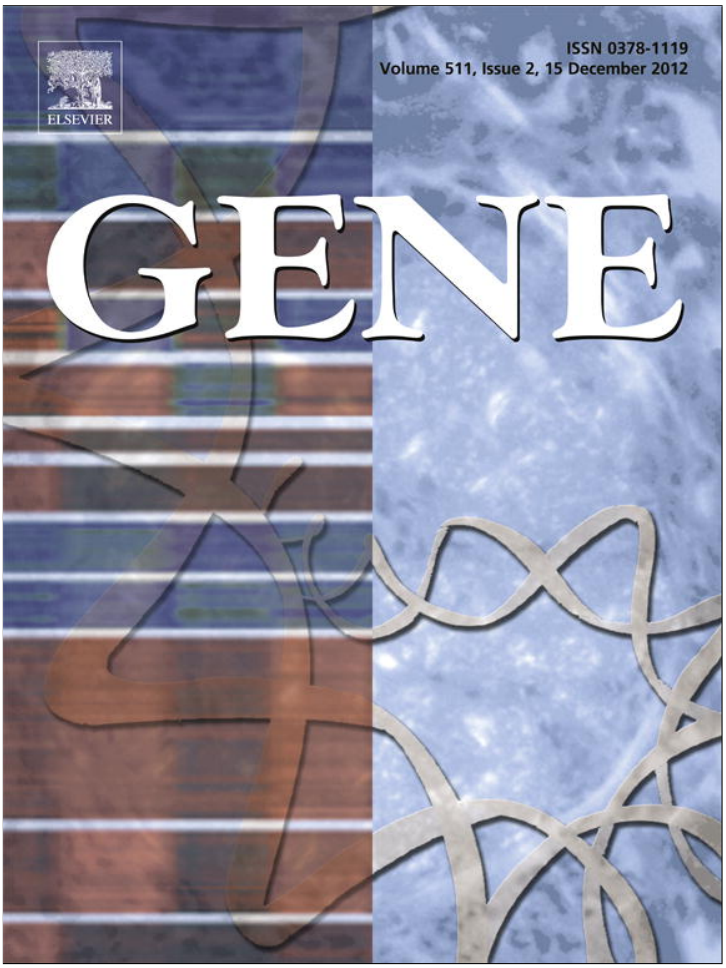

Gene 511 (2012) 195–201

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

First draft genome sequence of the Japanese eel, *Anguilla japonica*

Christiaan V. Henkel [a,1], Ron P. Dirks [a,*,1], Daniëlle L. de Wijze [b], Yuki Minegishi [b], Jun Aoyama [c], Hans J. Jansen [a], Ben Turner [d], Bjarne Knudsen [d], Martin Bundgaard [d], Kenneth Lyneborg Hvam [d], Marten Boetzer [e], Walter Pirovano [e], Finn-Arne Weltzien [f], Sylvie Dufour [g], Katsumi Tsukamoto [c], Herman P. Spaink [b], Guido E.E.J.M. van den Thillart [b]

[a] *ZF-screens B.V., Niels Bohrweg 11, 2333 CA Leiden, The Netherlands*
[b] *Institute of Biology, Leiden University, P.O. Box 9505, 2300 RA Leiden, The Netherlands*
[c] *Ocean Research Institute, University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba, 277–8564, Tokyo, Japan*
[d] *CLC bio, Finlandsgade 10–12, Katrinebjerg, DK-8200 Aarhus N, Denmark*
[e] *BaseClear B.V., Einsteinweg 5, 2333 CC Leiden, The Netherlands*
[f] *Norwegian School of Veterinary Science, P.O. Box 8146, 0033 Oslo, Norway*
[g] *UMR BOREA, CNRS 7208, Museum National d'Histoire Naturelle, 7, rue Cuvier, 75231 Paris Cedex 05, France*

ARTICLE INFO

Article history:
Accepted 11 September 2012
Available online 29 September 2012

Keywords:
Eel
Anguilla
Assembly pipeline
Illumina sequencing
De novo genome
Hox genes

ABSTRACT

The Japanese eel is a much appreciated research object and very important for Asian aquaculture; however, its genomic resources are still limited. We have used a streamlined bioinformatics pipeline for the *de novo* assembly of the genome sequence of the Japanese eel from raw Illumina sequence reads. The total assembled genome has a size of 1.15 Gbp, which is divided over 323,776 scaffolds with an N50 of 52,849 bp, a minimum scaffold size of 200 bp and a maximum scaffold size of 1.14 Mbp. Direct comparison of a representative set of scaffolds revealed that all the Hox genes and their intergenic distances are almost perfectly conserved between the European and the Japanese eel. The first draft genome sequence of an organism strongly catalyzes research progress in multiple fields. Therefore, the Japanese eel genome sequence will provide a rich resource of data for all scientists working on this important fish species.

1. Introduction

The first draft genome sequence of an organism has a major impact on research approaches in multiple fields, including molecular biology, physiology, ecology and evolution. Publication of the draft genomes of teleosts such as the Japanese pufferfish *Takifugu rubripes* (Aparicio et al., 2002), the green spotted pufferfish *Tetraodon nigroviridis* (Jaillon et al., 2004) and the medaka *Oryzias latipes* (Kasahara et al., 2007) has opened the window to the investigation of the entire genetic architecture of their biology. Combinations of complete genome information and other genome-wide scanning techniques, such as amplified fragment length polymorphisms (AFLP), quantitative traits loci (QTL) mapping and microarray or mRNA-Seq mediated transcriptome mapping, enable detailed investigation of the genetic mechanisms of specific phenomena like growth, osmoregulation, immune responses, stress tolerance and local adaptations. Those tools thus enable to "genomically" address central biological questions such as how fishes adapt to various water environments – including pollution – and how their remarkable morphological and ecological diversity has been controlled and evolved.

