MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2017) **17**, 806–811 doi: 10.1111/1755-0998.12608

Draft genome of the American Eel (*Anguilla rostrata*)

SCOTT A. PAVEY,*† MARTIN LAPORTE,* ERIC NORMANDEAU,* JÉRÉMY GAUDIN,* LOUIS LETOURNEAU,‡ SÉBASTIEN BOISVERT,§ JACQUES CORBEIL,§ CÉLINE AUDET¶ and LOUIS BERNATCHEZ*

*Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Pavillon Charles-Eugène-Marchand, Québec, QC G1V 0A6, Canada, †Department of Biological Sciences and Canadian Rivers Institute, University of New Brunswick, Saint-John, NB E2L 4L5, Canada, ‡McGill University and Génome Québec Innovation Centre, Montréal, QC H3A0G1, Canada, §Faculty of Medicine, CHUL Research Center, Université Laval, Québec, QC G1V4G2, Canada, ¶Institut des sciences de la mer de Rimouski, Université du Québec à Rimouski, Rimouski, Québec G5L 3A1, Canada

Abstract

Freshwater eels (*Anguilla* sp.) have large economic, cultural, ecological and aesthetic importance worldwide, but they suffered more than 90% decline in global stocks over the past few decades. Proper genetic resources, such as sequenced, assembled and annotated genomes, are essential to help plan sustainable recoveries by identifying physiological, biochemical and genetic mechanisms that caused the declines or that may lead to recoveries. Here, we present the first sequenced genome of the American eel. This genome contained 305 043 contigs (N50 = 7397) and 79 209 scaffolds (N50 = 86 641) for a total size of 1.41 Gb, which is in the middle of the range of previous estimations for this species. In addition, protein-coding regions, including introns and flanking regions, are very well represented in the genome, as 95.2% of the 458 core eukaryotic genes and 98.8% of the 248 ultra-conserved subset were represented in the assembly and a total of 26 564 genes were annotated for future functional genomics studies. We performed a candidate gene analysis to compare three genes among all three freshwater eel species and, congruent with the phylogenetic relationships, Japanese eel (*A. japanica*) exhibited the most divergence. Overall, the sequenced genome presented in this study is a crucial addition to the presently available genetic tools to help guide future conservation efforts of freshwater eels.

Keywords: American eel, *Anguilla rostrata*, conservation, freshwater eels, full-genome sequencing, functional genomics

Received 12 March 2016; revision received 12 October 2016; accepted 13 October 2016

Introduction

Freshwater eels (*Anguilla* sp.) have large economic, cultural, ecological and aesthetic importance worldwide (ASMFC 2000; COSEWIC 2012; ICES 2013; Jacoby & Gollock 2014; Arai 2016). They are phylogenetically situated in the least derived teleostei lineage (*Elopomorpha*; Betancur *et al.* 2013) and their behaviours, reproductive strategy and physiology are unusual and extreme compared with other fishes (*i.e.* catadromous, inhabiting drastically different environments and species panmixia; Als *et al.* 2011; Côté *et al.* 2013; Gong *et al.* 2014; Pavey *et al.* 2015). The high demand for human consumption coupled with (i) habitat destruction, (ii) water pollution, (iii) arrival of new diseases and iv) the inability for commercial production to complete the life cycle (except for a small scale success with Japanese eel: *Anguilla japonica*; Tanaka *et al.* 2003) has resulted in a more than 90% decline in global stocks over the past decades (Dekker *et al.* 2003; Busch & Braun 2014). Functional genomics studies may help to plan sustainable recoveries by identifying physiological, biochemical and genetic mechanisms that led to declines (Savolainen *et al.* 2013). Therefore, proper genetic resources such as sequenced, assembled and annotated genomes are required to help those species.

Recently, draft genomes of European Eel (*A. anguilla*) and Japanese eel have been assembled (Henkel *et al.* 2012a,b). This has facilitated several recent functional genomics studies, including within-generation selection in European eel (Pujolar *et al.* 2014, 2015), transcriptomics response to pollutants (Baillon *et al.* 2015) and further investigation on osmotic response and fatty acid biosynthesis in Japanese eel to optimize their farming (Henkel *et al.* 2012b; Tse *et al.* 2014; Wang *et al.* 2014). In addition, comparative genomics between American eel (*A. rostrata*) and European eel were made possible by aligning high-throughput reads from both species to the

Correspondence: Scott A. Pavey; E-mail: scott.pavey@unb.ca

European eel genome (Jacobsen *et al.* 2014a). However, the genome of the American eel is still missing and its availability would allow finer functional genomics analyses and better support of the conservation effort on this species.

