MOLECULAR ECOLOGY

Molecular Ecology (2017) **26**, 178–192 doi: 10.1111/mec.13816

SPECIAL ISSUE: THE MOLECULAR MECHANISMS OF ADAPTATION AND SPECIATION: INTEGRATING GENOMIC AND MOLECULAR APPROACHES

Standing chromosomal variation in Lake Whitefish species pairs: the role of historical contingency and relevance for speciation

ANNE-MARIE DION-CÔTÉ,*[1] RADKA SYMONOVÁ,† FABIEN C. LAMAZE,‡ ŠÁRKA PELIKÁNOVÁ,§ PETR RÁB§ and LOUIS BERNATCHEZ*

Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, 1030, Avenue de la Médecine, Québec, Québec, Canada G1V 0A6, †Research Institute for Limnology, University of Innsbruck, Mondseestraße 9, A-5310 Mondsee, Austria, ‡Ontario Institut for Cancer Research, MaRS Centre, 661 University Avenue, Suite 510, Toronto, Ontario, Canada M5G 0A3, §Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, AS CR, vvi, Liběchov 277 21, Czech Republic,

Abstract

The role of chromosome changes in speciation remains a debated topic, although demographic conditions associated with divergence should promote their appearance. We tested a potential relationship between chromosome changes and speciation by studying two Lake Whitefish (*Coregonus clupeaformis*) lineages that recently colonized postglacial lakes following allopatry. A dwarf limnetic species evolved repeatedly from the normal benthic species, becoming reproductively isolated. Lake Whitefish hybrids experience mitotic and meiotic instability, which may result from structurally divergent chromosomes. Motivated by this observation, we test the hypothesis that chromosome organization differs between Lake Whitefish species pairs using cytogenetics. While chromosome and fundamental numbers are conserved between the species ($2n$ = 80, NF = 98), we observe extensive polymorphism of subtle karyotype traits. We describe intrachromosomal differences associated with heterochromatin and repetitive DNA, and test for parallelism among three sympatric species pairs. Multivariate analyses support the hypothesis that differentiation at the level of subchromosomal markers mostly appeared during allopatry. Yet we find no evidence for parallelism between species pairs among lakes, consistent with colonization effect or postcolonization differentiation. The reported intrachromosomal polymorphisms do not appear to play a central role in driving adaptive divergence between normal and dwarf Lake Whitefish. We discuss how chromosomal differentiation in the Lake Whitefish system may contribute to the destabilization of mitotic and meiotic chromosome segregation in hybrids, as documented previously. The chromosome structures detected here are still difficult to sequence and assemble, demonstrating the value of cytogenetics as a complementary approach to understand the genomic bases of speciation.

Keywords: *Coregonus*, cytogenetics, polymorphism, salmonids, speciation, standing genetic variation

Received 30 November 2015; revision received 8 August 2016; accepted 11 August 2016

Correspondence: Anne-Marie Dion-Côté, Fax: (418) 656 7176; E-mail: anne-marie.dion-cote.1@ulaval.ca

[1]Present address: Molecular Biology and Genetics Department, Biotechnology building, Cornell University, 526 Campus Road, Ithaca, NY, 14853, USA

Introduction

Understanding the role of genetic and chromosomal changes associated with divergence is a major focus in evolutionary biology (Brown & O'Neill 2010; Marie Curie SPECIATION Network 2012). Thanks to the

advent of massive parallel sequencing technologies, considerable progress has been made in the past decade to decipher the genetic basis of adaptation and speciation (Seehausen *et al.* 2014). However, repetitive regions such as centromeres, constitutive heterochromatin and associated repetitive elements remain challenging to sequence, assemble and characterize (Kim *et al.* 2014), hindering our understanding of their role in population divergence and speciation. Cytogenetic techniques specifically targeting these regions are complementary to next-generation sequencing approaches. They can help to reveal how genomes are structurally organized into chromosomes, how they are shaped by the interplay of evolutionary forces and how chromosome structure changes contribute to speciation (Brown & O'Neill 2010; Faria & Navarro 2010). Such integrative approaches in plants, yeast, mammals and fishes begin to reveal that chromosome structure changes (in addition to well-documented inversions) are associated with divergence and reproductive isolation (Symonová *et al.* 2013; Charron *et al.* 2014).

Some heterochromatic and other poorly assembled repetitive regions are involved in gene expression regulation, meiotic recombination, chromosome segregation and genome stability (Grewal & Jia 2007; Hoskins *et al.* 2007; Brown & O'Neill 2010; Cioffi & Bertollo 2012; Altemose *et al.* 2014). Accordingly, these chromosome structures may greatly influence the genomic landscape of speciation, modulate recombination rate along chromosomes and impact hybrid fitness (Dernburg *et al.* 1996; Grewal & Jia 2007; Brown & O'Neill 2010). In addition, variations in heterochromatin and repetitive region distribution represent a substantial source of intraspecific variation (King 1993; Kidd *et al.* 2008; Britton-Davidian *et al.* 2012), which in turn may modulate the extent of reproductive isolation being achieved during the process of speciation (Cutter 2012). Therefore, more integrative studies are needed in order to understand the role of heterochromatic and repetitive regions in divergence and speciation.

Salmonids typically display substantial inter- and intraspecific chromosome rearrangement and polymorphism (Phillips & Ráb 2001; Sutherland *et al.* 2016), which may be the result of the plasticity of their genome conferred by their ancestral tetraploid state (Allendorf & Thorgaard 1984; Mable *et al.* 2011). Indeed, teleosts experienced a third whole genome duplication (3R – WGD) event preceding their diversification ~350 MYA and salmonids underwent an additional salmonid-specific WGD (4R) 60–90 MYA (Allendorf & Thorgaard 1984; Crête-Lafrenière *et al.* 2012; Macqueen & Johnston 2014). Cytogenetic studies have often revealed intraspecific polymorphism, typically resulting from Robertsonian fusion and fission of chromosomes, but also involving additions and deletions of heterochromatin (reviewed in Phillips & Ráb 2001). For example, the largest metacentric chromosome in *Coregonus* shows length polymorphism in the Lake Whitefish *C. clupeaformis* and the closely related European Whitefish *C. lavaretus*, possibly resulting from variable heterochromatin content (Ráb & Jankun 1992; Jankun *et al.* 1995; Phillips *et al.* 1996; Jankun & Ráb 1997). In Lake Trout (*Salvelinus namaycush*), large blocks of heterochromatin are heritable, as shown by inheritance studies, and polymorphic, in terms of presence/absence and band size (Phillips & Ihssen 1986). However, the role of these intrachromosomal changes has rarely been examined during the early stages of speciation.

The well-characterized phylogeography of the Lake Whitefish makes it a useful model to study potential chromosome changes in the context of divergence and speciation. During the Pleistocene glaciation, two Lake Whitefish lineages (the Atlantic and Acadian lineages) underwent geographical isolation ~60 000 YBP (or ~15–20 000 generations ago) in northeastern North America (Jacobsen *et al.* 2012). The Atlantic and Acadian lineages repeatedly came into secondary contact when they independently colonized newly formed lakes following the Laurentide ice sheet retreat ~12 000 YBP (3–4000 generations ago; Bernatchez & Dodson 1991). Competitive interactions and niche availability presumably contributed to the divergence of a derived, dwarf limnetic form from the ancestral normal benthic form in multiple lakes (Landry *et al.* 2007; Landry & Bernatchez 2010). Previous phylogeographic studies indicated that the dwarf form originated from the Acadian lineage, with variable levels of genetic divergence and admixture with the sympatric normal form of the Atlantic lineage (Lu & Bernatchez 1999; Renaut *et al.* 2012; Gagnaire *et al.* 2013). We subsequently refer to the normal and dwarf Lake Whitefish as distinct species, acknowledging the considerable level of divergence and reproductive isolation between them.

