype associations were analysed with LFMM. This technique can uncover subtle changes in allele frequencies (such as expected in polygenic selection) that are not detected in traditional outlier analyses (Rellstab et al. 2015). LFMM was run with the same parameters stated in the previous paragraph with ten repetitions including the p-value adjustment, an FDR of 0.05 and a correction for population structure based on the admixture analyses. The phenotypic variables were the principal components scores for each individual that explain most of the variation for head and body shape based on the broken stick method.

Gene ontology

Loci potentially under selection detected by either of the different approaches (Bayescan and LFMM) were blasted against the Rainbow Trout genome (*Oncorhynchus mykiss*) (Berthelot et al. 2014) to determine possible functions with the following parameters: an e-value threshold of 1e-6, a word size of 11 bp and a max target of 100 bp. Resulting loci were filtered based upon three criteria; the number of similar hits, the bit-score and sequence length. First, loci with only one hit and having≥ 50 bp long were kept. Second, loci with multiple hits having the first best hit≥ 20 bit -score higher than the second best hit with sequence length ≥ 50bp were kept.

RESULTS

Ecotype identification and morphometric analyses

Based upon consensus analysis between head, body and visual identification, an ecotype was assigned to each fish. First, the best model for each site that distinguished, at least, between leans and siscowets with BIC values and mean uncertainties was used for ecotype assignment (Table S1). For each site, fish to be genotyped were chosen from the consensus identification (Table 1). If some ecotypes were not distinguished by the morphometric analysis from either the head or body shape, expert visual identification from these fish was used based upon the presence of life history traits divergence as stated in the Methods section (Muir et al. 2015). Based on the broken stick method, the first four PCs were retained for body shape and the first six PCs were retained for head shape corresponding to 70% and 81% of total

variance respectively to conduct the multivariate analysis of variance (MANOVA). First, the overall shape difference between ecotypes was assessed by pooling similar ecotypes from the four sites. For the head shape, the ecotypes, the sites and the sex were significantly different ($p < 0.001$). Interactions between ecotypes and sex ($p < 0.01$) or sites ($p < 0.001$) were also significant. Similar results were observed for body shape ($p < 0.001$) (Table 2). The group-PCA revealed the same pattern for the head and body shape except that no difference between sexes was detected (Table S2, Supporting information). The first axes of the group-PCA for head shape explained 56.2% of the variance and discriminated siscowets from leans, whereas the second axis explaining 15.9% of the variance discriminated humpers from redfins (Figure 2a). For body shape, the first two axes of the group-PCA explained 65.5% and 14.5% of the variance and mainly distinguished leans from siscowets (Figure 2b). In both head and body analyses, the third and fourth axes discriminated Lake Trout more by sampling sites than ecotypes (Figure. 2c-2d). Ecotypes were not significantly different within all sites, either based on morphometric analyses of head or body shape but yet could be differentiated by visual inspection (Figure 2a-2b, Figure 3b, Table S3, Supporting information). Within Big Reef, only leans differed from siscowets and redfins in terms of both head and body shapes ($p < 0.05$). Within Isle Royale, head and body shape differed between all four ecotypes ($p < 0.05$) with two exceptions; humpers did not differ from leans in body shape and leans did not differ from redfins in head shape. Within Stannard Rock, leans differed from siscowets and redfins in both head and body shape ($p < 0.05$). Finally, within Superior Shoals, siscowet body shape differed from all other ecotypes ($p < 0.05$), except for head shape which was not different from humper's. In addition, leans head shape was different from redfin's ($p < 0.05$). In some cases, similar ecotypes from different sites had significant different head and/or body shapes (Figure 2c-2d, Figure 3b, Table S3, Supporting information). Indeed, siscowets head shape differed among sites whereas body shape was not different between Big Reef and Stannard Rock. Body shape of Superior Shoals leans differed from other leans except Isle Royale's whereas Isle Royale leans differed from Stannard Rock's. On the other hand, Isle Royale lean heads differed from both Stannard Rock and Big Reef leans. Redfins from Isle Royale differed in head and body shape from all other redfins and lastly humpers were not different from site to site. Despite the fact that not all ecotypes from all sites were morphologically different based on morphometric analyses, we conserved this grouping for the genetic analysis based on the visual inspection of other traits (e.g. size of mature fish, body or fin colours).

Sequencing and SNP calling

Raw reads cleaning and demultiplexing resulted in a total of 1.6 billion reads with an average of 3.2 million reads per individual and a relatively small mean coefficient of variation (CV) of 0.23. The assembly resulted in a catalog containing 1,052,664 loci and a total of 212,804 SNPs (49,399 loci) after

the population module. Fifteen individuals having more than 40% missing genotype were removed from the analysis. After custom filtration 6822 high quality SNPs were retained for subsequent analysis (Table 3).

Genetic diversity and differentiation among sites and ecotypes

Genetic statistics revealed modest but significant F_{ST} among some ecotypes within each sampling site (Mean F_{ST} = 0.0055) (Figure 3a, Table S4, Supporting information). Mean F_{ST} among ecotypes within sites were as follow: Big Reef 0.0047 [0.000; 0.012], Isle Royale 0.0087 [0.001; 0.017], Stannard Rock 0.0053 [0.002; 0.012] and Superior Shoals 0.0032 [0.000; 0.006]. No trend in patterns of genetic diversity was observed between ecotypes within each site (Table 4). That is, there was no evidence that diversity in some ecotypes tended to be higher than in others. On the other hand, F_{ST} among sites were on average three times higher than observed among ecotypes within site (Mean F_{ST} = 0.016). For instance, F_{ST} between sites (all four ecotypes pooled) were all significant: Big Reef ↔ Isle Royale 0.017, Big Reef ↔ Stannard Rock 0.009, Big Reef ↔ Superior Shoals 0.022, Stannard Rock ↔ Isle Royale 0.02, Superior Shoals ↔ Isle Royale 0.01 and Stannard Rock ↔ Superior Shoals 0.02. A lower value between Big Reef ↔ Stannard Rock and Isle Royale ↔ Superior Shoals site pairs was consistent with their closer geographic proximity. Also genetic diversity parameters tended to show greater differences between sites, than between ecotypes within sites (Table 4). Namely, genetic diversity, in terms of nucleotide diversity (π) and heterozygosity (H_o, H_e), was notably lower within Stannard Rock in comparison to the three other sites (Table 4). Overall, Superior Shoals ecotypes had the lowest effective population size (N_e) estimates while having, with Isle Royale, the highest inbreeding coefficient (G_{is}, F_h) whereas ecotypes from Big Reef had the highest effective population size (N_e) and the lowest inbreeding coefficient (G_{is}, F_h) while Stannard Rock showed intermediate indices. The more pronounced pattern of population differentiation between sites than between ecotypes was also evidenced by the AMOVA which revealed no net genetic variance explained by the ecotype grouping (F_{CT} = -0.002) compared to the net and significant genetic variance explained by sites grouping (F_{CT} = 0.011) (Table 5). Finally, no significant association was obtained between the F_{ST} matrix and either head (r = -0.1025 p_{value} = 0.862) or body (r = 0.1032 p_{value} = 0.115) shape Euclidean distances matrices (Figure 3).

