Ecotoxicology (2013) 22:938–957
DOI 10.1007/s10646-013-1083-8

Evolutionary change driven by metal exposure as revealed by coding SNP genome scan in wild yellow perch (*Perca flavescens*)

Sébastien Bélanger-Deschênes · Patrice Couture · Peter G. C. Campbell · Louis Bernatchez

Accepted: 30 April 2013 / Published online: 31 May 2013

Abstract Pollution can drive rapid evolutionary change in wild populations. This study targets functional polymorphisms of chronically metal-contaminated wild yellow perch (*Perca flavescens*). A de novo transcriptome scan contrasted subsets of individuals from clean ($n = 16$) and contaminated ($n = 16$) lakes to identify 87 candidate annotated coding SNPs. Candidate genotypes and liver [metal] were obtained in 10 populations ($n = 1{,}052$) and a genome scan distinguished outliers: one nuclear (cyclin G1 gene) and two mitochondrial (cytochrome b and NADH dehydrogenase subunit 2 genes) also displaying allelic correlation to mean population [cadmium]. Whole mtDNA and 17 kb surrounding cyclin G1 were characterised through 454 sequencing thus revealing two non-synonymous substitutions involving dissimilar amino acids. Based on associated functions and inter-population differentiation, contaminated perch may have been selected for fast life cycle completion (p53 pathway) and memorization impairment mitigation (long-term potentiation pathway). In accordance with predicted evolutionary trajectory for stressed and energy deprived organisms, adapted perch would not compensate for repair mechanism inhibition, instead reallocating energy towards growth and favouring inexpensive impairment mitigation adaptations over costly detoxification. Overall, 85 years of selection could have driven rapid, potentially adaptive evolution by selecting alleles increasing perch fitness in polluted environments.

Keywords Evolutionary ecotoxicology · Ecotoxicogenomics · SNP · Contemporary evolution · Pollution · Yellow perch · Cadmium

Introduction

The impact of human activities is global and affects all ecosystems to some extent (Vitousek et al. 1997). Airborne pollution is of particular concern as it spreads across the planet, forcing the need to address the associated negative impacts through world-wide cooperation (Akimoto 2003). To this end, thorough documentation of the dynamics and consequences of pollution in an ecological framework is crucial (Chapman 2002).

Moreover, it is well documented that certain classes of anthropogenic pollutants pose various fitness issues to wild populations, potentially resulting in strong directional selection pressure (Posthuma and van Straalen 1993). In such cases, microevolution can lead to local adaptation to new environmental conditions (Hendry and Kinnison 2001; Reznick and Ghalambor 2001). Instances of human induced rapid evolution are indeed reported in a growing number of studies (reviewed in Smith and Bernatchez 2008; Tobler et al. 2010; Williams and Oleksiak 2011). However, although understanding the underlying evolutionary mechanisms is key to ecological risk assessment (ERA) as well as restoration strategies, they are still poorly understood (van Straalen and Timmermans 2002).

Electronic supplementary material The online version of this article (doi:10.1007/s10646-013-1083-8) contains supplementary material, which is available to authorized users.

S. Bélanger-Deschênes (✉) · L. Bernatchez
Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, 1030 Avenue de la Médecine, Québec, QC G1V 0A6, Canada
e-mail: sebastien.belanger-deschenes.1@ulaval.ca

P. Couture · P. G. C. Campbell
Institut national de la recherche scientifique, Centre Eau Terre et Environnement, 490 de la Couronne, Québec, QC G1K 9A9, Canada

The nascent field of evolutionary ecotoxicology precisely seeks this knowledge and is increasingly empowered by the fast paced ''-omics'' technological improvements (Rokas and Abbot 2009). The discipline is thus in the midst of a paradigm shift that tends to focus studies on functional targets of pollution-driven selection in wild populations. Genome and transcriptome wide analyses by means of next generation sequencing (NGS) are the main drivers of new approaches allowing large-scale marker mining without the need for prior target candidates. Yet, few studies to date have bridged the gap between outlier detection and functional characterization of putative targets.

The region of Rouyn-Noranda (Québec, Canada) presents a highly relevant context to study local environmental impacts of the mining industry. For the last 85 years (since 1927), local dominant winds have blown metal emissions from a copper smelter towards the north-east, creating a polymetallic gradient of contamination in nearby lakes (Couillard et al. 2004). Moreover, the cadmium (Cd) and copper (Cu) that deposit in lacustrine environments are responsible for profound, well documented ecosystemic perturbations of the benthic niche (Kovecses et al. 2005), planktonic communities (Cattaneo et al. 2008) and freshwater molluscs (Perceval et al. 2006).

Numerous studies have provided evidences that indigenous yellow perch populations suffer from these alterations (Kovecses et al. 2005; Sherwood et al. 2002). Notably, perch bioaccumulate metals in various tissues, making them prone to several physiological and metabolic impairments that may affect their fitness in natural environments (Couture and Pyle 2008; Couture et al. 2008a, b; Pyle et al. 2008). Bourret et al. (2008) reported a negative correlation between Cd contamination and genetic diversity, thus revealing evolutionary consequences of metal contamination in these populations. Previous studies did not show that yellow perch from contaminated environments have superior detoxification capacities (Campbell et al. 2005). Indeed, when resources are scarce in contaminated environments, detoxification may not be the first trait to evolve due to its high energetic cost as a general defense mechanism (Posthuma and van Straalen 1993). Instead, especially if mortality is high, traits allowing for fast life-cycle completion may be more important targets of selection (Sibly and Calow 1989). Metal-contaminated yellow perch from the Rouyn-Noranda area may have taken this evolutionary trajectory since they display various negative physiological effects while tending to grow fast, reproduce and die at early age (Couture and Pyle 2008).

This study aims to identify putative adaptive polymorphisms in nucleotide sequences and to set the basis for in silico functional characterization of metal contamination-induced selection in wild yellow perch populations. This was achieved following a four-step approach: (i) perform a transcriptome scan from de novo sequencing to identify putative candidate SNPs located in annotated genes; (ii) develop SNP assays for a subset of these markers to genotype 1,052 individuals from 10 populations spanning a wide range of Cd and Cu contamination; (iii) assess the association between allele frequency of outliers and mean liver metal concentrations using different regression methods; (iv) resequence large amplicons to investigate the nature and functional basis of polymorphisms surrounding these outliers.

Materials and methods

Sampling

Six lakes were selected for their varying degree of Cd and Cu contamination and previous knowledge on the system (Bourret et al. 2008; Pierron et al. 2009) (Fig. 1). Sampling was conducted in June and July 2010 using beach seine nets in shallow waters. A minimum of 50 yellow perch were sampled in each lake (Table 1). Each fish was measured, weighed and sexed. Fin clips were collected and stored in 95 % EtOH for subsequent DNA extraction. Since it is the most suitable organ to evaluate bioaccumulation of Cd and Cu in these populations (Giguere et al. 2004), liver was sampled and snap-frozen in liquid nitrogen within 15 min following sacrifice. In addition, samples from Bourret et al. (2008) collected in 2005 were also used for genotyping. These provided samples for four additional lakes (total number of lakes = 10) and provided two temporal replicates for six lakes. Overall, six contaminated and four clean lakes were studied (Table 1 for lake classification).

Metal quantification

Fragments of liver samples were freeze-dried in 1.5 mL tubes that had previously been washed in 15 % HNO_3 baths to remove any trace metals. Cold digestion was conducted according to Pierron et al. (2009). Certified reference material provided by the National Research Council of Canada (TORT2) and blanks were also included for certifying analytical accuracy and monitoring metal recovery. Concentrations of Cd and Cu were measured using an inductively coupled plasma mass spectrometer (ICP-MS, Thermo Elemental, Model X-7). TORT-2 and blank samples were analyzed every 12 samples to control for contamination and assess reproducibility.

Coding gene SNP development and genotyping

We used whole transcriptome sequences from Pierron et al. (2011) for SNP identification. Briefly, total RNA was

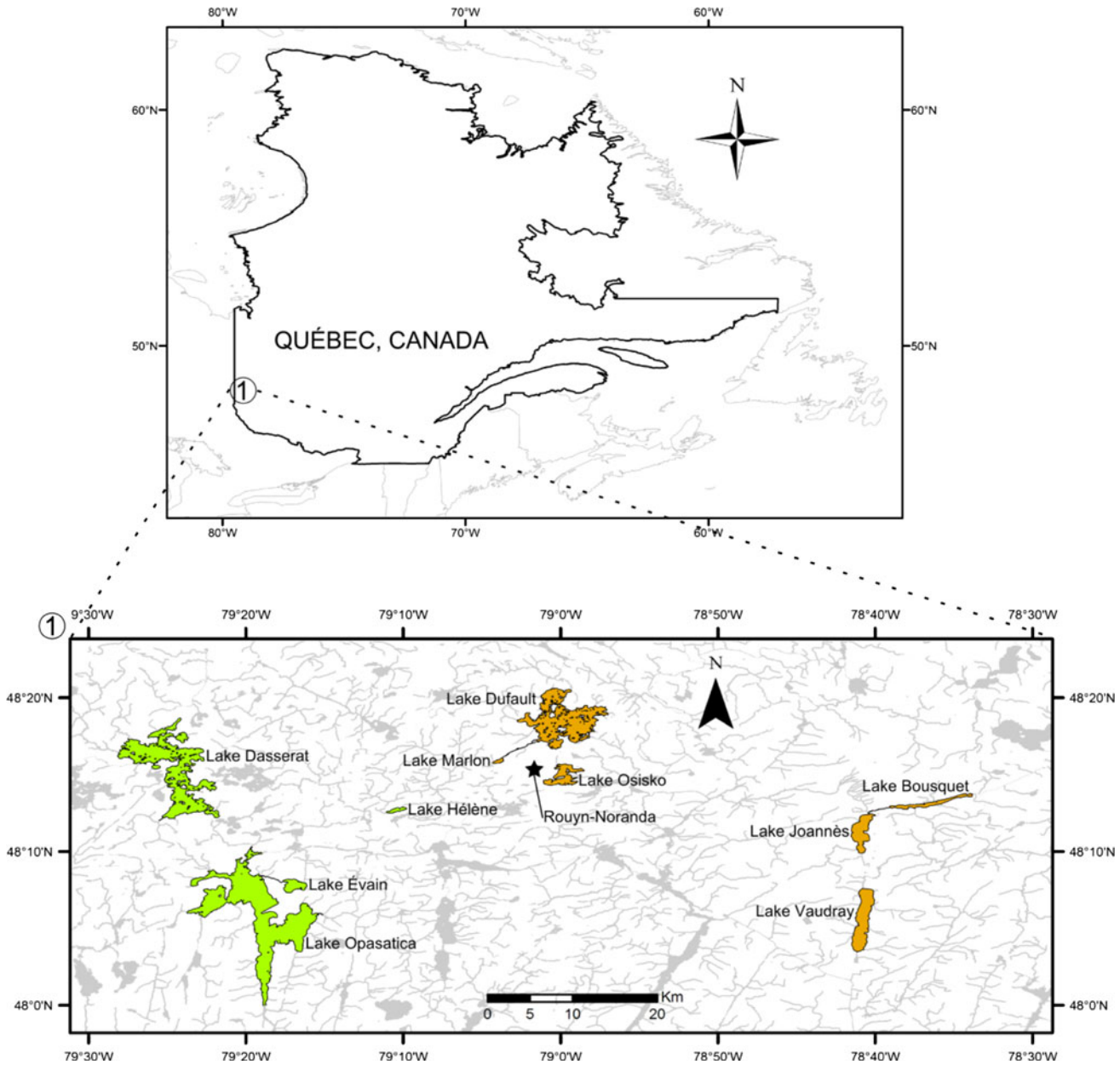

Fig. 1 Map of study area and sampled lakes. In *green* control lakes, in *orange* contaminated lakes. The *star* indicates Rouyn-Noranda from where metal emissions take place

extracted from 16 individuals equally sampled in clean lakes Opasatica and Adeline (Adeline population was not included in this study) as well as 16 individuals from contaminated lakes Dufault and Marlon. The retro transcribed cDNA was individually marked using MIT tags and the 32 individuals were sequenced with a Roche GS-FLX 454 DNA Sequencer. CLC Genomics Workbench 3.7 was used to identify SNPs using the Neighborhood Quality Standard method (NQS; Brockman et al. 2008). The Beaumont and Nichols (1996) selection detection method implemented in Arlequin 3.5 (Excoffier and Lischer 2010) was used to identify putative targets of directional selection induced by metal exposure. A subset of 130 SNPs including outliers potentially under directional selection and putatively neutral markers were both retained on the basis of their F_{ST} for subsequent genotyping in large numbers of individuals from all studied populations.