Earlier genetic and transcriptomic studies provided no evidence for differentially fixed mutations nor substantial gene expression differences between normal and dwarf Lake Whitefish (e.g. Campbell & Bernatchez 2004; Renaut *et al.* 2009; Gagnaire *et al.* 2013; Hébert *et al.* 2013; Dion-Côté *et al.* 2014). Yet pronounced postzygotic reproductive isolation has been documented, thus qualifying dwarf and normal Lake Whitefish as distinct species under a relaxed interpretation of the biological species concept. F1-hybrids and backcrosses suffer from a much higher embryonic mortality rate relative to pure parental forms (Lu & Bernatchez 1998; Rogers & Bernatchez 2006; Renaut *et al.* 2009). In backcrosses, hybrid breakdown involves the appearance of a characteristic malformed phenotype [including

reduction in head and eyes size and deformed tail, as described by Renaut & Bernatchez (2011)], gene expression deregulation and transposable element derepression (Renaut & Bernatchez 2011; Dion-Côté *et al.* 2014). Moreover, healthy and malformed backcrosses experience mitotic instability and meiotic breakdown respectively (Dion-Côté *et al.* 2015), suggesting a role for a chromosomal component to reproductive isolation in Lake Whitefish. Nevertheless, it remains unknown whether chromosome rearrangements or structure changes occurred between normal and dwarf Lake Whitefish, nor is it known whether these potential changes played a role in reproductive isolation, given that only modest genetic differences have been documented between them.

In this study, we report detailed cytogenetic characterization of three sympatric pairs of normal and dwarf Lake Whitefish, specifically targeting intrachromosomal markers associated with heterochromatin and repetitive sequences. We test the hypothesis that divergence in the Lake Whitefish system is accompanied by differentiation at the chromosomal and subchromosomal levels. Such structural divergence may result from ancestral allopatry, demographic processes associated with postglacial lake colonization such as founder effect, divergence in sympatry or a combination of these factors. While the basic karyotype remained stable among all populations examined, we observe that intrachromosomal accumulations of heterochromatin and repetitive regions are highly polymorphic. By applying multivariate analyses to these cytogenetic markers, we observe divergence between glacial lineages, among lakes and to some extent between species, and further identify markers associated with this divergence. We discuss how these observations support the presence of *standing chromosomal variation* at the time of lake colonization, and whether this type of polymorphism may be associated with reproductive isolation in this system.

Materials and methods

Sampling and chromosome suspension preparation

We sampled individuals from three lakes with sympatric populations of the St. John River basin: Cliff Lake (ME, USA), Témiscouata Lake (Québec, Canada) and East Lake (Québec, Canada), which are part of a long-term research programme on Lake Whitefish (Bernatchez *et al.* 2010). Glacial lineage assignment (Atlantic or Acadian) relies on earlier phylogeographic studies, which showed variable levels of admixture between glacial lineages within each lake (Lu *et al.* 2001; Pigeon *et al.* 1997). Species assignment (dwarf or normal) was done visually and any individual presenting an ambiguous phenotype was excluded from the study. In Cliff Lake, there is no admixture between dwarf and normal species and they are thus considered as being of pure Acadian and Atlantic origin, respectively. At the other end of the spectrum, normal individuals from Témiscouata Lake are of ~50% Acadian origin while dwarf individuals are of ~60% Acadian origin, likely resulting from a higher level of admixture following secondary contact. Recent analyses based on RADseq genotyping and historical demography inferences suggest that secondary contact also occurred in East Lake (Rougeux *et al.*, in preparation), where previous analyses of only a few available markers had predicted that these individuals originated from a single origin (Acadian). Yet we still labelled East Lake individuals as being of 'Acadian' origin, owing to excessively asymmetrical admixture. To summarize, in Cliff and Témiscouata Lakes dwarf and normal individuals are assigned to Acadian and Atlantic lineages, respectively, while dwarf and normal individuals from East Lake were both assigned to the Acadian lineage. In total, 29 individuals were sampled among the three lakes [see Table 1 for a summary and Table S1 (Supporting information) for more details].

Chromosome suspensions were prepared as described by Fujiwara *et al.* (2001) with some modifications (Dion-Côté *et al.* 2015). Between 0.2 and 2 mL of fresh blood was sampled with heparinized syringes and kept on ice for no more than 12 h. White blood cells were transferred to 5 mL of freshly prepared cell culture media [media 199 (Life technologies), 10% FBS (Sigma), 0.01% LPS (Sigma), 60 μg/mL kanamycin (Sigma), 18 μg/mL phytohemagglutinin (Sigma), 0.5× antibiotic antimycotic (Sigma) and 1.75 μL of 10% β-mercaptoethanol per 100 mL of media]. Cells were incubated for 6 days at 20 °C with gentle mixing every 24 h. Colchicine (25 μL of a 1% solution) was added to the cell suspension 45 min before collection. Cells were hypotonized for 20 min in 2 mL of 0.075 M KCl at room temperature and then fixed by the addition of an equal volume of fresh fixative (3:1 methanol:acetic acid). Three washes with fixative were performed before dropping the suspensions on slides (SuperFrost quality).

Giemsa, Chromomycin A_3 and C-band stainings

Metaphase spreads were stained for 10 min in 3% Giemsa-Romanowski (Dr. Kulich Pharma, Hradec Králové, Czech Republic) in phosphate buffer (pH 6.8–7.0) and then rinsed thoroughly with dH_2O. Chromosomes were sequentially stained with Chromomycin A_3 (CMA_3) and C-banding (with DAPI as a counter stain) according to Rábová *et al.* (2015). CMA_3 stains GC-rich

Table 1 Number of individuals analysed per lake and species with their average phenotypic characteristics (average ± standard deviation)

Lake	Species	Lineage	*n* individuals	Average weight (g)	Average length (cm)	Average Fulton	Average *n* gill rakers
Témiscouata	Dwarf	Acadian	5	82.00 ± 9.90	21.04 ± 0.90	0.88 ± 0.04	22.60 ± 3.51
	Normal	Atlantic	5	197.42 ± 111.41	26.56 ± 5.00	0.97 ± 0.06	22.80 ± 0.45
East	Dwarf	Acadian	4	48.98 ± 22.98	17.68 ± 2.18	0.84 ± 0.14	21.50 ± 2.12
	Normal	Acadian	4	310.38 ± 251.40	29.80 ± 6.49	1.01 ± 0.10	24.25 ± 2.50
Cliff	Dwarf	Acadian	5	138.68 ± 21.97	23.86 ± 1.61	1.02 ± 0.10	24.75 ± 1.26
	Normal	Atlantic	6	427.33 ± 25.42	33.97 ± 0.89	1.10 ± 0.03	24.00 ± 1.41

Fulton: Fulton condition index.

DNA regions, which often colocalize with heterochromatin and major rDNA genes, while C-banding stains constitutive heterochromatin (i.e. chromatin that remains compacted through interphase) presumably associated with repeats, including centromeres (Comings 1978). Chromosomes were examined using a Provis AX70 Olympus microscope, and images taken with a CCD camera (DP30W Olympus) equipped with standard filters. To reduce technical artefacts, at least 10 metaphases per individual were examined, and only consistent signals among metaphases were scored.

Fluorescent in situ hybridization (FISH)

We amplified the whole 5S rDNA and adjacent non-transcribed DNA segments (~170 bp) using previously published primers (5S-A: TACGCCCGATCTCGTCC GATC, 5S-B: CAGGCTGGTATGGCCGTAAGC, Pendás *et al.* 1995). A ~240-bp fragment of the 28S rDNA was also amplified using published primers (28S-C1: ACCCGCTGAATTTAAGCAT, Dayrat *et al.* 2001; 28S-D2: TCCGTGTTTCAAGACGGG, Chombard *et al.* 1998). PCR product identities were confirmed by Sanger sequencing (Macrogen Inc., the Netherlands). PCR products were purified on agarose gel and labelled with biotin-dUTP or digoxigenin-dUTP using the Roche Nick Translation kit according to the manufacturer's instructions (Roche, Mannheim, Germany). Chromosomes were prepared for hybridization according to Cremer *et al.* (2008) following a minimal ageing of 3 h at 37 °C. Labelled probes were hybridized for 24 h at 37 °C. Cy-3-Streptavidin (Invitrogen, San Diego, USA) and anti-digoxigenin–fluorescein (Roche, Mannheim, Germany) were used to detect biotin-dUTP- and digoxigenin-dUTP-labelled probes, respectively.

Multiple factor analyses

We used multivariate analyses to (i) test whether chromosome changes are associated with explanatory variables and (ii) identify chromosome markers associated with these explanatory variables. The input data set included categorical explanatory variables (Sex, Lineage, Lake, Species, Lake:Species), continuous explanatory variables (weight, length, Fulton's condition index [$K = W/L^3 \times 100$; where K = Fulton's condition index, W = weight in g, L = length in cm] and the number of gill rakers) and cytogenetic markers (Table S1, Supporting information). We included phenotypic variables because they may better reflect the proportion of dwarf or normal ancestry of each individual, considering a certain level of gene flow in all three lakes. Therefore, continuous explanatory variables were considered in the imputation of missing data.