Clustering analysis

The Admixture program identified two groups (best K) corresponding to pairs of sites: Big Reef and Stannard Rock vs. Isle Royale and Superior Shoals (Figure 4a). No migrants from Isle Royale and Superior Shoals were detected in the Big Reef/Stannard Rock cluster while results suggested the occurrence of migrants and admixed individuals in the Isle Royale/Superior Shoals cluster with a tendency

for a greater proportion of migrants in Superior Shoals (Figure 4a). At K3-K4 Big Reef individuals tended to cluster separately from those of Stannard Rock although lean trout from Big Reef tended to be more similar to Stannard Rock leans. At K5, Isle Royale could be discriminated from Superior Shoals. Lastly, at K6 all four sites differed and some additional within-site distinctions began to appear. Within Big Reef, leans were distinct from other ecotypes, being more similar to the lean/redfin cluster from Stannard Rock. Within Isle Royale, siscowets were distinct from the other ecotypes while no obvious difference emerged between ecotypes within Superior Shoals. In addition, some siscowets from Stannard Rock seemed to be similar to Isle Royale siscowets. The NJ population tree mainly grouped ecotypes from different sites together with pair of sites closer geographically also clustering more closely in the tree (Figure 4b). In addition, as observed in the Admixture analysis, leans from Big Reef were closer to leans from Stannard Rock than from the other ecotypes within Big Reef. Admixture also showed that siscowets from Isle Royale were most distinct from the other three ecotypes within this site (Figure 4b).

Population assignment

Assignment success to sampling sites, based on the 6822 SNPs, was high, being 85% on average and up to 95% for Isle Royale and 93% for Stannard Rock (Figure 5a). Miss-assigned individuals from Big Reef were only assigned to Stannard Rock and vice-versa. Superior Shoals had a lower assignment success (78%) and had miss-assigned individuals to the three other sites. On the contrary, assignment success to ecotypes was low, being 41% on average, ranging from 12% for humpers up to 61% for siscowets (Figure 5b). Ecotype assignment success within each sampling site was highly variable, being highest on average within Isle Royale (55%) and lowest within Superior Shoals (21%) and in fact similar to random expectation while Big Reef (33%) and Stannard Rock (40%) showed intermediate results (data not shown). Assignment success within Isle Royale was 76% for siscowets, 62% for leans, 52% for humpers and 33% for redfins, whereas assignment success within Superior Shoals was 46% for siscowets, 23% for redfins, 18% for leans and 0% for humpers. Within Big Reef, individuals were assigned either to siscowets or leans whatever their current ecotype was. For instance, assignment success for siscowets was 82%, 51% for leans and 0% for humpers or redfins. Stannard Rock showed a similar pattern, where assignment success for siscowets was the highest (76%) followed by leans (74%) while the assignment for humpers (10%) and redfins (0%) was low.

Outlier detection and phenotype-genotype associations

Bayescan identified a total of 52 outliers from which 49 occurred between the four sites and three between the four ecotypes (Figure 6a-6b). No outliers were detected between ecotypes within each site. In addition, no outliers were common between sites and ecotype comparisons. For LFMM, the p-values were adjusted

using a lambda of 0.55 (ג) and population structure was corrected for each analysis using the number of ancestral groups (K) identified by Admixture for the overall dataset or within each site separately. According to the Admixture results, a K of five was used for between sites and between pooled ecotype comparisons while for within site comparisons, a K of two was used for Big Reef and Superior Shoals and a K of three was used for Isle Royale and Stannard Rock. LFMM identified a total of 670 unique outliers: 554 between sites, and 116 between ecotypes in which 20 were common to both comparisons (Figure 6a-6b). Thus, the number of outliers between ecotypes was lower than that observed between sites. For within site comparisons between ecotypes, 359 outliers were detected within Big Reef, 131 within Isle Royale, 360 within Stannard Rock and 120 within Superior Shoals. Overall, up to 27 outliers were common between some sites but none were common to all sites (Figure 6c). No outliers detected among ecotypes were common between LFMM and BAYESCAN but eight were common among sites.

Based on the broken stick method, the first four principal components were selected to represent head shape and the first six principal components were selected to represent body shape, for a total of 10 shape variables. Briefly, the p-values were adjusted using a lambda of 0.55 (ג) and population structure was corrected using a K of five. A total of 915 unique associations were detected with an FDR of 0.05 in which several were common between variables (Table S5, Supporting information). Four of these associated SNPs were common with BAYESCAN outliers (one with the between ecotype comparison and three with the between site comparison) (Figure 6a-6b). In addition, 349 of these associated SNPs were common with the previous LFMM analysis. Briefly, 71 were in common with between site comparison, 20 with between ecotypes comparison, 85 with within Big Reef comparison, 48 with within Isle Royale comparison, 91 with within Stannard Rock and 34 with within Superior Shoals comparisons.

Annotation

A total of 2056 loci detected either by BAYESCAN or LFMM between sites, between ecotypes or in association with phenotypic variation were blasted against the Rainbow Trout genome. After quality filtering, 258 loci that had an annotation in genes were retained (Table S6). From those with a known biological function, markers linked to lipid transport and metabolism as well as visual development and perception were of particular interest given previously documented phenotypic characteristics differentiating Lake Trout ecotypes (see Discussion).

DISCUSSION

This is the first study to combine genomic and morphometric analyses from all four Lake Trout ecotypes from several different locations in Lake Superior. This provided the unique opportunity to investigate

among and within site variation and the extent of parallelism, both at the phenotypic and genomic level. Both morphometric and genomic analyses revealed within-site morphological and genetic differences between ecotypes, but in general, genetic differences were more pronounced among sites than among ecotypes, even when comparing populations of the same ecotype. Similarly, we observed that values of demographic and genetic diversity parameters generally varied more by site than by ecotype. Moreover, the extent of both morphological and genetic differences among ecotypes observed within site varied from one location to the other, thus creating a continuum of differentiation. In addition, genome scans and association tests identified several loci potentially implicated in local adaptation and phenotypic divergence among ecotypes, among which loci linked to lipid metabolism and transport as well as visual acuity and development are of particular interest (see Discussion below). The relatively large number of outlier loci identified, which globally showed relatively modest levels of genetic differentiation among sites or ecotypes, suggests a polygenic origin of both local adaptation between sites and ecotypic differentiation. We discuss the implications of these results for the understanding of the biological processes responsible for the emergence of the different ecotypes of Lake Trout as well as for their management.

Parallel evolution of Lake Trout ecotypes?

Parallel evolution, the repeated evolution of similar phenotypic traits, has been documented in many populations within the same species (reviewed in Elmer and Meyer 2011). Shared phenotypic traits that evolved independently are generally believed to indicate parallel adaptive evolution in the face of shared environmental pressures between locations driving changes to similar optimum (Butlin et al. 2013). The evolution of these similar traits can be underlained by similar or different genome architecture (Elmer and Meyer 2011; Ralph and Coop 2014; Bernatchez 2016). Here, similar ecotypes corresponding to previously published descriptions were identified among all sampling sites. That is, a greater proportion of morphological variance, explaining 14.5% to 65.5%, clustered individual by ecotypes (first and second components of the PCAs, Figure 2a-2b) thus revealing parallelism in morphology between ecotypes from the four sampling sites. It is noteworthy that head shape better discriminated ecotypes than body shape (Figure 2a, Figure 3b), as reported previously in other Lake Trout studies both from the Great lakes and elsewhere (Chavarie et al. 2013; Moore and Bronte 2001; Alfonso 2004; Moore and Bronte 2007; Magalhaes et al. 2009). Moreover, a greater proportion of markers identified as outliers or associated with phenotypic differentiation was found for head shape compared to whole body shape. More pronounced ecotypic differentiation of head shape could suggest a predominant role for feeding ecology compared to other factors (e.g. locomotion) as the main driver for these morphological differences (Chavarie et al. 2013; Magalhaes et al. 2009; Jonsson and Jonsson 2001).