A total of 1,052 individuals were genotyped by KBioscience (Hoddeson, UK, http://www.kbioscience.co.uk) who also developed the genotyping assays (48 validated polymorphic markers out of 87 genotyped and 130 submitted; see Results) using the KASP chemistry (Table S1 for genotyping primers sequences) which is based on the specific hybridization of fluorescence-marked primers on

Table 1 Population means of hepatic cadmium (Cd) and copper (Cu) concentrations calculated from individual values measured as an average of three ICP-MS measurements per sample

Population	2005 sampling [Cd] (μg/g dlw) ± SD (sample size)	2010 sampling [Cd] (μg/g dlw) ± SD (sample size)	2005 sampling [Cu] (μg/g dlw) ± SD (sample size)	2010 sampling [Cu] (μg/g dlw) ± SD (sample size)
Bousquet	10.87 ± 3.65 (99/99)	11.53 ± 3.22* (52/52)	18.41 ± 7.05 (99/99)	13.40 ± 2.93 (52/52)
Dasserat	2.13 ± 1.13 (44/50)	–	15.77 ± 3.60 (44/50)	–
Dufault	36.46 ± 13.59* (93/93)	17.66 ± 6.48* (79/80)	81.04 ± 99.55* (93/93)	24.53 ± 12.71 (79/80)
Évain	6.48 ± 2.33 (50/50)	–	20.96 ± 5.68 (50/50)	–
Hélène	1.01 ± 0.37 (50/50)	1.18 ± 0.45 (47/50)	18.99 ± 6.70 (50/50)	11.87 ± 3.37 (47/50)
Joannès	14.42 ± 5.51* (92/93)	–	24.67 ± 9.42 (92/93)	–
Marlon	26.86 ± 13.26* (50/50)	27.21 ± 11.65* (52/52)	35.32 ± 17.33 (50/50)	23.22 ± 15.77 (52/52)
Opasatica	2.53 ± 1.26 (48/50)	2.53 ± 1.44 (81/81)	20.80 ± 10.79 (48/50)	12.54 ± 4.16 (81/81)
Osisko	20.80 ± 10.24* (99/100)	–	73.69 ± 44.48* (99/100)	–
Vaudray	27.66 ± 12.28* (50/50)	25.23 ± 6.66* (49/52)	28.96 ± 15.45 (50/50)	19.22 ± 5.60 (49/52)

Sample sizes show the number of values used for average calculation versus the total number of sampled individuals. "–" Unsampled populations

SD standard deviation, *dlw* dry liver weight

* Average values above significant contamination thresholds (11 μg/g dlw for Cd and 38.8 μg/g dlw for Cu) set by Couture and Pyle (2008)

SNPs. Several factors can explain the fact that only 48 out of 130 markers identified from cDNA sequence data were validated as true SNPs in genomic DNA, including alignment of paralogs, occurence of exon–exon boundaries and PCR efficiency (Sauvage et al. 2012).

Candidate gene library preparation for large amplicon 454 sequencing

Reference genomes from the phylogenetically closest relatives of *Perca flavescens* were retrieved from NCBI (www.ncbi.nlm.nih.gov) and Ensembl Genome Browser (Flicek et al. 2012; http://useast.ensembl.org) in order to build scaffolds on which the transcriptomic sequences of Pierron et al. (2011) were aligned with CLC. These scaffolds were used for primer design (Primer3 v0.4.0; Rozen and Skaletsky 2000; Suppl. mat 1 for primer sequences) in order to amplify and analyse further three genes that stood out as being the most relevant targets based on outlier and association tests (see below). Using an adapted version of the approach described by Jacobsen et al. (2012), the long-range PCR chemistry (Invitrogen Platinum® Taq DNA Polymerase High Fidelity kit) was used to amplify large fragments (up to 10.3 kb) containing the complete candidates genes along with intergenic regions and neighbouring genes.

For the candidate gene library preparation, we used an adapted version of the protocol described in Pierron et al. (2011) (Table S2 for PCR details). A total of 40 individuals (20 from clean Lake Opasatica and 20 from contaminated Lake Dufault) were selected to encompass the whole spectrum of possible genotype combinations at the three candidate loci. Genotype proportions were selected to reflect their natural occurrence, but were reequilibrated for better representing the rarest genotypes. The 160 fragments (i.e. 40 individuals × 4 amplicons) were then purified with AMPure XP magnetic beads (Agencourt). The DNA concentration of each sample was determined using a Nanodrop 2000 spectrophotometer (Thermo Scientific). Subsequently, approximately equal amounts of copies (ranging from 12.2E9 to 68.1E9) of each amplicon were combined in pools for each individual separately. A minimum of 30 μL of this material was then sheared in cold water using a Virsonic 475 sonicator (Virtis) set to 15 % and pulsing for 1 min 30 s (cycles of 30 s ON and 60 s OFF). Sheared DNA was sized following the Rapid Library Preparation Method Manual (May 2011, Roche) and using 50 μL of AMPure XP beads. The sizing solution was made in-lab by combining 8 % polyethylene glycol 8000 and 5 M NaCl. Proper sizing was ensured by visualisation on 1.5 % agarose gel and smears in the range of 1.0–1.3 kb were considered ideal for the 454 Titanium GS-FLX+ technology. Fragmented amplicon pools were polished to make them blunt ended and phosphorylated at the 5′ end through a reaction consisting of 3 units of T4 DNA polymerase, 3 units T4 polynucleotide kinase in T4 DNA ligase buffer, 100 μg/mL BSA and 10 mM dNTP (NEB). This mix was incubated at 12 °C for 15 min followed by 25 °C for 15 min and finally at 75 °C for 20 min in a Biometra T1 thermocycler. All samples were then purified using AMPure XP beads and resuspended in 1× TE pH 8. Afterwards, fragments were ligated with the adaptors (10 μM) using 400 units of T4 DNA ligase (NEB) and incubation temperatures of 25 °C for 15 min followed by 65 °C for 20 min. Adaptors are the standard 454 B and A primers supplemented with a 10 bp long barcode at the 3′

end (described in Roche #005-2009 technical bulletin). The next step was another AMPure bead purification to remove excess reagents. Each fragment pool was then PCR amplified (Table S2 for PCR details) using primer A and primer B (5′ labelled with biotin). This product was purified and sized one last time using the previously described method. Fragment size range was verified on agarose gels and the pools were combined in equal amounts into two separate libraries of 20 individuals (10 from each lake) based on their dsDNA concentration (Quant-iT PicoGreen dsDNA kit, Invitrogen). Each library was sequenced on a separate eighth plate of 454 Roche GS-FLX Titanium+ following the protocol described by Margulies et al. (2005).

Descriptive statistics

Except for mitochondrial markers, departure from Hardy–Weinberg equilibrium (HWE) was tested locus by locus in each populations using the exact test (Guo and Thompson 1992) implemented in Arlequin 3.5 (Excoffier and Lischer 2010). Bonferroni correction for multiple comparisons was performed to maintain the α-value at 0.05. Linkage disequilibrium (LD) between all pairs of loci in each population was tested using the log likelihood ratio test statistic implemented in Genepop 4.0.10 (Raymond and Rousset 1995; http://genepop.curtin.edu.au) and Bonferroni correction ($\alpha = 0.05$) was applied. Arlequin 3.5 (Excoffier and Lischer 2010) was used to compute allelic frequency, heterozygosity, and inter-population pairwise F_{ST} values.

Temporal stability

We first tested the temporal stability of genetic composition using Arlequin 3.5 (Excoffier and Lischer 2010) to compute pairwise F_{ST} (20,000 permutations) between the two sampling years of each population. We thus performed a locus by locus AMOVA (20,000 permutations) to examine the extent of variation explained by inter-annual variation in relation to other organisation levels (i.e. among and within populations). Genetic composition was temporally stable in all populations (see Results) and temporal replicates were therefore pooled for all analyses.

Outlier detection

In order to identify markers potentially under directional selection, we used the fully Bayesian approach implemented in BayeScan 2.0 (Foll and Gaggiotti 2008). The model accounts for different effective population sizes (N_E) (see Bourret et al. 2008 for N_E estimates) and migration rates between subpopulations (Foll and Gaggiotti 2008). The analysis was performed using all markers and populations. The reversible jump Markov chain Monte Carlo (RJ-MCMC) determining the posterior distribution was submitted to a burn-in of 100,000 to ensure convergence and a thinning interval of 20 to avoid auto-correlation bias. The number of iterations was 10,000 (30 pilot runs and 10,000 iterations for the proposal distribution of the reversible jump). The prior odd for the neutral model was set to 1 which is ideal for small numbers of markers and expresses the assumption that the neutral model is as likely as the model with selection (Bayescan 2.0 User Manual). A Bonferroni correction ($\alpha = 0.05$) was applied, resulting in a posterior probability threshold of 0.90. According to Jeffrey's scale of evidence for Bayes factor (Jeffreys 1961), values of 0.76–0.91 reflect substantial support for selection, 0.91–0.97 are strong support, 0.97–0.99 are very strong support and values over 0.99 are decisive support.

Genetic association with metal contamination

Three different regression methods were performed in order to compare allelic frequencies to population average concentrations of Cd and Cu. The contamination averages for 2005 were used for this purpose because they constitute the only snapshot comprising measurements for all lakes.

First, logistic regressions were conducted with MatSAM 2 beta (Joost et al. 2007) on all markers. This uses the presence or absence of an allele as the response variable and the explanatory variable which was the population level of Cd or Cu contamination. The significance of each model was assessed through Wald tests performed with the MatSAM supplied spreadsheet. Univariate regressions compared the mean parameter of the explanatory variable to its own error. A p value threshold of 1E−13 for the Wald tests was applied to determine association significance (critical value after Bonferroni = 5.21E−16). A low threshold was favoured because the model is very sensitive and does not correct for population stratification. It is also noteworthy that these logistic regressions are independent of any genetic model and are not compromised by departures from HWE (Joost et al. 2007).

