MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 4785–4798 doi: 10.1111/mec.12896

Genomic footprints of speciation in Atlantic eels (*Anguilla anguilla* and *A. rostrata*)

MAGNUS W. JACOBSEN,* JOSE MARTIN PUJOLAR,* LOUIS BERNATCHEZ,† KASPER MUNCH,‡ JIANBO JIAN,§ YONGCHAO NIU§ and MICHAEL M. HANSEN*

**Department of Bioscience, Aarhus University, Ny Munkegade 114, Aarhus C DK-8000, Denmark, †Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Pavillon Charles-Eugène-Marchand, 1030 Avenue de la Médecine, Québec QC, G1V 0A6, Canada, ‡Bioinformatics Research Centre (BiRC), Aarhus University, C. F. Møllers Allé 8, Aarhus C DK-8000, Denmark, §BGI-Shenzhen, Beishan Industrial Zone, Main Building, Yantian District, Shenzhen 518083, China*

Abstract

The importance of speciation-with-geneflow scenarios is increasingly appreciated. However, the specific processes and the resulting genomic footprints of selection are subject to much discussion. We studied the genomics of speciation between the two panmictic, sympatrically spawning sister species; European (*Anguilla anguilla*) and American eel (*A. rostrata*). Divergence is assumed to have initiated more than 3 Ma, and although low gene flow still occurs, strong postzygotic barriers are present. Restriction-site-associated DNA (RAD) sequencing identified 328 300 SNPs for subsequent analysis. However, despite the presence of 3757 strongly differentiated SNPs (F_{ST} > 0.8), sliding window analyses of F_{ST} showed no larger genomic regions (i.e. hundreds of thousands to millions of bases) of elevated differentiation. Overall F_{ST} was 0.041, and linkage disequilibrium was virtually absent for SNPs separated by more than 1000 bp. We suggest this to reflect a case of genomic hitchhiking, where multiple regions are under directional selection between the species. However, low but biologically significant gene flow and high effective population sizes leading to very low genetic drift preclude accumulation of strong background differentiation. Genes containing candidate SNPs for positive selection showed significant enrichment for gene ontology (GO) terms relating to developmental processes and phosphorylation, which seems consistent with assumptions that differences in larval phase duration and migratory distances underlie speciation. Most SNPs under putative selection were found outside coding regions, lending support to emerging views that noncoding regions may be more functionally important than previously assumed. In total, the results demonstrate the necessity of interpreting genomic footprints of selection in the context of demographic parameters and life-history features of the studied species.

Keywords: gene ontology, genomics, linkage disequilibrium, RAD sequencing, selection, speciation-with-gene-flow

Received 24 April 2014; revision received 14 July 2014; accepted 17 July 2014

Introduction

Speciation is the ultimate outcome of evolution, and unravelling the underlying processes is therefore central in evolutionary biology (Coyne & Orr 2004). Allopatric speciation (geographic isolation and no gene flow) has historically been favoured over parapatric (incomplete restrictions on gene flow) and sympatric speciation scenarios (continuous gene flow over the entire range) (Mayr 1963; Gavrilets 2004) as gene flow may counteract divergence by homogenization of the gene pool rendering completion of the speciation process difficult (Mayr 1942, 1963; Felsenstein 1981; Coyne & Orr 2004).

Correspondence: Magnus W. Jacobsen, Fax: +45 87150201; E-mail: mwj@biology.au.dk

However, recent theoretical work shows that even sympatric speciation is possible (Kondrashov & Kondrashov 1999; Via 2001; Gavrilets 2003; Pinho & Hey 2010) and speciation–with-gene flow may be more common than previously assumed (Mallet 1995; Nosil 2008).

Reproductive isolation and thus speciation may be favoured through reinforcement due to, for example, the build-up of genetic incompatibilities (Servedio & Noor 2003), which may result from genetic drift or long-term accumulation of different mutations during an initial allopatric phase. However, in cases with continuous gene flow like parapatrically or sympatrically evolving populations, directional selection is the dominant force. In sympatry, ecological speciation can occur when linkage disequilibrium builds up between genes encoding ecologically selected traits and traits involved in assortative mating (Kondrashov & Kondrashov 1999). This can theoretically involve one gene (coding for both ecologically important traits and mating preferences) (Smith 1966) or multiple independent genes, for example controlling one or several quantitative traits (Kondrashov 1983, 1986; Kondrashov & Mina 1986). As a consequence of diversifying selection, genomic differences will slowly build up between the genomes of the diverging incipient species. In the case of speciation-with-gene flow, four temporal phases have been proposed (Feder *et al.* 2012a). Direct selection (DS) on few genes constitutes the first initial phase where overall gene flow is high. In the second phase, differentiation builds up between the species around the selected loci through the process of divergence hitchhiking (DH). This is a consequence of lower gene flow around these regions compared with the rest of the genome, which subsequently will allow neutral or adaptive changes to get established, leading to 'genomic islands of divergence'. In the third phase, gene flow further decreases leading to genomic hitchhiking (GH) and genomewide differentiation, due to the establishment of new mutations of modest and weak effect. Finally, when gene flow throughout the genome decreases to a minimum, postspeciation divergence occurs (Feder *et al.* 2012a). Although some empirical studies show good agreement with this model (Via & West 2008; Hohenlohe *et al.* 2012; Nadeau *et al.* 2012; Gagnaire *et al.* 2013), most studies have been theoretically based and there is a general lack of empirical investigations into the generality of the different phases (Feder *et al.* 2012a). In particular, the importance of DH is controversial, as theoretical studies have shown that DH may only be important under specific demographic conditions (Feder & Nosil 2010; Feder *et al.* 2012b). Thus, GH may occur without DH, for example when multiple directly selected sites reduce genomewide gene flow (Feder *et al.* 2012b). Furthermore, patterns similar to DH may arise by genomic rearrangements (Yeaman 2013), reduced gene flow due to inversions or differences in genomic recombination rates (Feder *et al.* 2012a; Renaut *et al.* 2013) or genetic hitchhiking (Barton 2000) (the latter defined as change in frequency of an allele in a population due to it being carried along at a higher (or lower) frequency with other genes under selection, as opposed to genomic hitchhiking, a between-species process where divergent selection reduces average effective migration rate globally throughout the genome). These factors may all be of importance for the specific genetic and genomic footprint of speciation-with-gene flow.

Here, we focus on the genomic footprints of speciation in the two species of Atlantic eels, European (*Anguilla anguilla*) and American eel (*A. rostrata*). The exact mode of speciation, that is sympatric or allopatric, is not known, and either possibility would be difficult to rule out given extensive contemporary overlap of spawning areas (Munk *et al.* 2010) and a potential for historical gene flow even during time periods with less spatial overlap (Jacobsen *et al.* 2014). However, empirical studies support a speciation with geneflow scenario for Atlantic eels, even if allopatric phases have occurred (Jacobsen *et al.* 2014). Both species are assumed to spawn in the thermal fronts of the southern Sargasso Sea where their spawning areas show extensive overlap in space and to some extent in time, with February–April being the spawning time for American and April–June for European eel (McCleave *et al.* 1987; Tesch 2003). After hatching, the larvae (leptocephali) are advected by the Gulf Stream and other currents towards the respective continents of North America (American eel) and Europe/North Africa (European eel) (Schmidt 1923). Here, they metamorphose into so-called glass eels and enter freshwater or coastal habitats. After a period of 4–20 years, yellow eels metamorphose into silver eels that complete the life cycle by returning to the Sargasso Sea to spawn and later die (Tesch 2003).