The fact that different ecotypes within sites were generally genetically more similar than different populations of the same ecotype among sites suggests that parallel evolution is implicated in the origin and maintenance of ecotypes. Moreover, while both explanations are not exclusive, we cannot refute the possibility that more pronounced genetic similarity within sites might also reflect higher gene flow among ecotypes within sites than among population of a same ecotype among sites. This would also reflect less pronounced reproductive isolation among ecotypes within sites than among populations of a same ecotype among sites. It is also noteworthy that, although to a lesser extent some morphological components (explaining 3% to 11.5% of variance) could discriminate Lake Trout by sampling sites (third and fourth components of the PCAs, Figure 2c-2d). In some cases, such as siscowets, leans, and redfins different populations of a same ecotype from particular sites were morphologically divergent, indicating some dissimilarity in morphology. Such inter-site differences within ecotype have previously been reported by Bronte and Moore (2007) for siscowet and these were interpreted as either the presence of different reproductive populations and/or a plastic response to different environmental conditions among sites.

Both outlier detection methods (Bayescan and LFMM) differentiated more outlier markers among sampling sites than among ecotypes, again supporting the view that spatial variables (e.g. different environmental conditions or random genetic changes) may be more important than ecotypic differentiation in explaining the observed pattern of population structure in Lake Superior. Moreover, LFMM uncovered markers potentially under selection among ecotypes within all four sampling sites, none being common to all sites. These results also suggest that phenotypic parallelism in Lake Trout ecotypes is not accompanied by parallelism at the genome level, as reflected by the lack of association between the genetic and phenotypic divergence matrices, whereby the expression of a given ecotype in different sites is controlled by a different genetic architecture. Hypothetically there may have been random genetic differentiation (drift, founder effects in different parts of the lakes such that subsequent selection driving adaptive changes may have been acting on somewhat different gene pools in different parts of the lakes, This would lead to apparent non parallelism at the genome level. The absence of parallelism between phenotypic and genotypic differentiation has been reported in many species, including mice (*Peromyscus maniculatus*), cichlids, cavefish (*Astyanax mexicanus*), stickleback (*Gasterosteus* spp.), as well as ciscoes and whitefish (*Coregonus* spp.) (reviewed in Elmer and Meyer 2011; Bernatchez 2016). For instance, ciscoes in Lake Nipigon exhibit four morphological and ecological different species without evidence of corresponding neutral genetic differentiation (Turgeon et al. 1999). Similarly, ciscoes from several inland lakes showed variable levels of phenotypic differentiation which was not correlated to genetic divergence (Turgeon et al. 2016). Also, Laporte et al. (2015) recently documented a clear pattern of phenotypic parallelism in body shape between dwarf and normal sympatric pairs of lake whitefish with similar

genomic architecture underlying these traits being observed between some pairs but different genome architecture in others.

Genetic origin of ecotypes

Generally speaking, we found very limited support for a shared genetic origin among populations of the same ecotype. That is, we generally observed fewer genetic differences among ecotypes within sites than among populations (sites) for the same ecotype. The exception to this general pattern was for the lean ecotype for which we observed more genetic similarity between populations from Big Reef and Stannard Rock than between leans and other ecotypes from these locations. Similar results were previously reported by Ihssen (1988) and Dehring et al. (1981) who showed based on allozymes that Lake Trout of the lean ecotype from four different locations differed in allele frequencies. Different markers identified as being under selection among sites provide further support for the independent origin of ecotypes within each site. Alternatively, we cannot rule out that this may also reflect the presence of different genetic architecture underlying phenotypic variation within sites, or that markers under parallel selection were not detected because of insufficient coverage of the genome. Taken together, the combined results obtained for "neutral" and potentially "adaptive" markers highlight the contribution of both spatial isolation and local adaptation in shaping ecotypic variation within each sampling site.

Here, relatively large geographic distances between these sites, known for the relatively high occurrence of the four ecotypes separated by ranges of much lower abundance, may have contributed to reduce genetic exchange between spatially isolated populations. Thus, localized movements have been reported for Lake Trout based on tagging studies where an average movement of approximately 40 km has been reported (Kapuscinski et al. 2005; Eschmeyer 1955). Considering that the closest sites in this study are separated by about 69 km (Big Reef/Stannard Rock) to 87 km (Isle Royale/Superior Shoals) and that sites that are the farthest are separated by 98 km (Superior Shoals/Stannard Rock) to 212 km (Isle Royale/Big Reef), the presence of spatially genetically differentiated stocks is consistent with this observation. Spatial isolation could also have been exacerbated by historical water level fluctuations. Lake Superior has a very diverse bathymetric habitat covered by peaks and valleys, thus creating geographical barriers particularly when water levels fluctuated. This situation is thought to have occurred 8000 years ago, which could have triggered the spatial pattern of genetic divergence seen today (Bronte and Moore 2007).

Our data also revealed a continuum in the extent of both genetic and phenotypic divergence underlying the observed ecotypes ranging from intra-population polymorphism to clear genetically distinct populations within a sampling location. The extent of morphological differentiation in both head and body shape was also variable depending on the site being examined. Although the explanations for this pattern of continuum in morphological divergence are only hypothetical at this time, this could reflect different

levels of trophic polymorphism associated with different selective pressures (e.g. competitive interactions), as reported for other species, including Lake Whitefish (*Coregonus clupeaformis*) (Lu and Bernatchez 1999; Gagnaire et al 2013), Arctic Charr (*S. alpinus*) (Gislason et al. 1999), or Three-spined Stickleback (*Gasterosteus aculeatus*) (Hendry et al. 2009). The extent of genetic divergence between ecotypes was also variable depending on the site examined suggesting that different levels of reproductive isolation accompany different levels of phenotypic divergence (see references above, also reviewed by Hendry 2009).

In contrast to our general observation of higher genetic differentiation among sites than among ecotypes within site, a recent study conducted in Great Bear Lake, found more pronounced genetic differentiation among Lake Trout ecotypes than among sampling sites (Harris et al. 2014). These authors hypothesized that stronger genetic and morphological differentiation in Great Bear Lake could be due to its more pristine environment and limited human impact compared to Lake Superior where these factors may have altered the original pattern of population structuring. For instance, considerable stocking and fishery harvest has occurred in Lake Superior, which could certainly have had an impact on the extent of population admixture (Guinand et al. 2003) compared to Great Bear Lake, which has not been stocked and has only been subject to minor fishery harvest. However, it is noteworthy that stocking has been done essentially for the lean ecotype (Page et al. 2004). Consequently, it seems unlikely that this could explain the general pattern of structuring we documented for other ecotypes, although it could possibly explain why leans from different locations were more similar in some cases, as explained above.