As suggested by Dobigny *et al.* (2004), cytogenetic markers were transformed into a presence/absence matrix where a homozygote for absence is coded as '0', a heterozygote is '1' and a homozygote for presence is '2' (Table S1, Supporting information). Coding these variables as continuous better reflects their bi-allelic nature: the heterozygote is more closely related to each homozygote, while each homozygote is less similar. Since acrocentric chromosomes cannot be readily distinguished from one another, we counted the number of chromosomes that had CMA_3 and rDNA 28S (by FISH) signals (numbered 0–6). Similarly, we coded the acrocentric markers (aCMA and a28S) as continuous, which provides a better index of relative similarity between homo- and heterozygotes (e.g. aCMA = 5 is more similar to aCMA = 3 than to aCMA = 0). Missing or ambiguous data were coded as 'NA'. The marker names refer to (i) the number of the chromosome (but with indistinguishable acrocentric 'numbered' as 'a'), (ii) the chromosome arm on which it is found, (iii) the staining technique by which it was evidenced and (iv) the band number on the chromosome arm, where 1 is the closest to the centromere (see Table S1, Supporting information). Markers 1p and 10p refer to the length of the *p*-arm of chromosome 1 and 10, respectively, 0 being the homozygote for short form and 2 being the homozygote for long form. The final data set comprised 39 polymorphic markers (present in >1 individual) in

29 individuals. There were 143 missing data points ('NA') out of 1131 entries (12.6%, Table S1, Supporting information).

Missing data were imputed with the function imputeMFA() from the MISSMDA package (version 1.7.3, Josse & Husson 2012) in R (version 3.2.4, R Core Team 2012). This method imputes missing data (cytogenetic markers) based on the means of other known variables among *similar* individuals (length, weight, Fulton's condition index, gill raker numbers and cytogenetic markers). Without this preliminary step, downstream analyses would have replaced missing data by the mean of *all* of the individuals, thus potentially blurring signal. The imputation included four groups of variables (continuous and categorical): (i) explanatory categorical variables: glacial lineage (Atlantic or Acadian, *sensu* Bernatchez & Dodson 1991), lake (Cliff, East or Témiscouata), species (dwarf or normal) and the combination of lake and species (Lake:Species); (ii) explanatory continuous variables: weight, length, Fulton's condition index and the number of gill rakers; (iii) multichromosomic markers on indistinguishable acrocentric chromosomes; (iv) bi-allelic chromosome markers.

We used the function estim_ncpPCA() from the MISSMDA package (version 1.10, Josse & Husson 2016) to determine the number of principal components (PC) to be used to impute missing data, that is the number of PCs leading to the smallest mean square error of prediction. A function to estimate the number of PCs in MFA has not yet been implemented, but this function gives an approximate alternative (Julie Josse, personal communication). Three principal components were thus used for the imputation (ncp = 3) using the 'Regularized' method (see the MISSMDA package documentation for more details). The resulting complete data set (Table S2, Supporting information) was used for subsequent analyses.

We then performed a Multiple Factorial Analysis (MFA) on the chromosome markers using the MFA() function from the FACTOMINER package (version 1.32, Lê *et al*. 2008). Variables from groups 1 and 2 (explanatory categorical and continuous variables) were coded as supplementary, to test a potential relation with the chromosome markers, and thus did not contribute to defining the dimensions. The chromosome markers were coded in two groups: (i) multichromosomic aCMA3 and a28S markers ('a' stands for 'acrocentric', from 0 to 6 sites) and (ii) bi-allelic markers (from 0 to 2 for zygosity). The function dimdesc() from the FACTOMINER package was used to retrieve supplementary variables (factors and factor levels) significantly linked to dimensions constructed by the MFA. This function applies an ANOVA model with one factor for each dimension. *F*-tests were used to detect the association of each explanatory variable with the dimension (Sex, Lineage, Lake, Species, Lake:Species), followed by *t*-tests for each factor level (Atlantic, Acadian; Témiscouata, East, Cliff; Normal, Dwarf; Témiscouata:Dwarf, Témiscouata:Normal, East:Dwarf, etc.). The dimdesc() function also provides the estimate of the barycentre position (centroid) for each factor level with a significant association with MFA dimensions. The barycentre position can be interpreted as the average position of the individuals with this characteristic in the multivariate space. The function coord.ellipses() from the FACTOMINER package was used to calculate 95% confidence ellipses around the barycentre position.

Results

Karyotypes are stable among Lake Whitefish species pairs

Conventional Giemsa staining confirmed that the Lake Whitefish karyotype is of the salmonid type A *sensu* Phillips & Ráb (2001) and is conserved in all six Lake Whitefish populations from the three lakes ($2n = 80$, NF = 98). This karyotype includes 10 pairs of metacentric/submetacentric chromosomes, one pair of large acrocentric chromosomes and 29 pairs of subtelo/acrocentric chromosomes of decreasing size (Booke 1968; Phillips *et al*. 1996; Dion-Côté *et al*. 2015). Because chromosome 10 is a small submetacentric chromosome (almost subtelocentric) with variable length of the very small *p*-arm, we counted only two arms for this chromosome (and not four as for other submetacentric/metacentric chromosomes), following Phillips & Ráb (2001) guidelines, resulting in the NF = 98 instead of 100.

Subtle karyotype polymorphism was revealed and subsequently included in the multivariate analysis. As shown in Fig. 1, the length of the *p*-arm of chromosome 1 was polymorphic among individuals from all lakes. The short form was near fixation in dwarf fish compared to normal fish from Cliff Lake, while the opposite trend was found in fish from Témiscouata Lake (marker '1p', Table S1, Supporting information). The *p*-arm of chromosome 10 was also polymorphic among individuals from all three lakes (marker '10p', Table S1, Supporting information). In some instances, it was clearly submetacentric, while in others it was subtelomeric/acrocentric (Fig. 1). The *p*-arms of chromosome 1 and 10 are both heterochromatic, although of different nature as suggested by DAPI and C-banding staining, respectively, which may explain their length polymorphism.

In addition, we identified a B or supernumerary chromosome in one dwarf individual from Cliff Lake (Figure S1, Supporting information). B chromosomes,

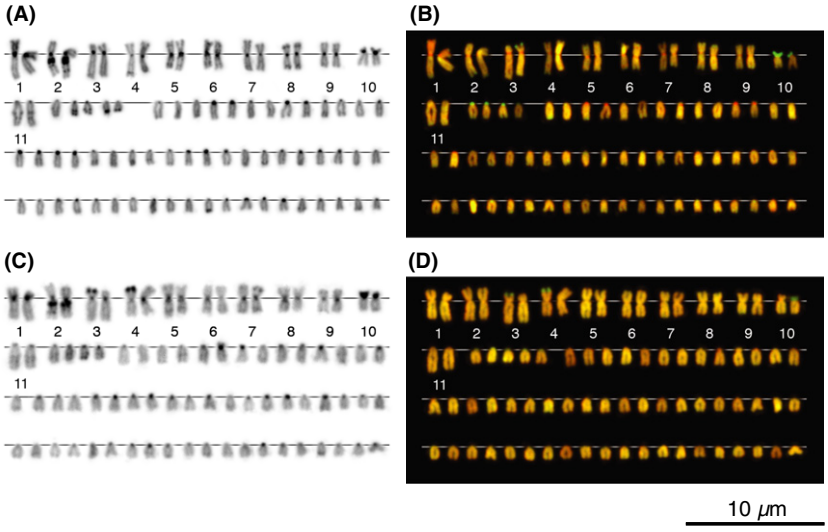

Fig. 1 Karyotypes of a normal and a dwarf individual from East Lake (upper and lower, respectively), sequentially stained with Giemsa (not shown), Chromomycin A_3 (CMA_3) and C-bands to exemplify polymorphism of heterochromatin blocks. Identifiable chromosomes are numbered, followed by acrocentric chromosomes with markers scored, and then remaining acrocentric chromosomes by decreasing size. (A) C-banded karyotype of a normal individual from East Lake (EN27). Note the long form of chromosome 1. (B) CMA_3 karyotype of the same normal individual from East Lake. (C) C-banded karyotype of a dwarf individual from East Lake (ED13). Note the short form of chromosome 1. (D) CMA_3 karyotype of the same dwarf individual from East Lake.

which are usually derived from A chromosomes, occur in some individuals of a population and do not segregate in a Mendelian fashion (Jones 1995; Camacho 2005). This bi-armed B chromosome was not present in all cells examined and was also never found in more than one copy. Additionally, it was positively stained with CMA_3 and C-bands, consistent with the presence of heterochromatin blocks and the presence of repeated elements, a common feature of B chromosomes (Ziegler *et al.* 2003; Camacho 2005; Valente *et al.* 2014).