Molecular studies have shown the two species to be sister species within the genus *Anguilla* (Tsukamoto & Aoyama 1998; Lin *et al.* 2001; Tsukamoto *et al.* 2002; Minegishi *et al.* 2005; Teng *et al.* 2009), and both of them are considered panmictic as evidenced by microsatellite analyses (Palm *et al.* 2009; Als *et al.* 2011; Côté *et al.* 2013) and RAD (Restriction-site-associated DNA) sequencing (Pujolar *et al.* 2014b) (but see Baltazar-Soares *et al.* (2014)) for a study reporting much higher differentiation than other studies. Recent studies employing RAD sequencing (Pujolar *et al.* 2013) and mitogenome sequencing (Jacobsen *et al.* 2014) suggest historical effective population sizes (N_e) in the order of hundreds of thousands to millions, although microsatellite-based studies have estimated lower contemporary and historical N_e, around 4000–10 000 (Wirth & Bernatchez 2003;

Pujolar *et al.* 2011; Côté *et al.* 2013). Estimates of divergence time based on mitochondrial DNA analysis have ranged from 1.5 Ma to 10.5 Ma (Avise *et al.* 1986; Tsukamoto & Aoyama 1998; Minegishi *et al.* 2005), with the most recent estimate based on mitogenome sequencing suggesting divergence ca. 3.38 Ma, coinciding with and possibly triggered by the closure of the Panama Gateway (Jacobsen *et al.* 2014).

Whereas the two species show two reciprocally monophyletic mitochondrial DNA lineages, genetic differentiation at the nuclear level is low, with reported average F_{ST} values at microsatellite loci between 0.018 and 0.09 and 0.0685 for AFLP markers (Mank & Avise 2003; Wirth & Bernatchez 2003; Gagnaire *et al.* 2009; Als *et al.* 2011). Levels of gene flow have consequently been proposed to be high (Gagnaire *et al.* 2009). Indeed, hybridization occurs and hybrids have been observed at low frequency among larvae sampled in the Sargasso Sea (Als *et al.* 2011; Pujolar *et al.* 2014a), but among glass and adult eels, they are almost exclusively observed in Iceland (Avise *et al.* 1990; Albert *et al.* 2006; Gagnaire *et al.* 2009; Pujolar *et al.* 2014a). Using 86 species-diagnostic SNPs, a recent study showed that hybrids in Iceland were primarily restricted to F1-, second- and third-generation backcrosses (Pujolar *et al.* 2014a), whereas RAD sequencing identified only four admixed individuals (American eel admixture proportions ranging from 0.03 to 0.05) among 225 eels from continental Europe and North Africa (Pujolar *et al.* 2014b). This suggests strong postzygotic barriers between the species, particularly when also considering results from a study based on RNA sequencing that showed positive selection and possible cytonuclear incompatibility between the mitochondrial ATP6 gene and its nuclear interactors (Gagnaire *et al.* 2012).

The species are morphologically highly similar, with average numbers of vertebrae representing the best diagnostic morphological trait (Tesch 2003). However, pronounced life-history trait differences also exist. These relate to the substantial differences in distances from the Sargasso Sea to the North American (ca. >2000 km) and European/North African continents (>5000 km), respectively. Hence, the spawning migration by European eel is longer and expectedly more energetically demanding than for American eel. Also, the larval stage is considerably longer for European eel (~2 years) as opposed to American eel (~7–9 months) (Tesch 2003) although still debated (Tesch 2003; Zenimoto *et al.* 2011). Genetic differences associated with energetics and developmental time of the larvae therefore could underlie speciation (Avise *et al.* 1990; Jacobsen *et al.* 2014). It is possible that some of the adaptive differentiation resides in regulatory rather than coding regions, as shown in three-spine stickleback (*Gasterosteus aculeatus*) (Jones *et al.* 2012). This is supported by a transcriptome study of larvae of the two species sampled in the Sargasso Sea that showed significant differences in timing of gene expression during early development (Bernatchez *et al.* 2011).

In this study, we employed RAD sequencing (Baird *et al.* 2008; Hohenlohe *et al.* 2010) to analyse the genomic footprints of speciation in Atlantic eels. We investigated the overall genomic patterns of putative selection by conducting sliding window analyses of F_{ST} between the two species along chromosomal regions. Based on the presumed age of the species and the assumption of limited gene flow, we expected a genomic pattern corresponding to divergence or genomic hitchhiking (Feder *et al.* 2012a). We also estimated linkage disequilibrium to assess whether regions of elevated differentiation were potentially the result of selective sweeps. Finally, we identified F_{ST} outliers between species to identify genes with a putative role in the speciation process. We specifically wanted to test 1) whether genes involved in energetics and larval development were overrepresented among outliers, as expected from differences in life history between the two Atlantic eel species and 2) whether selection was mainly observed within the coding parts of the genes, directly affecting protein function, or outside genes in, for example, regulatory regions. Overall, our study contributes to a better understanding of the selective processes involved in speciation and the maintenance of species boundaries in the pelagic marine environment, where few obvious physical barriers are present and speciation-by-geneflow processes may be common (Ward *et al.* 1994; Vega & Wiens 2012; Bernardi 2013).

Material and methods

Samples and sequencing

A total of 30 American and 30 European eels, sampled as glass or yellow eels from six different localities between 1999 and 2010 (Fig 1, Table 1), were analysed by RAD sequencing (Baird *et al.* 2008; Hohenlohe *et al.* 2010), conducted by Beijing Genomics Institute (BGI, Hong Kong, China). In short, genomic DNA was digested with the restriction enzyme EcoRI, and following the preparation of libraries, 10 individuals were sequenced per lane on an Illumina Genome Analyzer II, using paired-end sequencing encompassing 90 nucleotides as detailed in Pujolar *et al.* (2013). Two eels sampled in North America were found to be hybrids based on a preliminary STRUCTURE (Pritchard *et al.* 2000) analysis. As we assume that strong postzygotic selection acts against hybrids, we omitted these individuals from subsequent analyses to avoid potential bias.

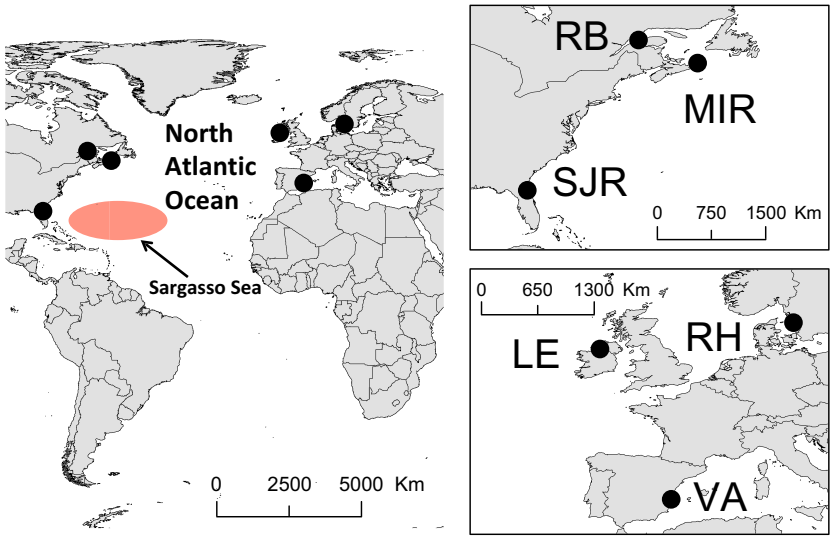


Fig. 1 Sampling locations of European and American eels (*Anguilla anguilla and A. rostrata*). The red ellipse shows the approximately location of the Sargasso Sea.

Table 1 Information about the sampled populations, location, year of sampling and sampling size

Population (abbreviation)	Country	Year of sampling	Life stage	Sample size
Riviere Blanche (RB)	Canada	2007	Y/G	15/5
Mira River (MIR)	Canada	2007	G	8
St. Johns River (SJR)	USA	1999	Y	2
Ringhals (RH)	Sweden	2008	G	10
Lough Erne (LE)	Ireland	2008	G	10
Valencia (VA)	Spain	2008	G	10

Y, yellow eel; G, glass eel.