In sum, the combined genomic and morphological data support the hypothesis that ecotypic differentiation among Lake Trout ecotypes from different geographic locations within Lake Superior can be arrayed along a continuum from quasi-panmixia to relatively pronounced reproductive isolation, mimicking the inter-specific pattern described by Hendry et al. (2009) among lacustrine north temperate freshwater fishes. Consequently, variation along this continuum might profitably be used for studying factors, outlined by Hendry et al. (2009), which can promote or constrain progress toward ecological speciation, including plasticity, natural selection, mate choice, geography, or historical contingency. However, the present study cannot rule out the possibility that different anthropogenic impacts among sites could have also contributed to the observed pattern of genomic and phenotypic variation. Indeed, a recent study conducted by Baillie et al. (2016) highlighted substantive losses of genetic diversity and genetic distances in lean, siscowet and humper trout from post-collapse recovery (1995-1999) compared to contemporary period (2004-2013). This homogenisation could be the result of overexploitation, intensive stocking and invasions of non-native species which could have led to the overlap in breeding or foraging area thus increasing hybridisation. Biodiversity losses and speciation reversal caused by anthropogenic activities

have been recorded in several freshwater species such as Lake Victoria cichlids (Seehausen et al. 2008), Great Lakes ciscoes (*Coregonus spp.)* (Todd and Stedman 1989) as well as the European whitefish (*Coregonus spp.)* (Bhat et al. 2014; Hudson et al. 2013).

Evidence of local adaptation and functional annotation

In both spatial and ecotypic differentiation however, a much larger proportion of markers potentially under selection were detected by the LFMM method compared to Bayescan, the former known to be more sensitive to polygenic effects, suggesting that weak or polygenic selection might be responsible for the observed pattern of "adaptive differentiation", both spatially and between ecotypes (Rellstab et al. 2015). Of the 258 loci for which successful annotation could be retrieved, several were of particular interest and linked to ecotypic differences observed in the present system. Two loci were linked to visual development and acuity of the retina; retinal guanylyl cyclase 2 and retinitis pigmentosa 1-like 1 protein, and one to visual perception; peripherin-2-like. Both markers linked to visual development and acuity were found in significant association with the second component of the head shape analysis from which the highest loading was for the eye position (landmark number 26). Changes in size, location and sensitivity of the eyes have been associated to adaptation to low light environment (Von der Emde et al. 2004). Indeed, larger eyes with a predominance of rods are known to increase visual acuity (Von der Emde et al. 2004). Large eyes, close to the snout have been reported in other deepwater, salmonid morphs similar to the siscowet and humper ecotypes in Lake Superior, potentially reflecting adaptation to low light condition (Eshenroder 2008; Skoglund et al. 2015; Moore and Bronte 2001).

Annotated markers of interest were also linked to lipid binding, transport, regulation and metabolism. A total of three annotated markers were linked to lipid binding; the spectrin beta non-erythrocytic 4-like isoform x1 and 1-like isoform x2 (SPTBN4, SPTBN1) and the calcium-dependant secretion activator 1 (CADPS), and one marker was linked to transport; the lipid phosphate phosphatase-related protein type 4-like (LPPR4) (http://genecards.org). SPTBN4 was found in significant association with the first component of the body shape analysis which was linked to belly curvature whereas SPTBN1 and CADPS were found in significant association with among ecotype analyses and LPPR4 in significant association with head depth. High lipid levels in the muscle of the deepwater siscowet ecotype have long been described and suggested to facilitate vertical migration in the water column by providing hydrostatic lift (Henderson and Anderson 2002; Eschmeyer and Phillips 1965).

Also, Goetz et al. (2010) showed that differences in lipid levels between the lean and the siscowet ecotype persist when reared under identical conditions. They also found several differentially expressed genes in controlled conditions between these two ecotypes linked to lipid metabolism. Interestingly, four of the annotated markers in the present study were also found to be differentially expressed by Goetz et al.

(2010), further suggesting that our study identified some candidate genes involved in the differentiation between these ecotypes. The four markers in common are the alpha-tectorin-like protein, the fk506-binding protein 5-like isoform x3, the galactosamine (n-acetyl)-6-sulfate sulfatase and the Peroxisome proliferator activated receptor. Functional descriptions of most of these genes are still lacking, therefore mechanistic links between these markers and Lake Trout ecotypic adaptations remains unknown. In sum, the hypothesis of genetically based adaptation in Lake Trout is supported by at least a few divergent annotated genes that are linked to biological functions (e.g., vision, lipid metabolism). These same genes are believed to play roles in the local adaptation to different water depths and trophic resource use (Goetz et al. 2010).

Limitations

Admittedly, we must also consider the possibility that several alternative factors could explain the pattern of continuum in ecotypic divergence observed here. Namely, sample sizes were small in some cases, especially for the humper ecotype, which could have limited our power to detect genetic divergence, namely between humper and redfin ecotypes. Also, our capability of assigning Lake Trout to different ecotypes based on their morphology varied among sites, which may have created artificially admixed groups of individuals resulting in lower level of differentiation among them. In such a case however, the clustering analysis performed with Admixture should have detected such groups of admixed individuals from different populations, which was not the case here. Instead, Admixture revealed homogeneous groups of individuals, independent of their ecotype in locations where we observed very weak or no genetic differentiation. Arguably, our results do not rule out a role for phenotypic plasticity induced by exposure to different environmental conditions, which will require further common garden studies of other ecotypes (humper and redfins ecotypes) from other locations, as performed by Goetz et al. (2010). In fact, phenotypic plasticity may have played an important role in the diversification of Lake trout ecotypes within Lake Superior. Indeed, the presence of environmentally induced (plastic) polymorphism within population has been hypothesized to facilitate the process of divergence (Adams and Huntingford 2004; Pfennig et al. 2010). Thus, phenotypic plasticity can promote the emergence of divergent phenotypes on which selection can act (Pfennig et al. 2010). In addition, trophic polymorphism may be an effective way to promote speciation by resource use because it may trigger reproductive isolation (Pfennig et al. 2010; Smith and Skúlason 1996). Finally, studies on sympatric ecotypes such as cichlids, whitefish and arctic charr have shown, using common-garden experiments, that some morphological characters were plastic and others heritable, thus demonstrating the role of phenotypic plasticity in shaping divergence (Adams and Huntingford 2004; Magalhaes et al. 2009; Lundsgaard-Hansen et al. 2013). Finally, when using methods of reduced genome representation such as RADseq, it is important to keep in mind that only a

small subsample of the whole genome variation has been screened. Consequently, some important targets of selection are most likely missed in such studies and results must be interpreted cautiously and accordingly. Here, this means that the interpretations of observed differences with a reduced genome representation are conservative.

