MOLECULAR ECOLOGY

Molecular Ecology (2015) doi: 10.1111/mec.13466

SPECIAL ISSUE: DETECTING SELECTION IN NATURAL POPULATIONS: MAKING SENSE OF GENOME SCANS AND TOWARDS ALTERNATIVE SOLUTIONS

RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels

M. LAPORTE,*[1] S. A. PAVEY,*†‡§¶[1] C. ROUGEUX,* F. PIERRON,†‡ M. LAUZENT,†‡ H. BUDZINSKI,†‡ P. LABADIE,†‡ E. GENESTE,†‡ P. COUTURE,§ M. BAUDRIMONT†‡ and L. BERNATCHEZ*

*Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Pavillon Charles-Eugène-Marchand, Québec, QC G1V 0A6, Canada, †UMR EPOC CNRS 5805, Université de Bordeaux, allée Geoffroy Saint-Hilaire, CS 50023, 33615 Pessac, France, ‡CNRS, EPOC, UMR 5805, allée Geoffroy Saint-Hilaire, CS 50023, 33615 Pessac, France, §Institut National de la Recherche Scientifique (INRS), Centre Eau Terre et Environnement, 490, rue de la Couronne, Québec, QC G1K 9A9, Canada, ¶Department of Biology, University of New Brunswick, PO Box 5050, Saint-John, NB E2L 4L5, Canada

Abstract

Measuring the effects of selection on the genome imposed by human-altered environment is currently a major goal in ecological genomics. Given the polygenic basis of most phenotypic traits, quantitative genetic theory predicts that selection is expected to cause subtle allelic changes among covarying loci rather than pronounced changes at few loci of large effects. The goal of this study was to test for the occurrence of polygenic selection in both North Atlantic eels (European Eel, *Anguilla anguilla* and American Eel, *A. rostrata*), using a method that searches for covariation among loci that would discriminate eels from 'control' vs. 'polluted' environments and be associated with specific contaminants acting as putative selective agents. RAD-seq libraries resulted in 23 659 and 14 755 filtered loci for the European and American Eels, respectively. A total of 142 and 141 covarying markers discriminating European and American Eels from 'control' vs. 'polluted' sampling localities were obtained using the Random Forest algorithm. Distance-based redundancy analyses (db-RDAs) were used to assess the relationships between these covarying markers and concentration of 34 contaminants measured for each individual eel. PCB153, 4′4′DDE and selenium were associated with covarying markers for both species, thus pointing to these contaminants as major selective agents in contaminated sites. Gene enrichment analyses suggested that sterol regulation plays an important role in the differential survival of eels in 'polluted' environment. This study illustrates the power of combining methods for detecting signals of polygenic selection and for associating variation of markers with putative selective agents in studies aiming at documenting the dynamics of selection at the genomic level and particularly so in human-altered environments.

Keywords: distance-based redundancy analysis, landscape genomics, polygenic selection, RAD sequencing, Random Forest algorithm

Received 18 May 2015; revision accepted 6 November 2015

Correspondence: Martin Laporte, Fax: +1 418 656-7176; E-mail: uni.mlaporte@gmail.com and Scott A. Pavey, Fax: +1 506-648-5811; E-mail: scottapavey@gmail.com

[1]These two first co-authors contributed equally to the manuscript

Introduction

Anthropogenic activities affect all ecosystems to some extent (Vitousek *et al.* 1997), and it is now widely accepted that they may impact evolutionary processes

(Smith & Bernatchez 2008). Water pollution in particular is one of the five major causes of the alarming decline of freshwater biodiversity (Dudgeon *et al.* 2006). It can impact organismal abundance (Ribeiro & Lopes 2013; Laporte *et al.* 2014; Mussali-Galante *et al.* 2014), modify behaviour (Weis & Candelmo 2012), alter gene transcription (Bozinovic & Oleksiak 2011; Pierron *et al.* 2009, 2011; Mehinto *et al.* 2012; Bougas *et al.* 2013; Sutherland *et al.* 2014; Rodriguez-Jorquera *et al.* 2015), induce mutations, damage DNA leading to carcinogenesis (Belfiore & Anderson 2001; Tabrez *et al.* 2014) and impose strong selection that can be detected at the genome level (Bélanger-Deschênes *et al.* 2013; Mussali-Galante *et al.* 2014). Because contaminations of aquatic ecosystems typically include many pollutants that present different mechanisms of action and toxic effects on organisms (Matthiessen *et al.* 1988), a complex selective response at multiple genes is expected. Therefore, documenting the effect of selection imposed by human-driven environment alteration is crucial to better understand the impact of such pollution on wild populations.

Given the polygenic basis of most phenotypic traits, quantitative genetic theory predicts that selection will cause subtle allelic changes among covarying loci more than pronounced changes at few loci of large effects (McKay & Latta 2002; Pritchard *et al.* 2010; Le Corre & Kremer 2012; Messer & Petrov 2013; Bourret *et al.* 2014). Searching for such polygenic selection should thus be prioritized when complex selective response at multiple genes is expected. However, standard approaches to detect selection (*i.e.* genome scans) are largely based on the classical hitchhiking model (Maynard Smith & Haigh 1974). This infers the process of 'selective sweeps' whereby a new advantageous mutation spreads to fixation, increasing linkage disequilibrium of the locus with neighbouring sites (Pritchard *et al.* 2010; Messer & Petrov 2013). Therefore, these approaches are best suited to detect individual loci showing strong allele frequency differentiation in comparison with all sampled loci. Because these approaches do not search for covariation among loci, they may not be able to detect complex and recent selective response at multiple loci (Pritchard *et al.* 2010; Messer & Petrov 2013; Kemper *et al.* 2014; Haasl & Payseur 2016).

The two North Atlantic eel species, the European Eel (*Anguilla anguilla*) and the American Eel (*A. rostrata*), have a catadromous life cycle. After spawning in the Sargasso Sea, the larvae are transported by currents to the Atlantic coasts of Europe and North America (Scott & Crossman 1973; Kottelat & Freyhof 2007). Estimates of population structure based on neutral markers revealed a total absence of population structure within both species, thus indicating two cases of panmixia (Als *et al.* 2011; Côté *et al.* 2013). However, spatially varying selection (i.e. within-generation signatures of local genetic differentiation caused by disparity of *in situ* mortalities) has been documented for both species (Gagnaire *et al.* 2012; Côté *et al.* 2014; Ulrik *et al.* 2014; Pavey *et al.* 2015), making them good models to test the efficiency of method searching for covarying loci to detect the effects of polygenic selection in comparison with standard genome scan approaches.

European and American Eels are, respectively, considered 'critically endangered' and 'endangered' by the IUCN Red List of Threatened Species (IUCN 2014). Recent studies showed that exposure to organic compounds and metals can alter their DNA methylation and gene expression (Pierron *et al.* 2014; Baillon *et al.* 2015) and reduce their survival (Couillard *et al.* 1997). Moreover, fat burning releasing stored contaminants during migration to Sargasso Sea most likely increase toxicity and mortality before reproduction (Van Ginneken & Van den Thillart 2000; Pierron *et al.* 2008; Geerarts & Belpaire 2010). Yet, the impact of water pollution as a within-generation selective agent is poorly understood. Because North Atlantic Eels are panmictic, local genetic drift and mutagenesis can be excluded as potential processes differentiating loci among sampling locations. Any relationship between genetic variation and contamination should thus be considered as a selective process occurring within a generation, once accounted for stochastic effects (e.g. sampling errors).

The main objective of this study was to test for the occurrence of polygenic selection in response to environmental pollution in both North Atlantic Eels, using a method searching for covarying loci able to discriminate eels between 'control' and 'polluted' environments (Random Forest algorithm). A relationship between individual genotypes of the covarying loci previously found and individual contaminant concentrations was subsequently assessed using the distance-based redundancy analysis (db-RDA). Such relationship is expected if local genetic differentiation is produced by contamination leading to differential mortality among localities. When possible, putative function of these genes was also examined and compared among species to get insight into the molecular mechanisms underlying resistance to pollutants. Finally, traditional genome scan (i.e. F_{ST} outlier analysis) approach was performed and compared to the analysis of polygenic selection performed with the Random Forest algorithm.

Material and methods

Sampling and genotyping

A total of 90 European Eels were sampled at the yellow eel stage (somatic growth phase) in France, in localities

along the Gironde aquatic system and in the Arcachon Bay (Table 1; Fig. 1). In Québec (Canada), a total of 89 yellow eels of American Eel at sampling sites located in the St Lawrence River drainage (Table 1; Fig. 1). For both species, two sampling sites were classified as 'control' (Certes and Dordogne in France; St Jean and Sud Ouest in Quebec) and two others as 'polluted' (Garonne and Gironde in France; St Pierre and St François in Quebec) according to available environmental pollutants data (Lee *et al.* 1999; Durrieu *et al.* 2005; Pierron *et al.* 2008; Tapie *et al.* 2011). All fish were collected between May and June of 2011 using trawl, fyke net or electrofishing.

DNA was isolated from each individual sample using the salt extraction method (Aljanabi & Martinez 1997) followed by RNase treatment (QIAGEN, Valencia, CA, USA). Quality of extracted DNA was verified with an agarose gel. RAD sequencing libraries were prepared as in Pavey *et al.* (2015). Briefly, extracted DNA samples were quantified with PicoGreen, fragmented by sonication (Diagenode, BioRuptor Sonication System) and digested with the *Eco*RI enzyme. The samples were labelled with individual barcodes and pooled into groups of 24 individuals.

RAD-seq pools were single-end-sequenced with 100-bp reads on eight HISeq Illumina lanes producing around 1.8 billion sequences in total. The STACKS pipeline (version 1.21) was used to generate genotypes for each species separately (Catchen *et al.* 2013). After the process RADtags step, all sequences were trimmed to 80 bp with FASTX toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/) to remove the 5′ end of the reads where quality tends to decrease. The draft genome of the American Eel (S.A. Pavey, E. Normandeau, J. Gaudin, L. Letourneau, L. Bernatchez, unpublished data) was used to align the reads for both study species using BWA (Li & Durbin 2010) with a maximum mismatch of 5. The minimum stack depth was specified as 5× coverage per fish locus. In the 'rx' step of STACKS, we used a bounded SNP model with α set to 0.1, the ε upper bound set to 0.1 and log likelihood set to −10. Following this pipeline, additional filtering was manually performed. SNPs with an F_{IS} value below −0.3 in any one sampling locality were discarded in an attempt to remove paralogs (typically causing heterozygous excess). Only SNPs genotyped in at least 16 individuals per sampling locality were retained. Additionally, only SNPs with a minor allele frequency of 0.02 over all individuals in each species were retained with vcftools in order to minimize the inclusion of false SNPs caused by sequencing errors (Danecek *et al.* 2011). Only the first SNP in each stack was kept for subsequent analyses as other SNP within a same read would obviously be in strong linkage disequilibrium.

