

Full Paper

# Highly dense linkage maps from 31 full-sibling families of turbot (*Scophthalmus maximus*) provide insights into recombination patterns and chromosome rearrangements throughout a newly refined genome assembly

F. Maroso<sup>1,\*</sup>, M. Hermida<sup>2</sup>, A. Millán<sup>1</sup>, A. Blanco<sup>2</sup>, M. Saura<sup>3</sup>,  
A. Fernández<sup>3</sup>, G. Dalla Rovere<sup>4</sup>, L. Bargelloni<sup>4</sup>, S. Cabaleiro<sup>5</sup>,  
B. Villanueva<sup>3</sup>, C. Bouza<sup>2</sup>, and P. Martínez<sup>2,\*</sup>

<sup>1</sup>GENEAQUA S.L, 2007 Lugo, Spain, <sup>2</sup>Departamento de Zoología, Genética y Antropología Física, Facultad de Veterinaria, Universidade de Santiago de Compostela, 27002 Lugo, Spain, <sup>3</sup>Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), 28040 Madrid, Spain, <sup>4</sup>Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro, Padova, Italy, and <sup>5</sup>Cluster de Acuicultura de Galicia (Punta do Couso), Aguiño-Ribeira 15695, Spain

\*To whom correspondence should be addressed. Tel. +34 982 822423. Email: francesco.maroso@gmail.com (F.M.); Tel. +34982822425. Fax. +982822001. Email: paulino.martinez@usc.es (P.M.)

Edited by Dr. Toshihiko Shiroishi

Received 15 February 2018; Editorial decision 3 May 2018; Accepted 5 May 2018

## Abstract

Highly dense linkage maps enable positioning thousands of landmarks useful for anchoring the whole genome and for analysing genome properties. Turbot is the most important cultured flatfish worldwide and breeding programs in the fifth generation of selection are targeted to improve growth rate, obtain disease resistant broodstock and understand sex determination to control sex ratio. Using a Restriction-site Associated DNA approach, we genotyped 18,214 single nucleotide polymorphism in 1,268 turbot individuals from 31 full-sibling families. Individual linkage maps were combined to obtain a male, female and species consensus maps. The turbot consensus map contained 11,845 markers distributed across 22 linkage groups representing a total normalised length of 3,753.9 cM. The turbot genome was anchored to this map, and scaffolds representing 96% of the assembly were ordered and oriented to obtain the expected 22 megascaffolds according to its karyotype. Recombination rate was lower in males, especially around centromeres, and pairwise comparison of 44 individual maps suggested chromosome polymorphism at specific genomic regions. Genome comparison across flatfish provided new evidence on karyotype reorganisations occurring across the evolution of this fish group.

**Key words:** *Scophthalmus maximus*, 2b-RAD sequencing, linkage mapping, genome assembly, chromosome reorganisations

## 1. Introduction

The information provided by genomes has become invaluable for research fields that deal with animal production, conservation biology or evolution.<sup>1</sup> High-density (HD) genetic maps provide information on genome organisation by establishing the cartography of thousands of markers, which can aid to disentangle the genetic architecture of productive or evolutionary traits. Moreover, these maps represent an essential tool for comparative genomics and can facilitate whole genome assembly.<sup>2,3</sup> The advent of next generation sequencing (NGS) has dramatically reduced the cost and time required for genomic analysis,<sup>4,5</sup> thus contributing to a fast increase of genomic resources. Nevertheless, while DNA sequencing has become an automated task at platforms, genome assembling still relies on the hierarchical ordering of short DNA fragments (contigs, scaffolds), currently being facilitated by the novel long-read assembling approaches (Nanopore, PacBio, Dovetail, Bionano).<sup>6</sup> HD genetic mapping can assist the process of genome assembly through positioning and ordering scaffolds or validating bioinformatic assembly.<sup>7,8,9</sup> This is a particularly challenging task in teleosts, considering the ancestral teleost-specific whole genome duplication event in this group of fish.<sup>10</sup>