Heterochromatin revealed by CMA_3 and C-banding

Polymorphic accumulations of heterochromatin and repetitive sequences were characterized among lakes and species pairs. Several chromosomes showed polymorphic CMA_3 banding patterns, indicative of GC-rich DNA regions (Fig. 1). Telomeric CMA_3 signals were present on chromosomes 3, 4, 5, 9 and 10 and on the centromere/*p*-arm of one to six acrocentric chromosomes. A strong telomeric CMA_3 signal on acrocentric chromosomes was found in only two normal individuals, one from Témiscouata Lake and the other from Cliff Lake (marker 'aCMA-telo', Table S1, Supporting information). A large CMA_3-positive block on the *p*-arm of chromosome 4 was also found in a dwarf and a normal individual from Cliff Lake. The same individuals were also the only ones showing CMA_3 band on the q-arm of chromosome 4. Although there was no fixed association between these variations and specific glacial lineage, species or lake, differential representations of these variants contributed to the resolution of significant sex, lake and glacial lineage clusters when combined with other chromosomal structures in subsequent multivariate analyses.

Monomorphic and polymorphic C-bands, indicating constitutive heterochromatin, were found. Most

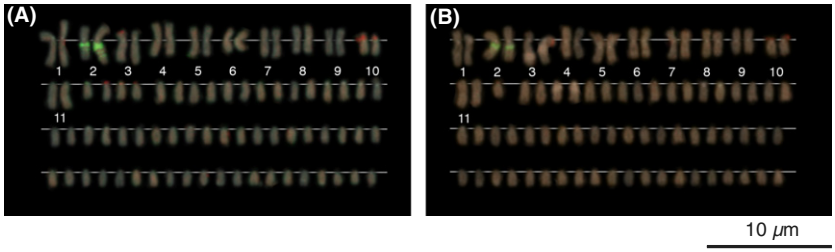

Fig. 2 Karyotypes of (A) a normal (EN17) and (B) a dwarf (ED13) individual from East Lake arranged from DAPI-stained chromosomes after FISH with 5S (green) and 28S (red) rDNA probes showing polymorphism of both rDNA sites.

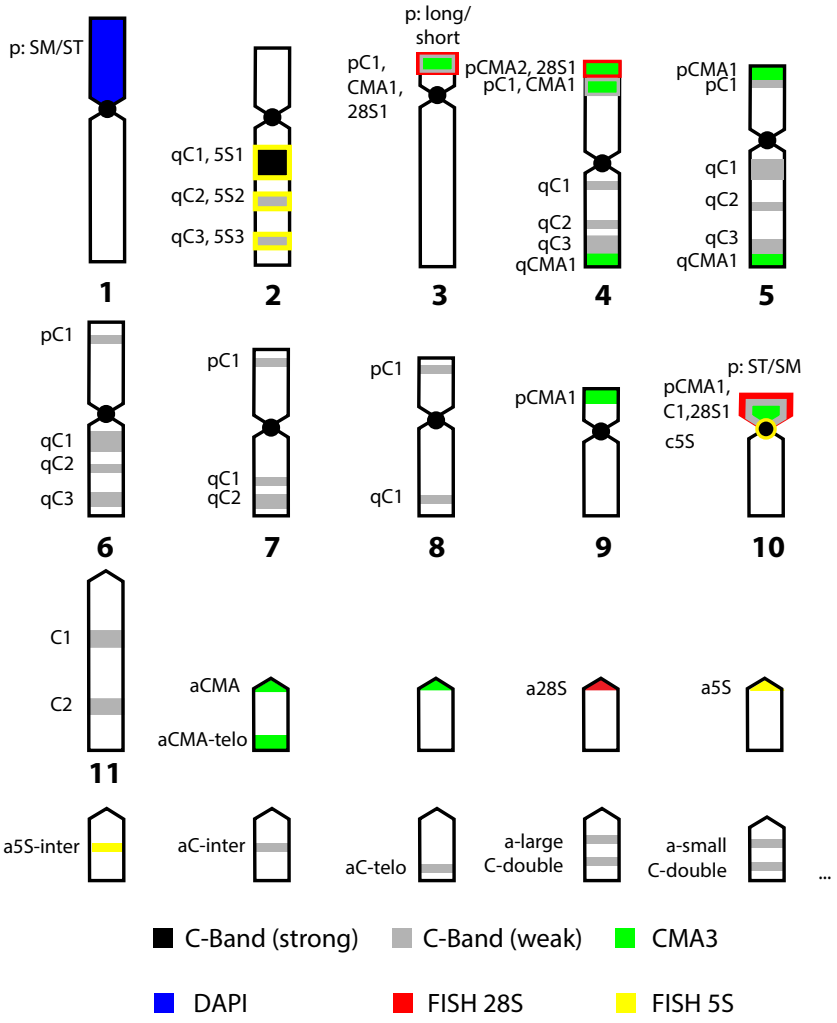

Fig. 3 Partial consensus ideogram for all three species pairs showing chromosome shape and all markers identified on chromosomes and scored. Markers are named according to the chromosome arm on which they are, technique used and distance from centromere (1 for the closest, then 2, etc.). The eleven readily identifiable chromosomes are numbered 1–11, followed by nine acrocentric chromosomes bearing markers (20 chromosomes missing).

chromosomes had centromeric C-bands in all individuals, although the staining was stronger in meta/submetacentric chromosomes compared to acrocentric chromosomes (Fig. 1). Several monomorphic heterochromatin blocks were found: (i) a large heterochromatin block on the q-arm of chromosome 2, close to the centromere; (ii) three bands on the q-arm of chromosomes 5 and 6, and (iii) a double interstitial C-band on the large acrocentric chromosome 11. The remaining bands were all polymorphic, and none were differentially fixed among sex, glacial lineages, lakes or species.

Polymorphism of 5S and 28S ribosomal RNA gene sites

Ribosomal RNA genes (rDNA) are organized as tandem repeats often associated with transposable elements (Cioffi *et al.* 2010; Symonová *et al.* 2013; Vergilino *et al.* 2013). Several sites located on different chromosomes hybridized with the 5S rDNA probe, most of which colocalized with C-bands (Fig. 2). There were three major 5S rDNA sites on chromosome 2: the two distal sites were polymorphic, similar to C-bands (Figs 1 and 2). The 5S rDNA signal was much weaker for other sites. One small interstitial band was found on chromosome 1, a centromeric signal on chromosome 10 (which almost colocalized with 28S rDNA), and centromeric and interstitial signals on different acrocentric chromosomes. Polymorphic 28S rDNA signals were also detected on the *p*-arms of chromosomes 3, 4 and 10. Finally, zero to six 28S rDNA signals were found on the *p*-arms/centromeres of acrocentric/subtelomeric chromosomes. These 28S rDNA sites tended to strongly colocalize with CMA_3 staining. All markers identified and described above are summarized in Fig. 3.