The freshwater eel, genus *Anguilla*, is a catadromous species, which migrates between growth habitats in freshwater and estuaries and spawning areas in the tropical and subtropical seas (Tesch, 2003). Previous studies have observed their migratory characteristics and revealed that anguillid species show various kinds and degrees of intra- and interspecific variation, such as migration scale, larval growth, spawning season and morphological characteristics (Aoyama, 2009); however, compared to e.g. salmon, little is known about the genetic backgrounds of eels (reviewed by Minegishi et al., 2012). Recently the draft genome sequence of the European eel has been determined (Henkel et al., 2012), and some mRNA-Seq studies of this species have been reported (Coppe et al., 2010) and are in progress (our unpublished data). In order to facilitate genome-wide studies of eels, comparative

Abbreviations: AFLP, amplified fragment length polymorphisms; MP, mate-pair; PE, paired-end; QTL, quantitative traits loci.

* Corresponding author at: ZF-screens B.V., Niels Bohrweg 11, 2333 CA Leiden, The Netherlands. Tel.: +31 71 3322270; fax: +31 84 8319969.
E-mail addresses: henkel@zfscreens.com (C.V. Henkel), dirks@zfscreens.com (R.P. Dirks), d.l.de.wijze@biology.leidenuniv.nl (D.L. de Wijze), y.minegishi@biology.leidenuniv.nl (Y. Minegishi), jaoyama@aori.u-tokyo.ac.jp (J. Aoyama), jansen@zfscreens.com (H.J. Jansen), bturner@clcbio.com (B. Turner), bknudsen@clcbio.com (B. Knudsen), mbundgaard@clcbio.com (M. Bundgaard), khvam@clcbio.com (K.L. Hvam), Marten.Boetzer@baseclear.com (M. Boetzer), Walter.Pirovano@baseclear.nl (W. Pirovano), Finn-Arne.Weltzien@nvh.no (F.-A. Weltzien), dufour@mnhn.fr (S. Dufour), ktpc@aori.u-tokyo.ac.jp (K. Tsukamoto), h.p.spaink@biology.leidenuniv.nl (H.P. Spaink), g.van.den.thillart@biology.leidenuniv.nl (G.E.E.J.M. van den Thillart).
[1] Contributed equally.

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.09.064

genomics will be necessary. The Japanese eel *Anguilla japonica* is widely distributed from the northern Philippines to North Japan (Ege, 1939). They spawn near Western Mariana Ridge, and eggs and larvae born there are transported to growth habitats in eastern Asian countries by the North Equatorial Current and then the Kuroshio Current (Shinoda et al., 2011; Tsukamoto, 2009). After metamorphosis to glass eels, they migrate upstream to enter their growth phase for several years, and migrate back to the spawning area located at 2000–3000 km away (Aoyama, 2009). In addition to migratory ecology, much attention has been paid to the physiology of the Japanese eel. Endocrinological changes and hormone dynamics of, especially reproduction and osmoregulation, have been investigated (e.g. reviewed by Kazeto et al., 2011; Miura and Miura, 2011) Although research on this species has been intensively conducted and a genetic linkage map has been recently developed (Nomura et al., 2011), it is still a long way to understand the genomic mechanisms of their biology.

Fresh water eels, especially Japanese, European and American eels (respectively *Anguilla japonica*, *A. anguilla* and *A. rostrata*), are also a very nutritious and popular food source and their high economic value has caused overfishing. Together with other potential causes such as habitat destruction, pollution and diseases, this has resulted in a more than 90% decline in global stocks over the past decades (Dekker et al., 2003). A considerable percentage of consumption eels are grown to a marketable size in specific eel farms; however, eel farming is still fully dependent on wild caught juvenile eels (glass eels). Attempts toward artificial reproduction of fresh water eels already started in the 1930s (Fontaine, 1936). Although artificial maturation has now been shown for multiple eel species (Lokman and Young, 2000; Ohta et al., 1997; Oliveira and Hable, 2010; Pedersen, 2003), complete artificial reproduction has so far only been achieved for the Japanese eel (Tanaka et al., 2003) and has recently resulted in the first artificial F2 generation of eels (Nomura et al., 2011). For the first time, selective breeding of fresh water eels is now within reach. In the near future, this will allow the selection of eel brood stocks based on multiple parameters, including growth rate, speed of maturation, resistance against diseases, taste and fat content of the final product, etcetera. Traditional selective breeding has much to gain from molecular biology and genetics. Molecular techniques will contribute to more precise breeding programs and reduced inbreeding. Therefore, proper genetic resources for fresh water eels will be indispensible for selective breeding. Vice versa, artificial breeding of the Japanese eel will allow the generation of improved genetic linkage maps, which will again result in improved quality of these genetic resources. Here we used a versatile *de novo* assembly protocol to generate the first draft genome of the Japanese eel, *Anguilla japonica*. The Japanese eel genome will be an invaluable source of information, both for future applications in aquaculture and for scientific purposes.