The American eel is the only freshwater eel present in North America (Tesch & Thorpe 2003; Froese & Pauly 2015). It spawns in the Sargasso Sea as a single panmictic population (Côté *et al.* 2013). Eggs hatch into leptocephali and drift in the North Atlantic current. Then, the eels metamorphose into glass eels, at which point they begin feeding in fresh, brackish or saltwater habitats. Depending on the rearing habitats, they exhibit drastically different phenotypes and differ in age of maturity, growth rate and sex ratios (Côté *et al.* 2015). Until recently, these differences were assumed to be plastic because no population structure has been detected using neutral markers (Côté *et al.* 2013). However, functional genetic differences were recently found and are either the result of within-generation selection and/or genotype-dependent habitat choice (Côté *et al.* 2014; Pavey *et al.* 2015). Furthermore, within-generation selection in relation to water temperature and anthropogenic pollution has also been observed, suggesting considerable functional genetic differentiation among eels inhabiting different sites (Gagnaire *et al.* 2012; Laporte *et al.* 2016). Together, this confirms the need for an American eel genome in order to better understanding the ecological genomics of the species. Here, we present a *de novo* assembly, scaffolding and annotation of the American eel genome. All of this work has been deposited at DDBJ/ENA/GenBank under the Accession no. LTYT00000000. The version described in this study is version LTYT01000000. We then demonstrate the usefulness of the genome by comparing three candidate genes using the genomes of the three freshwater eel species. This work adds another important tool for functional genomics of freshwater eels.

Materials and methods

Tissue sampling

A large individual from Lake Ontario was chosen for full-genome sequencing. Detailed methods for electrofishing and captive rearing can be found elsewhere (Côté *et al.* 2015). The individual was euthanized with an overdose of eugenol and liver tissue was immediately removed and frozen in liquid nitrogen.

Library preparation and sequencing

DNA was extracted with the phenol–chloroform method (Sambrook & Russell 2006). Tissues were placed in lysis buffer with a carbide bead and lysed on a mixer mill for 3 min at 25/sec. In addition to the steps of the extraction protocol, we truncated the tips for the 1-mL pipette in order to reduce shearing of high molecular weight DNA. DNA quality was verified on an agarose gel with a high molecular weight ladder. DNA quality was further assessed with a Bioanalyzer© (Agilent Technologies).

All library preparation, sequencing, assembly and scaffolding were performed at the Genome Québec McGill Innovation Centre. Six different libraries were produced for sequencing on seven HiSeq2000 Illumina lanes. Two Trueseq paired-end libraries were produced with an insert size of ~350 bp. Three different mate pair libraries were produced. Four standard mate pair with insert sizes of 3 and 5 kb (two for each) and a fifth using the Nextera library preparation protocol was produced with a 5-kb insert size.

The paired-end libraries were sequenced on two lanes each (four lanes total). For each standard mate pair library, the two libraries created for each insert size were combined in equimolar concentrations and sequenced on a single lane each. The Nextera mate pair library was sequenced on a single lane.

Genome assembly and scaffolding

All applications were used with the default parameters unless otherwise noted. Raw reads were assembled and initially scaffolded with RAY (ver. 2.3) (Boisvert *et al.* 2010) with a kmer value of 41, using Compute Canada resources in a highly parallel fashion (~2000 cores). Further scaffolding was performed with SSPACE (ver. 2.0 Basic) (Boetzer *et al.* 2011), specifying a minimum of three linkers to combine contigs and scaffolds, followed by the Gapcloser process of SOAPDENOVO (ver. 2.15) (Luo *et al.* 2012). The above two steps were then repeated. Finally, LRNA SCAFFOLDER (Xue *et al.* 2013) and BLAT (ver. 35.1) (Kent 2002) were used to further join scaffolds with *A. rostrata* (Gagnaire *et al.* 2012) and *A. anguilla* (Coppe *et al.* 2010) transcriptomic data. A 90% cut-off was chosen for the BLAT alignment. We used transcriptomic data in order to improve the assembly in and around the genes, thereby increasing the value of the tool for functional genomics studies.