Table 1 Sample characteristics: location, groups (control or polluted), geographic coordinates and number of eels (*N*) for both North Atlantic Eels. Levels of pollutants are provided in Tables 2 and 3

Location	Geographic coordinates	*N*
European Eel		
Control		
Certes	44°41′18″N – 01°01′39″W	22
Dordogne	44°48′05″N – 00°08′25″E	24
Polluted		
Gironde	45°12′07″N – 00°43′35″W	23
Garonne	44°43′51″N – 00°28′05″E	21
American Eel		
Control		
Saint-Jean	45°51′40″N – 64°28′47″W	21
Sud Ouest	48°22′27″N – 68°43′02″W	24
Polluted		
Saint-Pierre	45°09′18″N – 74°23′04″W	23
Saint-François	46°19′74″N – 74°32′30″W	21

Genotype distance matrix of covarying markers

For each species, we first searched for a group of covarying and/or interacting markers discriminating eels from 'control' and 'polluted' localities, using the *randomForest* function implemented in the 'randomForest' R package (Liam & Wiener 2002), with a total of 10 000 trees. The Random Forest algorithm is a tree-based ensemble machine learning tool that is well suited for 'large number of indicators and small sample size' problems (Goldstein *et al.* 2011; Chen & Ishwaran 2012). A clear and concise diagram explaining Random Forest algorithm is available in Boulesteix *et al.* (2012). This approach is well suited for genomic applications because these data contained large numbers of loci compared with the number of individuals. Random Forest also accounts for correlation and interactions among loci (Chen & Ishwaran 2012; Boulesteix *et al.* 2014), making it suitable to search for signals of polygenic selection. This approach is currently popular in the medicine and agriculture (Shi *et al.* 2005; Cordell 2009; Tang *et al.* 2009; Xu *et al.* 2011; Poland *et al.* 2012; Jarquín *et al.* 2014), but still infrequent in molecular ecology studies (but see Holliday *et al.* 2012; Brieuc *et al.* 2015; Laporte *et al.* 2015; Pavey *et al.* 2015). Given the total absence of population structuring in both species and the similar sample size for each site, we used the *na.roughfix* function implemented in the 'randomForest' R package (Liam & Wiener 2002) to fill missing data, a step required by the Random Forest algorithm. To select a set of covarying markers which we subsequently refer to as 'important markers' (Goldstein *et al.* 2011; Chen & Ishwaran 2012), we used the

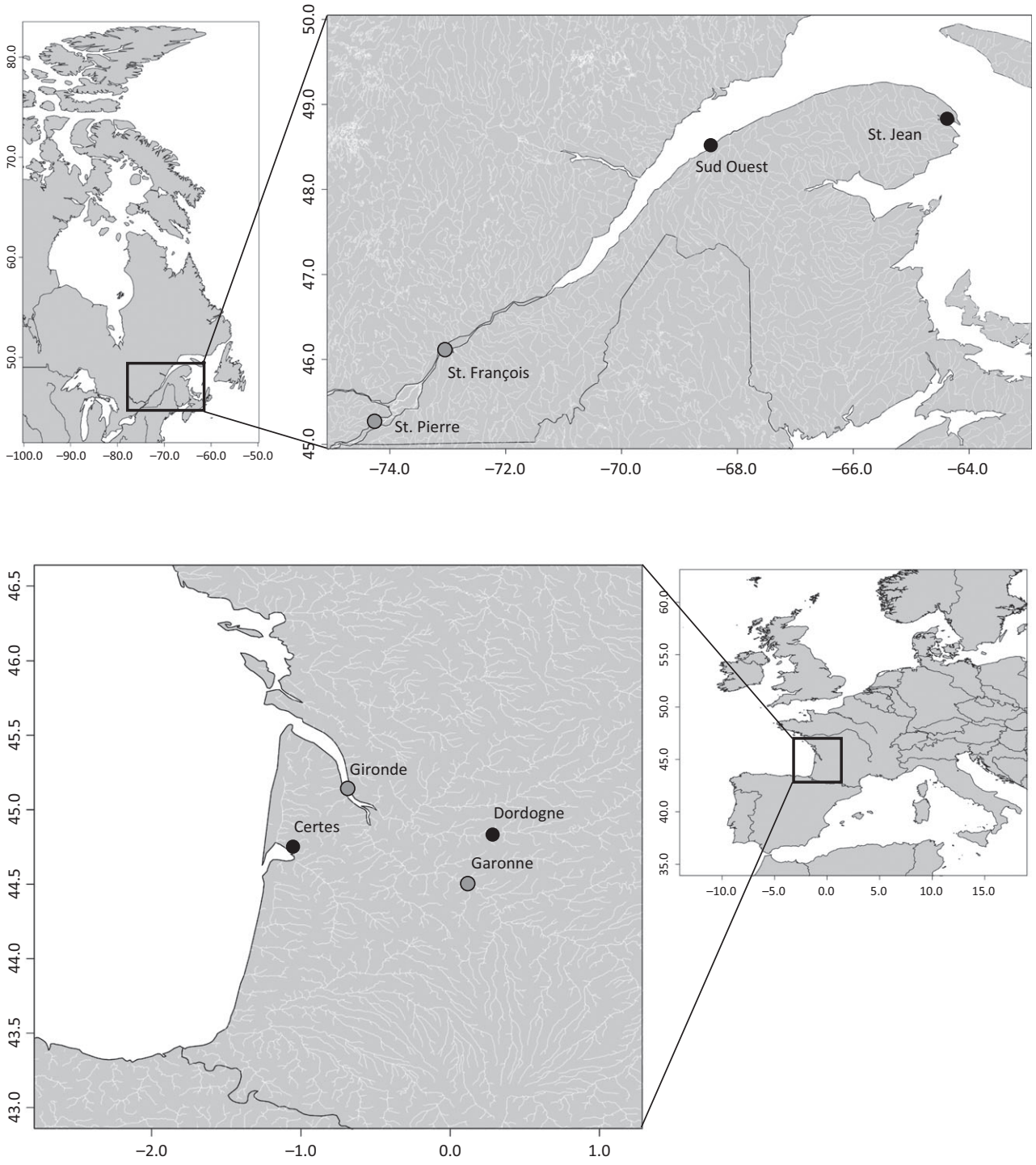

Fig. 1 Map of the American Eels (top) and European Eels (down) sampling locations. Black and grey circles represent, respectively, 'control' and 'polluted' sites.

'permuted importance' statistic (*i.e.* an indicator of how a marker in interaction with other markers will successfully classify an individual with respect to the response variable). As explained in Goldstein *et al.* (2011), Random Forest was rerun on markers having importance higher or equal to zero in the precedent run until we reached an 'out-of-bag error rate' (OOB-ER) under 5%. We chose this threshold because markers with importance under zero are less useful than the half of all markers to discriminate eels from

'control' vs. 'polluted' sites. Out-of-bag is a random subsample (33% of the sample) that is used to assess the classification success of a tree (Boulesteix *et al.* 2012). More precisely, the sample is divided in two groups, one to construct the tree (the training set) and another one to test it (the test set). The division of the sample differs for each tree (Boulesteix *et al.* 2012). The lower the OOB-ER, the higher the rank reliability of important markers is (Goldstein *et al.* 2011). As mentioned by Goldstein *et al.* (2011), there is no objective method to set the importance threshold between important markers and not important ones, and thus, we examined the distribution of importance of markers and used the upper end of the elbow as the cut-off to determine the important markers for subsequent analyses (Fig. 2).

Genotypes of important markers were used to produce a coefficient of similarity matrix among individuals. The similarity matrix was produced with a modified version of the multistate simple matching coefficient (S1) (Legendre & Legendre 1998). We computed similarity between two individuals as follows: (i) for a given locus, a score of '1', '0.5' or '0' was given if two individuals share both alleles, one or no allele, respectively, (ii) scores of all loci were added and (iii) then divided by the total number of loci for both individuals. The similarity matrix (S) was thereafter transformed into a distance matrix (D) using the formula $D = \sqrt{(1-S)}$ for maintaining Euclidian properties (Legendre & Legendre 1998). The distance matrix was used to perform a principal coordinate analysis (PCoA). PCoA is a variant of principal component analysis (PCA), which can use any distance matrix unlike PCA which requires a correlation or a covariance matrix. As a surrogate for multilocus genotypes of important markers, we used principal coordinates factors (PCo factors) showing eigenvalue higher than the broken-stick distribution (Legendre & Legendre 1998). The PCoA was produced and the broken-stick distribution was obtained with the function *pcoa* available in the APE package in R software v3.1.3.

Pollutant contamination matrix

Analysis of the seven indicator PCBs (polychlorinated biphenyl) (PCB50 + 28, PCB52, PCB101, PCB118, PCB138, PCB153 and PCB180), eight OCPs (organochlorine pesticide) (HCB, lindane, 2,4′DDE, 4,4′DDE, 2,4′DDD + (PCB154 + 77), 4,4′DDD, 2,4′DDT and 4,4′DDT) and nine PBDEs (polybrominated diphenyl ether) (PBDE28, PBDE47, PBDE49, PBDE99, PBDE100, PBDE153, PBDE154, PBPE183 and PBDE209) was performed on muscle samples, and analysis of ten metals (Ag, As, Cd, Cr, Cu, Ni, Pb, Se, Zn and Hg) was performed on liver. Samples of white muscle were dissected in a standardized area situated at 2 cm posterior to the anus and above the lateral line. PCBs, PBDEs and OCPs analyses were carried out on an HP 5890 series II gas chromatograph from Hewlett-Packard (Avondale, CA, USA) coupled to a 63Ni electron capture detector (ECD). A capillary column HP5-MS (Agilent Technologies, Massy, France) was used (30 m × 0.25 mm × 0.25 lm). Metal concentrations were measured by inductively coupled plasma-mass spectrometry (ICP-MS; Thermo Scientific XSeries 2), inductively coupled plasma-atomic emission spectrometry (ICP-AES; VarianVista AX) or both methods. Further methodological details are provided in Baillon *et al.* (2015). Average concentrations by locality for all pollutants are presented in Tables 2 and 3.

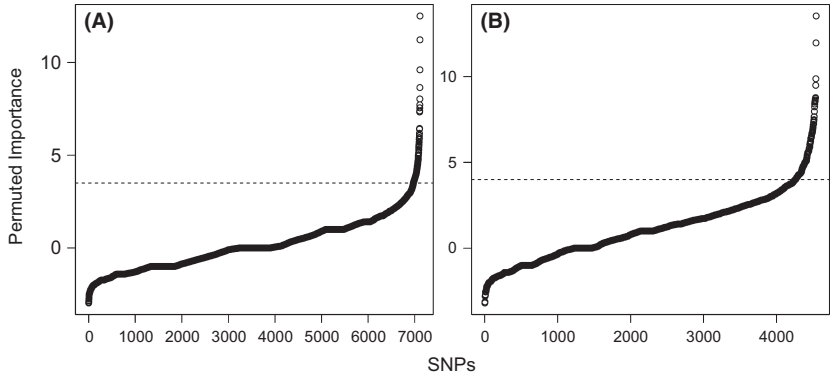

Fig. 2 Separation between important markers (covarying SNPs) and nonimportant markers using the end of the elbow of the permuted importance distribution as a cut-off (dashed line). The permuted importance estimation was obtained via Random Forest analyses (A for the European Eel and B for the American Eel). The number of SNPs (7083 and 4537, respectively, for the European and American eels) corresponds to the total SNPs with permuted importance equal or higher to 0 of the previous run of Random Forest (Random Forest algorithm was rerun until out-of-bag error reached a percentage under 5%). See Materials and Methods for details.