2. Materials and methods

2.1. Genomic DNA libraries

Genomic DNA was isolated from blood of an adult female Japanese eel using the Blood and Tissue DNeasy kit (Qiagen GmbH, Hilden) according to the manufacturer's description. Paired-end libraries were prepared from 5 μg of isolated gDNA using the Paired-End Sequencing Sample Prep kit (Illumina Inc., San Diego) according to the manufacturer's description. Either a 280 bp band or a 600 bp band was cut from the gel (libraries PE280 and PE600). After amplification for 10 cycles the resulting libraries were analyzed with a Bioanalyzer 2100 DNA 1000 series II chip (Agilent, Santa Clara). Mate-pair libraries were prepared from 10 μg of isolated gDNA using the Mate Pair Library Prep Kit v2 (Illumina Inc., San Diego) according to the manufacturer's description. Agarose slices containing ~2-kb, ~4-kb, ~6-kb, ~8-kb or ~10-kb DNA fragments were cut from gel (libraries MP2K, MP4K, MP6K, MP8K, and MP10K). After the first gel purification the fragment length was analyzed using a Bioanalyzer 2100 DNA 12000 chip (Agilent). After circularization, shearing, isolation of biotinylated fragments, and amplification, the 400–600 bp fraction of the resulting fragments was isolated from gel. Finally, the libraries were examined with a Bioanalyzer 2100 DNA 1000 series II chip (Agilent).

2.2. Illumina sequencing

All libraries were sequenced using an Illumina GAIIx or HiSeq 2000 instrument according to the manufacturer's description. Genomic paired-end libraries were sequenced with a read length of 2×76 nucleotides or 2×151 nucleotides, and genomic mate-pair libraries with a read length of 2×76 nucleotides. Image analysis and base calling were done by the standard Illumina pipeline.

2.3. Genome assembly

Sequencing reads from both paired-end libraries were used in building the initial contigs. Both sets were preprocessed to eliminate low quality reads and nucleotides, as well as adapter contamination (mainly caused by insert sizes smaller than the read length). Because of the small insert size of the PE280 library, many read pairs from this library overlap at their 3' ends. When possible, these pairs were merged into longer single reads. This preassembly procedure has the dual advantage of producing long reads (which improve the quality and efficiency of the subsequent assembly) and providing confirmation for the identity of the 3' ends of the reads (which are generally determined with lesser confidence). We merged read pairs that exhibited at least eighteen nucleotides of unambiguous sequence overlap. Using this criterion, 29.6% of pairs could be merged, resulting in single reads with a mean length of 242 nt.

For initial contig assembly, we employed the CLC Assembly Cell de novo assembler (versions 3.2 and 4.0beta, CLC bio, Aarhus, Denmark). This is an efficient implementation of a De Bruijn graph-based assembler, which enables the assembly of the Japanese eel genome on a dual quad-core Xeon workstation with 48 GB of RAM installed in approximately eight hours. In the present study we included an extensive beta test in which we compared the performance of version 3.2 and the latest version 4.0 of the Assembly Cell. A total of 25 complete *de novo* assembly rounds were performed with k-mer values ranging from 21 to 31 nucleotides, and using bubble sizes of 50, 100, 250 or 400 bp. The k-mer parameter sets the length in nucleotides of nodes in the assembly graph. All individual sequencing reads are split into overlapping k-mers and connections between these. Smaller k-mers are less likely to be unique in the genome than high ones, resulting in a more tangled graph and therefore shorter contig sequences; however, using smaller k-mers, a larger number of connections between them can be extracted from sequencing reads. Therefore, at low coverage or using short reads, a lower k-mer setting may still yield larger contigs than a high one. For example, a single 76 nt read can yield a graph with fifty-six 21-mers and 55 connections, or one with forty-six 31-mers and 45 connections. It is difficult to predict which settings will yield the globally least tangled graph. The bubble parameter, new in CLC Assembly Cell 4.0, sets the maximum length of ambiguities in the k-mer graph (e.g. small repeats) for which the algorithm will attempt to resolve a path using long reads or pairs. Since our data include sequence information over distances far exceeding the k-mer lengths (i.e. merged reads up to 284 nt, pair inserts up to 750 nt), this addition is expected to increase the contiguity of the output.

Initial contigs were oriented in larger supercontigs (scaffolds) using SSPACE (Boetzer et al., 2011). Briefly, SSPACE aligns paired reads to the contigs (using Bowtie), and combines contigs if they are connected by at least a specified number of pairs within the limits set for the insert size of the pair library. The insert size is then used to estimate the size of the gap between the contigs. In addition, the algorithm can be forced

to extend scaffolds with a contig only if the evidence for its unique placement is above a set threshold, or else abort growth for that scaffold. This allows contigs representing collapsed repeats to be either included or excluded from the final scaffolds. SSPACE was used to scaffold contigs in a hierarchical fashion, employing first links obtained from the PE600 library to generate intermediate supercontigs, which were used as input for subsequent runs with links from individual mate-pair libraries increasing in size. At each stage, a minimum of three non-redundant links was required to join two contigs. This procedure resulted in a final scaffold set with a total length of ~1.15 Gbp and an N50 of ~53 kb.

The draft assembly is available at www.eelgenome.org (login: zfscreens; password: h0x_genes). The NCBI accession numbers of the Japanese eel Hox gene scaffolds are as follows: JQ976895 (HoxAa), JQ976896 (HoxAa), JQ976897 (HoxAb), JQ976898 (HoxBa), JQ976899 (HoxBb), JQ976900 (HoxBb), JQ976901 (HoxBb), JQ976902 (HoxCa), JQ976903 (HoxCb), JQ976904 (HoxCb), JQ976905 (HoxDa), JQ976906 (HoxDb), and JQ976907 (HoxDb).