A huge number of markers and samples can now be analysed at very low cost using genotyping by sequencing (GBS) methods,<sup>4,5,9</sup> which enables to simultaneously identify and genotype thousands of single nucleotide polymorphisms (SNPs) across samples. A popular GBS method is Restriction Associated DNA sequencing (RADseq), a technique that combines the power of NGS with the simplification of genomes through restriction enzyme digestion.<sup>11–13</sup> RADseq is being increasingly used for the construction of HD genetic maps in fish comprising thousands of markers (e.g. Refs <sup>14–19</sup>). These maps have allowed the identification of genomic regions associated with productive traits<sup>20,21</sup> or with adaptation to specific environments.<sup>14,16</sup> The pattern of samples and families used in fish for map construction has been quite similar, and only very recently, the number of families has increased.<sup>19,22</sup> However, when many families have been analysed, the number of offspring per family has been much smaller (e.g. Tsai et al. 2016<sup>23</sup>: 60 families with less than 10 offspring each; Palaiokostas et al. 2016<sup>24</sup>: 75 families with 10 offspring on average). This is a way to achieve a consensus reference map at species level by averaging inter-individual variation, but the low family sampling size can result in suboptimal estimation of inter-marker distances and mapping order.<sup>25,26</sup>

The presence of genomic reorganisations and variation of mapping parameters across individual or sex-specific maps has been hardly addressed to date in fish despite its importance from both evolutionary and practical perspectives. For example, inversions might aid to maintain blocks of coadapted genetic variants at specific environments, such as reported in fish like cod<sup>27</sup> or stickleback.<sup>7,28</sup> Moreover, sex-specific mapping features are important when tracking markers associated to economically relevant traits since recombination patterns in males and females may differ and thus, they should be considered for accurate evaluations. NGS technologies now enable studying intra-specific variation on genome organisation and recombination with enough depth to improve the analysis of associations with traits of interest. This would aid for a more accurate mining to identify responsible genes and mutations and choose the best strategy for marker assisted selection.

The turbot (*Scophthalmus maximus*) is the main worldwide aquaculture flatfish,<sup>9</sup> People's Republic of China being currently the most important producer (>60,000 tons/year<sup>29</sup>). Genetic breeding

programs in this species began in the 90s. Nowadays, a broad battery of genetic markers and genomic tools are available in turbot, which can aid to understand the genetic basis of the main economically important traits (growth rate, disease resistance, sex control) and for applying a more efficient selection.<sup>30–35</sup> The first genetic maps in turbot, based on anonymous or gene-associated microsatellites and SNPs,<sup>36–38</sup> were later integrated in a consensus map.<sup>38</sup> More recently, a high-density map was reported by Wang et al.,<sup>21</sup> but the scarce information provided on genetic marker-associated sequences made it not feasible to integrate with previously reported maps.<sup>9</sup> Centromeres have been located in the turbot genetic map using half-tetrad analysis<sup>40</sup>, and recently, a high coverage draft genome has been assembled and anchored to the reported consensus map.<sup>36</sup> Moreover, the development of genomic resources for other important flatfish species<sup>41–43</sup> and their integration in a common framework has allowed comparative analyses to transfer genomic information on production-related traits between species and to look for insights into flatfish evolution.<sup>9,38</sup>

In this study, we genotyped 18,214 SNPs using the 2b-RAD technology in 1,268 individuals from 31 turbot full-sibling (sib) families with the aim of: (i) obtaining a robust consensus genetic map for the species; (ii) integrating the genetic and physical maps in a common framework to facilitate transference of segregation data into the physical genome; (iii) refining the turbot genome assembly and comparative mapping with other flatfish; and (iv) assessing the variation in genome organisation and recombination between sexes and individuals. Our results rendered much deeper information on intra-specific mapping variation than previously reported in fish and provided useful integrated genomic resources to understand the architecture of economically important traits for further applications in industrial breeding programs.