Table 2 Average concentration of 24 organic pollutants in muscle (ng/g, dry wet) and 9 metal pollutants in liver (μg/g, dry wet) of individual European Eels. The columns 'Control' and 'Polluted' correspond to the average (standard deviation) of all fish from the corresponding habitat category

Contaminants		Certes	Dordogne	Control	Gironde	Garonne	Polluted
Organic	HCB	2.07	2.41	2.24 (1.36)	3.22	2.74	2.99 (1.82)
	Lindane	0.62	9.87	5.44 (9.55)	0.47	2.63	1.50 (2.17)
	2,4′DDE	0.26	2.00	1.17 (1.36)	4.12	5.61	4.83 (4.76)
	4,4′DDE	3.26	45.45	25.27 (29.17)	114.84	108.08	111.61 ((60.73)
	2,4′DDD + (CB154 + 77)	0.07	0.71	0.40 (0.76)	1.10	0.21	0.67 (1.12)
	4,4′DDD	0.82	7.92	4.53 (6.18)	24.65	17.23	21.11 (12.54)
	2,4′DDT	0.08	0.10	0.09 (0.06)	0.08	0.13	0.10 (0.07)
	4,4′DDT	0.15	2.45	1.35 (1.59)	4.28	5.65	4.94 (3.02)
	PBDE 28	0.05	0.05	0.05 (0.01)	0.09	0.05	0.08 (0.11)
	PBDE 47	0.08	3.79	2.02 (2.63)	5.49	8.16	6.76 (4.64)
	PBDE 49	0.07	0.27	0.17 (0.20)	0.38	0.44	0.41 (0.31)
	PBDE 99	0.09	0.10	0.10 (0.04)	0.14	0.23	0.18 (0.17)
	PBDE 100	0.08	2.10	1.13	2.99	4.66	3.79 (2.18)
	PBDE 153	0.09	0.11	0.10 (0.04)	0.13	0.23	0.18 (0.14)
	PBDE 154	0.08	0.22	0.15 (0.16)	0.33	0.47	0.40 (0.26)
	PBDE 183	0.12	0.12	0.12 (0.00)	0.12	0.12	0.12 (0.00)
	PBDE 209	6.50	8.99	7.80 (6.53)	6.50	7.63	7.04 (3.57)
	PCB 50 + 28	0.74	2.14	1.47 (1.14)	9.01	8.25	8.65 (7.84)
	PCB 52	1.60	8.98	5.45 (5.62)	28.61	28.35	28.48 (27.64)
	PCB 101	2.03	18.64	10.70 (12.72)	49.35	54.38	51.75 (62.59)
	PCB 118	6.51	29.38	18.44 (19.44)	64.22	66.68	65.39 (35.92)
	PCB 153	20.41	283.44	157.64 (248.09)	521.71	307.72	419.58 (322.97)
	PCB 138	15.22	131.41	75.84 (91.51)	306.56	239.06	274.35 (237.28)
	PCB 180	6.44	71.87	40.58 (49.55)	240.80	160.00	202.34 (218.93)
Metal	Ag (silver)	0.27	0.35	0.31 (0.21)	1.14	0.39	0.78 (0.77)
	As (arsenic)	7.09	1.84	4.47 (3.46)	5.33	1.22	3.92 (3.36)
	Cd (cadmium)	0.36	3.62	1.99 (2.36)	6.56	2.62	4.68 (4.40)
	Cr (chrome)	3.14	1.67	2.40 (3.81)	0.74	0.94	1.62 (0.52)
	Cu (copper)	67.02	51.57	59.30 (24.05)	84.37	43.29	62.25 (39.29)
	Ni (nickel)	0.68	0.34	0.51 (0.68)	0.23	0.39	0.31 (0.46)
	Pb (lead)	0.21	0.53	0.37 (0.29)	1.06	0.62	0.85 (0.70)
	Se (selenium)	13.67	25.55	19.61 (10.22)	37.14	23.87	25.22 (16.79)
	Zn (zinc)	185.19	170.72	177.95 (54.88)	193.52	145.27	171.08 (71.60)
	Hg (mercury)	0.18	0.79	0.36 (0.37)	0.97	0.52	0.63 (0.55)

For each species, we performed a PCA to minimize colinearity among contaminant indicators, using a covariance matrix to preserve information on their variation (Legendre & Legendre 1998; Bourret *et al.* 2014). Because measurements were produced on two different organs (muscles for organic contaminants and liver for metals), we separated the organic contaminants and the metals to conserve the 'dimension homogeneity' of the variables. Indeed, the muscles and the liver could accumulate the contaminants differently (Türkmen *et al.* 2013). Therefore, a PCA for each organ was performed to test independently for relationship between bioaccumulation levels (organic or metal) and multilocus genotypes. In addition, only principal component factors (PC factor) with eigenvalues >1 (Kaiser–Guttman criterion) (Yeomans & Golder 1982) were kept to express multivariate composites of the 34 pollutants.

Genotype–pollutants associations

In order to test whether the important SNPs found between localities with different levels of pollution are related to the level of contamination of specific pollutants, we assessed the relationship between the important SNPs multilocus genotypes and contaminant concentrations measured for each individual. This was performed via a distance-based redundancy analysis (db-RDA), a variant of canonical correlation analysis (CCA) (Legendre & Legendre 1998; Legendre & Gallagher 2001). db-RDA has the advantage to give less weight to rare variables occurring in the data set (in our case, rare alleles that could be caused by genotyping error) than CCA, because of the characteristics of the chi-square distance used in the CCA (see Legendre & Gallagher 2001 for further precisions). Here, using individuals as subjects, we tested whether

Table 3 Average concentration of 24 organic pollutants in muscle (ng/g, dry wet) and 9 heavy metal pollutants in liver (μg/g, dry wet) in individual American Eel. The columns 'Control' and 'Polluted' correspond to the average (standard deviation) of all fish from the corresponding habitat category

Contaminants		St Jean	Sud Ouest	Control	St Pierre	St François	Polluted
Organic	HCB	1.87	0.77	1.38 (1.97)	2.29	2.31	2.30 (1.12)
	Lindane	0.16	0.19	0.17 (0.15)	0.21	0.19	0.20 (0.19)
	2,4′DDE	0.43	0.22	0.34 (0.55)	1.73	1.59	1.65 (1.78)
	4,4′DDE	11.10	6.86	9.21 (7.77)	52.98	29.32	40.61 (35.46)
	2,4′DDD + (CB154 + 77)	0.07	0.07	0.07 (0.00)	0.07	0.07	0.07 (0.00)
	4,4′DDD	1.11	0.43	0.80 (0.59)	6.28	5.07	5.65 (4.08)
	2,4′DDT	0.08	0.08	0.08 (0.00)	0.09	0.08	0.08 (0.04)
	4,4′DDT	1.50	0.42	1.02 (0.82)	2.61	0.75	1.64 (2.30)
	PBDE 28	0.09	1.43	0.68 (1.91)	0.36	0.11	0.23 (0.55)
	PBDE 47	1.95	0.86	1.47 (1.17)	28.90	9.94	18.99 (18.20)
	PBDE 49	0.24	0.07	0.16 (0.17)	2.54	1.43	1.96 (1.73)
	PBDE 99	0.09	0.09	0.09 (0.00)	1.06	0.44	0.75 (0.56)
	PBDE 100	0.47	0.93	0.68 (0.52)	7.47	2.87	5.06 (4.59)
	PBDE 153	0.09	0.09	0.09 (0.00)	1.06	0.44	0.74 (0.83)
	PBDE 154	0.09	0.08	0.09 (0.04)	1.45	0.98	1.21 (0.99)
	PBDE 183	0.12	0.12	0.12 (0.00)	0.12	0.12	0.12 (0.00)
	PBDE 209	6.50	6.50	6.50 (0.00)	6.86	6.50	6.60 (1.15)
	PCB 50 + 28	0.94	0.37	0.69 (0.52)	2.69	4.81	3.80 (2.54)
	PCB 52	2.63	1.80	2.26 (1.48)	7.37	8.73	8.08 (4.68)
	PCB 101	2.80	3.15	2.96 (1.85)	9.47	11.86	10.72 (7.05)
	PCB 118	4.02	2.58	3.38 (2.12)	17.44	21.25	19.43 (8.22)
	PCB 153	7.41	6.78	7.13 (3.02)	29.30	30.50	29.93 (14.32)
	PCB 138	4.79	4.29	4.57 (2.49)	27.03	29.86	28.51 (14.20)
	PCB 180	1.60	2.25	1.89 (0.88)	13.07	13.13	13.10 (7.05)
Metal	Ag (silver)	0.98	0.54	0.76 (0.44)	1.23	0.31	0.76 (1.04)
	As (arsenic)	3.81	2.72	3.30 (1.37)	1.85	1.25	1.54 (1.39)
	Cd (cadmium)	0.28	1.13	0.70 (0.65)	1.69	0.36	1.01 (1.87)
	Cr (chrome)	2.56	0.66	1.64 (4.21)	1.89	1.97	1.93 (3.78)
	Cu (copper)	70.71	32.16	51.93 (30.28)	149.75	109.14	128.97 (116.84)
	Ni (nickel)	0.55	0.23	0.39 (0.41)	0.14	0.17	0.15 (0.01)
	Pb (lead)	0.28	0.18	0.24 (0.11)	0.35	0.07	0.21 (0.33)
	Se (selenium)	14.62	12.20	13.44 (5.12)	27.46	30.79	29.17 (12.01)
	Zn (zinc)	205.83	122.66	165.13 (112.32)	218.47	213.00	215.67 (110.32)
	Hg (mercury)	0.17	0.53	0.34 (0.41)	0.87	1.72	1.38 (0.87)

the independent parameters (contamination PC factors) could predict the dependent parameters (important markers PCo factors). An analysis of variance (ANOVA; 1000 permutations) was then performed to assess the global significance of the db-RDA, and a marginal ANOVA (1000 permutations) was run to determine which contaminant PC factors were significantly correlated with PCo factors. We then identified the main contaminants correlated with the significant PC factors, using a minimum correlation threshold of 0.70. db-RDAs were computed using the function *rda* available in the VEGAN package in R software v3.1.3.

Population genetics parameters and genome scans

In order to quantify the extent of spatially varying selection between 'control' and 'polluted' localities for each species separately, genetic differentiation among sites was assessed using θ as the estimator of F_{ST} (Weir & Cockerham 1984) with all loci and then with important markers only, using GENODIVE v2.27 (Meirmans & Van Tienderen 2004). We also calculated the average major allele frequency change (delta p) of the important markers between 'control' and 'polluted' sampling sites as a surrogate for the selection strength acting on these SNPs (Bourret *et al.* 2014; Pavey *et al.* 2015).

In order to compare a genome scan approach with the Random Forest algorithm, we tested for loci under potential differential selection by habitat using BAYESCAN v2.1 (Foll & Gaggiotti 2008). For each species, we grouped individuals in two groups ('control' vs. 'polluted') and tested for SNPs potentially under divergent selection. BAYESCAN estimates locality-specific F_{ST} coefficients and uses a cut-off based on the mode of the

posterior distribution to detect SNPs under selection (Beaumont & Balding 2004; Foll & Gaggiotti 2008). SNPs with a posterior probability >0.95 were considered as outliers, after running 100 000 iterations on all samples together. As suggested by Lotterhos & Whitlock (2014) and Benestan *et al.* (2015), we specified a 'prior' odd of 10 000, which set the neutral model being 10 000 times more likely than the model with selection in order to avoid elevated false discovery rate. Because a 'prior' odd of 10 000 may be considered overly conservative, we rerun both analyses with a 'prior' odd of 10 (the default value).