After assembly and scaffolding, raw paired-end reads from the four runs were mapped back onto the preliminary genome. Scaffolds that had an average coverage of less than 10× coverage in at least three of the four lanes (~20× was expected based on overall coverage) were removed as likely assembly artefacts. Scaffolds less than 200 bp were removed as they are of limited use and also likely artefactual.

Assessment of protein-coding completeness

The application CEGMA (Parra *et al.* 2007) was used to assess the completeness of the 458 most conserved eukaryotic genes and also an ultra-conserved subset of 248. The vertebrate flag (-vrt) was included along with the default command to allow for longer intron lengths to be considered.

Annotation

The MAKER2 (Holt & Yandell 2011) application was used to annotate the genome. MAKER2 coordinates several other applications thus produces genome annotation from multiple lines of evidence. REPEATMASKER (Tarailo-Graovac & Chen 2009) was used to soft-mask repetitive elements with a teleostei-specific repeat database (Bao *et al.* 2015). Gene prediction models were developed with MAKER2 runs of previous versions of the genome and were used as an input hmm file for the present version in order to optimize the *ab initio* gene predictor SNAP (Korf 2004). *Ab initio* gene prediction was combined with EST evidence in the form of *A. anguilla* and *A. rostrata* transcriptomes (Coppe *et al.* 2010; Gagnaire *et al.* 2012; Baillon *et al.* 2015) and *A. anguilla* as well as proteomes of seven model species (*Fugu rubripes, Danio rerio, Xiphophorus maculatus*, *Latimeria chalumnae, Oreochromis niloticus*, *Petromyzon marinus, Mus musculus* and *Homo sapien* from Ensemble). We specified 200 bp for a maximum flank of evidence clusters for ab initio predictions and 10 000 bp for maximum intron length. Also, a minimum length of 100 bp was specified for single exon ESTs.

Following the preliminary genome annotation with MAKER2, the application INTERPROSCAN was used on all predicted proteins. Only proteins containing at least one known protein domain were retained for the final annotation.

Comparisons to A. Anguilla and A. japanica

In order to compare complete sequences in regions of potential functional interest, three candidate genes were selected to compare among the three freshwater eel genomes. Neuroplastin (*Nptn*) and myosin light chain kinase 3 (*Mylk3*) were chosen as genes found to be under selection in American eel between freshwater and brackish saltwater ecotypes (Pavey *et al.* 2015). Heat-shock protein 90 (*HSP90*) was chosen because of the large latitudinal distributions of the three species that is partially but not completely overlapping. Amino acid sequences were obtained on UniProt (Consortium, The Uniprot 2015) and tBLASTx (Camacho *et al.* 2009) was used to find corresponding CDS regions in each of the three species. MEGA 5 (Tamura *et al.* 2011) was used to compare the sequences among the three species and calculate synonymous and nonsynonymous polymorphisms.

Results and discussion

The total sequencing effort resulted in 238 billion base pairs in reads. The RAY assembly resulted in contigs totalling 1 187 377 314 base pairs. After scaffolding, the total genome size was 1 413 032 609 base pairs with an average of 169× coverage (Table 1). The total genome size is in the middle of estimates of *A. rostrata* (~1.0–1.6 Gb) from more direct methods (i.e. flow cytometry) (Ronchetti *et al.* 1995; Hardie & Hebert 2003). After assembly, scaffolding and filtering, the final genome contained 305 043 contigs and 79 209 scaffolds. The contig N50 is 7397 and the scaffold N50 is 86 604 (Table 2). A total of 1040 scaffolds were joined as a result of the transcriptome-assisted scaffolding step. Although this represents only a small improvement as a percentage of all of the scaffolds (~1.3%), it was retained for the final genome because the parts of the assembly that include the protein-coding genes are the most useful for functional genomics studies. All of that improvement was in or near protein-coding genes.