RAD data analysis and filtering

Sequenced reads were first sorted by their unique barcode tags and then subsequently quality-filtered using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Reads were trimmed to 75 nucleotides, and a minimum Phred score of 10 per nucleotide was chosen, meaning that the whole read was dropped if one nucleotide had a lower score (see Pujolar *et al.* (2013) for details). Due to restricted coverage of the paired-end reads (without the cut site), only the first reads were used in further analyses. All individual reads were aligned to the European eel draft genome (www.eelgenome.com) (Henkel *et al.* 2012) using the ungapped aligner BOWTIE version 0.12.8 (Langmead *et al.* 2009). A maximum of two mismatches between individual reads and the genome were allowed. Individual reads that aligned to multiple positions (≥2) were discarded to decrease the risk of paralogous sequences in the data set.

The aligned data were analysed using STACKS version 0.9995 (Catchen *et al.* 2011, 2013). First, PSTACKS was used to build stacks by aligning exactly matching sequences within each individual that were in turn merged to form putative loci. At each locus, SNPs were called at each position using a maximum-likelihood framework. A minimum stack depth of 10 was used to minimize sequencing errors. CSTACKS was then used to build a catalogue of loci, matching stacks by genome position. SSTACKS was used to map each individual data set back to the catalogue to report SNPs among all individuals. Finally, POPULATIONS was used to process all loci and SNPs using only RAD loci observed in >66.7% of the individuals within each species.

The generated data files were subsequently filtered according to three final criteria to remove paralogs and otherwise spurious loci. First, loci showing >2 haplotypes in any individual were removed from the total data set. Second, all loci containing SNPs with heterozygosity (H_{obs}) of 1 (all heterozygous) or 0 (all homozygous) within one species were excluded. Finally, the average coverage of individual loci was calculated and loci with exceptionally high coverage (>1000 per individual) were first removed. Subsequently, the mean and standard deviation were calculated and loci with coverage higher than the mean plus 2 standard deviations were excluded, corresponding to 95% confidence limits.

Linkage disequilibrium analysis

Linkage disequilibrium (LD) for each species was analysed using HAPLOVIEW (Barrett *et al.* 2005) based on the 30 longest scaffolds in the European eel genome (0.91–2.04 million bp). Only SNPs with a minor allele frequency (MAF) ≥0.2 were included. LD was calculated as pairwise r^2 or D' (Barrett *et al.* 2005). To assess the

decay of LD along the scaffolds, a local weighted regression line (LOESS) (Cleveland & Devlin 1988) was fitted to the pairwise r^2-values and boxplots were made for D'-values for four categories of pairwise distance between SNPs (in bp): 1–100; 101–1000; 1001–10 000; >10 000.

Genomewide divergence between species

To assess the overall pattern of differentiation between the two species, both pairwise and sliding window (100 000 bp) based F_{ST} (using Weir's (1996) unbiased estimator) was calculated for all SNPs using POPULATIONS in STACKS (Catchen *et al.* 2011, 2013). The sliding window analysis was restricted to the 30 longest scaffolds in the European draft genome. Due to a high number of rare alleles, the same analyses were also conducted using MAF ≥ 0.05.

Candidate SNPs for being under directional selection

Two groups of candidate SNPs for directional selection were identified. The first group encompassed SNPs found to be outliers using a F_{ST}-based outlier test implemented in the software LOSITAN (2008). Only SNPs with MAF > 0.05 in at least one species were used. LOSITAN uses a coalescence-based simulation approach to identify outliers based on the distributions of F_{ST} and heterozygosity (Beaumont & Nichols 1996). The analysis was conducted using 1000 000 simulations, and as recommended in Antao *et al.* (2008), a neutral mean F_{ST} was enforced by removing potentially non-neutral loci after calculating an initial mean F_{ST}. This neutral F_{ST} was then used in a second run to establish confidence intervals and calling outliers. Confidence intervals of 0.995 and a false discovery rate (FDR) of 0.1 were assumed. The second group of candidate SNPs was defined as those showing $F_{ST} > 0.8$.

Genomewide distribution of SNPs

All postfiltered SNPs were mapped to the European eel genome gene prediction file (www.eelgenome.com) (Henkel *et al.* 2012) to assess the numbers observed in exons, coding DNA sequences (CDSs; i.e. disregarding untranslated regions of exons), introns or upstream regions (defined as the regions <5000 upstream a gene, possibly harbouring regulatory regions). Subsequently, chi-square tests were conducted to test whether SNPs associated with outlier loci were randomly distributed across the genome. This was conducted by comparing the distribution of SNPs with F_{ST} equal to 1, >0.8, and LOSITAN outliers to the background represented by the remaining SNPs.

Patterns of nonsynonymous substitutions

All SNPs present in coding regions (CDS) were analysed in the program SNPeff (Cingolani *et al.* 2012) to assess whether they represented nonsynonymous substitutions. The eel genome was annotated following Henkel *et al.* (2012) with slight modifications (see Note S1, Supporting information), and an input SNP.vcf file was created using POPULATIONS in STACKS. Chi-square tests were used to assess whether the proportion of nonsynonymous and synonymous substitutions differed significantly between groups showing high F_{ST} values ($F_{ST} = 1$, >0.8 or LOSITAN outliers) compared to the background.

Gene ontology analysis of candidate genes

Evaluation of the general functions of genes including candidate SNPs ($F_{ST} > 0.8$ or LOSITAN outliers) was conducted using the Database for Annotation, Visualization and Integrated Discovery (DAVID) webserver v. 6.7 (Huang *et al.* 2009a,b). Both exons and introns were considered part of a gene. Additional analyses of SNPs in CDS and upstream regions <5000 bp were conducted separately. Prior to the analyses, European eel transcript sequences from the European eel genome project (www.eelgenome.com) (Henkel *et al.* 2012) were blasted to the Zebra fish (*Danio rerio*) Ensembl protein database (assembly Zv9; GCA_000002035) (Howe *et al.* 2013) using BLASTX (Camacho *et al.* 2009). For each transcript, the annotation with the lowest e-scores (all $< 1 \times 10^{-10}$) and sequence similarity >50% were chosen as representative of the transcript. Following annotations, matching transcripts with outlier SNPs were extracted to make files for subsequent analyses. The protein IDs were transformed to gene IDs using the BIOMART data mining tool (Kasprzyk 2011) in the ENSEMBLE website (http://www.ensembl.org/biomart/) (Flicek *et al.* 2013). Finally, gene ontology term (GO term) enrichment analyses of 'biological processes' were analysed in DAVID using the entire annotation data set as background. Significance level was set to 0.01 using the EASE score (Huang *et al.* 2009a).

Results

RAD data analysis and filtering

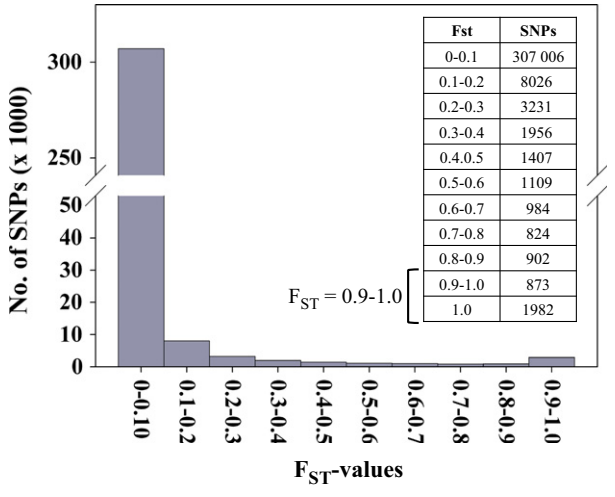
An average of 9.9 ± 3.2 and 9.3 ± 2.5 million reads per individual were sequenced for European and American eel, respectively. After quality filtering, 7.9 ± 2.1 (84.8%) and 8.1 ± 2.6 (82.5%) million reads were retained (Table S1, Supporting information). As expected, a higher number of European eel reads

aligned to the European eel draft genome (5.7 ± 1.8 million reads; 70.1%) in comparison with American eel reads (5.1 ± 1.4 million reads; 64.7%) (Table S2, Supporting information). Discarded sequences due to alternative alignments were similar in both species (ca. 4.3%) (Table S2, Supporting information).