Management implications

The maintenance of genetic diversity, and thus the potential of a species to evolve in the face of a changing environment is central in conservation genomics and fishery management (Toro and Caballero 2005). Improper management may lead to depletion of the resource and/or impaired resilience by decreasing genetic diversity or eroding local adaptations (Laikre et al. 2005; Zimmerman et al. 2009). Management units are groups of conspecific individuals among which connectivity is sufficiently low so that each group should be managed separately (Palsboll et al. 2007). The delineation of these management units is still debated and has usually been based upon the rejection of panmixia (Waples and Gaggiotti 2006) or the absolute amount of population divergence between populations (Palsboll et al. 2007). Thresholds above which populations should be considered distinct management and demographically independent units do not exist, but a dispersal level < 10% has been suggested (Palsboll et al. 2007). Based on our results, the primary basis to define management units in Lake Trout of Lake Superior should be the sampling sites rather than ecotypes since we observed pronounced levels of net genetic differentiation and high assignment success (varying between 74-95%) among sites compared to net genetic differentiation and very low assignment success (varying between 12-61%) among ecotypes. Yet, depending on locations, ecotypic differentiation must also be considered since ecotypes were also genetically distinct in some cases, such as Isle Royale in particular. Also, evidence of local adaptation was uncovered and therefore, caution must be taken within sites to avoid depletion of locally adapted traits by stocking or exploitation. Since the extirpation of Lake Trout from most of the Great Lakes other than Superior, stocking programs have been developed in some lakes without success (Page et al. 2003). Matching stocking sites with proper ecotype could increase reintroduction success (Zimmerman et al. 2009). Based on this study, we would advocate for reintroduction and translocation of Lake Trout from the least genetically differentiated site, namely Superior Shoals since this would provide the full range of ecotypic differentiation within a quasi-panmictic gene pool, a situation that would be reminiscent of the early stage of ecological speciation (Smith and Skúlason 1996; Hendry 2009). Moreover, such intra-population polymorphism may increase survival in a new environment while maintaining genetic diversity and potential for local adaptation (Wennersten and Forsman 2012). In addition, our results provided limited evidence for local adaptation associated with ecotypic differentiation at this location, which could improve survival in a different lake environment given that local adaptation is typically associated with

683 trade-offs wherein locally adapted individuals exhibit higher fitness in their local environment compared
684 with individuals from a different population and environment (Kawecki and Ebert 2004). However,
685 further studies on the extent of population differentiation throughout Lake Superior will be necessary not
686 only to better define boundaries to gene flow but also characterise potentially adaptive traits in other
687 localities.
688
689 **Acknowledgements**
690 We thank the Great Lakes Fishery Commission for the funding of the project and particularly to C Bronte,
691 C Krueger and M Hansen for their help in fish identification. We also thank the KIYI crew for the
692 opportunity to fish on the research boat on Lake Superior. We thank « Ressources Aquatique Québec
693 (RAQ) » for bursaries and monetary help in assisting conferences. We are also grateful to Laura Benestan
694 for graphical input and Martin Laporte for help in morphometric analysis, as well as Thierry Gosselin for
695 bioinformatics support. We are also grateful to Giacomo Bernardi and three anonymous reviewers for
696 their constructive and very helpful comments on an earlier version of the manuscript. This study was
697 supported by a research grant from the Great Lakes Fisheries Commission to L. Bernatchez and AM Muir,
698 and a research grant from Science and Engineering Research Canada (NSERC strategic grant program) to
699 L Bernatchez and Pascal Sirois.
700

701 LITERATURE CITED
702
703 Adams CE, Huntingford FA (2004) Incipient speciation driven by phenotypic plasticity? Evidence from
704 sympatric population of Arctic charr. *Biological Journal of the Linnean Society*, **81**, 611-618.
705
706 Ahrenstorff T, Hrabik TR, Stockwell JD *et al*. (2011) Seasonally dynamic diel vertical migrations of
707 *Mysis diluviana*, coregonine fishes, and siscowet Lake Trout in the pelagia of western Lake Superior.
708 *Transactions of the American Fisheries Society*, **140**, 1504–1520.
709
710 Alexander DH, November J, Lange K (2009) Fast model-based estimation of ancestry in unrelated
711 individuals. *Genome Research Cold Spring Harbor Lab*, **19**, 1655-1664.
712
713 Alfonso NR (2004) Evidence for two morphotypes of Lake charr, *Salvelinus namaycush*, from Great Bear
714 Lake, Northwest Territories, Canada. *Environmental Biology of Fishes*, **71**, 21-32.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692-4693.

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, **10**, 701-710.

April J, Hanner RH, Dion-Côté A *et al.* (2013) Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Molecular Ecology*, **22**, 409-422.

Arendt J, Reznick D (2008) Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology and Evolution*, **23**, 26-32.

Baillie SM, Muir AM, Scribner K *et al.* (2016) Loss of genetic diversity and reduction of genetic distance among Lake Trout *Salvelinus namaycush* ecomorphs, Lake Superior 1959 to 2013. *Journal of Great Lakes Research*, **42**, 1-13.

Bernardi G (2013) Speciation in fishes. *Molecular Ecology*, **22**, 5487-5502.

Bernatchez L (2016) On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*.

Bernatchez S, Laporte M, Perrier C *et al.* (2016) Investigating genomic and phenotypic parallelism between piscivorous and planktivorous ecotypes of Lake Trout (*Salvelinus namaycush*) by means of RADseq and morphometrics analyses. *Molecular Ecology*. **25**, 4773-4792.

Berthelot C, Brunet F, Chalopin D (2014) The Rainbow Trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications*, **5**.

Bhat S, Amundsen P, Knudsen R et al. (2014) Speciation reversal in European whitefish (*Coregonus lavaretus (L.)*) caused by competitor invasion. *PLoS ONE*, **9**, 1-10.

Blackie C, Weese D, Noakes D (2003) Evidence for resource polymorphism in the Lake charr (*Salvelinus*

namaycush) population of Great Bear Lake, Northwest Territories, Canada. *Ecoscience*, **10**, 509-514.

Bronte CR, Ebener MP, Schreiner DR et al. (2003) Fish community change in Lake Superior , 1970 – 2000. *Canadian Journal of Fisheries and Aquatic Sciences*, **60**, 1552-1574.

Bronte CR, Moore SA (2007) Morphological variation of siscowet Lake Trout in Lake Superior. *Transactions of the American Fisheries Society*, **136**, 509-517.

Bronte CR, Sitar SP (2008) Harvest and relative abundance of siscowet Lake Trout in Michigan waters of Lake Superior, 1929-1961. *Transactions of the American Fisheries Society*, **137**, 916-926.

Burnham-Curtis MK, Smith G (1994) Osteological evidence of genetic divergence of Lake Trout (*Salvelinus namaycush*). *Copea*, **4**, 843-850.

Butlin RK, Saura M, Charrier G *et al.* (2013) Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution*, **68**, 935-949.

Catchen JM, Amores A, Hohenlohe P *et al.* (2011) Stacks: Building and genotyping loci de novo from short-read sequences. *G3*, **1**, 171-182.

Chavarie L, Howland KL, Tonn WM (2013) Sympatric polymorphism in Lake Trout: The coexistence of multiple shallow-water morphotypes in Great Bear Lake. *Transactions of the American Fisheries Society*, **142**, 814-823.

Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928-1933.

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland, MA. 545 pp.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format VCFtools. *Bioinformatics*, **27**, 2156-2158.