Simulations

To assess whether the number of common important markers between the two species could be obtained by chance, we generated simulated data sets as follows. We created two vectors of 23 659 and 14 755 objects, respectively, using the R software v3.1.3, which correspond to the total number of filtered markers in each species (see Results section). In each group of markers, 142 SNPs were randomly picked without replacement using the *sample* function and compared, which corresponds to the number of important markers detected in each species (see Results section). For each iteration, we identified the number of common 'important markers' between the two data sets. A total of 1000 iterations were produced to generate a neutral distribution of matched 'important markers' number. We computed the probability that our observed number of common important markers between the two species obtained with Random Forest analysis resulted from chance only by first calculating the number of iterations with a number of matched 'important markers' equal or higher than the observed value and then divided this number by the total number of iterations.

Annotation of important markers

Important markers that were identified were annotated based on the American Eel draft genome (S.A. Pavey, E. Normandeau, J. Gaudin, L. Letourneau, L. Bernatchez, unpublished data). The genome has a total assembled size of 1.4 Gb with a contig N50 of 5818 and a scaffold N50 of 75 600. The genome has a very good coverage of the protein-coding genes, as the program CEGMA (Parra *et al.* 2007) indicated that 84.0% of the 458 core eukaryotic genes were completely or partially represented in the genome. The genome was annotated with MAKER2 (Holt & Yandell 2011), and 23 961 protein-coding genes could be attributed to a gene name via BLAST to the Swissprot database. The important markers that occurred within an annotated gene (exon or interior intron) were considered to represent that gene for the Gene Ontology (GO) enrichment analysis. GWAS with de novo assemblies of nonmodel organisms such as ours are largely limited to inferences of genes in the extremely local neighbourhood of the SNP because the scaffolds cannot be arranged into chromosomes without a genetic map. With these two species in particular, linkage decays quite rapidly, even detectable after 100 bp (Jacobsen *et al.* 2014). Therefore, we focused on inferring function only for SNPs that were within a gene. However, we made a second functional analysis using a flanking regions of 5000 bp before and after the first and last exon and produced a second GO analysis.

The GoMiner application was used for the GO enrichment analysis (Zeeberg *et al.* 2003). For each species separately, the gene names of the important markers that were within a gene (exon or interior intron) were used as the test list and all of the gene names annotated in the genome were used as the background. The human gene ontology was used to determine which GO categories were the most enriched for each species with a P-value <0.01.

Results

Genomic analyses

Following the first 'cstacks' step, we started with a catalog of 882 061 loci for the European Eel and 1 005 126 for the American Eel (Table 4). After filtering for the log-likelihood threshold, we retained 716 254 loci in the catalog for the European Eel and 835 393 for the American Eel (Table 4). After filtering for proportion of individuals genotyped, one SNP per stack, F_{IS} and MAF, we retained 23 659 SNPs for the European Eel and 14 755 SNPs for the American Eel (Table 4). These are the final genotype data sets used for all subsequent analyses (Table 4).

Using Random Forest analyses to genetically distinguish 'control' vs. 'polluted' sampling sites, we reached an OOB-ER of 2% for both species after 3 and 4 runs, respectively, for the European and American Eels. This means that 98% of individuals were reassigned correctly for both species and that the rank of markers based on their importance to assign the individuals correctly was reliable for subsequent important markers selection (Goldstein *et al.* 2011). A total of 141 and 142 important markers for the European Eel and for the American Eel were then identified using the upper end of the elbow of their importance value distributions (Fig. 2; Table 4). The mean delta P values of these important markers between 'control' and 'polluted' localities were 0.112 (SD 0.054) for the European Eel and 0.106 (SD 0.046) for American Eel (Fig. 3).

Reflecting the total absence of population structure in both species, pairwise F_{ST} values between all populations within species varied between 0.000 and 0.001 and were nonsignificant ($P > 0.05$) for all pairwise comparisons when using all loci in both species (Table 5). Similar F_{ST} values using important markers were obtained when comparing within 'control' or within 'polluted' localities in both species (between 0.000 and 0.001) (Table 5). However, F_{ST} values on important markers between 'control' and 'polluted' localities were much higher and all significant, varying between 0.057 and 0.066 for the European Eel and between 0.050 and 0.054 for the American Eel (Table 5).

Based on broken-stick distribution, two PCo factors were selected to represent the genetic variation of the important SNPs in each species. These PCo factors represent 10.9% of the important SNPs variation in European Eel and 10.1% in American Eel. All subsequent test of association (db-RDA) between important loci and pollutants concentration was also repeated with all PCo factors, in order to verify whether similar results could be obtained when all genetic variation is considered in the analyses. These results are not presented because they were similar to the one produced with informative axes only.

Table 4 Number of putative SNPs retained following reach filtering steps in STACKS. Random Forest values correspond to the number of important markers identified by this algorithm for both species. Bayescan values correspond to the number of outlier detected with a prior odds of 10 and 10 000

	SNP count	
From Reads to SNPs	European Eel	American Eel
Filtering		
Initial catalogs	882 061	1 005 126
Rx stack: log likelihood −10	716 254	835 383
Coverage 5X		
Proportion: >75% inside each population F_{IS} >0.3	105 472	73 018
MAF >0.02	23 659	14 755
SNPs under potential selection		
Random Forest	141	142
Bayescan (prior odds: 10)	0	2
Bayescan (prior odds: 10 000)	0	0

Association between genetic variation and contaminants

For European Eel, we retained 14 PCs for the 24 organic contaminant concentrations in muscle and 6 PCs for the 10 metal concentrations in liver, based on Kaiser–Guttman criterion. For American Eel, 10 PCs for the organic contaminants and 6 PCs for the 10 metals were retained, respectively. In all these four contaminant PCA (two types of contaminant X two species), more than 99% of the variation was represented by the PC axes that were retained.

Using the individual genetic variation obtained from important SNPs multilocus genotypes as subjects and the organic contamination PC factors as explanatory variables, two db-RDAs (one for each species) were performed on the two first PCo of important SNPs (Figs 4 and 5). Both db-RDAs were globally highly significant with a P-value <0.001 (ANOVAs; European Eel: $F = 4.26$ and American Eel: $F = 4.73$) with an adjusted coefficient of determination of 0.34 and 0.30, respectively (Figs 4 and 5). In European Eel, the two first db-RDA axes represented 44.3% of the variation (Fig. 4). The marginal ANOVA showed that PC factors 1 and 3 were both highly significant predictors of the genetic variation of the important SNPs with P-value <0.001 (ANOVA, $F = 16.11$ and 30.51). Contaminants with correlation loading >0.70 for each of these highly

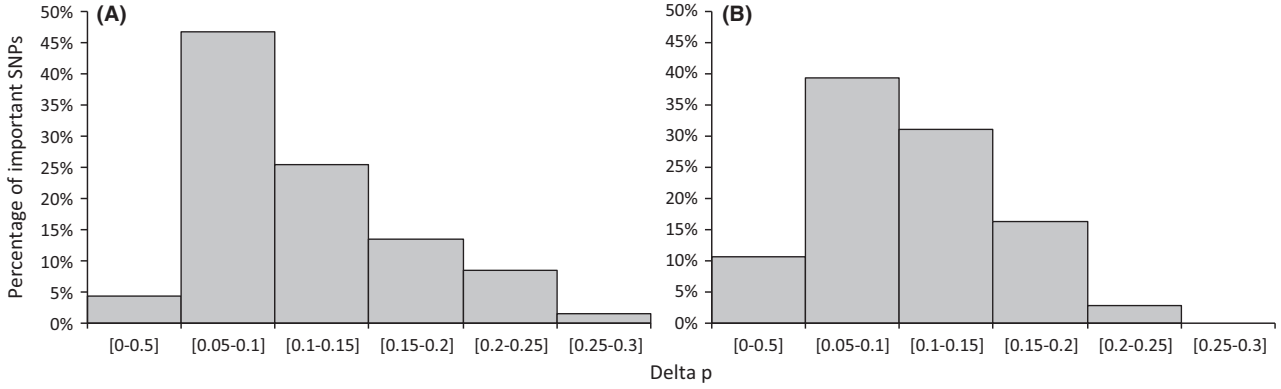

Fig. 3 Distribution of allelic changes (delta p) between eels from 'control' vs. 'polluted' sampling sites. (A) For the European Eel among 141 important SNPs identified with Random Forest. (B) For the American Eel among 142 important SNPs identified with Random Forest.

Table 5 Pairwise multilocus estimates of genetic differentiation based on F_{ST}. Lower triangle reports values based on all loci (23 659 loci for European Eel and 14 755 for American Eel), and upper triangle reports values for the important markers (141 for European Eel and 142 for American Eel)

European Eel	Certes	Dordognes	Gironde	Garonne
Certes	—	0.001	0.065***	0.066***
Dordogne	0.001	—	0.058***	0.057***
Gironde	0.000	0.000	—	0.000
Garonne	0.000	0.000	0.000	—
American Eel	**St Jean**	**Sud Ouest**	**St François**	**St Pierre**
St Jean	—	0.000	0.054***	0.053***
Sud Ouest	0.000	—	0.053***	0.050***
St François	0.000	0.000	—	0.000
St Pierre	0.000	0.000	0.000	—

***P-value <0.001; if no *P-value >0.05.

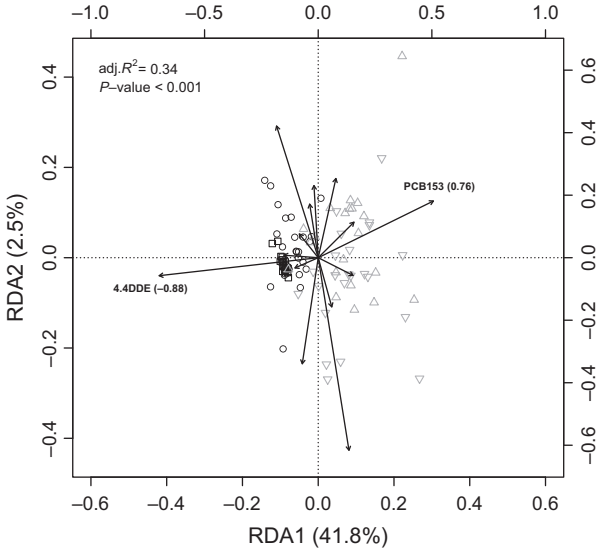

Fig. 4 Redundancy analysis Axis 1 (41.8% of variance) and Axis 2 (2.5% of variance) showing the position of European Eel individuals from a PCoA using a modified version of simple matching distance matrix (see Materials and Methods for details) and related organic contaminants concentration in muscle PC factors illustrated by black arrows. Eels from 'control' localities are in black (Dordogne: circle and Certes: square), and eels from 'polluted' localities are in grey (Garonne: up triangle and Gironde: down triangle). Positions of environmental PC factors are according to scales on top and right axes. Contaminants with correlation loading >0.70 were added on the highly significant PC factors (P-value <0.001).

significant PC factors were PCB153 for PC1 (correlation of 0.76) and 4′4′DDE for PC3 (−0.88) (Fig. 4). This means that PCB153 and 4′4′DDE concentration, respectively, increase and decrease in the direction of the

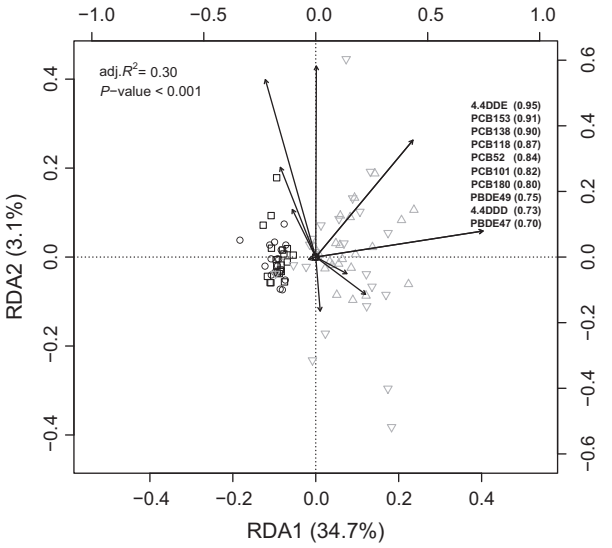

Fig. 5 Redundancy analysis Axis 1 (34.7% of variance) and Axis 2 (3.1% of variance) showing position of American Eel individuals from a PCoA using a modified version of simple matching distance matrix (see Materials and Methods for details) and related organic contaminants concentration in muscle PC factors illustrated by black arrows. Eels from 'control' localities are in black (Sud Ouest: circle and St Jean: square), and eels from 'polluted' localities are in grey (St François: triangle and St Pierre: inverted triangle). Positions of PC factors are according to scales on top and right axes. Contaminants with correlation loading >0.70 were added on the highly significant PC factors (P-value <0.001).

arrow of their correlated PC factor. Because PC1 and PC3 arrows point towards opposite direction (Fig. 4), both contaminants thus show an increase in concentration in individuals inhabiting 'polluted' localities. In American Eel, the two first db-RDA axes represented 37.8% of the variation (Fig. 5). The marginal ANOVA showed that only PC factor 1 was significant (P-value <0.001; ANOVA, F = 26.88). A total of 10 organic contaminants (including PCB153 and 4′4′DDE) had a correlation equal or higher to 0.70 with the PC factor 1 (Fig. 5). All these contaminants showed an increase in concentration in individuals inhabiting 'polluted' localities.