Increasing the number of allowed mismatches in BOWTIE, from two to three for American eel, led to an increase of ~3.8% of the number of called RAD loci. However, the total number of SNPs increased considerably more by ~11%, suggesting that alignment led to an increase of paralogous loci. Thus, downstream analyses were conducted using only the data set allowing two mismatches for both species.

After filtering, a total of 67 583 RAD loci were retained comprising ~20% of the prefiltering loci (Tables S3 and S4, Supporting information). The loci contained 328 300 SNPs, 75 337 with MAF > 0.05 in at least one species. American eel showed slightly higher levels of variation, both in terms of total number of SNPs (186 534 vs. 183 787, 42 002 shared [22.5 and 22.9%]) and SNPs with MAF > 0.05 (42 821 vs. 40 801, 8285 shared [19.3 and 20.3%]). Nucleotide diversity (π) also was higher for American eel (0.00 366 vs. 0.00 338 in European eel).

Using all SNPs, the average F_{ST} between species was estimated to 0.041. The distribution of F_{ST} across all the SNPs showed an L-shaped distribution with higher number of SNPs exhibiting low F_{ST} (Fig. 2). However, the last two categories (F_{ST} = 0.8–0.9 and 0.9–1) showed increased numbers of SNPs compared with the preceding categories, suggesting that selection rather than drift may have shaped differentiation at these loci. In the latter category, this was especially due to a high number of SNPs with differentially fixed alleles (N = 1982).



Fst	SNPs
0-0.1	307 006
0.1-0.2	8026
0.2-0.3	3231
0.3-0.4	1956
0.4.0.5	1407
0.5-0.6	1109
0.6-0.7	984
0.7-0.8	824
0.8-0.9	902
0.9-1.0	873
1.0	1982

Fig. 2 Distribution of F_{ST} for the 328 320 SNPs. The inserted table shows the exact number of SNPs in each category.

Linkage disequilibrium analysis

The plot of pairwise r^2-values revealed rapid decay of linkage disequilibrium (LD) over the 30 scaffolds analysed in both species (Fig. 3). This was also evident for the D′-measurements, where already the second category (101–1000 bp) showed reduced LD by decreases of the lower 25% quartiles in both species, thus indicating extremely rapid decay of LD within 1000 bp. For the D′ estimation, the '101–1000' category was represented by few observations. This was, however, expected given an estimated mean number of SNPs per RAD locus of 4.86 (328 300 SNPs/67 583 RAD loci) and a mean distance between RAD loci of 16 296 bp (1101 362 346 base positions in draft genome/67 583 RAD loci).

Candidate SNPs for selection

LOSITAN analysis resulted in 29 101 (8.86%) outlier SNPs that are candidates for marking chromosomal regions under directional selection between the two species. Using a criterion of F_{ST} > 0.8, a total of 3757 loci were identified as candidates for directional selection (Fig. 2).

Genomewide divergence between species

SNPs showing high F_{ST} values (F_{ST} > 0.8) as well as LOSITAN outliers were in general scattered across the scaffolds (Fig. 4). Sliding window F_{ST} values showed flat distributions across scaffolds and revealed no larger regions of increased differentiation (Fig. 4), in accordance with the rapid decay of linkage disequilibrium across scaffolds (Fig. 3). Using only SNPs with MAF > 0.05, sliding windows F_{ST} showed more variation but based on fewer data points (Fig. S1, Supporting information).

Genomewide distribution of SNPs and patterns of nonsynonymous substitutions

Candidate SNPs (defined by the criteria of F_{ST} = 1, >0.8 and LOSITAN outliers) were found slightly more often in CDSs, exons and introns compared to the distribution of the background categories (Table 2). However, none of the differences were significant as based on chi-square tests. Only the upstream regions demonstrated significant differences, where SNPs showing F_{ST} = 1 were significantly underrepresented (Table 2).

The percentage of nonsynonymous substitutions was slightly elevated in the F_{ST} = 1 and >0.8 categories compared to the backgrounds (Table 3) while slightly lower for the LOSITAN outliers. However, again the differences were not statistically significant.

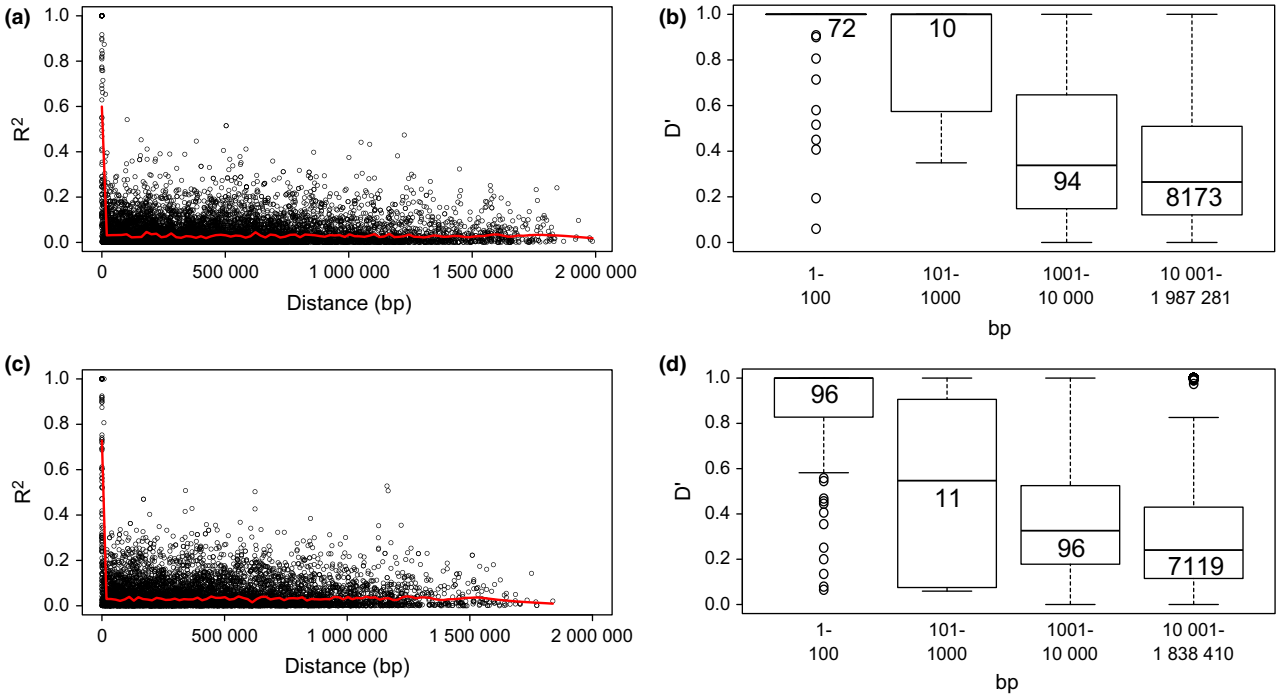


Fig. 3 Plot showing the decay of linkage disequilibrium estimated along the 30 longest scaffolds of the European eel draft genome for (a+b) European and (c+d) American eel. (a+c) r^2-values, (b+d) D'-values. The red line denotes the local weighted regression line. The white box denotes the 25% quartile, the error bar denotes the highest value excluding outliers, and the horizontal line represents the median values.