## 2. Materials and methods

### 2.1. Families for mapping

The broodstock of CETGA (Aquaculture Cluster of Galicia, Ribeira, Spain), a population representative of turbot from the Atlantic area, was used to found a set of 44 families within the goals of the FISHBOOST project (EU 613611). This project is focused on improving the efficiency and profitability of finfish European aquaculture by fostering selective breeding to a next level, including the application of genomic selection. This goal also involved the refinement and improvement of existing mapping resources and their integration in a common framework within the reported turbot genome assembly.<sup>34</sup> The experiment was performed under approval of the ethics committee of CETGA to adjust to animal welfare legislation.

A total of 31 full-sib families were selected among the 44 families considering both successful genotyping and informativeness (Fig. 1). The partly factorial mating of 22 males × 22 females performed made available 27, either maternal or paternal, half-sibs families (eight males and nine females contributed to more than one family). A total of 1,224 offspring were used for linkage analysis representing on average 39.5 individuals per family (range: 36–45).

### 2.2. SNP calling and genotyping

#### 2.2.1. DNA extraction and library preparation

Genomic DNA was extracted from fin clips using SSTNE buffer (a TNE buffer modified by adding spermidine and spermine) and a standard NaCl isopropanol precipitation.<sup>44</sup> Library preparation followed the 2b-RAD protocol<sup>45</sup> slightly modified to include replicated

M/F	005	007	168	169	178	183	186	188	189	208	230	232	279	282	329	330	334	343	350	351	353	356	
001						40																	
005			40																				
007				40																			
035												36											
040							40				40												
058													39										
064					40			39															
066									40														
085	40																						
087										39	40												
088	38																						
089		40	40														40						
100													40										
112								38															
131												40											
143														38									
145																	40						38
146																							41
152															40					45			
156																		38			39		
159																		38					
162							38																40

**Figure 1.** Families selected for linkage analysis in *S. maximus*. Male (M) × female (F) crosses and the number of offspring per family are indicated. Note the presence of several half-sib families sharing the father or the mother.

polymerase chain reactions (PCRs) per sample (to reduce PCR duplicate bias) and to quantify DNA sampling before pooling (to increase sample coverage homogeneity). Detailed information about the protocol can be found in [Supplementary Material](#) text. Briefly, DNA of each sample was separately digested using *AlfI* restriction enzyme, adaptors including specific sample barcodes ligated, and the resulting fragments amplified. After PCR purification, samples were merged in pools of 128 samples for sequencing and genotyping the offspring, while the 44 parents were genotyped in a single pool, to increase coverage and genotyping accuracy. A total of 11 pools (1 for the parents and 10 for the offspring) were sequenced on an Illumina NextSeq sequencer using a 50 bp single end protocol at BMR Genomics (Padova, Italy).

### 2.2.2. Raw data filtering and bioinformatics

Raw data were demultiplexed at the sequencing facility. All reads were trimmed at 36 bp and centred on the enzyme recognition site. Stacks' *process\_radtag* pipeline<sup>46,47</sup> was used to filter raw reads according to sequencing quality. All reads with an uncalled base were removed, and quality was checked using a sliding window of 25% of the sequence length. If the mean quality score of this window dropped below 20, the sequence was removed.

Filtered reads were mapped against the turbot genome<sup>34</sup> using Bowtie 1.2.0<sup>48</sup> and allowing a maximum of three mismatches in the first 33 bases. Only a unique match on turbot genome was permitted, otherwise reads were discarded. Mapped reads from parent samples were fed into Stacks' module *ref\_map.pl* using the bounded model with a significance threshold of 0.05 and an upper bound of 0.05. As a further filter, only SNPs genotyped in 80% of the parents with a minimum read depth of 10 reads were retained using the *populations* module. When a RAD-tag contained more than one SNP, only the first SNP from the 5' end was retained. The set of SNPs obtained