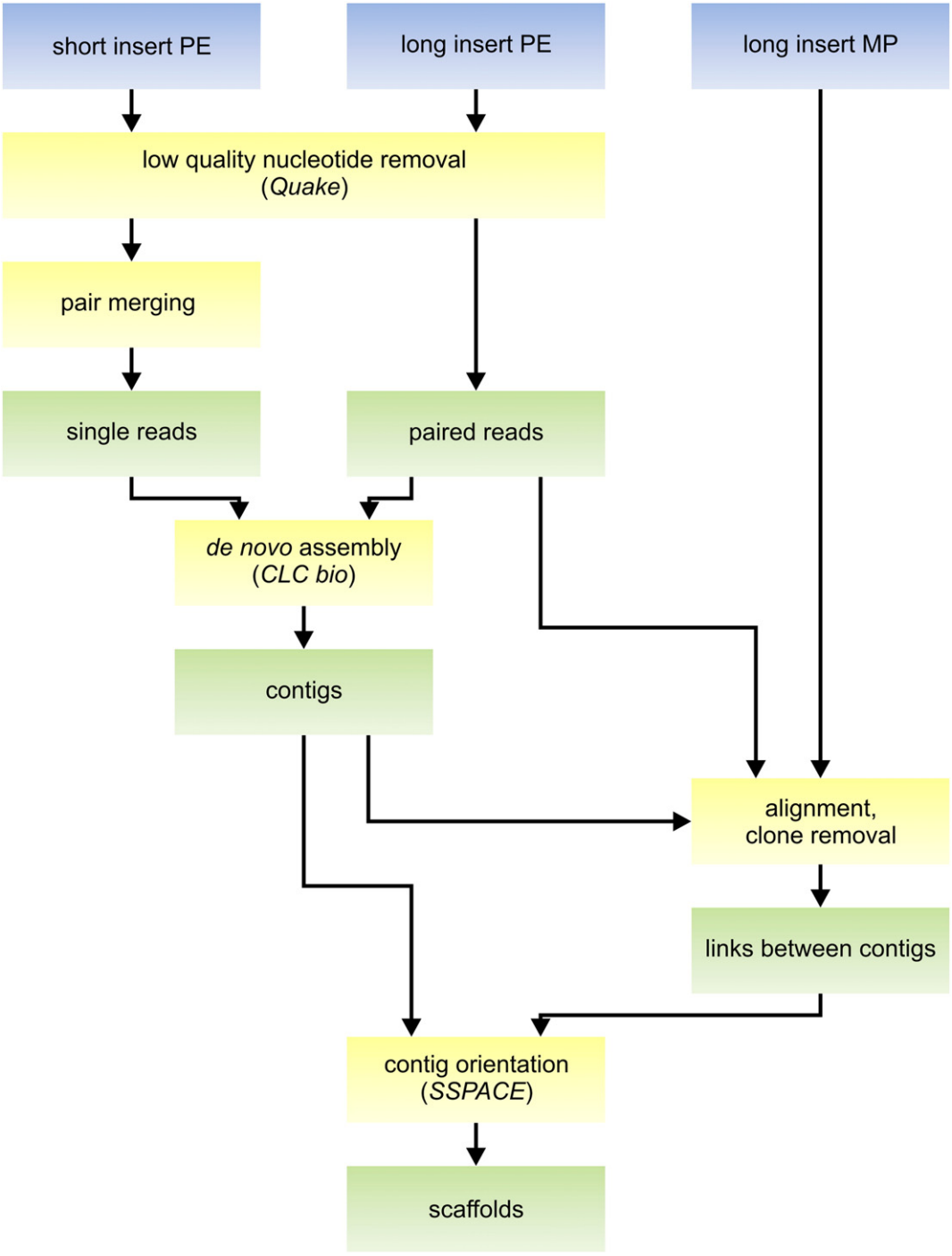

Fig. 1. Main stages of the genome assembly pipeline. Raw data are shown in blue boxes, processing in yellow, and processed data in green. See Materials and methods and Results and discussion sections for details. Quake was described by Kelley et al. (2010). SSPACE was recently developed in our group (Boetzer et al., 2011). CLC bio is a Denmark-based company that has developed the CLC Assembly Cell *de novo* assembly software.

Table 1
Illumina libraries, runs and raw sequence data.

Library	Protocol	Technology	Insert size (bp)	Reads	Raw sequence (Gbp)	Sequence used for assembly (Gbp)	Contig links
PE280	Paired-end	GAIIx, 2×151	180–290	103 M	31.0	22.4	N.A.
PE600	Paired-end	GAIIx, 2×151	525–750	35 M	10.7	7.0	6.3 M
PE600	Paired-end	HiSeq 2000, 2×76	525–750	85 M	13.0	11.5	
MP2K	Mate-pair	HiSeq 2000, 2×76	1550–2450	28 M	4.3	N.A.	2.2 M
MP4K	Mate-pair	HiSeq 2000, 2×76	3000–5600	33 M	5.0	N.A.	1.1 M

3. Results and discussion

The Japanese eel genome was assembled using a whole-genome shotgun sequencing strategy solely based on Illumina technology, as pioneered in the assembly strategy of the giant panda genome (Li et al., 2010). The assembly pipeline (summarized in Fig. 1) was split into the following subparts: design, construction and sequencing of Illumina libraries, quality control of the raw sequence data, *de novo* contig assembly and scaffold assembly.