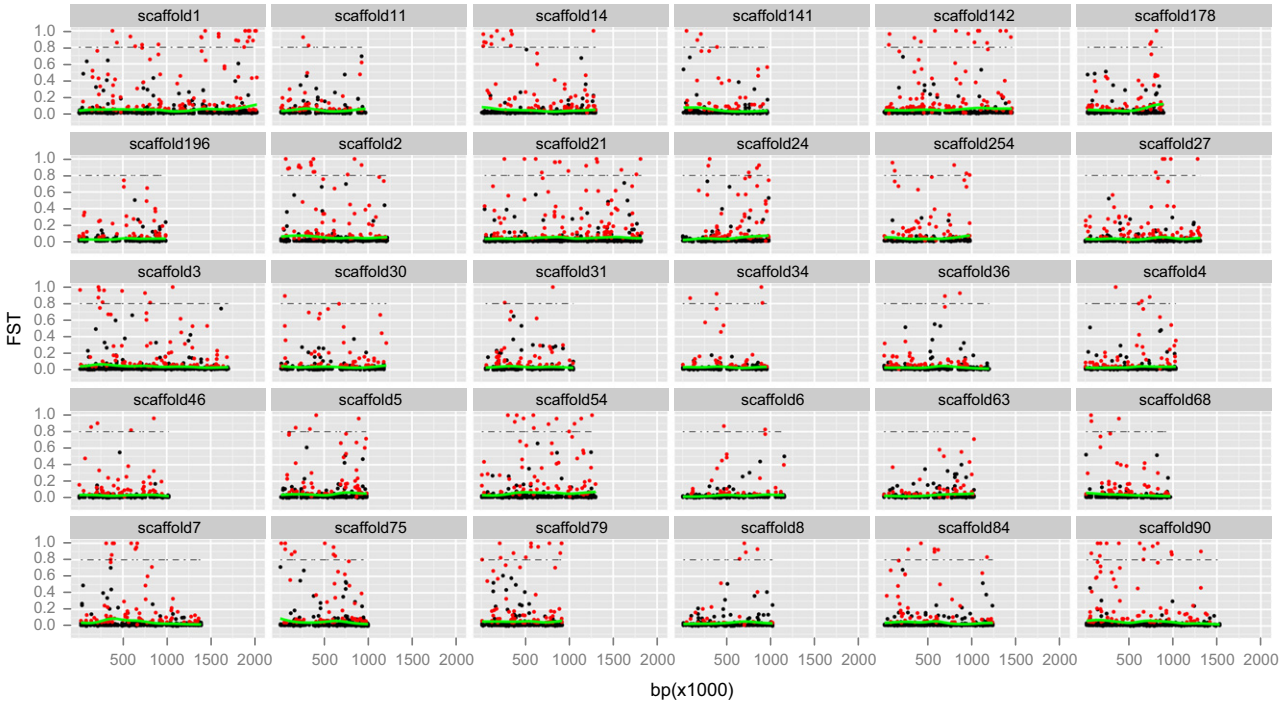


Fig. 4 Plots of F_{ST} across the 30 longest scaffolds. Dots represent pairwise estimates for individual SNPs, whereas the green lines are estimates using a sliding window of 100 000 bp. Red dots represent outliers detected using LOSITAN. Dashed lines indicate F_{ST} = 0.8.

Table 2 Distribution of SNPs and chi-square analyses comparing the distribution of the specific F_{ST}-categories to the background. The *P*-values are shown in the same row as the background tested

Groups	Total SNPs	In exon			In CDS			In Intron			Upstream (5000 bp)		
		SNPs	% of total	χ^2	SNPs	% of total	χ^2	SNPs	% of total	χ^2	SNPs	% of total	χ^2
$F_{ST} = 1$	1982	138	6.96	—	64	3.23	—	826	41.68	—	212	10.70	—
$F_{ST} < 1$	326 318	22 022	6.75	n.s.	9279	2.84	n.s.	132 026	40.46	n.s.	44 016	13.49	$P < 0.01$
$F_{ST} > 0.8$	3757	258	6.87	—	115	3.06	—	1552	41.30	—	476	12.76	—
$F_{ST} < 0.8$	324 542	21 902	6.75	n.s.	9228	2.84	n.s.	131 300	40.46	n.s.	4752	13.48	n.s.
LOSITAN	29 101	1973	6.78	—	837	2.88	—	12 026	41.33	—	3886	13.35	—
Rest	299 199	20 187	6.75	n.s.	8506	2.84	n.s.	120 826	40.38	n.s.	40 342	13.48	n.s.
All SNPs	328 300	22 160	6.75	—	9343	2.85	—	132 852	40.47	—	44 228	13.48	—
Genome*	—	—	5.45	—	—	2.97	—	—	32.01	—	—	—	—

*Calculated using the European draft genome gene prediction file (www.eelgenome.com) compared to the overall number of bases in the European eel draft genome.

Table 3 Distribution of nonsynonymous and synonymous substitutions and chi-square analyses comparing the distribution of the specific F_{ST}-categories to the background. The *P*-values are shown in the same row as the background tested

Groups	Total SNPs	Nonsynonymous					Synonymous		
		AA change	Stop gained/lost	Start lost	SNPs	% of total	SNPs	% of total	χ^2
$F_{ST} = 1$	64	32	3/0	0	35	54.69	29	45.31	—
$F_{ST} < 1$	9279	4350	240/32	8	4630	49.90	4649	50.10	n.s.
$F_{ST} > 0.80$	115	58	3/0	0	61	53.04	54	46.96	—
$F_{ST} < 0.80$	9228	4324	240/32	8	4604	49.89	4624	50.11	n.s.
LOSITAN	837	378	41 718	1	402	48.03	435	51.97	—
Rest	8506	4004	220/29	7	4263	50.12	4243	49.88	n.s.
All	9343	4382	243/32	8	4665	49.93	4678	50.07	—

Candidate genes and function

The GO enrichment analysis showed 37 enriched GO terms based on the LOSITAN outliers (intron and exons combined). Most enriched GO categories were related to one of three main categories: phosphorylation (12 GO terms), development (14 GO terms) and cell processes including transport (10 GO terms) (Table 4). Individual GO terms of developmental growth and phosphorus/phosphate metabolic processes were represented. The latter category contained two genes of the mitochondrial ATP synthase F1 complex part of the oxidative phosphorylation pathway: O subunit and delta subunit. These genes are among the candidates found to be under divergent selection although not identical to the genes reported by Gagnaire *et al.* (2012).

Gene ontology (GO) enrichment analysis of the genes containing SNPs with $F_{ST} > 0.8$ showed strong overlap with the LOSITAN outliers and yielded three significant results: 'protein amino acid phosphorylation', 'post-translational protein modification' and 'development processes' (Table 4). Here, the GO terms related to phosphorylation were dominated by protein kinases, whereas the categories relating to development primarily consisted of transcription factors, like homeobox genes and growth factors and kinases (see Tables S5 and S6, Supporting information for a list of genes). Upstream regions only showed two significant GO terms, both involving response to stimulus. CDS did not represent any significant GO terms, probably due to low sample size.

Discussion

Potential sources of bias

Our results suggest a pattern of genomic divergence between European and American eel, where background genetic differentiation is low, but at the same time, many independent regions show high or even fixed differences. However, before discussing the biological implications, it is important to consider whether technical or mutational sources of bias could have affected results.

Table 4 Gene ontology (GO)-term enrichment analysis on 'biological processes' performed in DAVID on the outlier SNPs found in LOSITAN. Significant GO terms from the smaller dataset comprising $F_{ST} > 0.8$ are denoted by a '*'. Significant GO terms from the upstream category are denoted by a '#'. Significance level equals $P < 0.01$. Fold-change denotes the increase of genes including the respective GO term in the outlier group compared to the background

GO-term	Fold-change	*P*-Value
Protein amino acid phosphorylation*	1.4	5.8E-6
Phosphorylation	1.3	9.6E-5
Regulation of Ras protein signal transduction	1.8	1.0E-4
Anatomical structure development	1.2	1.2E-4
Regulation of small GTPase mediated signal transduction	1.7	1.2E-4
Multicellular organismal process	1.2	3.0E-4
Regulation of Rho protein signal transduction	2.2	4.2E-4
System development	1.2	5.0E-4
Cell adhesion	1.5	5.7E-4
Biological adhesion	1.2	5.7E-4
Phosphorus metabolic process	1.2	6.6E-4
Phosphate metabolic process	1.2	6.6E-4
Post-translational protein modification*	1.2	1.2E-3
Multicellular organismal development	1.1	1.2E-3
Regulation of biological quality	1.4	1.5E-3
Developmental process*	1.1	1.6E-3
Regulation of cell communication	1.4	2.0E-3
Localization	1.1	2.1E-3
Organ development	1.2	2.2E-3
Central nervous system development	1.4	2.3E-3
Regulation of signal transduction	1.4	2.8E-3
Protein modification process	1.2	2.9E-3
Embryonic organ development	1.5	2.9E-3
Brain development	1.4	3.2E-3
Anatomical structure morphogenesis	1.2	3.2E-3
Cell motion	1.4	3.5E-3
Neurotransmitter transport	2.0	3.6E-3
Cell surface receptor-linked signal transduction	1.2	3.6E-3
Developmental growth	1.8	5.0E-3
Transmembrane receptor protein tyrosine kinase signalling pathway	1.8	5.0E-3
Nervous system development	1.3	5.1E-3
Enzyme-linked receptor protein signalling pathway	1.6	5.7E-3
Regulation of cellular component size	1.8	6.7E-3
Embryonic development	1.2	7.9E-3
Regulation of neuron differentiation	2.2	8.3E-3
Metal ion transport	1.3	9.8E-3
Monovalent inorganic cation transport	1.3	9.9E-3
Response to chemical stimulus #	1.7	2.0E-3
Response to stimulus #	1.3	6.5E-3