While the scaffold N50 is only modestly improved compared with the European eel genome (77 800), the contig N50 is 4.42 × greater (American eel: 7397; European eel: 1672; Henkel *et al.* 2012a), indicating that the

Table 1 Sequencing effort including insert length in base pairs and coverage for each type of library

	Insert (bp)	Coverage
Paired-end	350	101X
Mate pair (standard)	2 k	40X
Mate pair (standard)	4.5 k	18X
Mate pair (Nextera)	4.5	10X

Table 2 Assembly statistics comparison of American eel, European eel and Japanese eel including N50, largest and number of both contigs and scaffolds, as well as total assembled size

Statistic	American eel	European eel	Japanese eel
Contig			
N50	7397	1672	3885
Largest	83 791	53 800	81 595
Number	305 043	1 520 000	698 249
Scaffold			
N50	86 641	77 800	52 849
Largest	866 215	923 000	1 141 856
Number	79 209	186 000	323 740
Total assembled size (Gb)	1.41	0.996	1.15

present genome is substantially less fragmented (Table 2). In addition, both contig and scaffold N50 also show substantial improvement in comparison with the Japanese eel (contig: 4000 and scaffold: 52 849; Henkel *et al.* 2012b) (Table 2). The number of scaffolds in our genome is also much lower, despite the fact that we assembled more (1.41 vs. 0.996 Gb for *A. anguilla* and 1.15 Gb for *A. japanica*). The entire mitochondrial genome assembled into a single contig of the size expected for teleost fish (~16 kb).

Protein-coding regions including introns and flanking regions are very well represented in the genome, as 95.2% of the 458 core eukaryotic genes and 98.8% of the 248 ultra-conserved subset were represented in the assembly (Table 3). All of these metrics offer substantial improvement over the European eel genome.

Annotation with MAKER2 initially resulted in 41 103 protein-coding genes. After assessing protein domain content, 26 564 genes were retained. This is comparable to other recently annotated genomes such as the Japanese eel (18 121) (Liu *et al.* 2016), northern pike (*Esox lucius*; 19 601) (Rondeau *et al.* 2014) and three-spine stickleback (*Gasterosteus aculeatus*; 20 787 (Jones *et al.* 2012). We found less protein-coding genes than the rainbow trout (*Oncorhynchus mykiss)* genome (46 585 genes) (Berthelot *et al.* 2014), as would be expected given that all salmonids have undergone an additional whole-genome duplication approximately 60 MYA (Crête-Lafrenière *et al.* 2012).

All three candidate genes were present in all three freshwater eel genomes. The 384 bp of the CDS region for neuroplastin (*Nptn*) was identical in all three species. However, in the 2691 bp of the CDS region for myosin light chain kinase 3 (*Mylk3*), there was one nonsynonymous and five synonymous polymorphisms between *A. rostrata* and *A. anguilla*, versus five and 21 between *A. rostrata* and *A. japonica*. In the 825 bp of the CDS region for heat-shock protein 90 (*HSP90*), there was one nonsynonymous and two synonymous polymorphisms with *A. anguilla*, versus one and four with *A. japanica*. This exercise demonstrates that our genome is compatible with the other two freshwater eel genomes, showing a higher divergence within functional gene regions with *A. japonica* as expected based on the phylogenetic relationships. This demonstrates that having all three genomes available will be useful tools to facilitate comparative genomics among the three species.

Table 3 CEGMA results

Statistic	American (%)	European (%)	Japanese (%)
Partial 458 CEGs	95.2	85.8	94.8
Complete 248 CEGs	89.1	76.1	85.9
Partial 248 CEGs	98.8	90.3	96.0

Partial and complete coverage percentages for the 458 and 248 subset of the core eukaryotic genes for the three freshwater eel genomes.

Conclusion

This study presents the first genome of the American eel and, at present, the most complete genome for the *Elopomorpha* superorder. This genome fills a critical gap in freshwater eel genomic resources and is an essential tool to decipher the functional genomics of the endangered American eel. This genome has already proven to be a very useful tool for GWAS and RT-qPCR primer design (Pavey *et al.* 2015; Gaillard *et al.* 2016; Laporte *et al.* 2016). Future conservation genetics studies on this species will take advantage of this public genome available on GenBank (LTYT01000000). In addition, the American and European eels are two sister species (divergence occurred 3.38 million years ago; Jacobsen *et al.* 2014b) and the availability of both genomes opens the possibility of reciprocal mapping, which should be also beneficial for functional genomics studies of this other endangered species of *Anguilla*. Overall, the sequenced genome present in this study is a crucial addition to design future conservation efforts of the North Atlantic eels.