Multiple factorial analysis (MFA) reveals divergence between lineages and among lakes but no parallelism between species

To detect patterns of cytogenetic variation among all fish analysed, a multiple factorial analysis (MFA) was

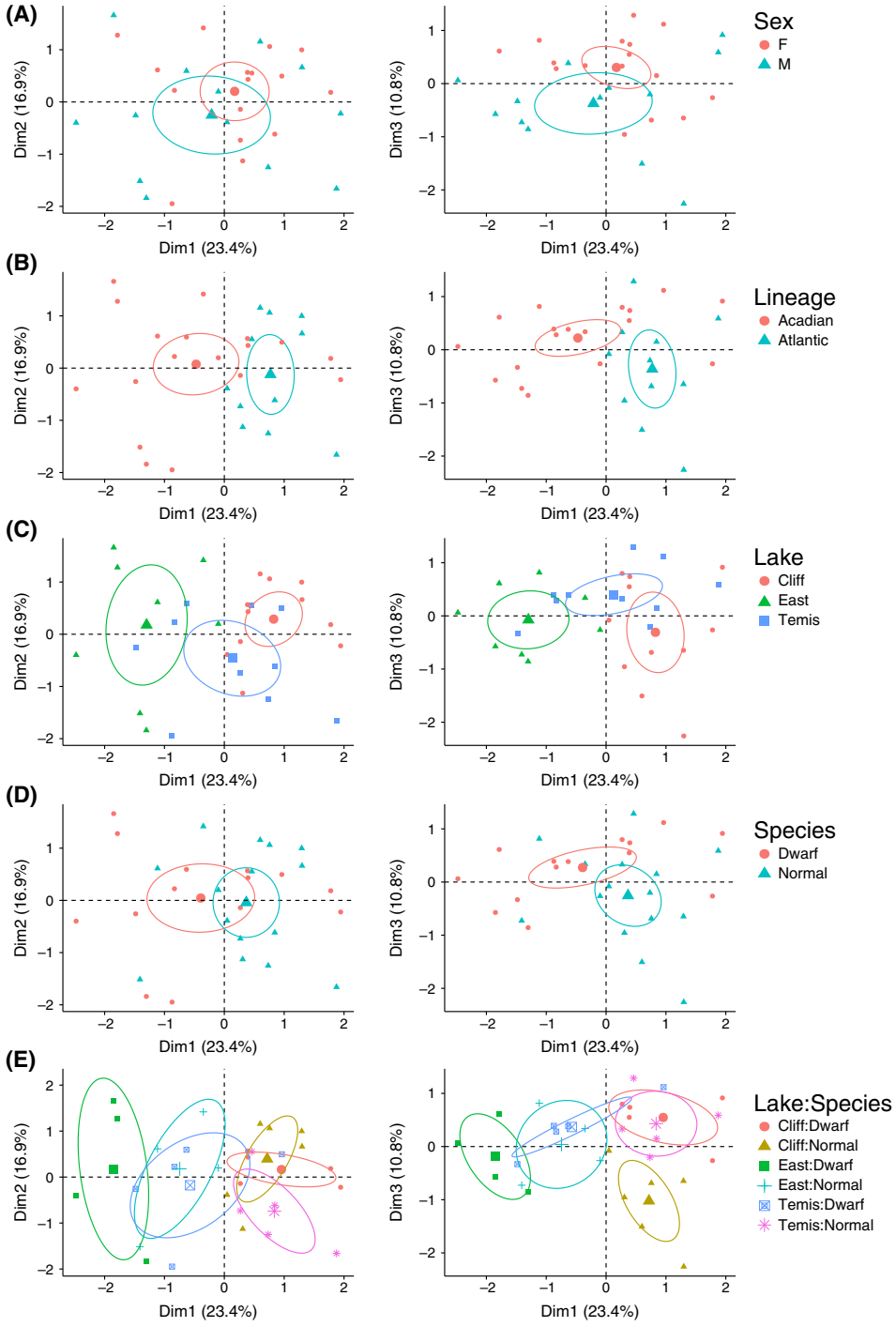

Fig. 4 Multiple factor analysis performed with FACTOMINER showing results for dimensions 1 and 2 (left) and 1 and 3 (right). Each data point represents a single individual in the multivariate space. (A) 95% confidence ellipses around the factor levels of 'Sex', (B) 95% confidence ellipses around the factor levels of 'Lineage', (C) 95% confidence ellipses around the factor levels of 'Species', (D) 95% confidence ellipses around the factor levels of 'Lake' and (E) 95% confidence ellipses around the factor levels of 'Lake:Species'.

applied to chromosome markers, after imputation of missing data (Table S2, Supporting information). Together, dimensions 1 and 3 (23.4% and 10.8% of the variance, respectively) revealed differentiation between glacial lineages, sex and among lakes (Fig. 4). Sympatric dwarf and normal whitefish within each lake also tended

Table 2 Factor effect on dimensions 1, 2 and 3 from the multiple factor analysis

Factor	R^2	*P*-value
Dimension 1		
Sex	0.0268	0.396
Lineage	0.2664	0.004
Lake	0.5352	<0.001
Species	0.1057	0.085
Lake:Species	0.7267	<0.001
Dimension 2		
Sex	0.0517	0.236
Lineage	0.0092	0.620
Lake	0.1148	0.205
Species	0.0019	0.822
Lake:Species	0.1474	0.564
Dimension 3		
Sex	0.1793	0.022
Lineage	0.127	0.058
Lake	0.1469	0.127
Species	0.1102	0.079
Lake:Species	0.5204	0.003

R^2 and *P*-values (*F*-test) were calculated by the dimdesc() function from the FACTOMINER package. The factor 'Lake:Species' denotes the combination of both characteristics (and not a statistical interaction). No factor had a significant effect on dimension 2.

to diverge (minimal overlap between ellipses), although not in parallel and not significantly. This can be seen by examining 95% confidence ellipses around the centroid position for each variable analysed (Fig. 4). The second dimension (16.9% of the variance) was not significantly associated with sex, glacial lineage, lake or species.

Dimension 1 correlated significantly with the variables 'Lake:Species', 'Lake' and 'Lineage' (R^2 = 0.73, 0.53, 0.27, respectively, *F*-test, all *P*-values <0.01; Table 2). Glacial lineage factor levels 'Atlantic' and 'Acadian' were significantly associated with dimension 1 (*t*-test, *P*-value <0.01; Table 3). Cliff and East Lakes were also differentiated by dimension 1 (*t*-test, *P*-value = 0.002 and <0.001, respectively; Table 3). East Lake and Cliff Lake dwarf whitefish were also significantly differentiated by dimension 1 (*t*-test, *P*-value <0.01 and = 0.046, respectively; Table 3). Finally, Fulton condition index was significantly correlated with dimension 1 (R^2 = 0.48, *q*-value = 0.05, *F*-test), but not length alone, weight alone or gill rakers number.

Dimension 3 correlated significantly with the variables 'Lake:Species' and 'Sex' (R^2 = 0.52, 0.18, respectively, *F*-test, *P*-values <0.05; Table 2). It is worth mentioning that the variables 'Lineage' and 'Species' almost reached significance but were more weakly cor-

Table 3 Barycentre (centroid) position estimates of the factor levels on dimensions 1, 2 and 3 from the multiple factor analysis

Factor level	Estimate	*P*-value
Dimension 1		
M	−0.1922	0.396
F	0.1922	0.396
Acadian	−0.6209	0.004
Atlantic	0.6209	0.004
Temis	0.246	0.673
East	−1.1847	<0.001
Cliff	0.9388	0.002
Dwarf	−0.3798	0.085
Normal	0.3798	0.085
Temis:Dwarf	−0.4615	0.244
Temis:Normal	0.9462	0.084
East:Dwarf	−1.7428	<0.001
East:Normal	−0.6338	0.182
Cliff:Dwarf	1.0649	0.046
Cliff:Normal	0.8271	0.098
Dimension 2		
M	−0.2271	0.236
F	0.2271	0.236
Acadian	0.0984	0.620
Atlantic	−0.0984	0.620
Dwarf	0.0434	0.822
Normal	−0.0434	0.822
Temis	−0.4621	0.075
East	0.175	0.570
Cliff	0.2871	0.235
Temis:Dwarf	−0.1776	0.672
Temis:Normal	−0.7397	0.071
East:Dwarf	0.177	0.716
East:Normal	0.18	0.711
Cliff:Dwarf	0.1654	0.698
Cliff:Normal	0.3948	0.293
Dimension 3		
M	−0.3382	0.022
F	0.3382	0.022
Acadian	0.2917	0.058
Atlantic	−0.2917	0.058
Dwarf	0.2638	0.079
Normal	−0.2638	0.079
Temis	0.394	0.052
East	−0.0796	0.766
Cliff	−0.3144	0.109
Temis:Dwarf	0.3391	0.271
Temis:Normal	0.4013	0.196
East:Dwarf	−0.2182	0.623
East:Normal	0.0115	0.916
Cliff:Dwarf	0.518	0.097
Cliff:Normal	−1.0516	<0.001

Position estimates and *P*-values (*t*-test) were calculated by the dimdesc() function from the FACTOMINER package. Note that no factor had a significant effect on dimension 2, although Témiscouata and Témiscouata:Normal were nearly significant.

related with dimension 3 (R^2 = 0.13, 0.11, respectively, *F*-test, *P*-value = 0.058 and 0.079, respectively; Table 2). The variable levels male ('M') and female ('F') had a significant effect on dimension 3 (*t*-test, *P*-value <0.05; Table 3). The variable level Cliff:Normal also had a significant effect on dimension 3 (*t*-test, *P*-value <0.001; Table 3).