Secondly, χ^2 tests were performed with Eigenstrat 3.0 in order to correct for population stratification (Price et al. 2006). This performs a principal component analysis (PCA) on the genotypes of individuals to document differences in ancestry (Patterson et al. 2006; Price et al. 2006). Genotypes and phenotypes (in our case, population levels of hepatic Cd and Cu bioaccumulation) are then continuously adjusted by amounts reflecting their ancestry through calculation of residuals of linear regressions (Price et al. 2006). Then, adjusted χ^2 statistics of association between markers and population degree of contamination are computed. For first documenting differences in ancestry

by means of PCA, we used a subset of 39 markers that excluded one marker of each pair that was in significant LD (Reich et al. 2008). The three markers that were most likely under directional selection were also removed for that step as recommended by Price et al. (2006). Subsequently, the PCA was carried out using all individual genotypes and considering 10 principal components while using the default five iterations for the outlier removal process. The full set of 48 polymorphic markers was inputted in Eigenstrat 3.0 to produce association statistics with the population degree of contamination of Cd and Cu separately while correcting for stratification due to ancestry. The third test of association was a series of univariate linear regressions (R 2.13.0, *R core development team*) that compared population allelic frequencies as a dependent variable to population level of Cd and Cu as the explanatory variable.

Sequence characterisation of candidate genes

Combining the results from the above three association tests with those of the genome scan identified three candidate genes (see Results) that were particularly relevant to be functionally characterized using 454 sequencing. Sequence reads were assembled de novo in two separate contigs comprising all sequenced individuals, one for the cyclin G1 region and one for the complete mitogenome. After quality check, the best consensus of each contig was used as a reference for all read mappings.

For the cyclin G1 gene primer development, Pierron's et al. (2011) cDNA reads were mapped on the phylogenetically closest available reference genome, *Gasterosteus aculeatus* (PRJNA13579). The sequence comprising cyclin G1 and its neighbouring genes is located on linkage group IV (11814311–11829375) and was retrieved using Ensembl Genome Browser (Flicek et al. 2012; http://ensembl.org). Using this mapping, two overlapping pairs of primers that targeted the middle of the cyclin G1 coding region and the coding regions of neighbouring septin 8 and GABRA-6 were designed (Fig. 2).

For the two mitochondrial candidate genes, the whole mitogenome was sequenced. Using the approach previously described, we mapped the yellow perch transcriptomic sequences on the closest available mitogenome (Sloss et al. 2004), *Percina macrolepida* (NC_008111.1). Since the mitochondrial control region is usually highly polymorphic across species which causes mapping problems (Loman et al. 2012), we used a publically available yellow perch control region sequence (FJ155949.1) on the mapping template. Two pairs of primers were designed on the cytochrome oxidase I (COI) and on the cytochrome b (Cytb) genes. Each primer pair was designed to produce an amplicon overlapping the other at both ends by at least 400 bp, allowing for whole mitogenome amplification with only two amplicons.

Global mappings comprising all individuals were used to discover biallelic SNPs with CLC Genomics Workbench's scoring function (minimum coverage = 40, minimum variant frequency = 0.05). These were visually inspected to remove pseudo-SNPs from the list of variable positions. The sequencing reads were sorted by tag and mappings were done separately for each individual and for both the mitogenome and the cyclin G1 genomic region. These 40 individual mappings were fed to SAMTools (Li et al. 2009) and genotypes were extracted using the '*mpileup*' command (minimum base quality = 10). To filter out unreliable data, a genotype quality threshold of 20, equivalent to 1 % chance of scoring the wrong genotype, was used.

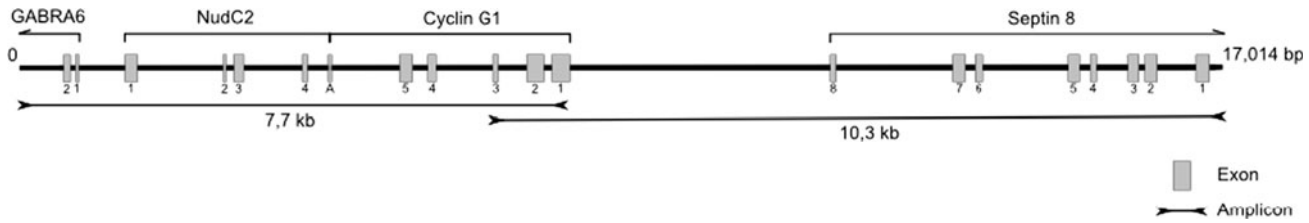

Fig. 2 Scaled representation of annotated exons for each gene (partial or complete) located in cyclin G1 genomic region. *Numbers* under *grey boxes* represent annotated exons under the figure. Exons labeled "*1*" are initiation exons with start codons (except for *septin 8*, which is partial). Polyadenylation signals (A) of cyclin G1 and NudC2 are side by side; none was found for *septin 8* and *GABRA6*'s signal is outside of the contig. Amplicon names/lengths can be used to retrieve corresponding primer sequences (Supplementary material 1). Annotation : Gamma-aminocutyric acid acid A receptor, alpha 6-like (GABRA 6; −strand; exon 1: 772–818; exon 2: 589–695), NudC domain containing protein 2-like (NUDC2; +strand; exon 1 (init): 1,461–1,649; exon 2: 2,861–2,909; exon 3: 3,005–3,156; exon 4 (term): 3,968–4,051; polyA signal: 4,373–4,378), Cyclin G1 (CCNG1; strand−; exon 1 (init): 7,508–7,771; exon 2: 7,155–7,408; exon 3: 6,686–6,764; exon 4: 5,457–5,864; exon 5 (term): 5,353–5,544; PolyA signal: 4,381–4,386), Septin 8 (SEPT8; −strand; exon 1: 16,627–16,823; exon 2: 15,908–16,094; exon 3: 15,653–15,814; exon 4: 15,121–15,217; exon 5: 14,816–14,984; exon 6: 13,489–13,621; exon 7: 13,171–13,361; exon 8: 11,466–11,543). Init = Contains initiation (START) codon. Term = Contains terminal (STOP) codon. PolyA signal = Polyadenylation signal. GABRA6 and septin 8 do not have initiation or terminal codons because they are partial on this contig or because they could not be inferred

Mitochondrial genotypes were then visually analyzed to identify all possible haplotypes. The same technique was applied to the cyclin G1 which was made possible by the fact that there was extremely low haplotypic diversity in that region (see Results for details). Both contigs were annotated in order to obtain exonic sequences and positions for each gene. GENSCAN (Burge and Karlin 1997) was used for the cyclin G1 contig and MITOS (http://mitos.bioinf.uni-leipzig.de) for the mitogenome. Putative coding sequence of each annotated gene was verified through nucleotide (*nr/nt, Blastn*) and translated nucleotide (*nr, Blastx*) blasts at NCBI (http://www.ncbi.nlm.nih.gov).

Using Consite (Sandelin et al. 2004), we analyzed the 10 kb surrounding the cyclin G1 gene to infer putative transcription factor binding sites (TFBS). This uses JASPAR's (Portales-Casamar et al. 2010) non-redundant and curated database to infer TFBS in the most conserved regions of orthologous pairs of genomic sequences. The threespine stickleback genome sequence from primer development was used as the ortholog and the conservation threshold was set to 70 % (minimum BIT score = 15, transcription factor score threshold = 80 %). These data were crossed with SNP loci positions to verify if polymorphisms could influence TFBS specificity.

Non-synonymous mutations were identified on both contigs by looking at every coding genetic variant in MEGA5 (Tamura et al. 2011). Refseq (Pruitt et al. 2009) annotations of homologous proteins were used to verify if the mutations affected binding or active domains. In addition, the hydropathy index (Kyte and Doolittle 1982), sidechain charge and polarity (Wimley et al. 1996) of amino acids (AAs) involved in the substitutions were taken into account. Alignments were performed using 20–30 (according to availability) protein sequences (NCBI, *Blastp*) that are homologous to the target genes. The sequences range from closely related (e.g. yellow perch, percids and other fishes) to very distant relatives (e.g. dog, rat, human). These alignments were used to assess ancestry and the degree of conservation of the non-synonymous sites and ultimately to formulate hypotheses about the evolutionary importance of the substitutions.

Finally, 3D models of each target protein were created using GeneSilico prediction metaserver (Kurowski and Bujnicki 2003; https://genesilico.pl/meta2). This combines a complete array of methods that search for available templates, performs alignments and assesses their quality. It retrieves data on primary, secondary, tertiary structure and other important features such as disulfide bonds and disordered regions. This information was then fed to FRankenstein3D (http://genesilico.pl/frankenstein/) which implements a fold recognition method and combines the data to produce a hybrid model (Kosinski et al. 2003). In general, it recombines the multiple alignments from the different templates to produce many alternative models for which quality is assessed. The best fragments are combined and the model is perfected through iterative quality checks and alternative alignments in non-consensus regions.

To verify the validity of the models, RAMPAGE (Lovell et al. 2003; http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) was used to produce updated Ramachandran plots. This structure validation method describes the deviation from expected AA backbone to sidechain angles. Major deviations indicate misfit conformations or inappropriate refinement restraints (Lovell et al. 2003). Once the produced models were validated, they were introduced in PyMol 1.3 (Schrodinger 2010) to produce a visual three-dimensional representation including marked polymorphic AAs.

Results

Metal quantification

Metal recoveries from the ICP-MS measurements amounted to 111.6 ± 1.6 % for Cd and 100.8 ± 1.7 % for Cu [mean ± standard error (SE)]. A total of 17 values (out of 685 in 2005 and 367 in 2010) were removed as outliers or could not be measured. Most Cd and Cu concentrations were stable in 2005 and 2010 except for Lake Dufault in which Cd contamination decreased by 51.6 %, from 36.5 to 17.7 µg/g of dry liver weight (dlw) and Cu concentration passed from 81.0 to 24.2 µg/g of dlw, a 70.10 % decrease (Table 1).

The thresholds established by Couture and Pyle (2008) for yellow perch were used to discriminate between clean and contaminated fish populations. These thresholds were developed using metabolic indicators that display significant impairment attributed to a given level of metal contamination. Thus, a liver Cd concentration in excess of 11.0 µg/g dlw indicates contaminated (i.e. impaired) fish 70 % of the time with reference fish above the contamination threshold only 10 % of the time. Liver Cu concentrations exceeding 38.8 µg/g dlw discriminate clean from contaminated fish 72 % of the time with reference fish above the threshold 10 % of the time.

Marker development

Whole transcriptome sequencing (Pierron et al. 2011) produced 1,197,973 sequences ≥100 bp that were aligned using CLC Genomics Workbench (v3.7) into 9,856 contigs representing 5,206 unique genes. A total of 10,571 SNPs with a minimal coverage of 15 sequences and four observations (or 5 % frequency) were detected in these unique genes. The genotypes of the 32 individuals were extracted and markers were ordered by F_{ST}. Overall, 4,959 of these putative SNPs were visually screened for quality or filtered out using

indications of paralogy, abnormal marker abundance or proximity to the polyadenylated tail. The sequences surrounding the selected SNPs were blasted in NCBI (*Blastn* and *Blastx*) to assign all markers to annotated genes using a minimal e-value criterion of 1^{-30}. Finally, 130 of the best putative SNPs, including all outliers detected by the Beaumont and Nichols (1996) selection detection method implemented in Arlequin 3.5 (Excoffier and Lischer 2010), were selected for population wide genotyping.

Out of these 130 putative SNPs, primers of 98 markers were successfully designed and genotyped on a test panel of 40 individuals distributed among the 10 lakes. Among these, 11 were dropped due to genotyping issues, 37 due to systematic monomorphism and two due to rare allele frequencies $\leq$0.5 %. The final dataset thus consisted of 48 high-quality fully annotated and validated polymorphic SNPs, including two mitochondrial loci (Table 2 for annotations).

Temporal stability

Pairwise F_{ST} calculated between the two sampling years in each population did not reveal any inter-sampling difference i.e. all values were not significantly different from 0 ($\alpha = 0.01$). The locus-by-locus AMOVA (Table S3) indicated that inter-annual variation was not significantly different from zero. Therefore, the genotypes of both sampling years were pooled for subsequent analyses.