Acknowledgements

Funding was from a Natural Sciences and Engineering Research Council of Canada (Strategic Partnership Program), Ressources Aquatiques Québec (RAQ) and Canada Research Chairs Program for LB, JC and SP). We thank Thomas Pratt, Guillaume Côté and Caroline Côté for coordinating sampling and rearing. We thank the sequencing and bioinformatics team at McGill University and Génome Québec Innovation Centre for sequencing and bioinformatics services and advice. We are also grateful to Chief Editor S. Narum, Associate Editor A. DeWoody and two anonymous reviewers for their comments on a previous version of the manuscript. We thank NCBI and GenBank for curating the genome, and the DRYAD digital repository for archiving the annotation.

References

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.

Arai T (2016) *Biology and Ecology of Anguillid Eels*. CRC Press, Boco Raton, FL.

ASMFC (Atlantic States Marine Fisheries Commission) (2000) Interstate fishery management plan for American Eel. Atlantic States Marine Fisheries Commission, Report **36**, Washington, DC

Baillon L, Oses J, Pierron F *et al.* (2015) Gonadal transcriptome analysis of wild contaminated female European eels during artificial gonad maturation. *Chemosphere*, **139**, 303–309.

Bao W, Kojima KK, Kohany O (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**, 11.

Berthelot C, Brunet F, Chalopin D *et al.* (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications*, **5**, 3657.

Betancur RR, Broughton RE, Wiley EO *et al.* (2013) The tree of life and a new classification of bony fishes. *PLOS Currents*, **5**, doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Boetzer M, Henkel CV, Jansen HJ *et al.* (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, **27**, 578–579.

Boisvert S, Laviolette F, Corbeil J (2010) RAY: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *Journal of Computational Biology*, **17**, 1519–1533.

Busch WDN, Braun DP (2014) A case for accelerated reestablishment of American Eel in the Lake Ontario and Champlain watersheds. *Fisheries*, **39**, 298–304.

Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 1–9.

Consortium, The Uniprot (2015) UNIPROT: a hub for protein information. *Nucleic Acids Research*, **43**, D204–D212.

Coppe A, Pujolar JM, Maes GE *et al.* (2010) Sequencing, de novo annotation and analysis of the first *Anguilla anguilla* transcriptome: EeelBase opens new perspectives for the study of the critically endangered European eel. *BMC Genomics*, **11**, 635.

COSEWIC (Committee on the Status of Endangered Wildlife in Canada). (2012) *COSEWIC Assessment and Status Report on the American Eel Anguilla rostrata in Canada*. Committee on the Status of Endangered Wildlife in Canada, Ottawa, Ontario. 109 pp.

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

Côté CL, Castonguay M, Kalujnaia SV *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.

Côté CL, Pavey SA, Stacey JA *et al.* (2015) Growth, female size and sex ratio variability in American Eel (*Anguilla rostrata*) of different origins in both controlled conditions and the wild: implications for stocking programs. *Transactions of the American Fisheries Society*, **144**, 235–245.

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

Dekker W, Casselman JM, Cairns DK *et al.* (2003) Worldwide decline of eel resources necessitates immediate action: Quebec declaration of concern. *Fisheries*, **28**, 28–30.

Froese R, Pauly D (2015) *FishBase*. World Wide Web electronic publication. www.fishbase.org (10/2015).

Gagnaire P-A, 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.

Gaillard MSA, Pavey CL Côté *et al.* (2016) Regional variation of gene regulation associated with storage lipid metabolism in American glass eels (*Anguilla rostrata*). *Comparative Biochemistry and Physiology Part A.*, **196**, 30–37.

Gong XL, Ren SJ, Cui ZK *et al.* (2014) Genetic evidence for panmixia of Japanese eel (*Anguilla japonica*) populations in China. *Genetic and Molecular Research*, **13**, 768–781.

Hardie DC, Hebert PDN (2003) The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome*, **46**, 683–706.

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

Henkel CV, Dirks RP, de Wijze DL *et al.* (2012b) First draft genome sequence of the Japanese eel, *Anguilla japonica*. *Gene*, **511**, 195–201.