Two db-RDAs were also performed for each species with metal contamination PC factors as explanatory variables (Figs 6 and 7). Both of these db-RDAs were globally highly significant with P-value <0.001 (ANOVAs; European Eel: F = 3.46 and American Eel: F = 8.87). In European Eel, the two first db-RDA axes represented 20.1% of the variation (Fig. 6). The marginal ANOVA showed that PC factor 3 was highly significant (P-value <0.001; ANOVA, F = 12.26) and only selenium (Se) was correlated with it (0.81) (Fig. 6). For American Eel, the two first db-RDA axes explained

39.3% of the variation (Fig. 7). Similar to the European Eel, Se was correlated (0.77) with the only highly significant PC factor (PC factor 3; *P*-value <0.001; ANOVA, $F = 23.20$) (Fig. 7). For both species, Se showed an increase in concentration in individuals inhabiting 'polluted' localities. Finally, all db-RDA (Figs 4–7) revealed genetic differentiation (on the first axis) at the important markers between individuals from 'control' (in black) and individuals from 'polluted' sampling sites (in grey).

Population genetics parameters and comparison with genome scan

Only four important markers matched the same genome contig between the two species, which was nevertheless higher than the value of 0.856 (SD 0.888) for the average common number of important markers obtained from the 1000 simulations. The probability of obtaining four common important markers by chance was 0.008, indicating that the match between the two species at these markers was not random, while being admittedly low.

A total of 30 and 32 important markers for European and American eels were found in annotated genes (these numbers increased to 56 and 49, respectively, when we added flanking regions of 5000 bp; see Appendix S1, Supporting information for further details). One of these was similar in both species, *ABCG5*. No annotated genes were found for the other three common scaffolds where important markers were found in both species. Gene ontology revealed a total of seven biological processes and molecular functions that converged between species across the 57 (European Eel) and 21 (American Eel) terms identified among the important markers (more than one function can be imputed to a same gene; for example, 57 functions are >30 markers detected in European Eel) (Table 6). These common processes and functions mainly comprised different sterol absorption and transport processes in addition to regulation of the digestive system (Table 6). Finally, BAYESCAN found no loci under potential divergent selection in either species with a 'prior' odd of 10 000 (Table 4). However, two loci under potential divergent selection were detected in the American Eel with a 'prior' odd of 10 (Table 4; Appendix S2 and S3, Supporting information). These two loci were also identified by the Random Forest analysis.

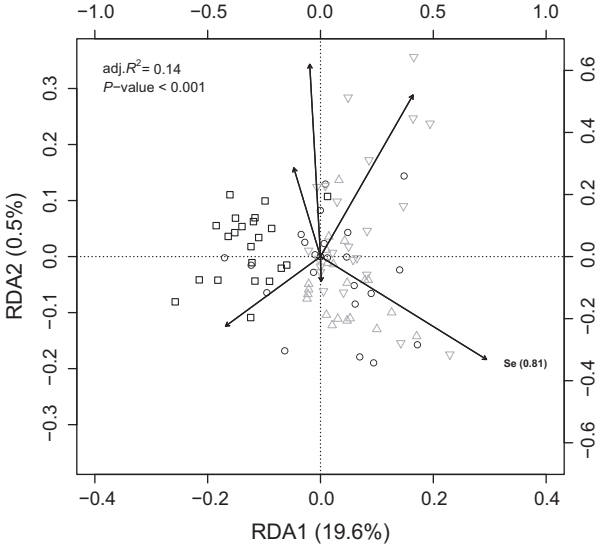

Fig. 6 Redundancy analysis Axis 1 (19.6% of variance) and Axis 2 (0.5% of variance) showing position of European Eel individuals from a PCoA using a modified version of simple matching distance matrix (see Materials and Methods for details) and related metal contaminants concentration in liver PC factors illustrated by black arrows. Eels from 'control' localities are in black (Dordogne: circle and Certes: square), and eels from 'polluted' localities are in grey (Garonne: triangle and Gironde: inverted triangle). Positions of PC factors are according to scales on top and right axes. Contaminants with correlation loading >0.70 were added on the highly significant PC factors (*P*-value <0.001).

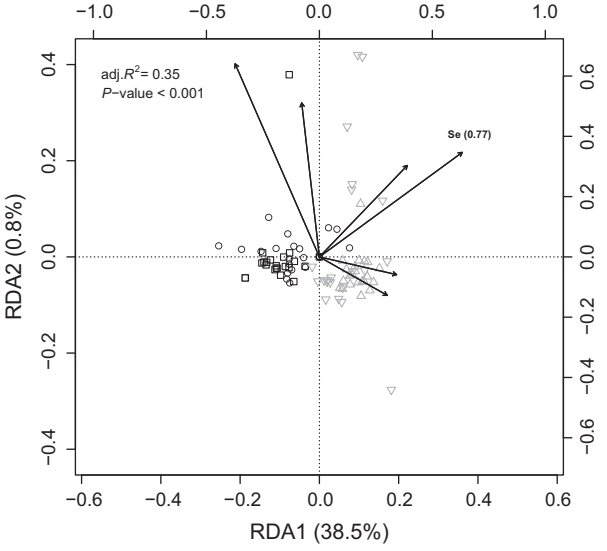

Fig. 7 Redundancy analysis Axis 1 (38.5% of variance) and Axis 2 (0.8% of variance) showing position of American Eel individuals from a PCoA using a modified version of simple matching distance matrix (see Materials and Methods for details) and related metal contaminants concentration in liver PC factors illustrated by black arrows. Eels from 'control' localities are in black (Sud Ouest: circle and St Jean: square), and eels from 'polluted' localities are in grey (St François: up triangle and St Pierre: down triangle). Positions of PC factors are according to scales on top and right axes. Contaminants with correlation loading >0.70 were added on the highly significant PC factors (*P*-value <0.001) only.

Table 6 Gene Ontology (GO) enrichment for 141 and 142 important markers in European and American eels that differentiate eels from 'control' vs. 'polluted' sampling sites. The columns represent the GO identifier enriched, the number of genes in the genome implicated in that term (Total), the number of important markers implicated in that term, the *P*-value and the name of the term. Terms from both biological process and molecular function with a *P*-value <0.01 are included (Fisher's exact test). First section corresponds to shared terms between both species, while the second section reports SNPs unique to European Eel and the third section those unique to American Eel

GO ID Both species	Total	SNPs	*P*-value European Eel	*P*-value American Eel	Term
10949	2	1	0.0052	0.0045	Negative regulation of intestinal phytosterol absorption
30300	2	1	0.0052	0.0045	Regulation of intestinal cholesterol absorption
45796	2	1	0.0052	0.0045	Negative regulation of intestinal cholesterol absorption
60457	2	1	0.0052	0.0045	Negative regulation of digestive system process
60752	2	1	0.0052	0.0045	Intestinal phytosterol absorption
32372	3	1	0.0078	0.0067	Negative regulation of sterol transport
32375	3	1	0.0078	0.0067	Negative regulation of cholesterol transport
European Eel	Total	SNPs	*P*-value		Term
35089	2	2	0.0000	—	Establishment of apical/basal cell polarity
61162	2	2	0.0000	—	Establishment of monopolar cell polarity
61339	2	2	0.0000	—	Establishment or maintenance of monopolar cell polarity
35088	6	2	0.0001	—	Establishment or maintenance of apical/basal cell polarity
61245	6	2	0.0001	—	Establishment or maintenance of bipolar cell polarity
9605	413	6	0.0004	—	Response to external stimulus
30010	13	2	0.0005	—	Establishment of cell polarity
7613	19	2	0.0011	—	Memory
6935	215	4	0.0018	—	Chemotaxis
42330	215	4	0.0018	—	Taxis
7402	1	1	0.0026	——	Ganglion mother cell fate determination
8052	1	1	0.0026	—	Sensory organ boundary specification
10160	1	1	0.0026	—	Formation of organ boundary
30859	1	1	0.0026	—	Polarized epithelial cell differentiation
34447	1	1	0.0026	—	very low-density lipoprotein particle clearance
34750	1	1	0.0026	—	Scrib-APC–beta-catenin complex
45198	1	1	0.0026	—	Establishment of epithelial cell apical/basal polarity
71896	1	1	0.0026	—	Protein localization to adherens junction
7163	30	2	0.0027	—	Establishment or maintenance of cell polarity
31175	258	4	0.0036	—	Neuron projection development
32501	1833	10	0.0050	—	Multicellular organismal process
2036	2	1	0.0052	—	Regulation of L-glutamate transport
5608	2	1	0.0052	—	Laminin-3 complex
10825	2	1	0.0052	—	Positive regulation of centrosome duplication
19896	2	1	0.0052	—	Axon transport of mitochondrion
30061	2	1	0.0052	—	Mitochondrial crista
30300	2	1	0.0052	—	Regulation of intestinal cholesterol absorption
90162	2	1	0.0052	—	Establishment of epithelial cell polarity
7155	292	4	0.0056	—	Cell adhesion
22610	292	4	0.0056	—	Biological adhesion
48666	292	4	0.0056	—	Neuron development
48468	479	5	0.0057	—	Cell development
71702	145	3	0.0057	—	Organic substance transport
22612	46	2	0.0062	—	Gland morphogenesis
7411	155	3	0.0069	—	Axon guidance
31012	156	3	0.0070	—	Extracellular matrix
1921	3	1	0.0078	—	Positive regulation of receptor recycling
5886	1283	8	0.0078	—	Plasma membrane
30229	3	1	0.0078	—	Very low-density lipoprotein receptor activity
32863	3	1	0.0078	—	Activation of RacGTPase activity