Descriptive statistics

HWE tests revealed that only four nuclear markers (SNP_PRS6A, SNP_SID4, SNP_RPL18A, SNP_F7P) significantly departed from equilibrium in at least one of the 10 populations after Bonferroni correction ($\alpha = 0.05$, critical value = 0.0012).

LD statistics revealed that five pairs of markers (including the mitochondrial pair) were under significant linkage in at least one of the populations after Bonferroni correction ($\alpha = 0.05$, critical value = 3.9E−6). One marker from each pair was dropped from the datasets used to perform analyses that may be biased by LD between markers.

All inter-population pair-wise F_{ST} (Table S4) were highly significant (all p values <0.0001) and F_{ST} values ranged from 0.009 (Opasatica-Évain) to 0.414 (Osisko-Vaudray). Heterozygosity values and allelic frequencies for each marker in each population separately are shown in (Table S5).

Outlier detection and association tests

BayeScan revealed 11 outlier markers, three of which were putatively under directional selection and eight under balancing selection (Table 2; Fig. 3). The three markers that displayed significant patterns of directional selection are located on the nuclear cyclin G1 gene (SNP_CCNG1), the mitochondrial Cytb (SNP_CYTB) and the mitochondrial NADH dehydrogenase subunit 2 (ND2) genes (SNP_ND2).

Logistic regressions conducted with MatSAM 2 (Joost et al. 2007) revealed that alleles from 14 different genes are significantly associated with mean population hepatic Cd concentrations, 18 with mean Cu and 8 with both Cd and Cu (Table S6). The three outlier markers were associated with both metals, but more significantly so for Cd than Cu. For the ancestry-adjusted χ^2 tests and after Bonferroni correction ($\alpha = 0.05$ critical value = 0.001), four SNPs displayed significant association with Cd and five with Cu (Table 3). Globally, the four SNPs, including the three selection outliers, were associated with both mean population Cd and Cu hepatic concentrations and in each case, the χ^2 values were again much more significant for Cd than for Cu. Linear regression tests of association appeared more stringent than the other two methods. In particular, the three outlier markers that are shown in Fig. 4 were not significant after Bonferroni correction (p value threshold of 0.001 for $\alpha = 0.05$). However, they generally showed a nearly significant trend, in particular for Cd for which the correlation between SNP_CCNG1 and [Cd] was significant before Bonferroni correction ($R^2 = 0.493$, p value = 0.024). Overall, three SNPs stood out as significantly responding to both main association tests (logistic regressions and χ^2 tests) and being outliers in the genome scan. These were thus retained for further characterization: SNP_CCNG1 (cyclin G1), SNP_ND2 (NADH dehydrogenase subunit 2) and SNP_CYTB (Cytb).

Mapping of candidate genes

Genomic and mitochondrial amplicon lengths were between 7.7 and 10.3 kb (Table S1 for all lengths). They were successfully amplified on all 40 individuals which produced 455,936 reads measuring between 20 and 1,539 bp with an average length of 339.26 bp after trimming. Globally, 204,018 reads contributed to the mappings, 87,587 to the global cyclin G1 mapping (average coverage = 2,418) and 116,431 to the global mitogenome mapping (average coverage = 2,772). The 80 individual mappings had average coverage between 27 and 123 for cyclin G1 and between 34 and 193 for the mitogenome.

Sequence analysis of cyclin G1

For cyclin G1, 17 quality SNPs were found on the 17,014 bp contig. We could manually rebuild the three haplotypes (Table S7) found in these populations since

Table 2 Outlier detection and gene annotation

Marker name (main text)	Marker name (data)	Complete gene name	Probability (PO)	α	F_{ST}
SNP_CCNG1	SNP901_759	Cyclin G1	0.975	0.842	0.345*
SNP_ND2	SNP1044_232	NADH dehydrogenase subunit 2	0.928	0.875	0.353*
SNP_CYTB	SNP9273_440	Cytochrome b	0.900	0.798	0.338*
SNP_GALM	SNP1124_320	Aldose 1-epimerase (galactose mutarotase)	0.509	0.267	0.246
SNP_RPS10	SNP147_1668	Ribosomal protein S10	0.291	0.004	0.204
SNP_FMVIA	SNP1562_919	Myosin VI isoform A (FMVIA)	0.333	0.069	0.214
SNP_NR0B1	SNP1658_356	Nuclear receptor subfamily 0, group B, member 1 (nr0b1)	0.305	0.048	0.211
SNP_A1AAO	SNP2092_809	Skin mucus antibacterial l-amino acid oxidase	0.345	0.002	0.205
SNP_FMO	SNP2196_690	Flavin monooxygenase	0.977	−1.112	0.093*
SNP_NPR2	SNP240_158	Atrial natriuretic peptide receptor 2	0.551	0.329	0.257
SNP_ACTR3	SNP2609_878	Actin-related protein 3	0.349	0.120	0.221
SNP_GRHPR	SNP3125_843	Glyoxylate reductase/hydroxypyruvate reductase	0.375	−0.053	0.199
SNP_ENO1	SNP3476_729	Enolase 1 alpha-like	0.428	0.190	0.234
SNP_SEPPB	SNP4056_756	Selenoprotein Pb	0.963	−1.047	0.098*
SNP_RCG	SNP4076_260	rCG52946-like	0.493	−0.246	0.174
SNP_4677	SNP4677_901	–	0.943	−0.975	0.104*
SNP_PMAP	SNP4995_1424	Adipocyte plasma membrane-associated protein	0.341	−0.059	0.196
SNP_CTH	SNP511_869	Cystathionine gamma-lyase	0.851	−0.777	0.122
SNP_EVLCFA	SNP5315_240	Similar to vertebrate elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast-like)	0.333	−0.062	0.196
SNP_TYMS	SNP5797_227	Thymidylate synthase	0.757	0.758	0.336
SNP_PRS6A	SNP581_333	26S protease regulatory subunit 6A	0.999	−2.280	0.0380*
SNP_LRH1	SNP5863_236	LRH-1	0.675	0.417	0.270
SNP_ADH4	SNP6352_739	Alcohol dehydrogenase Class VI	0.309	−0.029	0.200
SNP_FLEC	SNP6410_978	Putative F-type lectin	0.807	−0.654	0.132
SNP_BSEP	SNP650_1161	Bile salt export pump (sister of P-glycoprotein)	0.929	−0.905	0.109*
SNP_GST	SNP6732_738	Glutathione S-transferase	0.474	0.240	0.242
SNP_SID4	SNP7025_816	Secreted immunoglobulin domain 4	0.879	−0.809	0.118
SNP_PLG	SNP7034_207	Plasminogen	0.856	0.666	0.314
SNP_ACTB	SNP7215_1074	B-Actin	0.320	−0.029	0.200
SNP_DDX21	SNP7318_403	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	0.313	−0.043	0.198
SNP_UGT2B10	SNP7462_816	UDP glucuronosyltransferase 2B10-like	0.766	−0.580	0.139
SNP_RPL3	SNP7792_316	60S ribosomal protein L3	0.949	−1.012	0.101*
SNP_RPS19	SNP8018_162	40S ribosomal protein S19	0.362	−0.048	0.199
SNP_CDLC2	SNP8044_579	Dynein light chain 2, cytoplasmic-like isoform 2	0.310	−0.014	0.202
SNP_DPYS	SNP8475_1311	Dihydropyrimidinase	0.570	−0.322	0.166
SNP_FMO5	SNP8550_935	Flavin containing monooxygenase 5	0.926	−0.891	0.110*
SNP_PTPRC	SNP8816_89	Protein tyrosine phosphatase receptor-type C (CD45)	0.331	−0.068	0.195
SNP_RPL18A	SNP8852_100	Ribosomal protein L18A	0.993	−1.326	0.079*
SNP_RPS10	SNP8880_318	Ribosomal protein S10	0.317	−0.050	0.197
SNP_FTCD	SNP8891_798	Formimidoyltransferase cyclodeaminase	0.308	0.023	0.207
SNP_8896	SNP8896_108	–	0.312	0.052	0.211
SNP_EF1A	SNP9009_1475	Elongation factor 1-alpha	0.351	−0.091	0.192
SNP_RPSA	SNP9082_60	40S ribosomal protein SA	0.869	0.625	0.305
SNP_ADHA	SNP9106_1174	Alcohol dehydrogenase chain A	0.374	−0.116	0.189
SNP_PDLIM5	SNP915_427	PDZ and LIM domain 5 isoform 2	0.439	0.213	0.239

Table 2 continued

Marker name (main text)	Marker name (data)	Complete gene name	Probability (PO)	α	F_{ST}
SNP_F7P	SNP9272_173	Coagulation factor VII precursor	0.554	−0.298	0.169
SNP_TTC36	SNP9465_111	Tetratricopeptide repeat protein 36	0.473	0.221	0.238
SNP_PLG2	SNP9993_1663	Plasminogen	0.652	0.432	0.274

Output of outlier test performed with BayeScan, PO, α, F_{ST}. Marker names (main text) are abbreviations of the genes on which the markers are located. They represent usual abbreviations for each gene when possible. Marker names (data) have numbers referring to the contig number and position on which they are located in Pierron et al. (2009) assemblies. Complete gene names were found by blasting (*Blastn* and *Blastx*; NCBI) a large sequence surrounding the SNP and always correspond to results with e-values ≥1E−30

PO probability of being under selection compared to neutrality, α strength and directionality of selection, F_{ST} Bayesian *F* statistic

* Outlier F_{ST} values; positive α values indicate directional selection while negative values indicate balancing selection

58 % of the individuals were homozygous for all SNPs. These haplotypes were then compared to the heterozygous individuals, which could all be explained as a combination of two of the three haplotypes.

The GENSCAN (Burge and Karlin 1997) annotation (Fig. 2) revealed the exonic structure of two complete genes [cyclin G1 and NudC domain-containing protein 2 (NudC2)] and two partial genes [gamma-aminobutyric acid receptor subunit alpha-6 (GABRA6) and septin 8] located on the contig. TFBS inference (Table S8) revealed ten putative binding sites around cyclin G1, the most important being a strongly supported (BIT score = 19.53) p53 site reported in humans (Ohtsuka et al. 2004) suggesting that yellow perch and human cyclin G1 share expression regulation mechanisms and pathways. However, no SNPs were found on any of the inferred putative TFBSs. Three non-synonymous mutations were found, two on cyclin G1 and one on septin 8 (Table 4). In particular, one of these

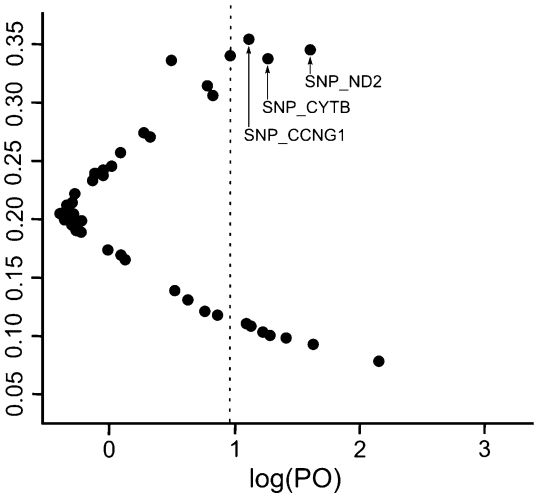

Fig. 3 Results of the BayeScan 2.0 outlier test. Posterior probability significance threshold (*vertical bar*) of 0.90 after Bonferroni correction (α = 0.05). Outliers most likely under directional selection are *labeled* (see Table 2 for complete gene names)

substitutions on AA 204 of cyclin G1, substantially altered the protein physicochemical properties by changing an alanine (non-polar and hydrophobic) to serine (polar and slightly hydrophilic), whereas the other two substitutions conferred very similar physicochemical properties (Table 4). In addition, the cyclin G1 protein alignment revealed that AA 204 was extremely conserved in all vertebrates which all shared an alanine at that site except for guppies for which it was replaced by a valine (ABN80454.1) (Fig. S1). Thus, a serine at AA 204 is unique to yellow perch and predominates in polluted lakes. In addition, the neighbouring AAs also were well conserved, making this a potential functional domain of cyclin G1. According to the Refseq annotation of homologs (Reference: XP_003455215.1), the mutation is not located on a known binding site of cyclin G1 but could be functionally important nonetheless (see Discussion).