from parents was used as a reference for offspring genotyping using the Stacks module *sstacks*. The obtained dataset (parents and offspring) was further filtered according to population parameters and Mendelian segregation using the 26 most informative families. Those markers showing consistent deviations from Mendelian segregation ( $P < 0.01$ ) across families or extreme departures from Hardy-Weinberg equilibrium (HWE) in the parental population ( $P < 0.001$ ) due to heterozygote excess ( $F_{IS} < -0.5$ ,  $P < 0.0001$ ; paralogous genes) or deficit ( $F_{IS} > 0.5$ ,  $P < 0.0001$ ; null alleles) were filtered out. Finally, RAD-tags containing SNPs were mapped against the turbot gene catalog<sup>38</sup> using Blastn (E-value  $< e^{-20}$ ) to obtain sequence annotations.

### 2.3. Construction of genetic maps

Genotypes of filtered SNP markers were properly coded as JOINMAP Cross Pollinator (CP) cross type with unknown linkage phase. Markers genotyped in less than 90% of the offspring were excluded for linkage analysis in each family. All genetic maps (individual, family, male consensus, female consensus and species consensus) were constructed using JoinMap 4.1.<sup>49</sup> First, markers were associated to their linkage groups (LGs) using the *Grouping* function of JoinMap based on a series of LOD scores (logarithm of the likelihood ratio that two genes are linked regarding that they segregate independently) increasing by one, from 3 to 10. The LOD score was then selected in each family based on the number of turbot chromosomes ( $n = 22$ ); in all cases, a  $LOD > 6.0$  was used. Unlinked markers and markers assigned to very small LGs exceeding the haploid number of 22 chromosomes were excluded from further analysis. Second, marker ordering was performed using the Maximum Likelihood algorithm with default parameters, and the recombination frequencies converted into map distances in centi-Morgans (cM) using the Kosambi mapping function. For each family, individual

female and male maps, as well as a consensus family map, were generated. Due to computational limitations, only the three most informative families were selected to construct the species consensus, the female consensus and the male consensus maps using MergeMap.<sup>50</sup> All maps constructed were drawn using MapChart ver. 2.3.<sup>51</sup> In order to investigate the heterogeneity of recombination frequency (RF) between sexes, RFs were extracted from JoinMap analysis. The pattern of RF in the father and the mother could not be straightforwardly compared within each family due to the different panel of informative SNPs; consequently, the set of markers shared between both parents was selected in each family for a detailed evaluation of RF variation between sexes across LGs within each family (heterozygous in both parents with 1:2:1 segregation).

#### 2.4. Genetic versus physical map: anchoring scaffolds to linkage groups

The consensus genetic map was used to refine the turbot genome assembly<sup>34</sup> and to obtain a physical map consistent on 22 megascaffolds, representing the complete sequence of each of the 22 LGs of turbot map.<sup>32,39</sup> This task implied (i) to assign previously unlinked scaffolds to specific LGs in the new high-dense (HD) genetic map; (ii) to correct scaffold misassemblies by comparing the genetic and physical maps; and (iii) to establish the correct orientation of scaffolds within each megascaffold. All contigs including at least one marker significantly assigned to a LG were anchored in the new map. When markers of a specific scaffold were assigned to more than one LG, the scaffold was split into fragments pertaining to different LGs or to the same LG but in different positions (see Results). To establish the most confident splitting of these scaffolds, we compared the sequence of the gap between the flanking markers assigned to different LGs with the two model fish genomes, *Gasterosteus aculeatus* and *Oryzias latipes*, which had demonstrated large synteny with turbot.<sup>34,38</sup> Then, we looked for the turbot genes in the gap assigned to different LGs in the model species in order to narrow down as much as possible the gap (Supplementary Fig. S1). Following this approach, the narrowest gap was finally cut in the middle and halves assigned to each LG. Reorganisations were only validated when information was consistent with both model species. To find out the correct order and orientation of scaffolds within LGs, we obtained the average map position of the markers belonging to each scaffold using the maps of the 10 most informative families. Then, we checked for the correlation between the physical position of markers and their genetic position. When the correlation was positive, the original orientation was maintained, and when negative, the reverse complementary sequence of the scaffold was retrieved for assembling.