Table 6 *Continued*

GO ID Both species	Total	SNPs	*P*-value European Eel	*P*-value American Eel	Term
34189	3	1	0.0078	—	Very low-density lipoprotein binding
44070	3	1	0.0078	—	Regulation of anion transport
5624	323	4	0.0080	—	Membrane fraction
45211	53	2	0.0081	—	Postsynaptic membrane
7611	55	2	0.0087	—	Learning or memory
71944	1310	8	0.0089	—	Cell periphery
5576	533	5	0.0090	—	Extracellular region
44421	334	4	0.0090	—	Extracellular region part
5626	336	4	0.0092	—	Insoluble fraction
5198	174	3	0.0094	—	Structural molecule activity
American Eel	Total	SNPs		*P*-value	Term
1729	1	1	—	0.0022	Ceramide kinase activity
31753	1	1	—	0.0022	Endothelial differentiation G-protein coupled receptor binding
31755	1	1	—	0.0022	Edg-2 lysophosphatidic acid receptor binding
47620	1	1	—	0.0022	Acylglycerol kinase activity
3951	2	1	—	0.0045	NAD+ kinase activity
7158	3	1	—	0.0067	Neuron cell–cell adhesion
15299	3	1	—	0.0067	Solute/hydrogen antiporter activity
51665	3	1	—	0.0067	Membrane raft localization
32553	908	6	—	0.0086	Ribonucleotide binding
32555	908	6	—	0.0086	Purine ribonucleotide binding
30299	4	1	—	0.0089	Intestinal cholesterol absorption
44241	4	1	—	0.0089	Lipid digestion
48147	4	1	—	0.0089	Negative regulation of fibroblast proliferation
22804	67	2	—	0.0094	Active transmembrane transporter activity

Discussion

The main objective of this study was to test for the occurrence of polygenic selection in response to environmental pollution in both North Atlantic eels. Polygenic selection is expected to produce a signal of subtle allelic differentiation at multiple loci between environments. Combining analyses at both the sampling site and individual scale allowed us to establish the relationships between individual bioaccumulation levels and their effects at the genome level. At the scale of the sampling site using the Random Forest algorithm, we found 141 and 142 covarying and/or interacting SNPs (*i.e.* important markers) in European and American eels, respectively, distinguishing individuals from 'control' and 'polluted' sampling sites. In order to identify what contaminants may be the most likely selective agents, we tested for the relationships between individual genetic variation of those loci and levels of contamination. Significant relationships between genetic variation and individual contamination by PCB153, 4′4′DDE and Se were found in both North Atlantic eels, using a db-RDA. Putative biological processes and molecular functions were also examined to understand the molecular mechanisms underlying resistance to pollutants. The convergence of such processes and functions between both species suggested that sterol regulation plays a major role in the differential survival of eels in polluted sites. Overall, these results support the presence of polygenic selection in both North Atlantic Eels, associated with anthropogenic pollution.

Combining random forest algorithm and db-RDA to detect selective agents

Signals of polygenic selection were found with Random Forest algorithm between 'polluted' and 'control' environments. However, to provide further support for such a claim, covarying loci found based on external exposure should be tested with internal doses to strengthen the association with contaminants (Mussali-Galante *et al.* 2014). Therefore, we used a db-RDA to test for an association between internal concentration of pollutants and genetic variation among individuals, using the covarying loci previously identified with the Random Forest algorithm. As mentioned above, for both North Atlantic Eels, we found that three contaminants (PCB153, 4′4′DDE and Se) significantly explained the genetic variation of

covarying loci among individuals. Eight additional organic contaminants also explained the genetic variation at important markers in the American Eel. Interestingly, the average level of contamination of Se was slightly higher in Dordogne (25.55 μg/g, dry wet; control site) than in Garonne (23.87 μg/g, dry wet; polluted site). While copper and zinc showed the most pronounced contamination variation and were strongly associated with the first PC factor of contamination (European Eel: zinc 0.99, copper 0.81; American Eel: zinc 0.94, copper 0.85), both of these first PC factors were not significantly associated with the variation of important markers. Together, these results demonstrate that db-RDA is a powerful tool to detect selective agents associated with genetic variation of covarying markers previously found despite (i) that the variation of the selective agent is slightly different than the pattern of locations used to identifying the covarying markers and (ii) that the selective agent is not the major source of variation inside all potential selective agent tested.

Admittedly, the analytical design adopted in this study was eased by the total absence of population structure in eels. In situations of pronounced population genetic structure, it would instead be better to use the 'regression' method of Random Forest with a correction for neutral genetic structure (see Zhao *et al.* 2012 for details) or investigate with other genetic–environment correlation methods such as the software matSAM (Joost *et al.* 2008) or LFMM (Frichot *et al.* 2013) (but see Stephan 2016). In addition, a large proportion of the genetic variation was not explained by the internal doses of contaminants in the db-RDAs. Therefore, we cannot exclude the possibility that other contaminants and environmental variables not measured in the present study may also explain part of the genetic variation of important markers. Variation in the accumulation of toxin through age (unknown in our sample) could also add noise in those relationships. Nevertheless, we argue that these results illustrate the power of combining methods for detecting signals of polygenic selection and for associating variation of loci with putative selective agents in order to document the dynamics of selection at the genome level.

Impact of pollutants on genetic variation of eels

Genetic differentiation caused by contaminant exposure has been reported in various aquatic organisms (Arthropods: Gardeström *et al.* 2008; Molluscs: Ma *et al.* 2000; Amphibians: Matson *et al.* 2006; Fish: Maes *et al.* 2005; Bourret *et al.* 2008) and using different molecular markers, namely allozymes (Haimi *et al.* 2006), microsatellites (Durrant *et al.* 2011), AFLPs (Lind & Grahn 2011) and coding genes SNP (Bélanger-Deschênes *et al.* 2013). To our knowledge, this is the first study to address the impact of an exposure to pollution on genetic variation of an aquatic organism, using a more 'genomewide' genotyping approach (see Table 1 in the review of Mussali-Galante *et al.* 2014 for a summary). In European Eel, the impact of metal pollution on genetic variation was previously evidenced by Maes *et al.* (2005), using microsatellite markers. These authors also observed a negative correlation between metal exposure and fitness measured in terms of condition and hepatosomatic index. Other studies showed that PCBs and metals have severe impacts on European Eel health, ultimately leading to death (Nowell *et al.* 1999; Robinet & Feunteun 2002; Corsi *et al.* 2005; Geerarts & Belpaire 2010; Pujolar *et al.* 2012, 2013). In the American Eel, lower survival has previously been linked with contamination by organic compounds (Couillard *et al.* 1997). Altogether, these studies support the hypothesis that organic pollution and metal pollution act as strong selective agents causing differential mortality, ultimately leading to a genetic differentiation among localities in both eel species. The present study confirms this association and, in addition, illustrates which pollutants are having the strongest selective effect on the genome.

Genome scan vs. polygenic approaches

Pollution of aquatic ecosystems typically comprises many pollutants producing dissimilar toxic effects on organisms (Matthiessen *et al.* 1988). According to quantitative genetics theory, such complex perturbation is expected to produce subtle allelic frequency changes at multiple loci (*i.e.* polygenic selection) (McKay & Latta 2002; Pritchard *et al.* 2010; Le Corre & Kremer 2012). In our study, BAYESCAN identified two loci under potential selection in the American Eel only. This could be explained by the classical hitchhiking model used in this approach, which is best suited to detect loci showing strong allelic frequency differentiation compared to the rest of the genome (Maynard Smith & Haigh 1974; Pritchard *et al.* 2010; Messer & Petrov 2013). In addition, it has been shown that linkage disequilibrium decays very quickly in both North Atlantic Eels (Jacobsen *et al.* 2014). Thus, even when comparing the genomes of both species, which hybridize naturally (Albert *et al.* 2006), no islands of divergence were detected, but rather high divergence at single nucleotide positions was identified (Jacobsen *et al.* 2014). Given that there are no islands of divergence between the species and the rapid decay of linkage that is even detectable at a 100-bp interval (Jacobsen *et al.* 2014), it would be unreasonable to expect such islands within species. In such a case of pronounced decay in LD, the genotyping of many more SNPs may have been necessary to identify outlier, highly divergent SNPs between control and polluted sites. In comparison, the Random Forest algorithm

found 141 and 142 loci under potential polygenic selection in the European and American eels, respectively. This could be explained by the fact that this approach specifically searches for covarying loci able to discriminate between factors (in our case 'polluted' vs. 'control' environments). Therefore, as observed in recent studies (Bourret *et al.* 2014; Brieuc *et al.* 2015; Laporte *et al.* 2015; Pavey *et al.* 2015), these results support the contention that selection on ecologically complex traits (such as phenotypes being involved in coping with distinct environments) may be more efficiently detected by means of methods designed to search for modest variation at covarying loci.

Anthropogenic vs. environmental polygenic selection

A recent study based on the same life stage (yellow eel) and using a similar approach as used here also revealed a footprint of polygenic selection between American Eels inhabiting fresh and salt water environments (Pavey *et al.* 2015). In both studies, no genetic differentiation was found with all loci or within similar environments with covarying loci only. In contrast, both studies revealed genetic differentiation between different environments at markers under selection which was sufficiently pronounced to allow classifying eels by habitats with high accuracy (90% and above). As a surrogate for selection strength acting on these SNPs, delta p was used in both studies. Delta p was higher between SNPs distinguishing 'control' and 'polluted' environment (delta $P = 0.112$ SD 0.054 in European Eel and 0.106 SD 0.046 in American Eel) than between 'fresh' and 'salt' water habitats (delta $P = 0.034$ SD 0.002; Pavey *et al.* 2015). This suggests that water pollution imposes stronger selection than that resulting from differential habitat use occurring naturally in American Eel. Bourret *et al.* (2014) also found a lower delta p (varying between 0.030 and 0.090) in wild Atlantic salmon populations, which was caused by selective mortality at sea. While the above comparisons must be interpreted cautiously, they suggest that spatially varying selection caused by human-driven environment alteration associated with pollution may rapidly impact on the genetic diversity and perhaps on the evolutionary potential of both Eel species.

Selective agents acting on the genetic diversity in North Atlantic Eels

Individual genetic variation in European Eel appeared to be especially affected by two organic contaminants (PCB153 and 4′4DDE). PCB153 is linked to hypoactivity (Johanson *et al.* 2011) and upregulation of sterol biosynthesis genes (Yadetie *et al.* 2014), whereas both PCB153 and 4′4′DDE are known to produce oxidative stress and cell apoptosis (Song *et al.* 2011; Gao *et al.* 2013). In addition, Se was also found to be associated with individual genetic variation at markers under selection between control and polluted sites in both species. Selenium is an essential element that plays a crucial defensive role against oxidative stress (i.e. a known mechanism of toxicity of PCB153 and 4′4′DDE) (Monteiro *et al.* 2009; Selvaraj *et al.* 2012). Selenium is also toxic at high exposure concentrations, with poisoning symptoms in fish including swelling of gill lamellae, elevated lymphocytes, anaemia (reduced haematocrit and haemoglobin), corneal cataract and exophthalmos (popeyes), pathology in several organs (including liver) as well as teratogenic deformity of the spine, head, mouth and fins (Lemly 2002). In comparison, PCB52, PCB101, PCB118, PCB138, PCB153, PCB180, 4′4′DDD, 4′4′DDE, PCBE 49 and PCBE47 were identified as the most likely organic selective agents in American Eel. PCB mixtures can interfere with neuroendocrine systems (Soontornchat *et al.* 1994), thyroid hormone synthesis (McKinney & Waller 1994), as well as having various endocrine-disrupting effects (Maria *et al.* 2006; Teles *et al.* 2007). It is noteworthy that the two contaminants with the highest correlation with the one significant PC factor in the db-RDA analysis are PCB153 and 4′4′DDE, as was observed with the European Eel, as was Se. The convergence between these three selective agents (PCB153, 4′4′DDE and Se) in both North Atlantic Eels could suggest their common particular sensitivity to these contaminants. Gene ontology analyses on important markers also indicated convergence of ontology terms between both species, mainly linked to regulation, absorption and transport of sterols (a subclass of steroids). This hypothetically implies a response to these organic compounds and metals in both species. Detoxification of most of the PCB pollutants can only occur through the monooxygenase system, which can be partially inhibited by steroid hormones (Geerarts & Belpaire 2010). This suggests that the regulation of sterol biosynthesis plays a major role in the differential survival of eels in contaminated sites. This hypothesis could be directly tested in future studies with control and treatment groups in a laboratory setting for a better understanding of the molecular mechanism that could link sterol regulation to a higher survival in contaminated environments.