Lastly, a 3D model yellow perch cyclin G1 protein was created (Fig. 5). The Ramachandran plot indicates the good quality of the model; 95.7 % of all residues are in favored regions (~98.0 % expected), 4 % are in allowed regions (~2.0 % expected) and 1 residue (0.4 %) is an outlier.

Sequence analysis of the mitogenome

The complete 16,537 bp yellow perch mitogenome was reconstructed and 13 SNPs were found among the 40 individuals (Table S9). Overall, the yellow perch mitogenome is slightly shorter than that of *P. macrolepida* (16,602 bp; DQ536430.1) and *Etheostoma radiosum* (16,600 bp; AY341348.2). Comparing the only control region haplotype identified here to Sepulveda-Villet et al. (2009) data set of 21 North-American yellow perch control regions showed that it is closer to the haplotype that predominates the Great Lakes (haplotype 1; FJ155931.1) from which it differed overall by 11 SNPs and one insertion (data not shown). Modeling of the two mitochondrial proteins was also carried out successfully (Fig. 5). The Ramachandran plot of the Cytb model revealed that 98.4 %

Table 3 Stratification-corrected χ^2 associations with population mean hepatic metal contamination

Marker name	χ^2 for Cd association	*P* value for Cd association	χ^2 for Cu association	*P* value for Cu association
SNP_CCNG1	101.847	5.997E−24*	67.260	2.379E−16*
SNP_ND2	95.137	1.776E−22*	35.555	2.479E−09*
SNP_CYTB	84.142	4.605E−20*	29.095	6.893E−08*
SNP_GALM	0.883	0.347	4.496	0.034
SNP_RPS10	9.553	0.002	4.874	0.027
SNP_FMVIA	0.063	0.801	1.890	0.169
SNP_NR0B1	2.281	0.131	8.990	0.003
SNP_A1AAO	0.325	0.568	2.803	0.094
SNP_FMO	8.232	0.004	6.688	0.010
SNP_NPR2	0.043	0.836	0.298	0.585
SNP_ACTR3	0.014	0.906	2.873	0.090
SNP_GRHPR	0.189	0.664	1.264	0.261
SNP_ENO1	0.429	0.512	0.442	0.506
SNP_SEPPB	0.915	0.339	1.099	0.294
SNP_RCG	6.201	0.013	8.898	0.003
SNP_4677	0.510	0.475	0.032	0.858
SNP_PMAP	1.579	0.209	<0.0001	>0.992
SNP_CTH	0.111	0.739	0.105	0.746
SNP_EVLCFA	1.116	0.291	4.147	0.042
SNP_TYMS	6.207	0.013	9.692	0.002
SNP_PRS6A	5.454	0.020	8.148	0.004
SNP_LRH1	7.180	0.007	3.926	0.048
SNP_ADH4	0.038	0.845	5.056	0.025
SNP_FLEC	0.648	0.421	2.876	0.090
SNP_BSEP	0.367	0.545	6.581	0.010
SNP_GST	30.693	3.023E−08*	44.079	3.155E−11*
SNP_SID4	8.552	0.003	0.069	0.793
SNP_PLG	0.083	0.774	0.072	0.788
SNP_ACTB	0.075	0.784	0.001	0.970
SNP_DDX21	3.988	0.046	1.843	0.175
SNP_UGT2B10	0.0002	0.989	7.518	0.006
SNP_RPL3	7.138	0.008	0.590	0.443
SNP_RPS19	0.155	0.694	0.170	0.680
SNP_CDLC2	0.168	0.682	0.796	0.372
SNP_DPYS	1.836	0.175	5.187	0.023
SNP_FMO5	0.018	0.893	0.516	0.472
SNP_PTPRC	0.015	0.903	2.131	0.144
SNP_RPL18A	9.673	0.002	6.695	0.010
SNP_RPS10	0.396	0.529	0.482	0.487
SNP_FTCD	4.516	0.034	0.019	0.891
SNP_8896	0.0003	0.986	0.161	0.689
SNP_EF1A	2.284	0.131	9.851	0.002
SNP_RPSA	5.584	0.018	3.549	0.060
SNP_ADHA	0.631	0.427	1.009	0.315
SNP_PDLIM5	7.347	0.007	6.533	0.011
SNP_F7P	0.087	0.768	0.640	0.424
SNP_TTC36	0.094	0.759	1.672	0.196
SNP_PLG2	0.003	0.955	10.843	0.001*

* Significant values after Bonferroni correction ($\alpha = 0.05$, critical value = 0.001)

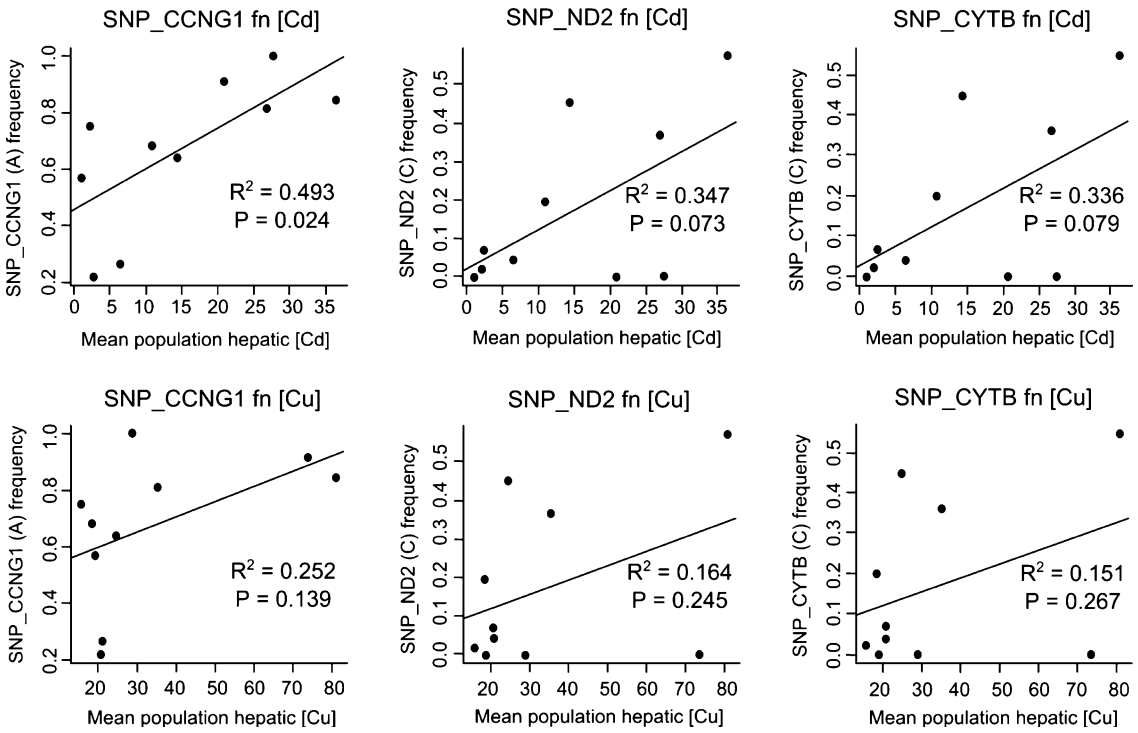

Fig. 4 Linear regressions between candidate SNP allelic frequencies in each population and mean population hepatic contamination. *P* value significance threshold of 0.001 after Bonferroni correction ($\alpha = 0.05$). Computed in the R 2.13.0 environment (*R core development team*)

of all residues are in favored regions (~98.0 % expected), 1.6 % are in allowed regions (~2.0 % expected) and 0 in outlier regions. The ND2 model was less reliable, 93.1 % of all residues are in favored regions, 5.2 % are in allowed regions and 6 residues (1.7 %) are outliers.

Only four non-synonymous SNPs were found; two on ND2 and two on Cytb (Table 4). According to Refseq annotations of the best matches (*Blastx*; NCBI), both Cytb (Reference: AAM19461.1) substitutions are outside functional and binding domains and involve AAs that have nearly identical properties (Table 4). The 3D visualization shows that the substitution is in an alpha helix in the outer region of the protein (Fig. 5). In contrast, the ND2 sequence has a threonine (slightly hydrophilic and polar) to alanine (hydrophilic and non-polar) substitution at position 278 (initial reference marker SNP_ND2) and another rare substitution found in two individuals (haplotype 5; isoleucine/valine substitution on AA 208 of ND2). The rare AA 208 substitution involves AAs with similar properties, whereas the AA 278 substitution confers very different ones (Table 4). In addition, the proteic alignment (Fig. S1) showed that the alanine (AA 278) that is found at higher frequency in contaminated populations is an ancestral AA that is highly conserved in fishes while being differentiated in mammals. ND2 is a subunit of respiratory complex I of the mitochondrion taking part in proton pumping (Gingrich et al. 2004) and the implications of the observed substitution are considered in the discussion.

Discussion

The main goal of this study was to test the hypothesis that adaptive evolution to metals has recently occured in the contaminated yellow perch populations of Rouyn-Noranda. To this end we: (1) developed a set of putative markers selected from transcriptomic data; (2) detected selection outliers along with marker frequencies associated with contamination; and (3) characterized candidate non-synonymous mutations in genetic regions displaying signals of selection and association with metals contamination. We were thus able to validate and genotype 48 annotated candidate SNPs on 1,052 individuals from 10 populations with various contamination levels. Using the genotypes in conjunction with hepatic [Cd] and [Cu], outlier detection and association with contamination tests were successfully performed. Comparing the results of each method, three SNPs were shown to be putatively under selection while displaying allelic frequencies mostly correlated to the mean populayion hepatic [Cd]. These markers defined candidate genetic regions for adaptation to metal contamination that were further characterized through the amplification of wide surrounding genetic regions. High-throughput sequencing of these regions allowed us to identify non-synonymous mutations in gene sequences. Using an array of sequence analysis methods, two candidate mutations that could potentially be responsible for adaptation to Cd contamination were identified.