Based on experience gained from the *de novo* assembly of the European eel genome (Henkel et al., 2012), paired-end libraries of short 280-bp (PE280) and long 600-bp (PE600) inserts and five mate-pair libraries of long 2-, 4-, 6-, 8- and 10-kb inserts (MP2K, MP4K, etc.) were designed. The genomic libraries were constructed using DNA isolated from a blood sample of a female *Anguilla japonica* specimen. The PE280 and PE600 libraries were sequenced in 2×76 nt and 2×151 nt runs on the Illumina GAIIx and HiSeq 2000 toward ~35–55-fold coverage of the predicted genome size of 0.99–1.63 Gbp as estimated for *Anguilla sp.* (Animal Genome Size Database [http://genomesize.com]). All mate-pair libraries were sequenced in a 2×76 nt run. An overview of the Illumina libraries and the amount of raw sequence data derived thereof is presented in Table 1.

Quality control of the raw sequence data consisted of two parts. First, the Quake package (Kelley et al., 2010) was used to correct substitution sequencing errors and remove low confidence base calls. This resulted in the correction of 9.8% of raw sequencing reads, and removal of low quality nucleotides corresponding to 23% of the original amount of raw data. These percentages are high, mainly as the result of below average quality of second reads from a single GAIIx run. Next, a considerable number (29.6%) of remaining pairs from the 2×151 nt PE280 Illumina runs exhibited an unambiguous overlap and could be merged into >240 nt-long single reads. The resulting quality-approved single and paired-end sequence reads were then used as input for the *de novo* assembly process.

De novo assembly of the Japanese eel genome was performed using the CLC Assembly cell as described in Materials and methods. We included an extensive beta test in which the performance of version 3.2 and the latest version 4.0 of the Assembly Cell were compared (Fig. 2). In order to determine the optimal settings for the assembly process, two parameters were calculated from the contig characteristics, namely the N50 and the total assembled genome size. The N50 value is the length-weighed median fragment length and an important indicator of the quality of the assembly process. Its value indicates that 50% of the assembled genome is made up of

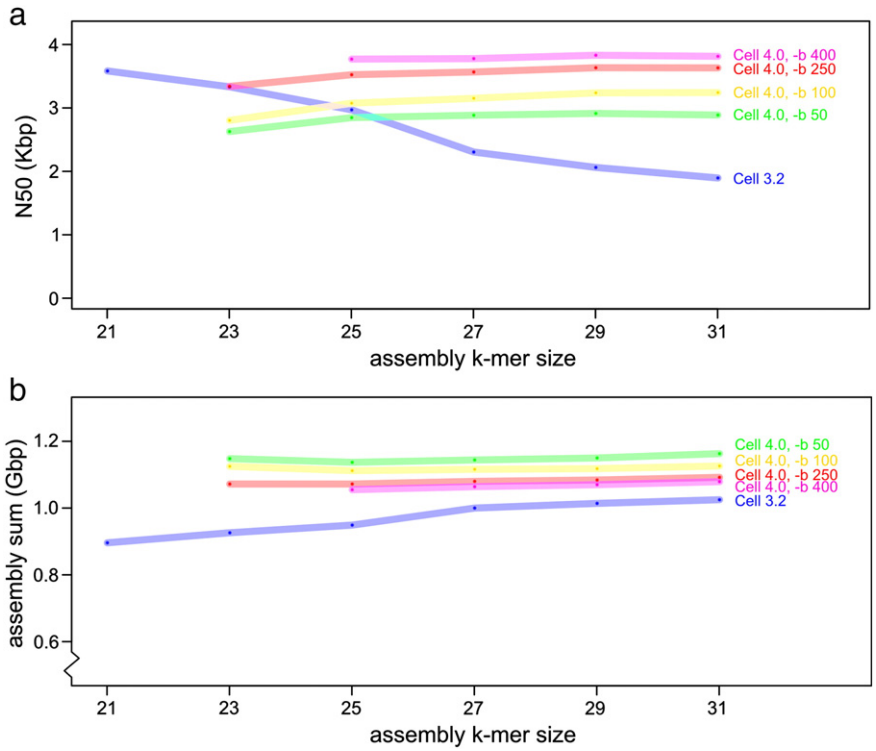

Fig. 2. *De novo* assembly using different generations of the CLC Assembly Cell algorithm. (a) Relationship between assembler settings and typical contig size (N50). Each dot represents a complete *de novo* assembly round (25 in total). The most contiguous assemblies are obtained with low k-mer sizes in version 3.2 of the Assembly Cell, or with high k-mer and bubble sizes (−b) in version 4.0. (b) Relationship between assembler settings and total output sequence length. Each dot represents a complete *de novo* assembly round (25 in total). At the k-mer setting at which version 3.2 achieves a high contig N50, the assembly appears incomplete. Version 4.0 outputs an assembly in the range of the expected genome size. Since resolved bubbles lead to only one sequence being output instead of two for unresolved bubbles, higher bubble size (−b) settings result in slightly smaller assembly sizes.

fragments of at least that size. The version 3.2 assembler reached the highest N50 at a k-value of 21 nt, whereas the version 4.0 Assembly Cell performed best at k-values above 25 nt and with increasing bubble size (Fig. 2a). The second important quality indicator is the size of the total assembled genome, which is the sum of all assembled contigs. The biggest assembly sum (1.15 Gbp) was obtained when the version 4.0 Assembly Cell was used at a bubble size of 50 bp, irrespective of the k-mer setting (Fig. 2b). In comparison with the version 4.0 Assembly Cell, the assembly sum obtained with the version 3.2 assembler was always lower, especially at k-mer settings of 25 nt or less; however, at high k-mer settings the N50 values obtained using the version 3.2 Assembler were much lower than those obtained using the version 4.0 assembler (Fig. 2a). In conclusion, the optimal combination of high N50 and high assembled genome size was obtained with the version 4.0 Assembly Cell at a bubble size of 400 bp and a k-mer size of 25–29 nt. The resulting contigs were then used as input for the scaffolding process (Fig. 3).