#### 2.5. Integrating genetic and physical maps with cytogenetic data

Genetic and physical maps were then integrated with the turbot cytogenetic map using previous information.<sup>32,40,52</sup> Starting from centromere positions<sup>40</sup> and considering the relationship between genetic and cytogenetic maps,<sup>32</sup> we established the correspondence between LGs, megascaffolds and chromosomes as far as possible. Then, we compared the consensus genetic map of each chromosome with its physical assembly (megascaffold) to assess their correspondence with centromere, telomere and interstitial heterochromatin locations. When no marker was available close to the centromeric region, the closest markers around the centromere were used as reference or, in a few cases, the farthest marker from the centromere was used for its

rough location at the other chromosome end, considering the acrocentric morphology of most turbot chromosomes.<sup>52</sup> Finally, this comparison was also carried out between sexes using the sex-specific consensus maps, in order to assess RF differences across chromosomes between males and females.

#### 2.6. Inter-individual genomic reorganisations in turbot

The availability of 31 family maps and 44 parental maps (Fig. 1), enabled an exhaustive comparison of genetic maps between individuals and sexes to assess inter-individual variation on RF across the genome. This variation, which might be related to chromosome polymorphisms existing in turbot populations, had not been previously addressed in fish with the number of markers and families here evaluated. In order to identify putative genomic reorganisation between individuals, we performed a pair-wise comparison between all parental maps by plotting the positions of each marker. Plots were created with R<sup>53</sup> and the different LGs were labelled with different colours for a better visualisation. In order to handle the amount of data obtained, 62 individual maps at 22 LGs (due to half-sib families; Fig. 1), each LG was divided in stretches of 50 SNPs representing ~1.6 Mb on average (e.g. 23 blocks for LG02, the longest LG with 1,132 SNPs). Then, the correlation between the physical and genetic maps was analysed for each block at each individual map in order to identify potential inversions. A heat map was generated to provide a more straightforward view of the results.

#### 2.7. Interspecific genomic reorganisations between turbot and other flatfish species

The improved turbot genome was further exploited to complete and refine previous information about genomic reorganisations and chromosome evolution of flatfish.<sup>9,34</sup> For this, we compared the improved turbot genome with those of Japanese flounder (*Paralichthys olivaceus*<sup>43</sup>) and half-smooth tongue sole (*Cynoglossus semilaevis*<sup>42</sup>) at chromosome level using LASTZ<sup>54</sup> (using options ‘-notransition -step = 400 -nogapped -format = rdotplot’) and compared each turbot chromosome with the orthologues of the other two species. Further, in order to assess the relation between flatfish megascaffolds to a model fish genome outside Pleuronectiformes, an additional LASTZ analysis was carried out with stickleback (*Gasterosteus aculeatus*) using the same parameters.

### 3. Results

#### 3.1. SNP calling, filtering and genotyping

A total of 3,946,354,960 reads (3.9 Gb) were produced at the BMR sequencing platform, corresponding to 1,282 individuals represented by 3,078,280 reads on average (range: 312 – 9,610,380). Fourteen offspring were removed due to the low number of usable reads (<75,000), leaving a total of 1,268 individuals (44 parents and 1,224 offspring) pertaining to 31 full-sib families for subsequent analyses.