Dehring T, Brown A, Daugherty C *et al.* (1981) Survey of the genetic variation among eastern Lake

781 Superior Lake Trout (*Salvelinus namaycush*). *Canadian Journal of Fisheries and Aquatic Sciences*, **38**,
782 1738-1746.
783
784 Do C, Waples RS, Peel D *et al.* (2014) NeEstimator V2 : re-implementation of software for the estimation
785 of comtempory effective population size (N_e) from genetic data. *Molecular Ecology Resources*, **14**, 209-
786 214.
787
788 Elmer KR, Meyer A (2011) Adaptation in the age of ecological genomics: Insights from parallelism and
789 convergence. *Trends in Ecology and Evolution*, **26**, 298-306.
790
791 Eshenroder R (2008) Differentiation of deep-water Lake charr *Salvelinus namaycush* in North American
792 lakes. *Environmental Biology of Fishes*, **83**, 77-90.
793
794 Eschmeyer PH (1955) The reproduction of Lake Trout in Southern Lake Superior. *Transactions of the*
795 *American Fisheries Society*, **84**, 47-74.
796
797 Eschmeyer PH, Phillips AM (1965) Fat content of the flesh of siscowets and Lake Trout from Lake
798 Superior. *Transactions of the American Fisheries Society*, **94**, 62-74.
799 Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and
800 codominant markers: A Bayesian perspective. *Genetics*, **180**, 977-993.
801
802 Fraley C, Raftery AE (2012) MCLUST Version 4 for R: normal mixture modeling for model-based
803 clustering, classification, and density estimation. *Technical Report no. 597, Department of Statistics,*
804 *University of Washington, June 2012*.
805
806 Franchini P, Fruciano C, Spreitzer M *et al.* (2013) Genomic architecture of ecologically divergent body
807 shape in a pair of sympatric crater lake cichlid fishes. *Molecular Ecology*, **23**, 1828-1845.
808
809 Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies.
810 *Methods in Ecology and Evolution*, **6**, 925-929.
811

Gagnaire P, Pavey S, Normandeau E *et al.* (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in Lake Whitefish species pairs assessed by Rad sequencing. *Evolution*, **67**, 2483-2797.

Gavrilets S, Vose A, Barluenga M *et al.* (2007) Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Molecular* Ecology, **16**, 2893-2909.

Gislason D, Ferguson MM, Skúlason S *et al.* (1999) Rapid and coupled phenotypic and genetic divergence in Icelandic Arctic Char (*Salvelinus alpinus*). *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 2229-2234.

Goetz F, Rosauer D, Sitar S *et al.* (2010) A genetic basis for the phenotypic differentiation between siscowet and lean Lake Trout (*Salvelinus namaycush*). *Molecular Ecology*, **19**, 176-196.

Goetz F, Sitar S, Rosauer D *et al.* (2011) The reproductive biology of siscowet and lean Lake Trout in Southern Lake Superior. *Transactions of the American Fisheries Society*, **140**, 1472-1791.

Gosselin T, Bernatchez L (2016). stackr: GBS/RAD data exploration, manipulation and visualization using R. R package version 0.2.1. https://github.com/thierrygosselin/stackr.

Guinand B, Scribner KT, Page KS *et al.* (2003) Genetic variation over space and time: Analyses of extinct and remnant Lake Trout populations in the Upper Great Lakes. *Proceedings of The Royal Society Biological sciences*, **270**, 425-433.

Hansen M, Nate N, Krueger C *et al.* (2012) Age, growth, survival, and maturity of Lake Trout morphotypes in Lake Mistassini, Quebec. *Transactions of the American Fisheries Society*, **141**, 1492-1503.

Hansen M, Nate N, Muir A *et al.* (2016) Life history variation among four Lake Trout morphs at Isle Royale, Lake Superior. *Journal of Great Lakes Research*, **42**, 421-432.

Harris LN, Chavarie L, Bajno R *et al.* (2014) Evolution and origin of sympatric shallow-water morphotypes of Lake Trout, *Salvelinus namaycush*, in Canada's Great Bear Lake. *Heredity*, **114**, 94-106.

Henderson B, Anderson D (2002) Phenotypic differences in buoyancy and energetic of lean and siscowet Lake charr in Lake Superior. *Environmental Biology of Fishes*, **64**, 203-209.

Hendry AP (2009) Speciation. *Nature*, **458**, 162-164.

Hendry AP, Bolnick DI, Bernier D *et al.* (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, **75**, 2000-2036.

Hrabik TR, Rothb BM, Ahrenstorff T (2014) Predation risk and prey fish vertical migration in Lake Superior: Insights from an individual based model of siscowet (*Salvelinus namaycush*). *Journal of Great Lakes Research*, **40**, 730–738.

Hudson AG, Vonlanthen P, Bezault E *et al*. (2013) Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish. *BMC Evolutionary Biology*, **13**, 108.

Ihssen PE, Casselman JM, Martin GW *et al.*(1988) Biochemical genetic differentiation of Lake Trout (*Salvelinus namaycush*) stocks of the Great Lakes region. *Canadian Journal of Fishery and Aquatic Sciences*, **45**, 1018-1029.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55-61.

Jonsson B, Jonsson N (2001) Polymorphism and speciation in Arctic charr. *Journal of Fish Biology*, **58**, 605-638.

Kalinowski ST (2009) How well do evolutionary trees describe genetic relationships between populations? *Heredity*, **102**, 506-513.

Kapuscinski K, Hansen S, Schram S (2005) Movements of Lake Trout in U.S. waters of Lake Superior, 1973–2001. *North American Journal of Fisheries Management*, **25**, 696-708.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225-1241.

Laikre L, Palm S, Ryman N (2005) Genetic population structure of fishes: Implications for coastal zone management. *Ambio*, **34**, 111-119.

Laporte M, Rogers S, Dion-Côté A *et al.* (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 | Genetics*, **5**, 1481-1491.

Lischer HEL and Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298-299.

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*, **53**, 1491-1505.

Lundsgaard-Hensen B, Matthews B, Vonlanthen P *et al.* (2013) Adaptive plasticity and genetic divergence in feeding efficiency during parallel adaptive radiation of whitefish (*Coregonus* spp.). *Journal of Evolutionary Biology*, **26**, 483-498.

Magalhaes IS, Mwaiko S, Schneider MV *et al.* (2009) Divergent selection and phenotypic plasticity during incipient speciation in Lake Victoria cichlid fish. *Journal of Evolutionary Biology*, **22**, 260-274.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10-12.

Mascher M, Wu S, St.Amand P *et al.* (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: A comparison of genetic and reference-based marker ordering in barley. *PLoS ONE*, **8**, 1-11.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792-794.

Mitteroecker P, Bookstein F (2011) Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology*, **38**, 100-114.

Moore S, Bronte C (2001) Delineation of sympatric morphotypes of Lake Trout in Lake Superior. *Transactions of the American Fisheries Society*, **130**, 1233-1240.

Moore S, Bronte C (2007) Morphological variation of siscowet Lake Trout in Lake Superior. *Transactions of the American Fisheries Society*, **136**, 509-517.

Muir AM, Hansen M, Bronte C *et al.* (2015) If Arctic charr *Salvelinus alpinus* is 'the most diverse vertebrate', what is the Lake charr *Salvelinus namaycush*? *Fish and Fisheries*.

Muir AM, Bronte C, Zimmerman MS *et al*. (2014) Ecomorphological diversity of Lake Trout at Isle Royale, Lake Superior. *Transactions of the American Fisheries Society*, **143**, 972-987.

Muir AM, Vecsei P, Krueger CC (2012) A Perspective on perspectives: methods to reduce variation in shape analysis of digital images. *Transactions of the American Fisheries Society*, **141**, 1161-1170.

Nosil P, Harmon L, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution*, **24**, 145-156.

Oksanen J, Blanchet FG, Kindt R *et al*. (2016) vegan: Community Ecology Package. R package version 2.3-3. http://CRAN.R-project.org/package=vegan

Page KS, Scribner KT, Bennett KR *et al.* (2003) Genetic assessment of strain-specific sources of Lake Trout recruitment in the Great Lakes. *Transactions of the American Fisheries Society*, **132**, 877-894.

Page KS, Scribner KT, Burnham-Curtis M (2004) Genetic diversity of wild and hatchery Lake Trout populations: Relevance for management and restoration in the Great Lakes. *Transactions of the American Fisheries Society*, **133**, 674-691.