Technical sources of bias would primarily involve sequencing errors scored as SNPs. We trimmed reads to 75 nucleotides to account for the Illumina Hiseq platform's propensity for sequencing errors in the tails of reads and conducted further filtering steps as detailed previously. Nevertheless, even if some sequencing errors remain, this is unlikely to have generated the distinct patterns of genomewide differentiation between the two species.

Mutations in restriction sites are likely to have occurred and can impose bias potentially affecting a range of summary statistics (Arnold *et al.* 2013). However, F_{ST} appears relatively robust towards such bias (Arnold *et al.* 2013), and it also cannot explain the observed patterns of genomewide genetic differentiation.

We find that the alignment of reads to the European eel genome (in the absence of a sequenced American eel genome) imposes the potentially most important bias. We allowed for a maximum of two mismatches between reads and reference genome, and this could filter out genomic regions where American eel is genetically highly divergent from European eel. However, when three mismatches were allowed, the number of RAD loci increased only marginally (by 3.8%), whereas the number of SNPs increased disproportionally by 11%. This suggests that the bias against highly differentiated regions is limited, and at the same time, allowing more mismatches might increase other problems, such as inclusion of paralogous loci. In total, when also considering differentiation along the longest scaffolds (Fig. 4), there does not appear to be long regions of missing data that could potentially include larger 'islands of genomic divergence'. We therefore conclude that the observed genomic patterns of differentiation are genuine and not the result of bias against highly differentiated regions.

Genomewide differentiation in Atlantic eels

The genomewide F_{ST} of 0.041 is in accordance with previous estimates using microsatellite markers (Mank & Avise 2003; Wirth & Bernatchez 2003; Gagnaire *et al.* 2009). The 8.86% of loci found to be outliers in LOSITAN generally correspond to other studies that find between 5% and 10% outliers (reviewed in Nosil *et al.* 2009). This includes an earlier AFLP study of the two Atlantic eel species that found 8.4% outliers (27 out of 321 loci), also using LOSITAN (Gagnaire *et al.* 2009). Among all outliers, the SNPs showing F_{ST} >0.8 are particularly strong candidates for being under selection, as confirmed by the distribution of F_{ST} values across all SNPs (Fig. 2).

Despite the many highly genetic differentiated SNPs, sliding window analyses of F_{ST} did not reveal larger

'genomic islands of divergence'. This pattern would seem congruent with the first phase of speciation with gene flow, where gene flow is high and direct selection at relatively few genes is the most important force (Feder *et al.* 2012a). However, contrary to expectations given in this scenario, we found many independent regions potentially under selection (marked by many highly differentiated SNPs). Although the distinct pattern of SNPs exhibiting high differentiation against a background of overall low differentiation could be biased due to the high proportion of rare alleles, the pattern was quite similar when using only SNPs with MAF > 0.05, although average F_{ST} increased (Fig. S1, Supporting information). Moreover, Gagnaire *et al.* (2012) observed a similar pattern in Atlantic eels where F_{ST} dropped to almost zero within 100–1000 bp of highly differentiated exons of the positively selected atp5c1 gene.

Although the observed genomic signature does not seem compatible with divergence (DH), it may be explained by genomic (GH) hitchhiking (Feder *et al.* 2012a). Pujolar *et al.* (2014a) reported evidence for strong postzygotic selection, but nevertheless also found one Icelandic individual that represented a late generation hybrid (beyond second-generation backcross). Similarly, a recent study based on RAD sequencing found four among a total of 225 European eels (excluding Iceland) and one among 30 American eels showing between 3 and 6% admixture (Pujolar *et al.* 2014b). If historical effective population size is high in the two species (Pujolar *et al.* 2013; Jacobsen *et al.* 2014), even low gene flow could have a strong effect on genetic differentiation (Pujolar *et al.* 2014a). If this is indeed the case, then the observed pattern of genomic divergence might in fact be considered in concordance with genomic hitchhiking; the two species show strong differentiation at multiple sites, presumably reflecting diversifying selection, whereas background differentiation increases extremely slowly due to the combined effects of low, but biologically significant gene flow and very low genetic drift. The SNPs identified as outliers in LOSITAN, many constituting none fixed differences between the two species (Fig. 4) may then be loci of modest or weak effect on fitness that now can increase in frequency in the face of a reduction of average genomic effective migration rate (Feder *et al.* 2012a).

A second potentially contributing factor to the observed pattern of differentiation involves the extremely low linkage disequilibrium (LD). If LD was also low during the early speciation process, then selection acting on standing genetic variation (Barrett & Schluter 2008) would lead to soft sweeps that would not result in strong building up of genetic differentiation (Hermisson & Pennings 2005). Even for sweeps involving *de novo* mutations, initial strong LD would be expected to erode over time due to the combined effects of new mutations, gene flow between species and recombination. This would apply for sweeps that took place during the early speciation phase, but not for more recent sweeps. However, the low sliding window F_{ST} and LD do not suggest the presence of recent selective sweeps in the surveyed genomic regions.

The genome-wide differentiation pattern observed in Atlantic eels may be quite common in other organisms with similar demographic features (sympatric or parapatric distribution, low gene flow, high N_e) and match results from other studies of species like *Drosophila melanogaster* (Turner *et al.* 2008), Pacific lamprey (*Entosphenus tridentatus*) (Hess *et al.* 2013), the mosquito *Anopheles gambiae* (Lawniczak *et al.* 2010) and different plants (Strasburg *et al.* 2012) also showing strong genetic differentiation confined to many small regions scattered throughout the genome. On the other side, there are also examples of marine fishes showing large and distinct genomic regions of high divergence, such as among different ecotypes of Atlantic cod (*Gadus morhua*) (Hemmer-Hansen *et al.* 2013; Karlsen *et al.* 2013). The precise mechanisms generating this pattern are presently unknown (Hemmer-Hansen *et al.* 2013), but could in addition to DH represent inversions or differences in genomic architecture (Feder *et al.* 2012a; Yeaman 2013).

Distribution of candidate SNPs in coding and noncoding regions

Given the very low observed LD, we would expect SNPs showing high differentiation to be very close to the actual targets of selection. One surprising result of our study therefore concerns the fact that outlier and nonoutlier SNPs in general did not show differential distribution across the genome (Table 2 and 3). If selection predominantly involved the coding parts of the genome, then a significantly higher proportion of the outliers would be expected in the CDS, with more being nonsynonymous substitutions. However, the proportion of nonsynonymous substitutions was not different between outlier and nonoutlier categories, and the only significant deviation from random genomewide distribution of SNPs was observed in the upstream regions, where the diagnostic SNPs showed a decreased representation compared to the background. This possibly reflects a conserved and thus important role of these regions in both species rather than being targets for directional selection.