Table 4 Characteristics of non-synonymous substitutions on candidate genes

Cyclin G1 contig				Mitogenome			
Position (ref/alt)	Localisation	Amino acid substitution	Substitution characteristics	Position (ref/alt)	Localisation	Amino acid substitution	Substitution characteristics
5813 (C/T)	Exon 4 of cyclin G1	217 Asp (GAC)/ Asn (AAC)	Asp = polar negative, hydropathy = −3.5 Asn = polar neutral, hydropathy = −3.5	4666 (A/G)	NADH dehydrogenase subunit 2	208 Ile (ATC)/ Val (GTC)	Ile = non-polar neutral, hydropathy = 4.5 Val = non-polar neutral, hydropathy = 4.2
5852 (C/A)	Exon 4 of cyclin G1	204 Ala (GCC)/ Ser (TCC)	Ala = non-polar neutral, hydropathy = 1.8 Ser = polar neutral, hydropathy = −0.8	4876 (A/G)	NADH dehydrogenase subunit 2	278 Thr (ACA)/ Ala (GCA)	Thr = polar neutral, hydropathy = −0.7 Ala = non-polar neutral, hydropathy = 1.8
15172 (C/T)	Exon 4 of septin8	198 Val (GTG)/ Met (ATG)	Val = non-polar neutral, hydropathy = 4.2 Met = non-polar neutral, hydropathy = 1.9	15465 (C/G)	Cytochrome b	361 Leu (CTT)/ Val (GTT)	Leu = non-polar neutral, hydropathy: 3.8 Val = non-polar neutral, hydropathy = 4.2
				15499 (T/C)	Cytochrome b	372 Val (GTT)/ Ala (GCT)	Val = non-polar neutral, hydropathy: 4.2 Ala = non-polar neutral, hydropathy = 1.8

All references are made in relation to a consensus contig (see Annex7 for haplotype details and accession numbers). Reference SNPs are underlined (position 4876 = reference SNP_ND2; position 15465 = reference SNP_CYTB; reference SNP_CCNG1 is located on an untranslated region)

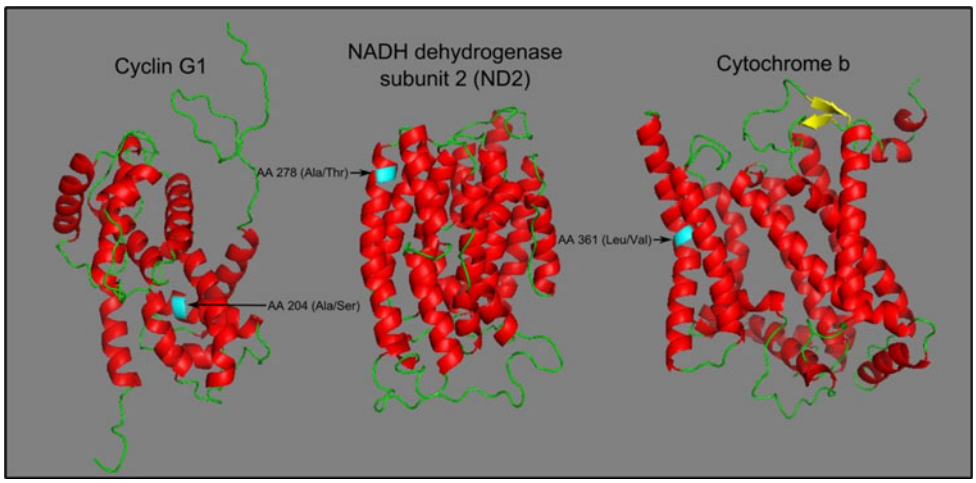

Fig. 5 Candidate proteins 3D models. *Red* alpha helixes, *yellow* beta sheets, *green* loops, *cyan* candidate SNP for each characterized protein. Models were created in conjunction between GeneSilico prediction metaserver (Kurowski and Bujnicki 2003) and FRankenstein3D (Kosinski et al. 2003), visual output produced with PyMol 1.3 (Schrodinger 2010)

From these results in conjunction with previous knowledge of this system, we propose that Cd represents a selective agent driving an evolutionary change in contaminated yellow perch populations and that energetic limitations could have driven the evolution of the faster life cycle that is characteristic of these populations. The following sections discuss the evidence that links the two candidate mutations found in the cyclin G1 and ND2 genes to this evolutionary setting as well as the applied implications of our findings.

Previous studies concerning the effects of smelter contamination on the yellow perch populations of Rouyn-Noranda mainly focused on the physiological impacts resulting from metal bioaccumulation. While the affected fish were not shown to have superior detoxification capacities when compared to clean fish, some of their life history traits differed significantly based on physiological measurements (Campbell et al. 2005). More specifically, contaminated fish populations systematically displayed a faster growth rate, earlier sexual maturation and shortened lifespan, presumably due to unrepaired physiologic damage (Couture and Pyle 2008). Interestingly, these life history characteristics are predicted to evolve in stressed animal population experiencing energetic limitations (Sibly and Calow 1989). Indeed, strong energetic limitations were shown to occur in studied contaminated perch populations due to an energetic bottleneck caused by a simplified food web (Sherwood et al. 2002). This implies that these differentiated life history traits could theoretically be partly attributed to adaptive evolution whereby energy is invested into growth and reproduction at the expense of life expectancy in order to maximize fitness in the face of a highly polluted and energy deprived environment. Conversely, in the absence of evolutionary data, life history patterns could solely be attributed to a plastic response. However, Bourret et al. (2008) showed that increasing Cd bioaccumulation was associated with a decrease of gene diversity in the contaminated populations, indirectly suggesting that directional selection for metal tolerance could have contributed to shaping this diversity pattern. Also, Couture and Pyle (2008) established contamination thresholds beyond which perch experience the physiologic effects of toxicity. Comparing these thresholds to measured population means of contamination (Table 1) indicated that Cd was a significant stressor in all contaminated populations whereas Cu only exceeded the threshold in two populations (Dufault and Marlon). Thus, Cd was the most plausible selecting agent for adaptive evolution of the traits observed in the contaminated populations of Rouyn-Noranda.

Adaptive importance of cyclin G1 for metal tolerance

All analyses performed here pinpoint cyclin G1 as a strong candidate for potential adaptive evolution for metal tolerance in contaminated perch populations. Indeed, the substitution on AA 204 was a strong outlier and the likeliest nuclear DNA target of directional selection. Its frequency varied with Cd concentration and involved AAs with very distinct physicochemical properties at a site otherwise highly conserved in vertebrates. In addition, the AA 204 substitution is located in an alpha helix, towards the center of the cyclin G1 protein. The localisation and properties of involved AAs make it theoretically possible for the substitution to provoke conformational changes or to alter an activation site as serines are phosphorylable (Cohen 2000).

Cyclin G1's function is partially understood (Ohtsuka et al. 2004) and its annotation has been done through prediction and homology methods (Pruitt et al. 2009) that might not have identified all important domains. As a consequence, the fact that the candidate substitution is not located on the protein's inferred functional domains does not negate the possibility that it affects an uncharacterized

one. While the list of cyclin G1 partners is partial, interactions with ARF and tumor suppressor p53 have been reported (Ohtsuka et al. 2003; Zhao et al. 2003). p53 provokes the transcriptional activation of cyclin G1 (Zhao et al. 2003) which in turn can recruit protein phosphatase 2A to dephosphorylate Mdm2, thereby downregulating p53 (Okamoto et al. 2002). Also, the binding of cyclin G1 to an unknown CDK partner would promote its own degradation (Piscopo and Hinds 2008) whereas the binding to ARF increases its half-life (Piscopo and Hinds 2008).

Despite limited comprehension of cyclin G1's functionality, p53 expression mediation makes consensus (Piscopo and Hinds 2008). In general, the p53 pathway is activated in reaction to DNA damage and triggers mechanisms that either seek to repair damage or, under strong stress signal, provoke controlled cell death i.e. apoptosis (Harris and Levine 2005). As a consequence, cyclin G1 is more expressed in situations of stress causing DNA damage (Harris and Levine 2005). Rouyn-Noranda's main toxicants, Cd and Cu, are indeed genotoxic stressors that can cause genetic damage (Pra et al. 2008; Schwerdtle et al. 2010). Part of Cd genotoxicity is attributed to its binding to p53 causing conformational changes that inhibit its ability to activate its transcriptional targets, ultimately leading to unrepaired DNA damage (Meplan et al. 1999). Moreover, recent work suggests that cyclin G1's expression is often associated with growth promotion (Piscopo and Hinds 2008). Thus, two main functions stand out for cyclin G1; growth regulation and p53 downregulation.

It is tempting to interpret the candidate substitution on AA 204 as an adaptation to reduce its indirect downregulation of p53, thus compensating for Cd inhibitory effects and allowing an appropriate stress response. However, contaminated Rouyn-Noranda perch do not display strong damage repair capacities, and conversely suffer from physiological impairments potentially causing early death (Couture and Pyle 2008). However, contaminated juveniles show faster growth rates and earlier sexual maturation than those from non-contaminated lakes despite lower trophic resources (Couture and Pyle 2008). This corroborates theory predicting the evolution of a faster life cycle (Sibly and Calow 1989) in constantly stressed and energy deprived populations (Sherwood et al. 2002).

Thus, considering the fast growth rates of contaminated juveniles (Couture and Pyle 2008), the most parsimonious adaptive hypothesis for AA 204 substitution involve it in a growth enhancement that would lead to a younger age of reproduction. This process could be achieved by an increase of cyclin G1 binding specificity for ARF or a decrease of specificity with a putative CDK partner since both mechanisms can hypothetically increase cyclin G1 half-life, allowing longer lasting growth promotion. Alternatively, the mutation could increase cyclin G1's aptitude to bind to the unknown effectors responsible for enhancing growth.

The proposed adaptive mechanisms would allow fish to grow fast in contaminated environments, thus more rapidly completing their life cycle, even though cyclin G1 expression is limited by the impairment of p53 by Cd. Thus, inhibition of p53 by Cd may have been an early historical factor contributing to the evolution of a fast life cycle strategy instead of detoxification and damage repair capacities. The possibility that contaminated perch reallocate energy towards growth instead of fixing the damage repair pathway may also partly account for the shorter lifespan observed in contaminated populations. Clearly, the hypothetical role of cyclin G1 in growth regulation in the context of metal contamination deserves further investigation.

The conservation pattern of AA 204 and those surrounding it throughout vertebrates is particularly striking, implying a potential functional role of the candidate substitution. At present, our data do not allow rigorous assessment whether the AA 204 polymorphism reflects standing genetic variation (Barrett and Schluter 2008) and upon which selection has been acting in contaminated lakes or instead, if the AA 204 substitution is a de novo mutation that appeared in contaminated lakes before migrating to neighboring clean lakes. However, cyclin G1 haplotypes that were identified differed by several fixed mutations (17 fixations out of about 17 kb sequenced), supporting the former hypothesis. Why such apparently strong directional selection has not driven AA 204 to fixation in contaminated lakes remains to be explained. One explanation could be the temporally cyclical nature of selection. Indeed, levels of contamination have historically fluctuated in lakes of Rouyn-Noranda (Couture et al. 2008a), such that fish may have experienced episodes of cleaner, less stressful conditions. Even if the candidate polymorphism was not maladaptive in clean conditions, this type of cyclical selection is expected to prevent adaptations from reaching fixation (Felsenstein 1976; van Straalen and Timmermans 2002). Another non exclusive explanation is that strong genetic drift associated with small effective population sizes in contaminated lakes (Bourret et al. 2008) may have opposed the effect of directional selection. Finally, possible gene flow between contaminated and clean lakes could also have prevented fixation throughout the system.

Adaptive importance of NADH dehydrogenase subunit 2 for metal tolerance

Our results also pinpointed a substitution on ND2 as a strong functional candidate potentially contributing to adaptive metal tolerance in contaminated perch. AA 278 substitution is putatively under directional selection based on genome scan analysis and its frequency is correlated with [Cd]. In addition, it involves dissimilar AAs that are

usually conserved in fishes, the ancestral AA being favored in contaminated populations. As for cyclin G1, selection acting on standing genetic variation is the likeliest mechanism that shaped the pattern of haplotype frequency observed at ND2. In contrast, non-synonymous mutations detected in Cytb gene seemed less likely to be the direct target of selection and may have hitchhiked along with the candidate substitution of ND2.