Identification of chromosomal markers associated with divergence

To identify the chromosome markers that were most correlated to differentiation, we retrieved markers that were most correlated to dimensions 1, 2 and 3 (Table S3, Supporting information). Namely, signals of CMA_3 and 28S rDNA on acrocentric chromosomes (a28S and aCMA) were positively correlated with each other and with the dimension 1 (R^2 = 0.90, $P < 0.001$; R^2 = 0.82, *q*-value <0.001, respectively). This dimension also resolved glacial lineages (Table 3, Fig. 4). This means that individuals from the Atlantic lineage tend to have more CMA_3 and 28S rDNA sites on the centromeres of their acrocentric chromosomes. Twenty-two different markers out of 39 were significantly correlated with dimensions 1 and 3 after *P*-value adjustment for multiple testing (Table S3, Supporting information). Therefore, many markers covary with each other and are associated with divergence between glacial lineages, among lakes and between dwarf and normal whitefish within lakes.

Discussion

The main objective of this study was to test the hypothesis that divergence in the Lake Whitefish system is accompanied by chromosomal and subchromosomal structure changes. Towards this goal, we investigated the relationship between chromosomal polymorphism and the rapid genetic and phenotypic divergence among Lake Whitefish species pairs, targeting cytogenetic markers associated with heterochromatin and repetitive DNA. Basic karyotype remained stable among lineages, lakes and species, thus rejecting a contribution of large-scale chromosomal rearrangement to divergence in this system. However, by implementing a multivariate statistical framework we found that intrachromosomal changes were modelled by historical contingency, being primarily associated with earlier allopatric divergence among glacial lineages and recent interlake divergence. In addition, dwarf and normal fish showed a trend towards divergence within lakes, although not in parallel among lakes. Together, these observations support the hypothesis that polymorphic subchromosomal traits are influenced by historical contingency and associated with divergence in the Lake Whitefish system.

A statistical multivariate strategy to analyse chromosomal polymorphism

Our statistical multivariate strategy helped to resolve patterns in this highly polymorphic data set and identify markers associated with these patterns of divergence. Importantly, this approach allows the use of discrete and continuous data, such as presence/absence of a specific marker or the number of rDNA sites. In addition, it is possible to include supplementary phenotypic measures (e.g. length, weight) or environmental data (e.g. lake) to test their association with cytogenetic patterns. A similar method was recently published based on principal coordinates analysis (PCoA, Peruzzi & Altınordu 2014). However, this method is based on continuous variables such as total haploid chromosome length and centromeric asymmetry, which are difficult to implement in nonmodel systems. In addition, this method does not allow handling of discrete data, such as presence and absence of cytogenetic markers. To our knowledge, ours is the first study for which MFA is applied to cytogenetic data. We have made our code readily available to the community (see Data S1, Supporting information).

High chromosomal polymorphism in the Lake Whitefish system

The karyotypes described herein are consistent with previous work in other populations of Lake Whitefish and show no large-scale chromosomal (e.g. inversions) or Robertsonian (i.e. fusions and fission) rearrangements within this species complex (Booke 1968; Phillips *et al.* 1996; Dion-Côté *et al.* 2015). Yet substantial intrachromosomal polymorphism was identified, mainly involving labile or rapidly evolving structures associated with heterochromatin and repetitive DNA. Specific characteristics of the Lake Whitefish system may contribute to this high level of polymorphism.

First, pronounced karyotype and intrachromosomal polymorphism is common in salmonids, including *Coregonus* (Phillips & Ráb 2001). For example, the *p*-arm of chromosome 1 in *Coregonus* shows variable length (Phillips *et al.* 1996; Jankun & Ráb 1997), and rDNA sites are also polymorphic among lineages of *Salmo trutta* (Caputo *et al.* 2009) and *Coregonus albula* (Jankun *et al.* 2003). These observations are consistent with karyotype and genetic flexibility in fishes in general and especially in salmonids following another round of genome duplication (Phillips & Ráb 2001; Ravi & Venkatesh 2008; Mable *et al.* 2011).

Also, we hypothesize that gene flow between dwarf and normal Lake Whitefish may contribute to spread and diversify chromosomal polymorphisms in Lake Whitefish. We have previously shown that Lake Whitefish hybrids experience genomic instability at several levels. Introgressive hybridization, which occurs in Lake Whitefish natural populations (Gagnaire *et al.* 2013), can promote genome reorganization in fishes, even with the same basic karyotype structure (Pereira *et al.* 2014, 2015). Hybridization between dwarf and normal Lake Whitefish is also associated with transposable element derepression (or reactivation) in hybrids (Renaut *et al.* 2010; Dion-Côté *et al.* 2014), which can promote genome rearrangements (Levin & Moran 2011). Lastly, we have found that aneuploidy, a severe form of genome instability that can promote chromosome rearrangements (Santaguida & Amon 2015), occurs in Lake Whitefish backcrosses (Dion-Côté *et al.* 2015).

Overall, the intrinsic properties of salmonid genomes, ongoing introgressive hybridization and complex historical biogeography of the Lake Whitefish are all expected to influence genome lability and consequently polymorphism. The sampling of purely allopatric populations and other sympatric species pairs, including from European lineages, should help to further disentangle the relative contributions of ancestral chromosomal polymorphisms and *de novo* intrachromosomal reorganization in relation to divergence.

Historical contingency and divergence acting on standing chromosomal variation

Heterochromatin and rDNA polymorphisms are mostly shared among the three species pairs examined, supporting the idea that ancestral Lake Whitefish population had high levels of polymorphism, or *standing chromosomal variation*. Nonetheless, multivariate analyses revealed three nested levels of divergence based upon cytogenetic markers, which are consistent with well-documented population genetic structure in the system (Pigeon *et al.* 1997; Lu *et al.* 2001; Campbell & Bernatchez 2004; Bernatchez *et al.* 2010; Renaut *et al.* 2011): (i) between glacial lineages, (ii) among lakes and (iii) between sympatric species pairs within lakes (albeit to a lesser extent and not in parallel).

Geographical isolation between the Atlantic and Acadian lineages either promoted divergence among ancestral chromosomal variants (*standing chromosomal variation*) and/or allowed for *de novo* remodelling of subchromosomal traits within a glacial lineage. Three nonmutually exclusive hypotheses may explain divergence among lakes. First, population bottleneck associated with lake colonization may have led to stochastic differentiation of ancestral chromosomal variants (Mayr 1954). Second, there may also have been *de novo* remodelling following lake colonization, considering the markers used are associated with repetitive sequences contained in heterochromatin and are extremely labile. Such rapid remodelling (<15 000 years) of intrachromosomal structures associated with ecological divergence has been documented in another young *Coregonus* species pair in Europe (Symonová *et al.* 2013). Third, the shared chromosomal variation between sympatric dwarf and normal fish may also be enhanced by ongoing gene flow (Gagnaire *et al.* 2013).

Finally, species within lakes showed a trend towards chromosomal divergence. However, this was not significant in most cases and did not occur in parallel among lakes despite extensive parallelism at the phenotypic level (Landry *et al.* 2007). In other words, cytogenetic markers of dwarf fish do not diverge from cytogenetic markers of normal fish similarly in all lakes. The absence of parallelism and incomplete differentiation between species within lake can be explained by a combination of factors: (i) the short time since divergence (~12 000 YBP or ~3000–4000 generations, (ii) the unique interaction and relative intensity (e.g. genetic drift vs. gene flow) of evolutionary forces in each lake (Lu & Bernatchez 1999; Gagnaire *et al.* 2013) and (iii) possible *de novo* intrachromosomal reorganization. While *de novo* subchromosomal changes following postglacial lake colonization may have contributed to the high level of polymorphism observed, our data suggest that there was a rather high level of *standing chromosomal variation* segregating between both Lake Whitefish lineages, as most polymorphism is shared among lakes.

How the observed levels of shared polymorphism have been maintained before secondary contact and in contemporary diverging populations calls for further studies. A possibility is that balancing selection acts on certain of these polymorphic chromosome traits (da Cunha *et al.* 1950). In particular, it has been recently documented that antagonistic sexual selection, a form of balancing selection, was involved in maintaining polymorphism at an important fitness-related gene in Atlantic salmon, *Salmo salar* (Barson *et al.* 2015). In the same manner, it could hypothetically be that here, some of the observed polymorphic chromosomal traits could be selectively favoured in one sex but negatively favoured in the other, which would lead to the maintenance of variation. Interestingly, we observed some subchromosomal differences between sexes (but no differentially fixed markers), suggesting a possibly functional, perhaps fitness-related impact of subchromosomal variation between sexes.