In the final part of the assembly process the program SSPACE (Boetzer et al., 2011) was used to orient and align the contigs into scaffolds. The reads from the PE600, MP2K and MP4K libraries were used in successive scaffolding rounds using the latest version of SSPACE. Runs from the MP6K, MP8K and MP10K libraries revealed that these mate-pair libraries did not contain sufficiently high numbers of unique reads for our stringent scaffolding process: only contigs that are linked via at least three independent mate-pair fragments are incorporated into scaffolds. Before scaffolding, paired reads were filtered by aligning them to the assembled contigs, and discarding all pairs that did not align to a unique position, did not bridge two contigs, or were clonal duplicates. The latest version of SSPACE has a more stringent quality control than the previous version, resulting in more reliable scaffolds, albeit with a slightly shorter N50 value. The best results were obtained when the data from individual libraries and overlapping libraries were used in five consecutive scaffolding rounds. The specifications of the final assembled version 1 draft of the Japanese eel genome are as follows: the total assembled genome size is 1.15 Gbp, which is divided over 323,776 scaffolds with an N50 of 52,849 bp, a minimum scaffold size of 200 bp and a maximum scaffold size of 1.14 Mbp.

We have recently determined the first draft genome of another fresh water eel, namely the European eel *Anguilla anguilla*, and found that it contains a remarkably high number of 73 Hox genes (Henkel et al., 2012). The European eel's Hox genes are located in eight clusters that could be mapped to only 10 *de novo* assembled scaffolds. Orthologues of all 73 Hox genes could be identified in the Japanese eel genome and we even identified the HoxC3b gene, which brings the total number of Hox genes that could be identified in the Japanese eel to 74 (Fig. 4). The 74 Hox genes are divided over 13 scaffolds, which indicates that the quality of the *de novo* assembled Japanese eel genome is sufficiently high for mapping large gene clusters and for basic comparative genomics. Distances between individual Hox genes have been almost perfectly conserved between the European and Japanese eel. The region between the HoxA4a and HoxA3a genes is longer than initially thought (Henkel et al., 2012) and also contains a non-Hox gene (TRIM35-like).

4. Conclusions

In this study we present the first draft genome sequence of the Japanese eel. To this end, we used a versatile assembly pipeline for *de novo* assembly of vertebrate genomes, starting from raw Illumina reads, via contigs with an N50 of ~4-kb, to scaffolds with an N50 of more than 50 kb. The goal of this project was to build a firm basis for genome-wide studies on the Japanese eel and for comparative

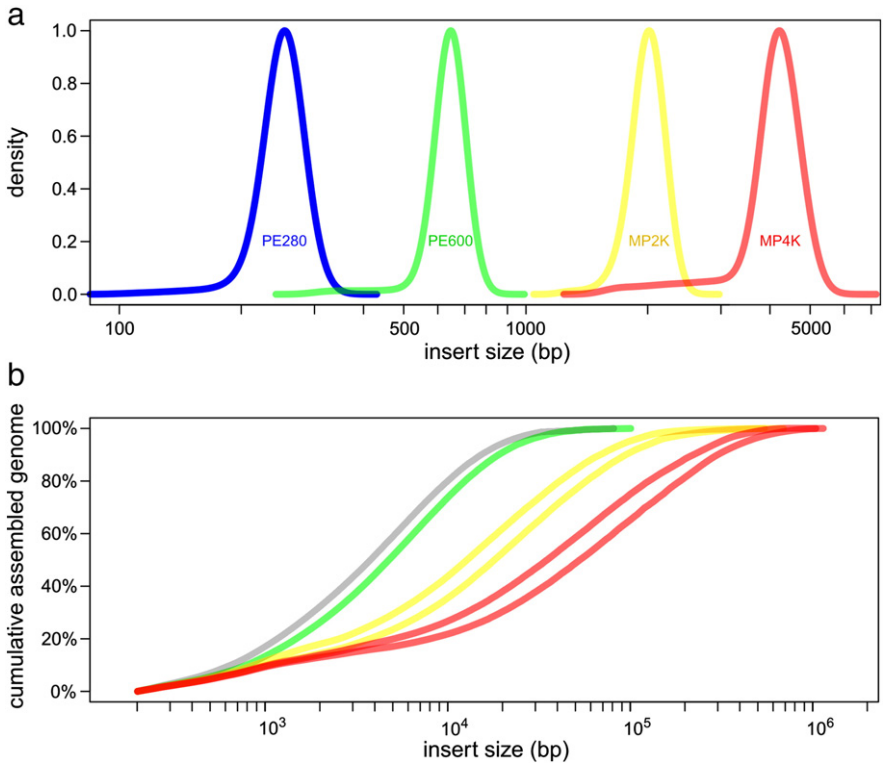

Fig. 3. Input and output of the scaffolding process. (a) Size distributions (log scale) for the four sequencing libraries used for scaffolding, as determined by alignment against a preliminary assembly. Sizes reflect fragment lengths, including sequencing reads from both ends. (b) Cumulative contig/scaffold size distributions at several stages of the hierarchical scaffolding process using SSPACE (Boetzer et al., 2011). The leftmost (gray) line represents the output of the *de novo* assembly, when most of the assembled genome is made up of contigs between 1000 and 10,000 bp in length. Green, yellow and red lines correspond to intermediate assemblies after scaffolding with the PE600, MP2K and MP4K libraries, respectively.