After quality filtering and alignment to the turbot genome, the average number of reads per sample was 6,663,809 (range: 2,601,955–8,946,621) for parents and 2,774,579 (range: 283,203–5,596,610) for offspring. After Stacks’ *populations* filter, 139,293 RAD-tags were retained, of which 25,511 (18.3%) contained at least one SNP. Since only one SNP was retained per RAD-tag, a total of 25,511 SNPs were considered after eliminating the remaining SNPs in the same RAD-tag (Supplementary Table S1). Among these 25,511

SNPs, 18,214 were finally retained after filtering loci by: (i) consistent deviations from Mendelian segregation (5,866 loci:  $P < 0.01$  in at least two families and representing more than 20% of the informative families); (ii) very low minimum allele frequency in the parental population ( $< 0.015$ ) suggestive of genotyping errors (991 SNPs); and (iii) extreme deviations from HWE suggesting either paralogous genes (364 loci) or high null allele frequency (91 loci) (Supplementary Table S1). This set of 18,214 SNPs represented a frequency of one SNP every 163 bp in the reduced portion of the genome sequenced with 2b-RAD, a frequency slightly lower than that reported in other studies in turbot,<sup>55,56</sup> and a density of one SNP every 31,130 bp in the whole turbot genome (567 Mb<sup>34</sup>). The average proportion of genotyping errors (Mendelian inconsistencies in families) of our 2b-RAD genotyping approach was 0.68% throughout all families in the filtered SNP dataset.

Out of the 18,214 SNPs finally retained, nearly half matched to turbot genes (9,028). Among those lying within exons (834), 400 corresponded to non-synonymous and 430 to synonymous allelic variants (Supplementary Table S1).

### 3.2. Construction of genetic maps

On average, 5,385 SNPs were used for map construction in each of the 31 families analysed (range: 2,079 in Fam55 to 6,757 in Fam30; Supplementary Table S2). Maps consisted of 22 LGs corresponding to the 22 chromosomes of the turbot haploid karyotype, averaging 140.5 cM and 245 markers per LG and family (Supplementary Tables S2 and S3). Note that LGs nomenclature follow previous publications<sup>32</sup> and accordingly, LG18 is missing due to its merging with LG08 (LG08 + 18, here LG08). LG22 was on average the shortest LG (94.9 cM; range: 28.1 cM in Fam55 to 223.5 in Fam19) and included the lowest number of markers (189.9; range: 56 in Fam16 to 257 in Fam39). LG02 was on average the longest LG (207.7 cM; range: 49.9 cM in Fam59 to 414.0 cM in Fam19) with the highest number of markers (341.7; range: 97 in Fam55 to 457 in Fam39). A total of 17,169 SNPs were successfully mapped at least in one family, 1,045 remaining unmapped. Of the last, 1,029 were located in scaffolds where other markers were anchored to the linkage map and could therefore be assigned to specific LGs. Only 16 SNPs could not be assigned to any LG. Marker interval across all families averaged 0.6 cM and ranged from 0.43 cM in Fam11 to 1.08 cM in Fam19. If unique positions instead of markers were accounted for, marker interval averaged 2.0 cM and ranged from 1.81 in Fam13 cM to 2.74 cM in Fam55.

The turbot consensus map constructed with the three most informative families using MergeMap included 11,845 markers distributed over 22 LGs representing a total length of 8,532.6 cM (Supplementary Table S3A; Fig. 2). LG length varied from 282.9 cM (LG22) to 588.7 cM (LG02), with an average length of 387.8 cM. The average marker interval was 0.72 cM and ranged from 0.57 cM in LG01 to 0.89 cM in LG4. Because of the inflated length observed for this consensus map, we also estimated the average LG lengths from the three families used to construct the consensus, as these values represent more realistic lengths than those rendered by the consensus MergeMap algorithm (see Discussion) (Table 1). These LG lengths sum up to 3,753.9 (i.e. an average marker interval of 0.32 cM).

Female and male consensus genetic maps differed significantly totalling 8,274.8 cM and 6,791.6 cM (F:M ratio = 1.22), with average LG length of 376.1 and 308.7 cM, respectively. In both females and males LG02 was the largest LG (545.8 vs 474.9 cM) and LG22 the

shortest one (292.1 vs 198.7 cM) (Supplementary Fig. S2). As for the species consensus map, Table 1 reports average LG lengths also for female and male consensus maps.