Impact of pollutants on the genome vs. transcriptome

Baillon *et al.* (2015) recently showed that the main contaminant affecting the level of gene transcription on the exact same fish analysed here were two metals (arsenic: As and cadmium: Cd) and one organic contaminant (lindane) that were not associated with individual genetic variation in the present study. Conversely, the

three contaminants (PCB153, 4′4′DDE and Se) associated with individual genetic variation in this study did not significantly cause transcriptional differences associated with levels of contaminants (Baillon *et al.* 2015). Hypothetically, this difference could first be explained by different levels of plasticity among genes affected by pollutants. Thus, a pronounced plastic response may buffer against the effect of selection (Pfennig *et al.* 2010; Thibert-Plante & Hendry 2011). Such buffering of selection by phenotypic plasticity at the transcriptome level and in link with pollution was recently partly supported in the yellow perch (*Perca flavescens*) (Pierron *et al.* 2009; Bélanger-Deschênes *et al.* 2013). Of the 65 genes that showed an altered pattern of gene transcription in relation to Cd and/or copper contamination (Pierron *et al.* 2009), only five of them showed a pattern of differential selection for the same contaminants (Bélanger-Deschênes *et al.* 2013). It is thus plausible that different contaminants could interfere differentially with different genes, either by altering their pattern of gene transcription, by selecting genetic variants or by both. Differences observed between studies performed at the transcriptome vs. genome level may also be explained by different genome representation between both types of studies. This would be exacerbated by the very low linkage disequilibrium reported in eels, which may have resulted in missing the target of selection associated with exposure to As, Cd and lindane contaminants (Jacobsen *et al.* 2014).

Conclusions

Polygenic selection caused by human-driven environmental selection was observed in both North Atlantic eels. Also for both species, subtle allelic frequency changes were found to be associated with water pollution indicating nonrandom mortality of individuals by two organic contaminants, PCB153 and 4′4′DDE, and one metal selenium. We thus propose that the regulation of sterol synthesis genes plays a major role in the differential survival of eels in polluted environments. Our results thus support the contention that fast and measurable selection can be produced by anthropogenic pollution during the lifetime of eels, which could potentially impact in the long term on the species' genetic diversity and ultimately their evolutionary potential. As such, this study provides another empirical demonstration of the potential impact of human-driven environmental change on the evolutionary potential of species in their natural environment. Finally, this study illustrates the power of combining suites of methods for detecting signals of polygenic selection and for associating genetic variation with putative selective factors to document the dynamics of selection at the genomic level.

Acknowledgements

We are grateful to L. Baillon, L. Benestan, B. Bougas, V. Bourret, A.R. Kusler, E. Normandeau and C. Senay for their invaluable logistic support and/or comments on earlier versions of the manuscript. We are also thankful to M. Foll and two anonymous reviewers for their constructive comments on a previous version of the manuscript. This research was supported by the ANR-NSERC IMMORTEEL (ANR-10-INTB-1704) led by P. Couture, L. Bernatchez, P.G.C. Campbell, M. Baudrimont, F. Pierron, P. Élie, H. Budzinski, B. Morin, G. Durrieu, P. Lambert, S. Dufour; and a Discovery research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) led by L. Bernatchez, who holds the Canadian Research Chair in Genomics and Conservation of Aquatic Resources. M. Laporte was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC). S.A.P. was supported by a postdoctoral fellowship from Ressources Aquatiques Québec (RAQ).

References

Albert V, Jónsson B, Bernatchez L (2006) Natural hybrids in Atlantic eels (*Anguilla anguilla*, *A. rostrata*): evidence for successful reproduction and fluctuating abundance in space in time. *Molecular Ecology*, **15**, 1903–1916.

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

Als T, Hansen MM, Maes G *et al.* (2011) All roads lead to home: Panmixia of European eel in the Sargasso Sea. *Molecular Ecology*, **20**, 1333–1346.

Baillon L, Pierron F, Coudret R *et al.* (2015) Transcriptome profile analysis reveals specific signatures of pollutants in Atlantic eels. *Ecotoxicology*, **24**, 71–84.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Bélanger-Deschênes S, Couture P, Campbell PGC *et al.* (2013) Evolutionary change driven by metal exposure as revealed by coding SNP genome scan in wild yellow perch (*Perca flavescens*). *Ecotoxicology*, **22**, 938–957.

Belfiore NM, Anderson SL (2001) Effects of contaminants on genetic patterns in aquatic organisms: a review. *Mutation Research*, **489**, 97–122.

Benestan L, Gosselin T, Perrier C *et al.* (2015) RAD-genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species; the American lobster (*Homarus americanus*). *Molecular Ecology* **24**, 3299–3315.

Bougas B, Normandeau E, Pierron F *et al.* (2013) How does exposure to nickel and cadmium affect the transcriptome of yellow perch (*Perca flavescens*) – results from a 1000 candidate-gene microarray. *Aquatic Ecotoxicology*, **142–143**, 355–364.

Boulesteix A-L, Janitza S, Kruppa J *et al.* (2012) Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Mining and Knowledge Discovery*, **2**, 493–507.

Boulesteix A-L, Janitza S, Hapfelmeier A *et al.* (2014) Letter to the editor: on the term 'interaction' and related phrases in the

literature on Random Forests. *Briefings in Bioinformatics*, **16**, 338–345.

Bourret V, Couture P, Campbell P *et al.* (2008) Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a polymetallic gradient. *Aquatic Toxicology*, **86**, 76–90.

Bourret V, Dionne M, Bernatchez L (2014) Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: Polygenic multilocus analyses surpasses genome scan. *Molecular Ecology*, **23**, 4444–4457.

Bozinovic G, Oleksiak MF (2011) Omics and environmental science genomic approaches with natural fish populations from polluted environments. *Environmental Toxicology and Chemistry*, **30**, 283–289.

Brieuc MSO, Ono K, Drinan D *et al.* (2015) Integration of Random Forest with population-based outlier analyses to provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, **24**, 2729–2746.

Catchen J, Hohenlohe PA, Bassham S *et al.* (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Chen X, Ishwaran H (2012) Random forests for genomic data analysis. *Genomics*, **99**, 323–329.

Cordell HJ (2009) Detecting gene–gene interactions that underlie human diseases. *Nature Review Genetics*, **10**, 392–404.

Corsi I, Mariottini M, Badesso A *et al.* (2005) Contamination and sub-lethal toxicological effects of persistent organic pollutants in the European Eel (*Anguilla anguilla*) in the Orbetello lagoon (Tuscany, Italy). *Hydrobiologia*, **550**, 237–249.

Côté CL, Gagnaire P-A, Bourret V *et al.* (2013) Population genetics of the American eel (*Anguilla rostrata*): *Fst* = 0 and North Atlantic Oscillation effects on demographic fluctuations of a panmictic species. *Molecular Ecology*, **22**, 1763–1776.

Côté CL, Castonguay M, Svetlana McWilliam K *et al.* (2014) In absence of local adaptation, plasticity and spatially varying selection rule: a view from genomic reaction norms in a panmictic species (*Anguilla rostrata*). *BMC Genomics*, **15**, 403.

Couillard CM, Hodson PV, Castonguay M (1997) Correlations between pathological changes and chemical contamination in American eels, *Anguilla rostrata*, from the St. Lawrence River. *Canadian Journal Fisheries and Aquatic Science*, **54**, 1916–1927.

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

Dudgeon D, Arthington AH, Gessner MO *et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Review*, **81**, 163–182.

Durrant CJ, Stevens JR, Hogstrand C *et al.* (2011) The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta* L.) residing in the River Hayle, Cornwall, UK. *Environmental Pollution*, **159**, 3595–3603.

Durrieu G, Maury-Brachet R, Girardin M (2005) Contamination by heavy metals (Cd, Zn, Cu, and Hg) of eight fish species in the Gironde estuary (France). *Estuaries*, **28**, 581–591.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for 681 both dominant and codominant markers: a bayesian perspective. *Genetics*, **180**, 977–993.

Frichot E, Schoville SD, Bouchard G *et al.* (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–1699.

Gagnaire PA, Normandeau E, Côté C *et al.* (2012) The genetic consequences of spatially varying selection in the panmictic American Eel (*Anguilla rostrata*). *Genetics*, **190**, 725–736.

Gao M, Wu N, Song Y *et al.* (2013) PCB153-induced oxidative stress and cell apoptosis on cultured rat Sertoli cells. *Toxicology Research*, **2**, 173.

Gardeström J, Dahl U, Kotsalainen O *et al.* (2008) Evidence of population genetic effects of long-term exposure to contaminated sediments: a multi-endpoint study with copepods. *Aquatic Toxicology*, **86**, 426–436.

Geerarts C, Belpaire C (2010) The effects of contaminants in European eel: a review. *Ecotoxicology*, **19**, 239–266.

Goldstein BA, Polley EC, Briggs FBS (2011) Random forests for genetic association studies. *Statistical Applications in Genetics and Molecular Biology*, **10**, 32.

Haasl RJ, Payseur BA (2016) Fifteen years of genome-wide scans for selection: trends, lessons, and unaddressed genetic sources of complication. *Molecular Ecology*, This issue.

Haimi J, Knott K, Selonen S *et al.* (2006) Has long-term metal exposure induced changes in life history traits and genetic diversity of the enchytraeid worm *Cognettia sphagnetorum*. *Environmental Pollution*, **140**, 463–470.

Holliday JA, Wang T, Aitken SN (2012) Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using Random Forest. *G3: Genes | Genomes | Genetics*, **2**, 1085–1093.

Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, **12**, 491.

IUCN (International Union for Conservation of Nature) (2014) The IUCN red list of threatened species. version 2014.3. http://www.iucnredlist.org..

Jacobsen MW, Pujolar JM, Bernatchez L *et al.* (2014) Genomic footprints of speciation in Atlantic eels (*Anguilla anguilla* and *A. rostrata*). *Molecular Ecology*, **23**, 4785–4798.

Jarquín D, Kocak K, Posadas L *et al.* (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics*, **15**, 740.

Johanson EB, Knoff M, Fonnum F *et al.* (2011) Postnatal exposure to PCB 153 and PCB 180, but not to PCB 52, produces changes in activity level and stimulus control in outbred male Wistar Kyoto rats. *Behavioural and Brain Functions*, **7**, 18.

Joost S, Kalbermatten M, Bonin A (2008) Spatial Analysis Method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources*, **8**, 957–960.

Kemper KE, Saxton SJ, Bolormaa S *et al.* (2014) Selection for complex traits leaves little or no classic signatures of selection. *Genomics*, **15**, 246.

Kottelat M, Freyhof J (2007) *Handbook of European Freshwater Fishes*. Kottelat, Cornol and Freyhof, Berlin.