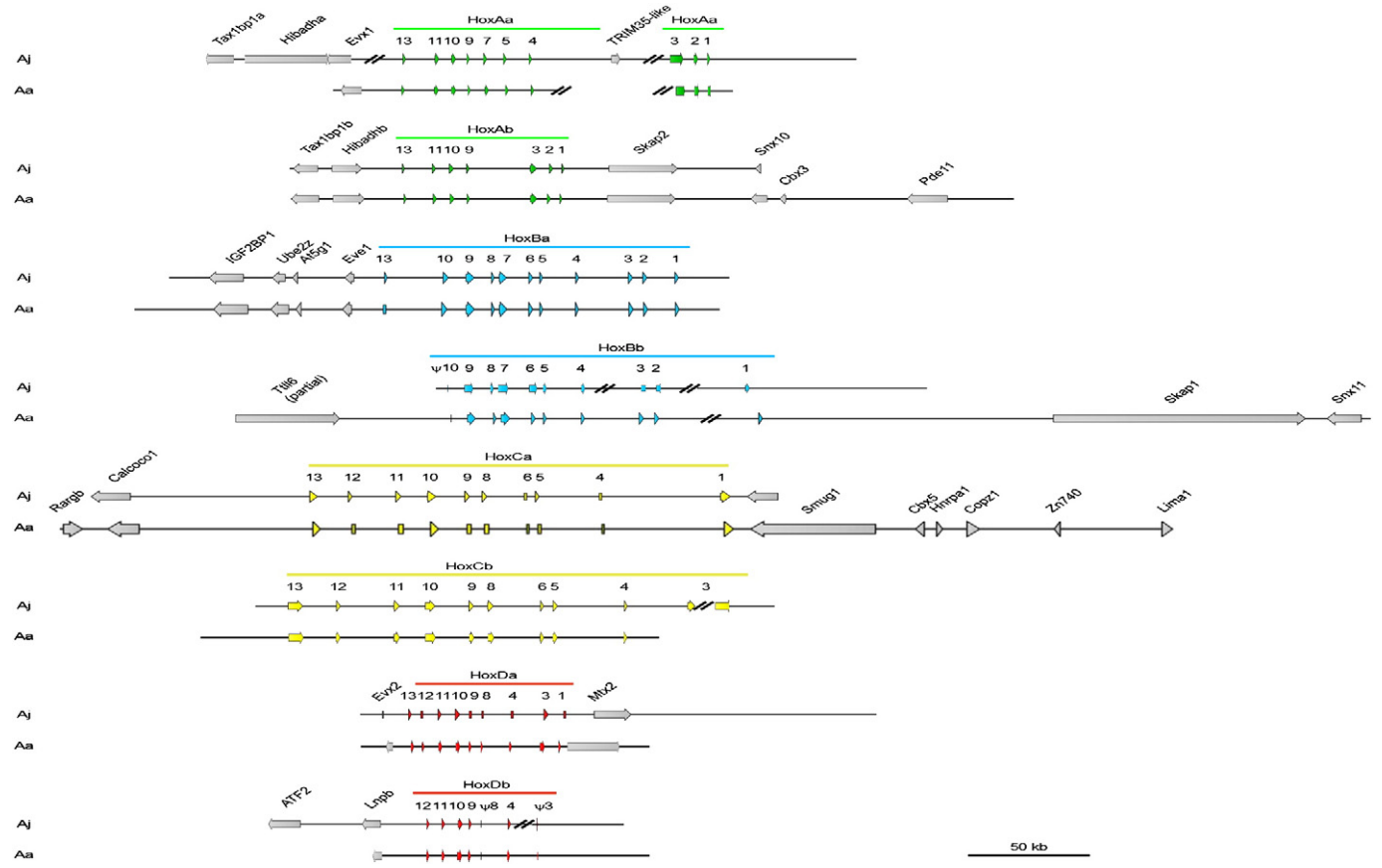

Fig. 4. Hox genes of the Japanese and European eel. The Hox genes of *Anguilla japonica* (Aj) were mapped to genomic scaffolds by using the Hox genes of *Anguilla anguilla* (Aa) as a reference. Scaffolds are indicated by horizontal black lines. Two forward strokes indicate gaps between scaffolds. Hox genes are indicated by colored arrows that are numbered according to their paralogous groups. Pseudogenes are indicated by the symbol ψ. Neighboring genes are indicated by grey arrows.