Most of the SNPs being candidates for directional selection were located outside known exons and CDSs, which suggests that a major part of adaptive divergence between the species involves variation in noncoding

regions. This is in accordance with the emerging view that a higher proportion of noncoding genomic regions may be functional than previously assumed. It has recently been suggested that 80.4% of the human genome contain elements linked to biochemical functions (Dunham *et al.* 2012). In three-spine stickleback (*Gasterosteus aculeatus*), up to 83% of SNPs diagnostic for distinct ecotypes are assumed to have a regulatory role (Jones *et al.* 2012), and in general, many introns host enhancers and promoters (Rose 2008; Chorev & Carmel 2012; Ecker *et al.* 2012; Hebert *et al.* 2013). The high numbers of candidate SNPs in noncoding regions of the Atlantic eel genomes are therefore likely to reflect regulatory differences between the species.

An alternative explanation could be that at least some of the fixed differences between the species reflect incompatibilities, for example due to inversions or other chromosomal rearrangements, rather than selection (Bierne *et al.* 2011). It is not possible to investigate this possibility using our RAD sequencing data of the present study, but whole genome sequencing of three-spine sticklebacks revealed that several genomic regions involved in ecotype divergence in fact represented inversions (Jones *et al.* 2012).

Candidate genes and function

Gene ontology (GO) enrichment analysis of 'candidate genes', that is genes containing outlier SNPs in introns and exons, both supported previous results and provided further insights into functional genomic divergence between the Atlantic eel species. We observed outliers in genes related to the mitochondrial ATP synthase F1 complex, which is part of the oxidative phosphorylation pathway and thus energy (ATP) production. This is in accordance with other studies suggesting positive selection at genes of this complex (Gagnaire *et al.* 2012; Jacobsen *et al.* 2014). Selection may be driven by differences in energetic requirements due to a considerable longer spawning migration experienced by European eel (>5000 km) compared to the American eel (>2000 km) (Tesch 2003).

GO categories related to phosphorylation and development were enriched in both outlier data sets (defined by LOSITAN analysis and a criterion of $F_{ST} > 0.8$), thus representing strong support for differential selection acting on genes within these functional categories. The genes included protein kinases, transcription factors and growth factors or other proteins involved in cellular growth. These genes mainly serve regulatory roles either by controlling protein activity and function (Cohen 2000) or transcription (Vaquerizas *et al.* 2009). This matches our initial expectations as timing of gene expression has been shown to differ between the species during early development. This could be the underlying mechanism determining larval phase duration (Bernatchez *et al.* 2011), assumed to be considerably longer for European than American eel (Tesch 2003). Moreover, given the distribution of candidate SNPs for selection, it is possible that selection is mainly acting on regulation and expression rather than functional changes. For the genes related to development processes, this matches a previous transcriptome study that observed significant differences of expression of individual genes related to 'cell cycle and development' between larvae of two species of Atlantic eel (Bernatchez *et al.* 2011). However, genes related to 'protein synthesis and RNA processing' showed the most pronounced differentiation between size groups of the two species (Bernatchez *et al.* 2011). This group was mainly represented by ribosomal RNAs, but this gene family is likely to be under-represented in the present study, as we deliberately filtered out sequences that aligned to more than one place in the genome.

Finally, although Atlantic eels exhibit features making it likely that they speciated in the face of gene flow (e.g. overlapping spawning area and spawning time and very low genetic differentiation suggesting historical gene flow), it is also possible that speciation was originally initiated in allopatry. If this is the case, then traits relating to different life histories and their underlying genetic basis could also have evolved in allopatry, causing incompatibilities following secondary contact. Under this scenario, the high differentiation of genes belonging to certain functional categories would represent a product rather than being the driver of speciation. Whereas this possibility cannot be entirely dismissed, it nevertheless appears less realistic, as the very low interspecific genetic differentiation is difficult to reconcile with an allopatric phase without gene flow between the species.

Conclusions

Our results demonstrate the power of next-generation sequencing methods such as RAD sequencing for analysing patterns and processes underlying speciation. The study highlights the complexity of interpreting patterns of genomic footprints in cases of speciation with gene flow. The many instances of very small genomic regions exhibiting high differentiation between the species, but without any appreciable increase of background genomic differentiation, suggest a phase of genome hitchhiking where drift and gene flow lead to only very slow neutral divergence. These findings stress the importance of considering genomic footprints of speciation-with-gene-flow in conjunction with demographic parameters such as effective population size.

Our study additionally suggests functional differences at genes involved in development and phosphorylation as playing an important role in speciation of Atlantic eels. This is in accordance with biological knowledge of differences in larval phases and migratory distances of the two species (Tesch 2003). A major role must, however, also be ascribed to regulatory processes, and our results adds to the growing body of results finding considerable amounts of seemingly adaptive divergence in noncoding genomic regions.

Acknowledgements

The authors acknowledge the Danish Council for Independent Research, Natural Sciences (grant 09-072 120 to MMH) and Elisabeth og Knud Petersen's Foundation (grant to MWJ) for funding. We thank Annie Brandstrup for technical assistance, Eric Normandeau for bioinformatics advice, Michael Glad for keeping computers running, Virginia Settepani for drawing the sampling map, Thomas D. Als, Håkan Wickström, Russell Poole and Javier Lobon-Cervia for providing samples, and the subject editor Shawn Narum and three anonymous referees for constructive comments and suggestions.

References

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

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

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179–3190.

Avise JC, Helfman GS, Saunders NC, Hales LS (1986) Mitochondrial DNA differentiation in North Atlantic eels: population genetic consequences of an unusual life-history pattern. *Proceedings of the National Academy of Sciences of the United States of America*, **83**, 4350–4354.

Avise JC, Nelson WS, Arnold J *et al.* (1990) The evolutionary genetic status of Icelandic Eels. *Evolution*, **44**, 1254–1262.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Baltazar-Soares M, Biastoch A, Harrod C *et al.* (2014) Recruitment collapse and population structure of the european eel shaped by local ocean current dynamics. *Current Biology*, **24**, 104–108.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Barton NH (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **355**, 1553–1562.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences*, **263**, 1619–1626.

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

Bernatchez L, St-Cyr J, Normandeau E *et al.* (2011) Differential timing of gene expression regulation between leptocephali of the two Anguilla eel species in the Sargasso Sea. *Ecology and Evolution*, **1**, 459–467.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST plus: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and Genotyping Loci De Novo From Short-Read Sequences. *G3-Genes Genomes Genetics*, **1**, 171–182.

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

Chorev M, Carmel L (2012) The function of introns. *Front Genet*, **3**, 55.

Cingolani P, Platts A, Wang LL *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly*, **6**, 80–92.

Cleveland WS, Devlin SJ (1988) Locally weighted regression – an approach to regression-analysis by local fitting. *Journal of the American Statistical Association*, **83**, 596–610.

Cohen P (2000) The regulation of protein function by multisite phosphorylation - a 25 year update. *Trends in Biochemical Sciences*, **25**, 596–601.

Côté CL, Gagnaire PA, 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.

Coyne JA, Orr HA (2004) *Speciation*. Sinnauer Associates, Sunderland, Massachusetts.

Dunham I, Kundaje A, Aldred SF *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

Ecker JR, Bickmore WA, Barroso I *et al.* (2012) Genomics: ENCODE explained. *Nature*, **489**, 52–55.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.

Feder JL, Egan SP, Nosil P (2012a) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Feder JL, Gejji R, Yeaman S, Nosil P (2012b) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 461–474.

Felsenstein J (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals. *Evolution*, **35**, 124–138.

Flicek P, Ahmed I, Amode MR *et al.* (2013) Ensembl 2013. *Nucleic Acids Research*, **41**, D48–D55.

Gagnaire PA, Albert V, Jonsson B, Bernatchez L (2009) Natural selection influences AFLP intraspecific genetic variability and introgression patterns in Atlantic eels. *Molecular Ecology*, **18**, 1678–1691.

Gagnaire PA, Normandeau E, Bernatchez L (2012) Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Molecular Biology and Evolution*, **29**, 2909–2919.

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

Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.

Gavrilets S (2004) *Fitness Landscapes and the Origin Of Species*. Princeton University Press, Princeton, New Jersey.

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

Hemmer-Hansen J, Nielsen EE, Therkildsen NO *et al.* (2013) A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, **22**, 2653–2667.