Palsbøll PJ, Bérubé M, Allendorf FW (2007) Identification of management units using population genetic data. *Trends in ecology & evolution*, **22**, 11-16.

Pfennig DW, Wund MA, Snell-Rood EC *et al.* (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, **25**, 459-467.

Ralph PL, Coop G (2014) Convergent evolution during local adaptation to patchy landscape. *PLOS Genetics*, **11**.

Rellstab C, Gugerli F, Eckert AJ *et al.* (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348-4370.

Riley SC, He J.X, Johnson JE *et al.* (2007) Evidence of widespread natural reproduction by Lake Trout *Salvelinus namaycush* in the Michigan waters of Lake Huron. *Journal of Great Lakes Research*, **33**, 917-921.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372-380.

Seehausen O, Takimoto G, Roy D *et al.* (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30-44.

Skoglund S, Siwertsson A, Amundsen P *et al.* (2015) Morphological divergence between three Arctic charr morphs - the significance of the deep-water environment. *Ecology and Evolution*, **15**, 3114-3129.

Smith TB, Skúlason S (1996) Evolutionary significance of resource polymorphisms in fishes, amphibians, and birds. *Annual review of ecology and systematic*, **27**, 111-133.

Stafford CP, McPhee MV, Eby LA *et al.* (2014) Introduced Lake Trout exhibit life history and morphological divergence with depth. *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 10-20.

Taylor EB (1999) Species pairs of north temperate freshwater fishes: Evolution, taxonomy, and conservation. *Reviews in Fish Biology and Fisheries*, **9**, 299-324.

Tittes S, Kane N (2014) The genomics of adaptation, divergence and speciation: A congealing theory. *Molecular Ecology*, **23**, 3938-3940.

Todd TN, Stedman RM (1989). Hybridization of ciscoes (*Coregonus spp.)* in Lake Huron. *Canadian Journal of Zoology*, **67**,1679- 1685.

Toro MA, Caballero A (2005) Characterization and conservation of genetic diversity in subdivided populations. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **360**, 1367-1378.

Turgeon J, Estoup A, Bernatchez L (1999) Species flock in the North American Great Lakes: Molecular ecology of Lake Nipigon ciscoes (Teleostei: Coregonidae: *Coregonus*). *Evolution*, **53**, 1857-1871.

Turgeon J, Reid SM, Bourret A *et al*. (2016) Morphological and genetic variation in Cisco (*Coregonus artedi*) and Shortjaw Cisco (*C. zenithicus*): multiple origins of Shortjaw Cisco in inland lakes require a lake-specific conservation approach. *Conservation Genetics*, **17**, 45-56.

Vander Zanden MJ, Shuter BJ, Lester NP et al. (2000) Within- and among-population variation in the trophic position of a pelagic predator, Lake Trout (*Salvelinus namaycush*). *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 725-731.

Von der Emde G, Mogdans J, Kapoor B (2004) *The senses of fish: adaptations for the reception of natural stimuli*. Springer, India.

Wagner C, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlids adaptive radiation. *Molecular Ecology*, **22**, 787-798.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419-1439.

Webster M, Sheets HD (2010) A practical introduction to landmark-based geometric morphometrics. In: Alroy J, Hunt G, editors. Quantitative Methods in Paleobiology Paleontological Society Papers. **16**,163–188.

Weir BS, Cockerham CC (1984) Estimating f-statistics for the analysis of population structure. *Evolution*, **38**, 1358–70.

Weissing F, Edelaar P, Van Doorn G (2011) Adaptive speciation theory: a conceptual review. *Behavioral Ecology and Sociobiology*, **65**, 461-480.

Wennersten L, Forsman A (2012) Population-level consequences of polymorphism, plasticity and randomized phenotype switching: A review of predictions. *Biological Reviews*, **87**, 756-767.

Wilson CC, Mandrak NE (2004) History and evolution of lake trout in shield lakes: past and future challenges. In:Boreal Shield Watersheds: Lake Trout Ecosystems in a Changing Environment (eds Gunn JM, Steedman RJ, Ryder RA), CRCPress, Boca Raton, Florida.

Zimmerman MS, Schmidt SN, Krueger CC *et al*. (2009) Ontogenetic niche shifts and resource partitioning of LakeTrout morphotypes. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 1007–1018.

Zimmerman MS, Krueger CC (2009) An ecosystem perspective on re-establishing native deepwater fishes in the Laurentian Great Lakes. *North American Journal of Fisheries Management*, **29**, 1352-1371.

Data accessibility

Individuals raw sequences are available at the Sequence Read Archive (SRA) (Study accession SRP096183) and necessary data for genomic and morphometric analyses are available at Dryad doi: 10.5061/dryad.k713n .

Author contributions

A. Muir and F. Goetz provided samples and visual identification. L. Bernatchez and A. Muir conceived the study, and A. Perreault did the laboratory work, analysed the data and wrote the paper. C. Perrier assisted in data analysis and writing the manuscript. E. Normandeau helped for bioinformatic analysis. P. Sirois and Louis Bertnatchez provided the funding for the present study and helped editing of the manuscript. All authors approved and edited the manuscript.

Figures

Figure 1. Map of Lake Superior sampling sites; Isle Royale, Superior Shoals, Stannard Rock and Big Reef. Circles correspond to sampling locations for each site.

Figure 2. Between-group PCA on partial warps of 501 Lake Trout. **(a)** First and second principal components for head shape representing 56.2% and 15.9% of the variance respectively distinguishing the four ecotypes. **(b)** First and second principal components for body shape representing 65.5% and 14.5% of the variance respectively that distinguish leans from siscowets based mainly on belly curvature. **(c)** Third and fourth principal component for head shape representing 11.5% and 7.2% of the variance respectively distinguishing the four sites. **(d)** Third and fourth principal components for body shape representing 8.8% and 3.0% of the variance respectively distinguishing the four sites. The colored points refer to the mean scores for each ecotype in each site. The sites are: Big Reef (black), Isle Royale (blue), Stannard Rock (red) and Superior Shoals (green). Ecotypes are: Siscowet (FT), Humper (HT), Lean (LT) and Redfin (RF). Under each ecotype are drawn the consensus shapes of all four ecotypes (gray) with the outline of the ecotype in question (black). The shaded ellipses have been drawn for clarity.

Figure 3. Heatmaps of (a) calculated F_{ST} values, and (b) calculated Euclidean distances between groups averages for body (below diagonal) and head (above diagonal) shape for the four ecotypes and four sampling sites. Ecotypes are: Siscowet (FT), Humper (HT), Lean (LT) and Redfin (RF).

Figure 4. Population structure analysis of Lake Trout. **a)** Admixture plot based on 486 individuals and 6822 SNPs (including outliers) for different values of K. Individuals are shown by sites and ecotypes. **b)** Neighbour joining tree based on 486 individuals and 6822 SNPs including outliers. Yellow circles represent Big Reef, orange circles Stannard Rock, blue circles Isle Royale and green circles Superior Shoals. Bootstrapping support is indicated on each branch. The four ecotypes are represented for each site; Lean (LT), Humper (HT), Redfin (RF) and Siscowet (FT).