genomics between eels and other teleosts. Since the Japanese and European eels are the most extensively and well studied species among the genus, the comparisons between these two species will potentially provide more meaningful and direct information on the genetic background of various genetic and biological characteristics. Moreover, since the drastic population decline of these two species has been reported (Dekker et al., 2003), the genetic diversity should be taken into account for conservation, stock management and aquaculture. To this end, the Japanese eel genome is very important to assess its genetic diversity, which has been rarely examined. The *de novo* assembled Japanese eel genome can be immediately applied at multiple levels: e.g. gene prediction software such as the web-based AUGUSTUS tool (Stanke et al., 2004) and annotation software such as BLAST2GO (Götz et al., 2008) can be used to quickly map the protein-coding part of the genome. This will provide a very important resource for all scientists working on this fish species with its remarkable life cycle. The quality of the genome is exemplified by the manual annotation of the 74 Hox genes. Looking at the individual Hox genes we can conclude that the eel is a typical representative of the early genome duplication which occurred in teleost fish. As compared with other fish species such as salmon (Mungpakdee et al., 2008), that has even more Hox genes as a result of an extra genome duplication, the eel has retained more Hox genes during evolution. The Japanese eel is a typical representative of the superorder of the *Elopomorpha*, for which no other genome sequences have yet been published. Since many researchers are intrigued by the complex life cycle of the members of the *Elopomorpha*, this genomic resource will be of great value for future research.

Acknowledgements

We thank Pieter Slijkerman for project financial management. SD received funding from Centre National de la Recherche Scientifique and L'Agence Nationale de la Recherche (grant 08-BLAN-0173), and F-AW from the Norwegian School of Veterinary Science and the Research Council of Norway (grant 184851).

References

Aoyama, J., 2009. Life history and evolution of migration in catadromous eels (genus *Anguilla*). Aqua-BioSci. Monogr. 2, 1–42.
Aparicio, S., et al., 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297, 1301–1310.
Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., Pirovano, W., 2011. Scaffolding preassembled contigs using SSPACE. Bioinformatics 27, 578–579.
Coppe, A., 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.
Dekker, W., Casselman, J.M., Cairns, D.K., Tsukamoto, K., Jellyman, D., Lickers, H., 2003. Worldwide decline of eel resources necessitates immediate action: Quebec declaration of concern. Fisheries 28, 28–30.
Ege, V., 1939. A revision of the genus Anguilla Shaw: a systematic, phylogenetic and geographical study. Dana Rep. 16, 1–256.
Fontaine, M., 1936. Sur la maturation complète des organes génitaux de l'anguille male et l'émission spontanéede ses produits sexuels. C. R. Acad. Sci. 202, 1312–1314.
Götz, S., et al., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–3435.
Henkel, C.V., et al., 2012. Primitive duplicate Hox clusters in the European eel's genome. PLoS One 7, e32231.
Jaillon, O., et al., 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431, 946–957.

Kasahara, M., et al., 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447, 714–719.
Kazeto, Y., Tosaka, R., Matsubara, H., Ijiri, S., Adachi, S., 2011. Ovarian steroidogenesis and the role of sex steroid hormones on ovarian growth and maturation of the Japanese eel. J. Steroid Biochem. Mol. Biol. 127, 149–154.
Kelley, D.R., Schatz, M.C., Salzberg, S.L., 2010. Quake: quality-aware detection and correction of sequencing errors. Genome Biol. 11, R116.
Li, R., et al., 2010. The sequence and de novo assembly of the giant panda genome. Nature 463, 311–317.
Lokman, P.M., Young, G., 2000. Induced spawning and early ontogeny of New Zealand freshwater eels (*Anguilla dieffenbachii* and *A. australis*). N. Z. J. Mar. Freshwater Res. 34, 135–145.
Minegishi, Y., Henkel, C.V., Dirks, R.P., van den Thillart, G.E.E.J.M., 2012. Genomics in eels - towards aquaculture and biology. Marine Biotechnol. 14, 583–590.
Miura, C., Miura, T., 2011. Analysis of spermatogenesis using an eel model. Aqua-BioSci. Monogr. 4, 105–129.
Mungpakdee, S., Seo, H.C., Angotzi, A.R., Dong, X., Akalin, A., Chourrout, D., 2008. Differential evolution of the 13 Atlantic salmon Hox clusters. Mol. Biol. Evol. 25, 1333–1343.
Nomura, K., et al., 2011. A genetic linkage map of the Japanese eel (*Anguilla japonica*) based on AFLP and microsatellite markers. Aquaculture 310, 329–342.
Ohta, H., Kagawa, H., Tanaka, H., Okuzawa, K., Iinuma, N., Hirose, K., 1997. Artificial induction of maturation and fertilization in the Japanese eel, *Anguilla japonica*. Fish Physiol. Biochem. 17, 163–169.
Oliveira, K., Hable, W.E., 2010. Artificial maturation, fertilization, and early development of the American eel (*Anguilla rostrata*). Can. J. Zool. 88, 1121–1128.
Pedersen, B.H., 2003. Induced sexual maturation of the European eel *Anguilla anguilla* and fertilization of the eggs. Aquaculture 224, 323–338.
Shinoda, A., et al., 2011. Evaluation of the larval distribution and migration of the Japanese eel in the western North Pacific. Rev. Fish Biol. Fisher. 21, 591–611.
Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B., 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32, W309–W312.
Tanaka, H., Kagawa, H., Ohta, H., Unuma, T., Nomura, K., 2003. The first production of glass eel in captivity: fish reproductive physiology facilitates great progress in aquaculture. Fish Physiol. Biochem. 28, 493–497.
Tesch, F.-W., 2003. The eel. Blackwell Science, Oxford (UK).
Tsukamoto, K., 2009. Oceanic migration and spawning of anguillid eels. J. Fish Biol. 74, 1833–1852.