Heterochromatin architecture divergence and reproductive isolation

Michalak (2009) suggested that heterochromatin structure and function may be involved in rapid divergence and hybrid breakdown. Heterochromatin plays pivotal roles in transcriptional regulation and chromosome segregation (Grewal & Jia 2007) and may influence crossover localization during meiosis (John & King 1985; Ramachandran *et al.* 1985; Cioffi & Bertollo 2012). Several indirect observations suggest that heterochromatin divergence between normal and dwarf whitefish, and disruption in their hybrids, may occur. We previously found that DNMT1, an enzyme involved in heterochromatin maintenance (DNA methylation specifically), is downregulated in malformed backcross embryos (Dion-Côté *et al.* 2014). We also reported global transcriptional deregulation, transposable element derepression and noncoding RNA upregulation, consistent with heterochromatin disruption in these malformed backcrosses (Dion-Côté *et al.* 2014). Moreover, we have previously shown that Lake Whitefish hybrids suffer from mitotic and meiotic instability (Dion-Côté *et al.* 2015). Here, we observed that cytogenetic markers associated with heterochromatin and repetitive DNA tend to differ between sympatric species pairs. As a first step to directly test the role of heterochromatin in the divergence of Lake Whitefish species pairs, we are currently studying DNA methylation patterns in dwarf and normal Lake Whitefish, and inheritance patterns in reciprocal hybrids.

In conclusion, by combining conventional and molecular cytogenetic techniques, we have uncovered extensive intrachromosomal polymorphism despite constant basic karyotype structure and chromosome number in three Lake Whitefish sympatric species pairs that evolved independently. Polymorphic chromosomal markers were correlated with geographical isolation, lake colonization and sympatric species divergence, albeit to a lower extent. Barbara McClintock predicted more than thirty years ago that rapid genome reorganization would accompany speciation (McClintock 1984). A large body of work has shown that natural selection drove ecological, morphological and physiological divergence in the Lake Whitefish system; polymorphic intrachromosomal traits may have contributed to consolidate reproductive isolation in the system. Clearly, cytogenetic tools can highlight important chromosome traits from an evolutionary standpoint, which cannot readily be evidenced even with the best sequencing techniques currently available. Consequently, we echo Valente *et al.* (2014) in claiming that such cytogenetics still have much to offer in the postgenomics era.

Acknowledgements

We thank Martin Laporte, Clément Rougeux, Anne C. Dalziel, Ben J.G. Sutherland and Daniel Zinshteyn for comments on an earlier version of this manuscript. Philipp Schlüter and three anonymous reviewers also provided careful and valuable feedback. We are grateful to Guillaume Côté and members of the Bernatchez laboratory for fieldwork, Alex Bernatchez for counting gill rakers, as well as Jana Čechová and Petra Šejnohová for technical help and support. This work was supported by a Natural Science and Engineering Research Council of Canada (NSERC) discovery grant and a Canadian Research Chair in Genomics and Conservation of Aquatic Resources to LB, a NSERC postgraduate scholarship to AMDC and a postdoctoral fellowship to FCL from the Fonds de Recherche du Québec – Santé (FRQS). RS and PR were supported by the project 14-02940S of the Czech Science Foundation. AMDC also received a stipend from the FRQ-NT for international training and from Québec-Océan. This is a contribution to the Québec-Océan research programme.

References

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (ed. Turner BJ), 1–53. Plenum Press, New York.

Altemose N, Miga KH, Maggioni M, Willard HF (2014) Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLoS Computational Biology*, **10**, e1003628: 1–14.

Barson NJ, Aykanat T, Hindar K *et al.* (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, **528**, 405–408.

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

Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of Lake Whitefish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1783–1800.

Booke HE (1968) Cytotaxonomic studies of coregonine fishes of the Great Lakes, USA: DNA and karyotype analysis. *Journal of the Fisheries Research Board of Canada*, **25**, 1667–1687.

Britton-Davidian J, Cazaux B, Catalan J (2012) Chromosomal dynamics of nucleolar organizer regions (NORs) in the house mouse: micro-evolutionary insights. *Heredity*, **108**, 68–74.

Brown JD, O'Neill RJ (2010) Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annual Review of Genomics and Human Genetics*, **11**, 291–316.

Camacho JPM (2005) B chromosomes. In: *The Evolution of the Genome* (ed. Gregory TR), pp. 223–286. Academic Press, Burlington, MA.

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Caputo V, Giovannotti M, Nisi Cerioni P, Splendiani A, Olmo E (2009) Chromosomal study of native and hatchery trouts from Italy (*Salmo trutta* complex, Salmonidae): conventional and FISH analysis. *Cytogenetic and Genome Research*, **124**, 51–62.

Charron G, Leducq J-B, Landry CR (2014) Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Molecular Ecology*, **23**, 4362–4372.

Chombard C, Boury-Esnault N, Tillier S (1998) Reassessment of homology of morphological characters in tetractinellid sponges based on molecular data. *Systematic Biology*, **47**, 351–366.

Cioffi MB, Bertollo LAC (2012) Chromosomal distribution and evolution of repetitive DNAs in fish. *Genome Dynamics*, **7**, 197–221.

Cioffi MB, Martins C, Bertollo LA (2010) Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. *BMC Evolutionary Biology*, **10**, 1–9.

Comings DE (1978) Mechanisms of chromosome banding and implications for chromosome structure. *Annual Review of Genetics*, **12**, 25–46.

Cremer M, Grasser F, Lanctôt C *et al.* (2008) Multicolor 3D fluorescence *in situ* hybridization for imaging interphase chromosomes. *Methods in Molecular Biology (Clifton, NJ)*, **463**, 205–239.

Crête-Lafrenière A, Weir LK, Bernatchez L (2012) Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLoS ONE*, **7**, e46662.

da Cunha AB, Burla H, Dobzhansky T (1950) Adaptive chromosomal polymorphism in *Drosophila willistoni*. *Evolution*, **4**, 212–235.

Cutter AD (2012) The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends in Ecology & Evolution*, **27**, 210–219.

Dayrat B, Tillier A, Lecointre G, Tillier S (2001) New clades of euthyneuran gastropods (Mollusca) from 28S rRNA sequences. *Molecular Phylogenetics and Evolution*, **19**, 225–235.

Dernburg AF, Sedat JW, Hawley RS (1996) Direct evidence of a role for heterochromatin in meiotic chromosome segregation. *Cell*, **86**, 135–146.

Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L (2014) RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young Lake Whitefish species. *Molecular Biology and Evolution*, **31**, 1188–1199.

Dion-Côté A-M, Symonová R, Ráb P, Bernatchez L (2015) Reproductive isolation in a nascent species pair is associated with aneuploidy in hybrid offspring. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20142862.

Dobigny G, Ducroz J-F, Robinson TJ, Volobouev V (2004) Cytogenetics and Cladistics. *Systematic Biology*, **53**, 470–484.

Faria R, Navarro A (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in Ecology & Evolution*, **25**, 660–669.

Fujiwara A, Nishida-Umehara C, Sakamoto T *et al.* (2001) Improved fish lymphocyte culture for chromosome preparation. *Genetica*, **111**, 77–89.

Gagnaire P-A, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in Lake Whitefish species pairs assessed by RAD sequencing. *Evolution*, **67**, 2483–2497.

Grewal SIS, Jia S (2007) Heterochromatin revisited. *Nature Reviews Genetics*, **8**, 35–46.

Hébert FO, Renaut S, Bernatchez L (2013) Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric Lake Whitefish species pair (*Coregonus clupeaformis*). *Molecular Ecology*, **22**, 4896–4914.

Hoskins RA, Carlson JW, Kennedy C *et al.* (2007) Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science (New York, NY)*, **316**, 1625–1628.

Jacobsen MW, Hansen MM, Orlando L *et al.* (2012) Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). *Molecular Ecology*, **21**, 2727–2742.

Jankun M, Ráb P (1997) Multiple polymorphism of chromosome no. 1 in the karyotype of whitefish, *Coregonus lavaretus* (Salmonidae) from lake system Saimaa, Finland. *Caryologia*, **50**, 185–195.