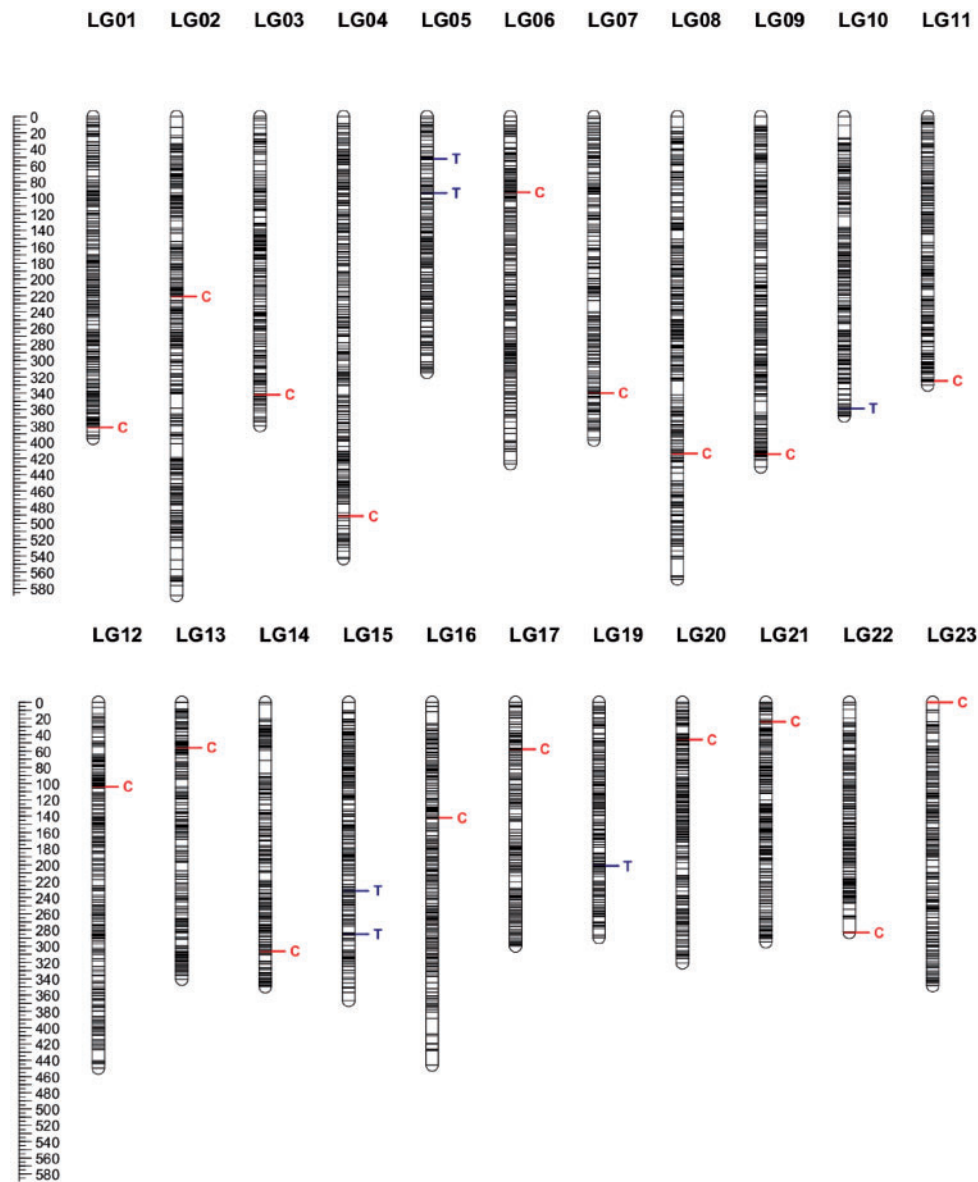
Individual female and male genetic maps (62 in total; Supplementary Tables S2 and S3) were also constructed for a more detailed comparison of RF between individuals and sexes, and to look for putative chromosome polymorphisms in the turbot population (44 parents). The length of female genetic maps was higher than that of male maps in all families and for all LGs (average F:M ratio = 1.36). For a more detailed comparison, 10,168 markers shared between sexes and mapped in at least one family were selected (Supplementary Table S4). As expected, the average inter-marker distance for the same pair of markers was significantly higher in females than in males (2.2 and 1.7 cM, respectively;  $P < 0.001$ , *t*-test for related samples). When F:M intervals were compared at each LG, the difference was significant for all groups excluding LG02, LG17 and LG22 ( $P > 0.05$ ) (Supplementary Table S4). When F:M intervals were compared across all LGs within each family, again inter-marker distances were higher and significant in females in all but four families (families 16, 22, 47 and 55,  $P > 0.05$ ; Supplementary Table S4).