Comparison of protein sequence with homologous RefSeq annotations indicate that AA 278 substitution does not affect any domain that would be functionally relevant for the proton pumping function of ND2. However, the protein also plays an alternative, and potentially important, role outside of the respiratory complex I. It serves as an adaptor anchoring Src to the *N*-methyl-D-aspartate receptor (NMDAR) in brain synapses (Gingrich et al. 2004). The AA 278 substitution precisely affects the ND2 domain responsible for this anchoring mechanism.

The brain is thought to be the major target of acute Cd poisoning in rat (Shah and Pant 1991) and cerebral function integrity is obviously essential to survival. Interestingly, ND2 is an adaptor anchoring Src and NMDAR together in excitatory glutaminergic synapses of the brain (Gingrich et al. 2004). Gingrich et al. (2004) demonstrated that its adaptor domain is located on the AAs 239–321, thus covering the candidate site (AA 278). In essence, the candidate substitution could influence the binding specificity among components of the NMDAR complex. Indeed, NMDAR opening up regulation by ND2-Src interactions is thought to be essential to long-term potentiation (LTP) (Ali and Salter 2001; Gingrich et al. 2004). Also, LTP is the main mechanism explaining learning and memory in animals (Kandel 2001). More specifically, synaptic changes allowing LTP are triggered by a metabolic cascade that begins after NMDAR allows calcium (Ca) into the post-synaptic spine following specific depolarization conditions (Nakazawa et al. 2004). In the particular case of the LTP pathway, Cd is known to block NMDAR and may diffuse through it (Mayer et al. 1989). In turn, blocked NMDARs will cause LTP inhibition (Schiller et al. 1998) and impair temporal and spatial memorization as it was demonstrated on NMDAR knockout mice (Shimizu et al. 2000).

Moreover, Cd ions are known to affect the activity of another protein of the LTP pathway, calmodulin (CaM) (Vig et al. 1989). CaM generally becomes activated via the binding of Ca at four specific sites (Chin and Means 2000). In neurons, activated CaM triggers the molecular cascade of the LTP pathway in reaction to calcium influx caused by NMDAR opening (Kandel 2001). Interestingly, Cd was shown to binds to CaM in rat brains, provoking conformational changes that inhibit its activity (Vig et al. 1991). Overall, Cd can block NMDAR and inhibit CaM, both mechanisms that will in turn inhibit LTP.

It is thus plausible that LTP impairment trough Cd–NMDAR binding and Cd–CaM occurs in perch populations suffering from toxic levels of Cd bioaccumulation. This could potentially decrease their fitness, for instance by reducing the ability to memorize the location and timing of resource availability and predation-associated hazards. The mutation in the ND2 adaptor sequence (AA 278) could affect the threshold at which NMDAR opens. Lowering the threshold would allow more Cd into neurons and favor NMDAR blockage, leading to LTP inhibition (Shimizu et al. 2000) or even severe adverse effects such as cell death (Kim et al. 2000). The most parsimonious hypothesis dictates that the NMDAR opening threshold would be increased through less ND2 adaptor specificity, which would in turn lead to less frequent Ca and Cd influx. Perhaps this higher threshold may result in LTP only being triggered under the most intense NMDAR excitatory conditions, effectively limiting NMDAR blockage and CaM inactivation by Cd. Less frequent NMDAR openings could also allow neurons to turn over Cd-inactivated proteins at a pace that would allow LTP to be at least partially functional.

This hypothetical mechanism could reflect a tradeoff between memorization capacity, brain damage and growth. As stated before, contaminated perch populations display a fast growth rate and a limited capacity for damage repair and detoxification (Couture and Pyle 2008). In addition, energetic limitations (Sherwood et al. 2002) and inhibition of the p53 stress response pathway by Cd indicate that the most probable evolutionary outcome consists of maximum energy allocation towards growth and sexual maturation (Sibly and Calow 1989). In essence, the proposed adaptive hypothesis for substitution 278 of ND2 would produce a partially functional LTP, thus compensating for incomplete detoxification for which compensation is unviable given its presumably high energetic cost. While this costless mitigation of LTP inhibition would be adaptive by itself, it could also decrease the costs of intraneuronal Cd management, freeing energy to be reallocated towards the fast growth that is characteristic of contaminated perch thus fitting the proposed evolutionary scenario. Finally, and as in the case of cyclin G1, AA 278 in ND2 was not fixed, except in two contaminated populations and one clean population, despite apparently strong directional selection operating in contaminated lakes. The same hypothetical explanations discussed for cyclin G1 can also be invoked here and their confirmation must await further studies.

Implications for conservation biology and impact assessment

Adaptive evolution is intimately linked to genetic diversity (van Straalen and Timmermans 2002) and should consequently be considered in any diversity conservation strategy (Medina et al. 2007; van Straalen and Feder 2012).

Yet, ERAs are still largely derived from surveys of biological response to physiological pollution-induced damage, with only limited genetic diversity inference. For instance, traits such as survival, development and fecundity may be used as fitness indicators for ERA (Medina et al. 2007). Since local adaption to pollution may affect the value of these traits, ERAs based on their assessment may be biased, potentially resulting, for example, in the false conclusion that a population is not affected by pollution. Indeed, fast adaptation to pollution is often accompanied by decreased genetic diversity and adapted individuals may experience negative pleiotropic effects stemming from their new adaptations (Medina et al. 2007). These two mechanisms may reduce the evolutionary potential of affected populations, potentially disappearing as a consequence of future environmental change to which they cannot adapt (van Straalen and Timmermans 2002).

While adaptive evolution should be integrated to ERA, its documentation is not straightforward as it often involves large scale de novo genomic studies. In our view, to ease characterization of fast evolution, one needs to acquire knowledge concerning the sets of underlying conditions and a sense of possible evolutionary outcomes. In essence, an early stage of ERA could evaluate the plausibility of adaptation to pollution by looking at (i) documented evolutionary outcomes in similarly contaminated populations (ii) thresholds of pollution duration and intensity at which adaptation was observed and (iii) classes of candidate genes that may be under selection by different pollutants. This procedure would help to evaluate if microevolution is plausible in a given setting (implying the need for an evolutionary ecotoxicology study) and provide lists of target candidate genes that may have been selected. This information could orient evolutionary characterization towards the most likely selection targets of a given metal. For instance, our results suggest that 85 years of Cd contamination at the bioaccumulation levels observed in this study may have been sufficient to drive evolution for a faster life cycle leading to early sexual maturation as well as the mitigation of Cd inhibition of at least one Ca-dependent pathway. Using this information, a study on a similar system could initially focus on testing selection patterns in genomic areas that harbor energy metabolism, growth regulation and Ca-dependent pathway genes. While admittedly in its infancy, thorough documentation of pollution-driven evolution really has the potential to aid the establishment and success of future integrated ERA studies. For instance, we can readily suggest that, in the Rouyn-Noranda system, an ERA that would base its fitness estimates on growth rates and reproductive success may produce false conclusions concerning the health status of contaminated fish. We can also infer that the fast growth rates that may have evolved in contaminated perch population could be maladaptive in clean environments. This implies a need for genetic monitoring of the return to clean conditions in contaminated lakes since it could increase extinction risk of some populations that lost genetic variation and became adapted to contamination. For instance, analogous considerations for consequences of return to "pristine states" have been given in the context of dam removals for salmonid species in the Pacific Northwest (Angilletta et al. 2008).

Conclusion and perspectives

This study delves somewhat outside of the classes of genes normally studied in ecotoxicology using a discovery method that does not bias initial target gene selection on the basis of previous knowledge. Indeed, research for adaptive polymorphisms in the particular case of pollutant-driven selection has focused on a few classic detoxification genes for a long time. This bias underlines the adaptive importance of certain genes such as cytochrome p450 (Snyder 2000) and metallothionein (Posthuma and van Straalen 1993), but prevents the expansion of knowledge towards a more complete portrait of adaptation to pollution. The approach that we used offers the power to identify new targets of metal driven selection, not only for detoxification, but also for life history and general fitness related traits. This should contribute to the diversification of candidate targets and approaches in evolutionary ecotoxicology.

The adaptive hypotheses that we have proposed regarding the importance of the substitution on AA 204 of cyclin G1 for enhancing growth could be verified through laboratory exposures designed to examine the correlation between cyclin G1 variants, contamination intensity and growth rates. Immunoprecipitation methods could also be used to test the hypothesis that cyclin G1's binding specificity for ARF or other effectors is affected by the candidate substitution.

Similarly, evidence that links the ND2 AA 278 substitution to adaptive LTP inhibition mitigation may also be further tested in the laboratory. Genetic variants could be studied under controlled Cd contamination conditions to investigate hypothetical differences in memorization capacity. Also, neuron cultures may be used to determine the in vivo NMDAR opening thresholds associated with different ND2 genotypes and under various Cd contamination levels. This may be achieved via the measurement of neural action potentials on cultures perfused with Cd concentrations reflecting those of the extracellular brain fluids of Rouyn-Noranda contaminated yellow perch in a fashion similar to Mayer et al. (1989).

Globally, our results suggest that the studied yellow perch populations may have indeed evolved adaptations in order to compensate for the negative impacts of contamination. The

sources of variation that were the targets of metal-driven selection in the Rouyn-Noranda lakes involve a mixed role of standing genetic variation (ND2) and de novo mutations (cyclin G1). The most parsimonious evolutionary strategy to have evolved in contaminated perch populations involves low cost impact mitigation (ND2) and growth increase (cyclin G1) promoting early sexual maturity and thus, successful life cycle completion in a race against reduced survival induced by pollution. While these conclusions might help to refine the understanding of fundamental concepts in field of evolutionary ecotoxicology, they may also serve to establish better ERA, conservation and restoration strategies in similar systems across the world.

Acknowledgments We would like to thank all members of the yellow perch project team from P.C. and L.B. laboratories and, more specifically : J. Grasset, F. Lamaze, E. Normandeau, B. Bougas, F. Pierron, C. Perrier, P.-A. Gagnaire, H. Maaroufi, B. Boyle, 755 S. Larose, J. Laroche, S. Pavey, G. Ouellet-Cauchon, G. Côté, C. Tingaud, D. Ferguson, L. Papillon, as well as M. Dutton (Vale Ltd) for his valuable inputs. We are grateful for the services of IBIS's genomic analysis platform (http://pag.ibis.ulaval.ca/seq/fr/), which was responsible for the 454 sequencing and associated troubleshooting. We also acknowledge the suggestions for improvement received from V. Bourret for this article. Finally, we thank Associate editor Lee Raleigh Shugart and two anonymous referees for their constructive inputs. This work was funded by a Collaborative Research and Development (CRD) Grant (CRDPJ 379611-08) mutually funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada and by Vale Ltd. The Canada Research Chair program supports LB and PGCC.

Ethical standards All experiments comply with current Canadian law for animal research and testing.

Conflict of interest The authors declare no other conflict of interest.