Jankun M, Ráb P, Vuorinen J, Lucyznski M (1995) Chromosomal polymorphism in *Coregonus lavaretus* populations from two locations in Finland and Poland. *Archive Hydrobiology Special Issues Advance Limnology*, **46**, 1–11.

Jankun M, Ocalewicz K, Pardo BG *et al.* (2003) Localization of 5S rRNA loci in three coregonid species (Salmonidae). *Genetica*, **119**, 183–186.

John B, King M (1985) The inter-relationship between heterochromatin distribution and chiasma distribution. *Genetica*, **66**, 183–194.

Jones RN (1995) B chromosomes in plants. *New Phytologist*, **131**, 411–434.

Josse J, Husson F (2012) Handling missing values in exploratory multivariate data analysis methods. *Journal de la Société Française de Statistiques*, **153**, 79–99.

Josse J, Husson F (2016) missMDA: a package for handling missing values in multivariate data analysis. *Journal of Statistical Software*, **70**, 1–31.

Kidd JM, Cooper GM, Donahue WF *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature*, **453**, 56–64.

Kim KE, Peluso P, Babayan P *et al.* (2014) Long-read, whole-genome shotgun sequence data for five model organisms. *Scientific Data*, **1**, 140045.

King M (1993) *Species Evolution: The Role of Chromosome Change*. Cambridge University Press, Cambridge.

Landry L, Bernatchez L (2010) Role of epibenthic resource opportunities in the parallel evolution of Lake Whitefish species pairs (*Coregonus* sp.). *Journal of Evolutionary Biology*, **23**, 2602–2613.

Landry L, Vincent WF, Bernatchez L (2007) Parallel evolution of Lake Whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of Evolutionary Biology*, **20**, 971–984.

Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*, **25**, 1–28.

Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, **12**, 615–627.

Lu G, Bernatchez L (1998) Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of Lake Whitefish (*Coregonus clupeaformis* Mitchill). *Proceedings of the Royal Society of London B: Biological Science*, **265**, 1025–1030.

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.

Lu G, Basley DJ, Bernatchez L (2001) Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of Lake Whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, **10**, 965–985.

Mable BK, Alexandrou MA, Taylor MI (2011) Genome duplication in amphibians and fish: an extended synthesis. *Journal of Zoology*, **284**, 151–182.

Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132881: 1–8.

Marie Curie SPECIATION Network (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27–39.

Mayr E (1954) Change of genetic environment and evolution.

McClintock B (1984) The significance of responses of the genome to challenge. *Science (New York, NY)*, **226**, 792–801.

Michalak P (2009) Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity*, **102**, 45–50.

Pendás AM, Moran P, Martinez JL, Garcia Vazquez E (1995) Applications of 5S rDNA in Atlantic salmon, brown trout, and in Atlantic salmon brown trout hybrid identification. *Molecular Ecology*, **4**, 275–276.

Pereira CSA, Aboim MA, Ráb P, Collares-Pereira MJ (2014) Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). *Heredity*, **112**, 343–350.

Pereira CSA, Pazian MF, Ráb P, Collares-Pereira MJ (2015) Dynamics of Rex3 in the genomes of endangered Iberian Leuciscinae (Teleostei, Cyprinidae) and their natural hybrids. *Molecular Cytogenetics*, **8**, 1–10.

Peruzzi L, Altınordu F (2014) A proposal for a multivariate quantitative approach to infer karyological relationships among taxa. *Comparative Cytogenetics*, **8**, 337–349.

Phillips RB, Ihssen PE (1986) Inheritance of Q band chromosomal polymorphisms in lake trout. *Journal of Heredity*, **77**, 93–97.

Phillips RB, Ráb P (2001) Chromosome evolution in the Salmonidae (Pisces): an update. *Biological Reviews*, **76**, 1–25.

Phillips RB, Reed KM, Ráb P (1996) Revised karyotypes and chromosome banding of coregonid fishes from the Laurentian Great Lakes. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **74**, 323–329.

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

R Core Team (2012) R: a language and environment for statisti cal computing. Available from: http://www.R-project.org/.

Ráb P, Jankun M (1992) Chromosome studies of coregonine fishes: a review. *Polskie Archiwum Hydrobiologii*, **39**, 523–532.

Rábová M, Völker M, Pelikánová Š, Ráb P (2015) Sequential chromosome banding in fishes. In: *Fish Cytogenetic Techniques: Ray-Fin Fishes and Chondrichthyans* (eds Ozouf-Costaz C, Pisano E, Foresti F, de Almeida Toledo LF), pp. 66–73. CRC Press Inc., Enfield.

Ramachandran C, Brandenburg WA, Nijs den APM (1985) Infraspecific variation in C-banded karyotype and chiasma frequency in *Cucumis sativus* (Cucurbitaceae). *Plant Systematics and Evolution*, **151**, 31–41.

Ravi V, Venkatesh B (2008) Rapidly evolving fish genomes and teleost diversity. *Current Opinion in Genetics & Development*, **18**, 544–550.

Renaut S, Bernatchez L (2011) Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in Lake Whitefish species pairs (*Coregonus* spp. Salmonidae). *Heredity*, **106**, 1003–1011.

Renaut S, Nolte AW, Bernatchez L (2009) Gene expression divergence and hybrid misexpression between Lake Whitefish species pairs (*Coregonus* spp. salmonidae). *Molecular Biology and Evolution*, **26**, 925–936.

Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in Lake Whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology*, **19**(Suppl. 1), 115–131.

Renaut S, Nolte AW, Rogers SM *et al.* (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in Lake Whitefish species pairs (*Coregonus* spp.). *Molecular Ecology*, **20**, 545–559.

Renaut S, Maillet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during speciation: the Lake Whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 354–363.

Rogers SM, Bernatchez L (2006) The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the Lake Whitefish species complex (*Coregonus clupeaformis*). *Journal of Evolutionary Biology*, **19**, 1979–1994.

Santaguida S, Amon A (2015) Short- and long-term effects of chromosome mis-segregation and aneuploidy. *Nature Reviews Molecular Cell Biology*, **16**, 473–485.

Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.

Sutherland BJG, Gosselin T, Normandeau E *et al.* (2016) Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps reveals chromosome evolution in salmonids. *BioRxiv*, 039164. doi: http://dx.doi.org/10.1101/039164.

Symonová R, Majtánová Z, Sember A *et al.* (2013) Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evolutionary Biology*, **13**, 1–11.

Valente GT, Conte MA, Fantinatti BEA *et al.* (2014) Origin and evolution of B chromosomes in the Cichlid Fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Molecular Biology and Evolution*, **31**, 2061–2072.

Vergilino R, Elliott TA, Desjardins-Proulx P, Crease TJ, Dufresne F (2013) Evolution of a transposon in Daphnia hybrid genomes. *Mobile DNA*, **4**, 1–14.

Ziegler CG, Lamatsch DK, Steinlein C *et al.* (2003) The giant B chromosome of the cyprinid fish *Alburnus alburnus* harbours a retrotransposon-derived repetitive DNA sequence. *Chromosome Research*, **11**, 23–35.

A.M.D.C., R.S., P.R. and L.B. designed the project, which is part of L.B.'s long-term research programme on Lake Whitefish speciation. A.M.D.C. and S.P. performed laboratory work. A.M.D.C., R.S. and F.C.L. analysed data. P.R. and L.B. supervised the project. A.M.D.C. wrote the manuscript in collaboration with R.S., F.C.L., P.R. and L.B.

Data accessibility

Complementary data including 5S and 28S rDNA sequencing files and raw microscopy images for all techniques used (Giemsa, C-bands, CMA_3 and FISH) are available on Dryad (doi: 10.5061/dryad.tg0mt).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Summary of all individuals samples across three lakes, individual characteristics, phenotypic characteristics measured and markers scored.

Table S2 Summary of all individuals samples across three lakes, individual characteristics, phenotypic characteristics measured and imputed marker values.

Table S3 Chromosome markers significantly correlated (q-value < 0.05) to dimensions 1, 2 and 3 from the multiple factor analysis.

Fig. S1 Metaphase spreads of a dwarf individual from Cliff Lake (CD10) where a B chromosome was found (shown with arrow, also enlarged and framed). A) C-banding, B) CMA_3/DAPI staining and C) FISH with 28S (red) and 5S rDNA (green) probes.

Data S1 Sample code for Multiple Factor Analysis to be performed with FactoMiner.