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

ICES (International Council for the Exploration of the Sea). (2013). Widely distributed and migratory stocks, European Eel, advice for 2014. Available URL: http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2013/2013/eel-eur.pdf. Accessed 4 January 2014.

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

Jacobsen MW, Pujolar JM, Gilbert T *et al.* (2014b) Speciation and demographic history of Atlantic eels (*Anguilla anguilla* and *A. rostrata*) revealed by mitogenome sequencing. *Heredity*, **113**, 432–442.

Jacoby D, Gollock M (2014) *Anguilla japonica. IUCN Red List of Threatened Species.* International Union for Conservation of Nature, e.T166184A-1117791. http://dx.doi.org/10.2305/IUCN.UK.2014-1.RLTS.T166184A1117791.en. Accessed 31 October 2016.

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

Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Research*, **12**, 656–664.

Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics*, **5**, 59.

Laporte M, Pavey SA, Rougeux C *et al.* (2016) RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, **25**, 219–237.

Liu Y-C, Hsu S-D, Chou C-H *et al.* (2016) Transcriptome sequencing based annotation and homologous evidence based scaffolding of *Anguilla japonica* draft genome. *BMC Genomics*, **17**, 177–192.

Luo R, Liu B, Xie Y *et al.* (2012) SOAPDENOVO2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, **1**, 18.

Parra G, Bradnam K, and Korf I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**(9), 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.

Pujolar JM, Jacobsen MW, Als TD *et al.* (2014) Genome-wide single-generation signatures of local selection in the panmictic European eel. *Molecular Ecology*, **23**, 2514–2528.

Pujolar JM, Jacobsen MW, Bekkevold D *et al.* (2015) Signatures of selection between life cycle stages separated by metamorphosis in the European eel. *BMC Genomics*, **16**, 600.

Ronchetti E, Salvadori S, Deiana AM (1995) Genome size and AT content in Anguilliformes. *European Journal of Histochemistry*, **39**, 259–264.

Rondeau EB, Minkley DR, Leong JS *et al.* (2014) The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS ONE*, **9**, e102089.

Sambrook J, Russell DW (2006) *Purification of Nucleic Acids by Extraction with Phenol: Chloroform*. Cold Spring Harbor Protocols, **1**, pdb-prot4455.

Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Review in Genetics*, **14**, 807–820.

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

Tanaka H, Kagawa H, Ohta H *et al.* (2003) The first production of glass eel in captivity: fish reproductive physiology facilitates great progress in aquaculture. *Fish Physiology and Biocheestry*, **28**, 493–497.

Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, **25**, 4.10.1–4.10.14.

Tesch F-W, Thorpe JE (2003) *The Eel*. Blackwell Science, Oxford, UK.

Tse WKF, Sun J, Zhang H *et al.* (2014) iTRAQ-based quantitative proteomic analysis reveals acute hypo-osmotic responsive proteins in the gills of the Japanese eel (*Anguilla japonica*). *Journal of Proteomics*, **105**, 133–143.

Wang S, Monroig O, Tang G *et al.* (2014) Investigating long-chain polyunsaturated fatty acid biosynthesis in teleost fish: functional characterization of fatty acyl desaturase (Fads2) and Elovl5 elongase in the

catadromous species, Japanese eel *Anguilla japonica*. *Aquaculture*, **434**, 57–65.

Xue W, Li J-T, Zhu Y-P *et al*. (2013) L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics*, **14**, 1–14.

S.P. and J.G. performed the laboratory work and data analysis and wrote the manuscript. M.L. wrote the manuscript. EN performed data analysis and genome clean-up. L.L. assembled the genome. S.B. and J.C. optimized the assembly application RAY 2.0 for fish genomes and provided advice on further assembly steps. L.B. and C.A. provided team leadership and cowrote the manuscript.

Data accessibility

The genome fasta file with links to raw sequencing reads has been deposited at DDBJ/ENA/GenBank under the Accession no. LTYT00000000. The version described in this study is version LTYT01000000. The annotation files are available at DRYAD (Accession number doi:10.5061/dryad.n816r). All raw sequencing data is available at Sequence Read Archive (SRA) under BioProject PRJNA285069.