References

Akimoto H (2003) Global air quality and pollution. Science 302:1716–1719

Ali DW, Salter MW (2001) NMDA receptor regulation by Src kinase signalling in excitatory synaptic transmission and plasticity. Curr Opin Neurobiol 11:336–342

Angilletta MJ, Steel EA, Bartz KK, Kingsolver JG, Scheuerell MD, Beckman BR, Crozier LG (2008) Big dams and salmon evolution: changes in thermal regimes and their potential evolutionary consequences. Evol Appl 1:286–299

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. Trends Ecol Evol 23:38–44

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond Ser B Biol Sci 263:1619–1626

Bourret V, Couture P, Campbell PG, Bernatchez L (2008) Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a polymetallic gradient. Aquat Toxicol 86:76–90

Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, Russ C, Lander ES, Nusbaum C, Jaffe DB (2008) Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res 18:763–770

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Campbell PGC, Giguere A, Bonneris E, Hare L (2005) Cadmium-handling strategies in two chronically exposed indigenous freshwater organisms—the yellow perch (*Perca flavescens*) and the floater mollusc (Pyganodon grandis). Aquat Toxicol 72:83–97

Cattaneo A, Couillard Y, Wunsam S (2008) Sedimentary diatoms along a temporal and spatial gradient of metal contamination. J Paleolimnol 40:115–127

Chapman PM (2002) Integrating toxicology and ecology: putting the "eco" into ecotoxicology. Marine Pollut Bull 44:7–15

Chin D, Means AR (2000) Calmodulin: a prototypical calcium sensor. Trends Cell Biol 10:322–328

Cohen P (2000) The regulation of protein function by multisite phosphorylation—a 25 year update. Trends Biochem Sci 25:596–601

Couillard Y, Courcelles M, Cattaneo A, Wunsam S (2004) A test of the integrity of metal records in sediment cores based on the documented history of metal contamination in Lac Dufault (Quebec, Canada). J Paleolimnol 32:149–162

Couture P, Pyle G (2008) Live fast and die young: metal effects on condition and physiology of wild yellow perch from along two metal contamination gradients. Hum Ecol Risk Assess Int J 14:73–96

Couture P, Rajotte JW, Pyle GG (2008a) Seasonal and regional variations in metal contamination and condition indicators in yellow perch (*Perca flavescens*) along two polymetallic gradients. III. Energetic and physiological indicators. Hum Ecol Risk Assess 14:146–165

Couture P, Busby P, Gauthier C, Rajotte JW, Pyle GG (2008b) Seasonal and regional variations of metal contamination and condition indicators in yellow perch (*Perca flavescens*) along two polymetallic gradients. I. Factors influencing tissue metal concentrations. Hum Ecol Risk Assess 14:97–125

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Felsenstein J (1976) Theoretical population-genetics of variable selection and migration. Annu Rev Genet 10:253–280

Flicek P et al (2012) Ensembl 2012. Nucleic Acids Res 40:D84–D90

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian Perspective. Genetics 180:977–993

Giguere A, Campbell PGC, Hare L, McDonald DG, Rasmussen JB (2004) Influence of lake chemistry and fish age on cadmium, copper, and zinc concentrations in various organs of indigenous yellow perch (*Perca flavescens*). Can J Fish Aquat Sci 61:1702–1716

Gingrich JR, Pelkey KA, Fam SR, Huang YQ, Petralia RS, Wenthold RJ, Salter MW (2004) Unique domain anchoring of Src to synaptic NMDA receptors via the mitochondrial protein NADH dehydrogenase subunit 2. Proc Natl Acad Sci USA 101: 6237–6242

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg Proportion for Multiple Alleles. Biometrics 48: 361–372

Harris SL, Levine AJ (2005) The p53 pathway: positive and negative feedback loops. Oncogene 24:2899–2908

Hendry AP, Kinnison MT (2001) An introduction to microevolution: rate, pattern, process. Genetica 112–113:1–8

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

Jeffreys HS (1961) Theory of probability/by Harold Jeffreys. Clarendon Press, Oxford

Joost S, Bonin A, Bruford MW, Despres L, Conord C, Erhardt G, Taberlet P (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol 16:3955–3969

Kandel ER (2001) The molecular biology of memory storage: a dialogue between genes and synapses. Science 294:1030–1038

Kim MS, Kim BJ, Woo HN, Kim KW, Kim KB, Kim IK, Jung YK (2000) Cadmium induces caspase-mediated cell death: suppression by Bcl-2. Toxicology 145:27–37

Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM (2003) A "Frankenstein's monster" approach to comparative modeling: merging the finest fragments of fold-recognition models and iterative model refinement aided by 3D structure evaluation. Proteins Struct Funct Bioinform 53:369–379

Kovecses J, Sherwood GD, Rasmussen JB (2005) Impacts of altered benthic invertebrate communities on the feeding ecology of yellow perch (*Perca flavescens*) in metal-contaminated lakes. Can J Fish Aquat Sci 62:153–162

Kurowski MA, Bujnicki JM (2003) GeneSilico protein structure prediction meta-server. Nucleic Acids Res 31:3305–3307

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

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

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30: 434–439

Lovell SC, Davis IW, Adrendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C alpha geometry: phi, psi and C beta deviation. Proteins Struct Funct Genet 50:437–450

Margulies M et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Mayer ML, Vyklicky L, Westbrook GL (1989) Modulation of excitatory amino-acid receptors by group IIB metal-cations in cultured mouse *Hippocampal*-neurons. J Physiol Lond 415:329–350

Medina MH, Correa JA, Barata C (2007) Micro-evolution due to pollution: possible consequences for ecosystem responses to toxic stress. Chemosphere 67:2105–2114

Meplan C, Mann K, Hainaut P (1999) Cadmium induces conformational modifications of wild-type p53 and suppresses p53 response to DNA damage in cultured cells. J Biol Chem 274:31663–31670

Nakazawa K, McHugh TJ, Wilson MA, Tonegawa S (2004) NMDA receptors, place cells and hippocampal spatial memory. Nat Rev Neurosci 5:361–372

Ohtsuka T, Ryu H, Minamishima YA, Ryo A, Lee SW (2003) Modulation of p53 and p73 levels by cyclin G: implication of a negative feedback regulation. Oncogene 22:1678–1687

Ohtsuka T, Jensen MR, Kim HG, Kim KT, Lee SW (2004) The negative role of cyclin G in ATM-dependent p53 activation. Oncogene 23:5405–5408

Okamoto K, Li H, Jensen MR, Zhang T, Taya Y, Thorgeirsson SS, Prives C (2002) Cyclin G recruits PP2A to dephosphorylate Mdm2. Mol Cell 9:761–771

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2:2074–2093

Perceval O, Couillard Y, Pinel-Alloul B, Campbell PGC (2006) Linking changes in subcellular cadmium distribution to growth and mortality rates in transplanted freshwater bivalves (Pygan-odon grandis). Aquat Toxicol 79:87–98

Pierron F, Bourret V, St-Cyr J, Campbell PG, Bernatchez L, Couture P (2009) Transcriptional responses to environmental metal exposure in wild yellow perch (*Perca flavescens*) collected in lakes with differing environmental metal concentrations (Cd, Cu, Ni). Ecotoxicology 18:620–631

Pierron F, Couture P, Normandeau E, Defo MA, Campbell PGC, Bernatchez L (2011) Effects of chronic metal exposure on wild fish populations revealed by high-throughput cDNA sequencing. Ecotoxicology 20:1388–1399

Piscopo DM, Hinds PW (2008) A role for the cyclin box in the ubiquitin-mediated degradation of cyclin G1. Cancer Res 68:5581–5590

Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao XB, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res 38:D105–D110

Posthuma L, van Straalen NM (1993) Heavy-metal adaptation in terrestrial invertebrates—a review of occurrence, genetics, physiology and ecological consequences. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol 106:11–38

Pra D, Franke SIR, Giulian R, Yoneama ML, Dias JF, Erdtmann B, Henriques JAP (2008) Genotoxicity and mutagenicity of iron and copper in mice. Biometals 21:289–297

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Pruitt KD, Tatusova T, Klimke W, Maglott DR (2009) NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res 37:D32–D36

Pyle G, Busby P, Gauthier C, Rajotte J, Couture P (2008) Seasonal and regional variations in metal contamination and condition indicators in yellow perch (*Perca flavescens*) along two polymetallic gradients. II. Growth patterns, longevity, and condition. Hum Ecol Risk Assess 14:126–145

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Reich D et al (2008) Long-range LD can confound genome scans in admixed populations. Am J Hum Genet 83:132–135

Reznick DN, Ghalambor CK (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. Genetica 112–113:183–198

Rokas A, Abbot P (2009) Harnessing genomics for evolutionary insights. Trends Ecol Evol 24:192–200

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Sandelin A, Wasserman WW, Lenhard B (2004) ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res 32:W249–W252

Sauvage C, Vagner M, Derome N, Audet C, Bernatchez L (2012) Coding gene single nucleotide polymorphism mapping and quantitative trait loci detection for physiological reproductive traits in Brook Charr, *Salvelinus fontinalis*. G3 (Bethesda) 2:379–392

Schiller J, Schiller Y, Clapham DE (1998) NMDA receptors amplify calcium influx into dendritic spines during associative pre- and post-synaptic activation. Nat Neurosci 1:114–118

Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 1.3

Schwerdtle T, Ebert F, Thuy C, Richter C, Mullenders LHF, Hartwig A (2010) Genotoxicity of soluble and particulate cadmium compounds: impact on oxidative DNA damage and nucleotide excision repair. Chem Res Toxicol 23:432–442

Sepulveda-Villet OJ, Ford AM, Williams JD, Stepien CA (2009) Population genetic diversity and phylogeographic divergence patterns of the yellow perch (*Perca flavescens*). J Great Lake Res 35:107–119

Shah J, Pant HC (1991) Effect of cadmium on Ca^{2+} transport in brain microsomes. Brain Res 566:127–130

Sherwood GD, Kovecses J, Hontela A, Rasmussen JB (2002) Simplified food webs lead to energetic bottlenecks in polluted lakes. Can J Fish Aquat Sci 59:1–5

Shimizu E, Tang YP, Rampon C, Tsien JZ (2000) NMDA receptor-dependent synaptic reinforcement as a crucial process for memory consolidation. Science 290:1170–1174

Sibly RM, Calow P (1989) A life-cycle theory of responses to stress. Biol J Linn Soc 37:101–116

Sloss BL, Billington N, Burr BM (2004) A molecular phylogeny of the Percidae (Teleostei, Perciformes) based on mitochondrial DNA sequence. Mol Phylogenet Evol 32:545–562

Smith TB, Bernatchez L (2008) Evolutionary change in human-altered environments. Mol Ecol 17:1–8

Snyder MJ (2000) Cytochrome P450 enzymes in aquatic invertebrates: recent advances and future directions. Aquat Toxicol 48:529–547

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tobler M, Culumber ZW, Plath M, Winemiller KO, Rosenthal GG (2010) An indigenous religious ritual selects for resistance to a toxicant in a livebearing fish. Biol Lett 7(2):229–232

van Straalen NM, Feder ME (2012) Ecological and evolutionary functional genomics—how can it contribute to the risk assessment of chemicals? Environ Sci Technol 46:3–9

van Straalen NM, Timmermans MJTN (2002) Genetic variation in toxicant-stressed populations: an evaluation of the "genetic erosion" hypothesis. Hum Ecol Risk Assess 8:983–1002

Vig PJS, Bhatia M, Gill KD, Nath R (1989) Cadmium inhibits brain calmodulin—invitro and invivo studies. Bull Environ Contam Toxicol 43:541–547

Vig PJS, Ravi K, Nath R (1991) Interaction of metals with brain calmodulin purified from normal and cadmium exposed rats. Drug Chem Toxicol 14:207–218

Vitousek PM, Mooney HA, Lubchenco J, Melillo JM (1997) Human domination of Earth's ecosystems. Science 277:494–499

Williams LM, Oleksiak MF (2011) Evolutionary and functional analyses of cytochrome P4501A promoter polymorphisms in natural populations. Mol Ecol 20:5236–5247

Wimley WC, Creamer TP, White SH (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry 35:5109–5124

Zhao L, Samuels T, Winckler S, Korgaonkar C, Tompkins V, Horne MC, Quelle DE (2003) Cyclin G1 has growth inhibitory activity linked to the ARF-Mdm2-p53 and pRb tumor suppressor pathways. Mol Cancer Res 1:195–206