Henkel CV, Burgerhout E, de Wijze DL *et al.* (2012) Primitive duplicate hox clusters in the European Eel's genome. *PLoS ONE*, **7**, e32231.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–2352.

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR (2013) Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, **22**, 2898–2916.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD Tags. *Plos Genetics*, **6**, e1000862.

Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 395–408.

Howe K, Clark MD, Torroja CF *et al.* (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, **496**, 498–503.

Huang DW, Sherman BT, Lempicki RA (2009a) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4**, 44–57.

Huang DW, Sherman BT, Lempicki RA (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, **37**, 1–13.

Jacobsen MW, Pujolar JM, Gilbert MTP *et al.* (2014) Speciation and demographic history of Atlantic eels (*Anguilla anguilla* and *A. rostrata*) revealed by mitogenome sequencing. *Heredity*, **112**, in press.

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

Karlsen BO, Klingan K, Emblem A *et al.* (2013) Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, **22**, 5098–5111.

Kasprzyk A (2011) BioMart: driving a paradigm change in biological data management. *Database*, **2011**, bar049.

Kondrashov AS (1983) Multilocus Model of Sympatric Speciation. 2. 2 Characters. *Theoretical Population Biology*, **24**, 136–144.

Kondrashov AS (1986) Multilocus model of sympatric speciation. 3. Computer-Simulations. *Theoretical Population Biology*, **29**, 1–15.

Kondrashov AS, Kondrashov FA (1999) Interactions among quantitative traits in the course of sympatric speciation. *Nature*, **400**, 351–354.

Kondrashov AS, Mina MV (1986) Sympatric speciation – when is it possible. *Biological Journal of the Linnean Society*, **27**, 201–223.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.

Lawniczak MKN, Emrich SJ, Holloway AK *et al.* (2010) Widespread divergence between incipient anopheles gambiae species revealed by whole genome sequences. *Science*, **330**, 512–514.

Lin YS, Poh YP, Tzeng CS (2001) A phylogeny of freshwater eels inferred from mitochondrial genes. *Molecular Phylogenetics and Evolution*, **20**, 252–261.

Mallet J (1995) A species definition for the modern synthesis. *Trends in Ecology & Evolution*, **10**, 294–299.

Mank JE, Avise JC (2003) Microsatellite variation and differentiation in North Atlantic eels. *Journal of Heredity*, **94**, 310–314.

Mayr E (1942) *Systematics and the Origin of Species*. Columbio University Press, New York.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.

McCleave JD, Kleckner RC, Castonguay M (1987) Reproductive sympatry of American and European eels and implications for migration and taxonomy. *American Fisheries Society Symposium*, **1**, 286–297.

Minegishi Y, Aoyama J, Inoue JG *et al.* (2005) Molecular phylogeny and evolution of the freshwater eels genus Anguilla based on the whole mitochondrial genome sequences. *Molecular Phylogenetics and Evolution*, **34**, 134–146.

Munk P, Hansen MM, Maes GE *et al.* (2010) Oceanic fronts in the Sargasso Sea control the early life and drift of Atlantic eels. *Proceedings of the Royal Society Series B: Biological Sciences*, **277**, 3593–3599.

Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 343–353.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Palm S, Dannewitz J, Prestegaard T, Wickstrom H (2009) Panmixia in European eel revisited: no genetic difference between maturing adults from southern and northern Europe. *Heredity*, **103**, 82–89.

Pinho C, Hey J (2010) Divergence with Gene Flow: models and Data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215–230.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pujolar JM, Bevacqua D, Capoccioni F *et al.* (2011) No apparent genetic bottleneck in the demographically declining Euro-

pean eel using molecular genetics and forward-time simulations. *Conservation Genetics*, **12**, 813–825.
Pujolar JM, Jacobsen MW, Frydenberg J *et al.* (2013) A resource of genome-wide single-nucleotide polymorphisms generated by RAD tag sequencing in the critically endangered European eel. *Molecular Ecology Resources*, **13**, 706–714.
Pujolar JM, Jacobsen MW, Als TD *et al.* (2014a) Assessing patterns of hybridization between North Atlantic eels using diagnostic single nucleotide polymorphisms. *Heredity*, **112**, 627–637.
Pujolar JM, Jacobsen MW, Als TD *et al.* (2014b) Genome-wide single generation signatures of local selection in the panmictic European eel. *Molecular Ecology*, **23**, 2514–2528.
Renaut S, Grassa CJ, Yeaman S *et al.* (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, **4**, 1827.
Rose AB (2008) Intron-mediated regulation of gene expression. *Current Topics in Microbiology and Immunology*, **326**, 277–290.
Schmidt J (1923) The breeding places of the eel. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **211**, 179–208.
Servedio MR, Noor MAF (2003) The role of reinforcement in speciation: theory and data. *Annual Review of Ecology Evolution and Systematics*, **34**, 339–364.
Smith JM (1966) Sympatric speciation. *American Naturalist*, **100**, 637–650.
Strasburg JL, Sherman NA, Wright KM *et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 364–373.
Teng HY, Lin YS, Tzeng CS (2009) A new Anguilla species and a reanalysis of the phylogeny of freshwater eels. *Zoological Studies*, **48**, 808–822.
Tesch F (2003) *The Eel*. Blackwell Science Ltd, Oxford.
Tsukamoto K, Aoyama J (1998) Evolution of freshwater eels of the genus Anguilla: a probable scenario. *Environmental Biology of Fishes*, **52**, 139–148.
Tsukamoto K, Aoyama J, Miller MJ (2002) Migration, speciation, and the evolution of diadromy in anguillid eels. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 1989–1998.
Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in Drosophila melanogaster. *Genetics*, **179**, 455–473.
Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics*, **10**, 252–263.
Vega GC, Wiens JJ (2012) Why are there so few fish in the sea? *Proceedings of the Royal Society B-Biological Sciences*, **279**, 2323–2329.
Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381–390.
Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.
Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, fresh-water, and anadromous fishes. *Journal of Fish Biology*, **44**, 213–232.
Weir BS (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, Massachusetts.
Wirth T, Bernatchez L (2003) Decline of North Atlantic eels: a fatal synergy? *Proceedings of the Royal Society B-Biological Sciences*, **270**, 681–688.
Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E1743–E1751.
Zenimoto K, Sasai Y, Sasaki H, Kimura S (2011) Estimation of larval duration in *Anguilla* spp., based on cohort analysis, otolith microstructure, and Lagrangian simulations. *Marine Ecology-Progress Series*, **438**, 219–228.

M.M.H., M.W.J, L.B. and J.M.P. conceived and designed the study. M.W.J., J.M.P. and K.M. conducted bioinformatics and population genomics analyses. J.J. and Y.N. were involved in data generation. M.W.J. wrote the manuscript with contributions from M.M.H., J.M.P., L.B., K.M., J.J. and Y.N.

Data accessibility

Sequence reads have been deposited in the NCBI Sequence Read Archive under project number PRJNA195555 (European eel) and PRJNA230782 (American eel). For individual codes, see Table S7, Supporting information. Raw SNP data are available from the Dryad database (http://datadryad.org) under doi:10.5061/dryad.f2313.

Supporting information

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

Table S1 Information about the filtering process for the American eel and European eel samples.

Table S2 Information about alignment of filtered reads for the American eel and European samples.

Table S3 Summary information on reads after p-STACKS (p-stacks; $M = 10$) for American and European eel samples.

Table S4 Filtering on the final number of loci after POPULATIONS.

Table S7 Individual NCBI codes of the used RAD sequenced samples.

Note S1 Information about the annotation of the genome prior to analyses of nonsynonymous and synonymous substitutions in SNPeff (Cingolani *et al.* 2012).

Fig S1 Plots of F_{ST} across the 30 longest scaffolds using a minor allele frequency (MAF) > 0.05.

Table S5 Genes represented by one or several SNPs with $F_{ST} > 0.8$ in either the genes themselves (including both introns and exon) or ≤5000 bp upstream the gene.

Table S6 Genes or upstream regions represented by LOSITAN SNP outliers, candidates for positive selection.