Figure 5. Assignment success of individuals to their sampling sites **(a)** or ecotypes **(b)**. Percentage assignment is written below circles with the exact number of individuals assigned within brackets. Percentage of correct assignment to either sampling sites or ecotypes is in bold. Sites are: Big Reef (BR), Isle Royale (IR), Stannard Rock (SR), Superior Shoals (SS). Ecotypes are: Humper (HT), Siscowet (FT), Lean (LT) and Redfin (RF).

Figure 6. Venn diagrams of outliers detected by LFMM and BAYESCAN among sites, ecotypes or among ecotypes within sites. **a)** Outliers detected among the four sites by BAYESCAN and LFMM including outliers detected by LFMM using morphological PC scores. **b)** Outliers detected among the four pooled ecotypes by BAYESCAN and LFMM including outliers detected by LFMM using morphological PC scores. **c)** Outliers among ecotypes within each site detected by LFMM.

Tables

Table 1. Sampling site information and consensus analysis of body shape, head shape and visual identification. Number of fish sampled per sampling site (N), year of collection and coordinates is provided. Ecotypes were identified by consensus analysis of body shape (B) and/or head shape (H) and/or visual identification (V). Fish for subsequent genetic analysis were chosen based on ecotype consensus. Fish less than 430 mm long were removed prior to the analysis.

Table 2. MANOVA on body and head shape to investigate the effect of the ecotype, the site of origin, the sex and all interactions. Significant variables are in bold.

Table 3. Number of SNPs remaining after each filtration step. Allelic imbalance corresponds to the ratio of the number of sequences for the major allele on the number of sequences for the minor allele.

Table 4. Population statistics estimated with 6822 SNPs: the observed heterozygosity (Ho), the expected heterozygosity (He), the inbreeding coefficient (Gis), the effective population size (N_e) and confidence interval in brackets, and the number of polymorphic loci (N). Genome-wide diversity (π) and the increase in individual homozygosity relative to mean Hardy-Weinberg expected homozygosity (F_h) was calculated on the dataset prior to filtration. Effective population size for ecotypes with sample size < 15 individuals were not calculated (NA). Ecotypes are: Siscowet (FT), Humper (HT), Lean (LT) and Redfin (RF).

Table 5. Analysis of molecular variance (AMOVA) on 486 individuals and 6822 SNPs. Missing data has been replaced by random picking in the overall pool of allele frequency.

Sites	Year	Coordinates	N		Lean-like	Siscowet-like	Humper-like	Redfin-like	no consensus	Total
Big Reef	2014	46°46,545°N 86°28,378°W	132	Consensus	39[B, H,V]	46[B, H, V]	8[V]	17[V]	13	123
				Chosen	39	40	8	17		104
Stannard Rock	2013-2014	47°11,450°N 87°11,531°W	362	Consensus	63[B, H, V]	66[B, H, V]	24[V]	24[V]	40	217
				Chosen	40	40	24	24		128
Isle Royale	2013	47°21,550°N 88°30,497°W	214	Consensus	55[B, H, V]	37[B, H, V]	35[H, V]	33[H, V]	42	202
				Chosen	40	37	34	33		144
Superior Shoals	2013	48°02,464°N 87°07,536°W	394	Consensus	35[B, H, V]	133[B, H, V]	11[V]	62[B, V]	74	315
				Chosen	31	41	11	42		125

	Body						Head				
Variables	Df	Pillai	Approx F	Num Df	Den Df	Pr (> F)	Df	Pillai	Approx F	Num Df	Den Df	Pr (> F)
Ecotype	3	0.59616	28.7686	12	1392	**< 2.2 e-16**	3	0.79729	27.871	18	1386	**< 2.2 e-16**
Site	3	0.43694	19.7752	12	1392	**< 2.2 e-16**	3	0.90345	33.181	18	1386	**< 2.2 e-16**
Sex	1	0.10732	13.8857	4	462	**1.052e-10**	1	0.10040	8.556	6	460	**7.883 e-09**
Ecotype: Site	9	0.47419	6.9486	36	1860	**< 2.2 e-16**	9	0.52612	4.966	54	2790	**< 2.2 e-16**
Ecotype: Sex	3	0.05388	2.1214	12	1392	**0.01339**	3	0.11061	2.948	18	1386	**3.28 e-05**
Site : Sex	3	0.04189	1.6427	12	1392	0.07412	3	0.04268	1.111	18	1386	0.3344
Ecotype: Site : Sex	9	0.07982	1.0520	36	1860	0.38565	9	0.12499	1.099	54	2790	0.2891

Before filtration:	**Count**
Catalog	1,052,664 loci
After population module (≥ 70% individuals in ≥ 2 sites)	212,804 SNP
Filters:	
Genotype quality:	
Genotype likelihood (≥ 6) Allelic imbalance (≤ 5)	193,678 SNP
Hardy-Weinberg:	
Heterozygosity (≤ 0.6) Fis [-0.3; 0.3]	185,445 SNP
MAF:	
Global (≥ 0.01) and/or Local (site) (≥ 0.05)	17,812 SNP
Locus:	
Maximum number of SNP per locus (≤ 8)	13,984 SNP
Position:	
1th SNP per locus kept	6822 SNP

		Ho	He	π	N_e	Gis	F_h	N
Big Reef	**LT**	0.067	0.067	0.000319	415 [356; 498]	-0.007	-1.04E-07	4132
	FT	0.074	0.072	0.000353	925 [698; 1365]	-0.027	-1.29E-07	4564
	RF	0.073	0.071	0.000337	1341 [646; infinite]	-0.03	-1.44E-07	3238
	HT	0.072	0.069	0.000334	NA	-0.034	-1.84E-07	2022
Isle Royale	**LT**	0.075	0.077	0.000393	181 [171; 194]	0.027	-5.58E-08	4904
	FT	0.074	0.075	0.000358	122 [116; 128]	0.012	-5.52E-08	4751
	RF	0.079	0.080	0.000387	124 [118; 131]	0.015	-5.41E-08	4663
	HT	0.074	0.075	0.000366	192 [178; 209]	0.017	-5.36E-08	4510
Stannard Rock	**LT**	0.062	0.061	0.000284	487 [408; 603]	-0.011	-9.50E-08	3649
	FT	0.062	0.061	0.000285	159 [149; 171]	-0.007	-9.18E-08	3638
	RF	0.062	0.061	0.000281	250 [213; 301]	-0.018	-1.03E-07	2923
	HT	0.061	0.060	0.000276	199 [174; 230]	-0.016	-1.05E-07	2920
Superior Shoals	**LT**	0.081	0.082	0.000385	146 [137; 155]	0.01	-6.30E-08	4538
	FT	0.078	0.080	0.000378	133 [127; 139]	0.021	-5.14E-08	5022
	RF	0.076	0.077	0.000367	94 [91; 97]	0.007	-7.57E-08	5227
	HT	0.073	0.074	0.000367	NA	0.012	-7.57E-08	2727

Source of variation	% Variance	F-stat	F-value	Std.Dev.	c.i.2.5%	c.i.97.5%	p-value
Sites as group							
Among sites	0.011	F_{CT}	0.011	0.000	0.01	0.012	< 0.001
Among ecotypes within sites	0.004	F_{SC}	0.004	0.000	0.004	0.005	< 0.001
Ecotypes as group							
Among ecotypes	-0.002	F_{CT}	-0.002	0.000	-0.002	-0.002	0.95
Among siteswithin ecotypes	0.015	F_{SC}	0.015	0.000	0.014	0.016	< 0.001