Laporte M, Bertolo A, Berrebi P *et al.* (2014) Detecting anthropogenic effects on a vulnerable species, the freshwater blenny (*Salaria fluviatilis*): the importance of considering key ecological variables. *Ecological Indicators*, **36**, 386–391.

Laporte M, Rodgers SM, Dion-Côté A-M *et al.* (2015) RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in Lake Whitefish (*Coregonus clupeaformis*) species pairs. *G3: Genes | Genomes | Genetics*, **5**, 1481–1491.

Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.

Lee K, Nagler J, Fournier M *et al.* (1999) Toxicological characterization of sediments from Baie des Anglais on the St Lawrence Estuary. *Chemosphere*, **39**, 1019–1035.

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination species data. *Oecologia*, **129**, 271–280.

Legendre P, Legendre L (1998) *Numerical Ecology*, 2nd English edn. Elsevier Science BV, Amsterdam.

Lemly AD (2002) Symptoms and implication of selenium toxicity in fish: the Belews Lake case example. *Aquatic Toxicology*, **57**, 39–49.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.

Liam A, Wiener M (2002) Classification and regression by randomForest. *R News*, **2**, 18–22.

Lind EE, Grahn M (2011) Directional genetic selection by pulp mill effluent on multiple populations of three-spined stickleback (*Gasterosteus aculeatus*). *Ecotoxicology*, **20**, 503–512.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Ma X, Cowles D, Carter R (2000) Effect of pollution on genetic diversity in the Bay Mussel *Mytilus galloprovincialis* and the acorn bernacle *Balanus gladula*. *Marine Environmental Research*, **50**, 559–563.

Maes G, Raeymaekers J, Pampoulie C *et al.* (2005) The catadromous European eel *Anguilla anguilla* (L.) as a model for freshwater evolutionary ecotoxicology: relationship between heavy metal bioaccumulation, condition and genetic variation. *Aquatic Toxicology*, **73**, 99–114.

Maria VL, Pacheco M, Santos MA (2006) *Anguilla anguilla* L. biochemical and genotoxic responses after in situ exposure to freshwater wetland (Pateira de Fermentelos, Portugal). *Environment International*, **32**, 510–515.

Matson C, Lambert M, McDonald T *et al.* (2006) Evolutionary toxicology: population-level effects of chronic contaminant exposure on the marsh frog (*Rana ridibunda*) of Azerbaijan. *Environmental Health Perspectives*, **114**, 547–552.

Matthiessen P, Whale GF, Rycroft RF *et al.* (1988) The joint action of pesticide tank-mixes to rainbow trout. *Aquatic Toxicology*, **12**, 61–76.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetic Research*, **23**, 23–35.

McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology and Evolution*, **17**, 285–291.

McKinney JD, Waller CL (1994) Polychlorinated biphenyls as hormonally active structural analogues. *Environmental Health Perspectives*, **102**, 290–297.

Mehinto AC, Martyniuk CJ, Spade DJ *et al.* (2012) Applications of next-generation sequencing in fish ecotoxicogenomics. *Frontiers in Genetics*, **3**, 62.

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

Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, **28**, 659–669.

Monteiro DA, Rantin FT, Kalinin AL (2009) The effects of selenium on oxidative stress biomarkers in the freshwater characid fish matrinxã, *Brycon cephalus* (Günther, 1869) exposed to organophosphate insecticide Folisuper 600 BR® (methyl parathion). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **149**, 40–49.

Mussali-Galante P, Tovar-Sánchez E, Valverde M *et al.* (2014) Genetic structure and diversity of animal populations exposed to metal pollution. *Reviews of Environmental Contamination and Toxicology*, **227**, 79–106.

Nowell LH, Capel PD, Dileanis PD (1999) Pesticides in stream sediment and aquatic biota: distribution, trends, and governing factors. Boca Raton CRC Press, *Pesticides in the Hydrologic System series*, 4, 1040.

Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**, 1061–1067.

Pavey SA, Gaudin J, Normandeau E *et al.* (2015) RAD-sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American Eel. *Current Biology*, **25**, 1666–1671.

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

Pierron F, Baudrimont M, Dufour S *et al.* (2008) How cadmium could compromise the completion of the European eel's reproductive migration. *Environmental Science and Technology*, **42**, 4607–4612.

Pierron F, Bourret V, St-Cyr J *et al.* (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, Normandeau E, Amery DeFo M *et al.* (2011) Effects of chronic metal exposure on wild fish populations revealed by high-throughput cDNA sequencing. *Ecotoxicology*, **20**, 1388–1399.

Pierron F, Baillon L, Sow M *et al.* (2014) Effect of low-dose cadmium exposure on DNA methylation in the endangered European eel. *Environmental Science and Technology*, **48**, 797–803.

Poland J, Endelman J, Dawson J *et al.* (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome*, **5**, 103.

Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**, 208–215.

Pujolar JM, Marino IAM, Massimo M *et al.* (2012) Surviving in a toxic world: transcriptomics and gene expression profiling in response to environmental pollution in the critically endangered European eel. *BMC Genomics*, **13**, 507.

Pujolar JM, Milan M, Marino IAM *et al.* (2013) Detecting genome-wide gene transcription profiles associated with high pollution burden in the critically endangered European eel. *Aquatic Toxicology*, **132–133**, 157–164.

Ribeiro R, Lopes I (2013) Contaminant driven genetic erosion and associated hypotheses on allele loss, reduced population growth rate and increased susceptibility to future stressors: an essay. *Ecotoxicology*, **22**, 889–899.

Robinet TT, Feunteun EE (2002) Sublethal effects of exposure to chemical compounds: a cause for the decline in Atlantic eels? *Ecotoxicology*, **11**, 265–277.

Rodriguez-Jorquera IA, Kroll KJ, Toor GS *et al.* (2015) Transcriptional and physiological response of fathead minnows (*Pimephales promelas*) exposed to Urban waters entering into wildlife protected areas. *Environmental Pollution*, **199**, 155–165.

Scott WB, Crossman EJ (1973) *Freshwater Fishes of Canada*. Fisheries Research Board of Canada Bulletin, 184, Ottawa, Ontario.

Selvaraj V, Yeager-Armstead M, Murray E (2012) Protective and antioxidant role of selenium on arsenic trioxide-induced oxidative stress and genotoxicity in the fish hepatoma cell line PLHC-1. *Environmental Toxicology and Chemistry.*, **31**, 2861–2869.

Shi T, Seligson D, Belldegrun AS *et al.* (2005) Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. *Modern Pathology*, **18**, 547–557.

Smith TB, Bernatchez L (2008) Evolutionary change in human-altered environments. *Molecular Ecology*, **17**, 1–8.

Song Y, Shi Y, Yu H *et al.* (2011) p, p'-Dichlorodiphenoxydichloroethylene induced apoptosis of Sertoli cells through oxidative stress-mediated p38 MAPK and mitochondrial pathway. *Toxicology Letters*, **202**, 55–60.

Soontornchat C, Li MH, Cooke PS *et al.* (1994) Toxicokinetic and toxicodynamic influences on endocrine disruption by polychlorinated biphenyls. *Environmental Health Perspectives*, **102**, 568–571.

Stephan W (2016) Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, This issue.

Sutherland BJG, Poley JD, Igboeli OO *et al.* (2014) Transcriptomic responses to emamectin benzoate in Pacific and Atlantic Canada salmon lice *Lepeophtheirus salmonis* with differing levels of drug resistance. *Evolutionary Applications*, **8**, 133–148.

Tabrez S, Priyadarshini M, Priyamvada S *et al.* (2014) Gene-environment interactions in heavy metal and pesticide carcinogenesis. *Mutation Research*, **760**, 1–9.

Tang R, Sinnwell JP, Jia L *et al.* (2009) Identification of genes and haplotypes that predict rheumatoid arthritis using random forests. *BMC Proceedings*, **3**, 68.

Tapie N, Le Menach K, Pasquaud S *et al.* (2011) PBDE and PCB contamination of eels from the Gironde estuary: From glass eels to silver eels. *Chemosphere*, **83**, 175–185.

Teles M, Pacheco M, Santos M (2007) Endocrine of metabolic responses of *Anguilla anguilla* L. caged in freshwater wetland (Pateira de Fermentelos, Portugal). *Science of the Total Environment*, **372**, 562–570.

Thibert-Plante X, Hendry AP (2011) The consequences of phenotypic plasticity for ecological speciation. *Journal of Evolution Biology*, **24**, 326–342.

Türkmen M, Tepe Y, Türkmen A *et al.* (2013) Assessment of heavy metal contamination in various tissues of six ray species from İskenderun Bay, northeastern Mediterranean Sea. *Bulletin of Environmental Contamination and Toxicology*, **90**, 702–707.

Ulrik MG, Pujolar JM, Ferchaud AL *et al.* (2014) Do North Atlantic Eels show parallel patterns of spatially varying selection? *Journal of Evolutionary Biology*, **14**, 138.

Van Ginneken VJT, Van den Thillart GEEJM (2000) Eel fat stores enough to reach the Sargasso. *Nature*, **403**, 156–157.

Vitousek PM, Mooney HA, Lubchenco J *et al.* (1997) Human domination of Earth's ecosystems. *Science*, **277**, 494–499.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 822–831.

Weis JS, Candelmo A (2012) Pollutants and fish predator/prey behavior: a review of laboratory and field approaches. *Current Zoology*, **58**, 9–20.

Xu M, Tantisira KG, Wu A *et al.* (2011) Genome wide association study to predict severe asthma exacerbations in children using random forests classifiers. *BMC Medical Genetics*, **12**, 90.

Yadetie F, Karlsen OA, Eide M *et al.* (2014) Liver transcriptome analysis of Atlantic cod (*Gadus morhua*) exposed to PCB 153 indicates effects on cell cycle regulation and lipid metabolism. *BMC Genomics*, **15**, 481.

Yeomans KA, Golder PA (1982) The Guttman-Kaiser criterion as a predictor of the number of common factors. *Statistician*, **31**, 221–229.

Zeeberg BR, Feng W, Wang G *et al.* (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biology*, **4**, 28.

Zhao Y, Chen F, Zhai RH *et al.* (2012) Correction for population stratification in random forest analysis. *International Journal of Epidemiology*, **41**, 1798–1806.

M. L., S.A.P. and L.B. wrote the initial manuscript. M. L. and S.A.P. analysed the data. L.B., P.C., M.B. and F.P. designed the project and supervised field operation. C.R. prepared the genomic library. All these authors have contributed to editing and revising the manuscript. H.B., M. L., E.G. and P.L. participated in the data genesis of the European Eel contamination.

Data accessibility

Pollutants data, full genotype for all filtered markers and important markers only, distance matrix input for PCoA analyses all result of PCoA analyses results, FST and delta p for each markers and SNP annotation results are available on Dryad doi:10.5061/dryad.g3122. Raw demultiplexed sequences and alignment files are available on SRA, SRA study accession SRP063740.

Supporting information

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

Appendix S1. Gene Ontology (GO) enrichment for 141 and 142 important markers in European and American eels that differentiate eels from 'control' vs. 'polluted' sampling sites including flanking regions of 5000 bp (before the start of the first exon and after the finish of the final exon).

Appendix S2. Bayesian test for selection on individual SNPs in BAYESCAN v 1.21 for the European Eel with a prior odd of 10.

Appendix S3. Bayesian test for selection on individual SNPs in BAYESCAN v 1.21 for the American Eel with a prior odd of 10.