### 3.3. Anchoring turbot genetic and physical maps: megascaffolds

Scaffolds were assigned to the 22 LGs of the turbot consensus map, then ordered and oriented, thus constituting 22 megascaffolds. Series of 100 Ns were used in the megascaffold sequence to show the gap between adjacent scaffolds denoting missing information. Additionally, the short uncertain regions around the boundaries of split scaffolds were written in lowercase to highlight their lower confidence. The complete nucleotide sequence of the 22 megascaffolds with their correspondent genes located and annotated has been uploaded to NCBI (SUB3239805).

A total of 28 scaffolds were split and reallocated according to the very consistent mapping information in all the 31 families analysed (Supplementary Table S5). Most scaffolds (22) were split into two fragments and reallocated to different LGs. In three cases, scaffolds were split in three different LGs (sm5\_s00043, sm5\_s00056 and sm5\_s00084), and in other three cases, the two fragments of the scaffold were reallocated within the same LG (sm5\_s00006, sm5\_s00007 and sm5\_s00049). In all cases, the boundaries of the breakage point were assigned with notable accuracy (Supplementary Fig. S1), and uncertain gap regions averaged 38,010 bp, ranging from 549 bp (sm5\_s00086) to 160,155 bp (sm5\_s00084) (Supplementary Table S5). Rearranged scaffolds represented a total of 108.5 Mb, to say ~20% of the entire turbot genome.

A total of 17,169 SNPs corresponding to 558 scaffolds (after splitting 28) rendered a significant match to the turbot genetic map (Supplementary Table S1). On average each scaffold contained 32.7 SNPs ranging from 610 SNPs in sm5\_s00003\_17 (LG17) to one SNP in 204 scaffolds. The total length of the anchored sequences was 524,925,810 bp, representing 96.5% of the turbot genome assembly (544 Mb). The average megascaffold length was 23,860,264 bp, ranging from 14,917,408 bp (LG22) to 31,897,078 bp (LG02) (Table 1). Megascaffolds contained on average 827 SNPs (range 578–1144) and 25.4 scaffolds (range 14–46) (Table 1). All scaffolds assigned to chromosomes were consistently ordered within the megascaffold according to the mapping information across families; the orientation within each chromosome could be verified for those 336 that contained at least two informative markers. The remaining 204 scaffolds with only one marker could not be oriented, but they



**Figure 2.** Consensus genetic map of *S. maximus* using the three most informative full-sib families; genetic distance in cM in the left bar; the position of centromeres (C) or genetic markers far from centromere (T) in acrocentric chromosomes are indicated.

represented a very minor proportion of the total assembly (2.0%) because of their small size. Interestingly, they were mostly located at the end of chromosomes (telomeres), suggesting assembling problems caused by repetitive motifs.

### 3.4. Integrating cytogenetic information: chromosomes vs LGs vs megascaffolds

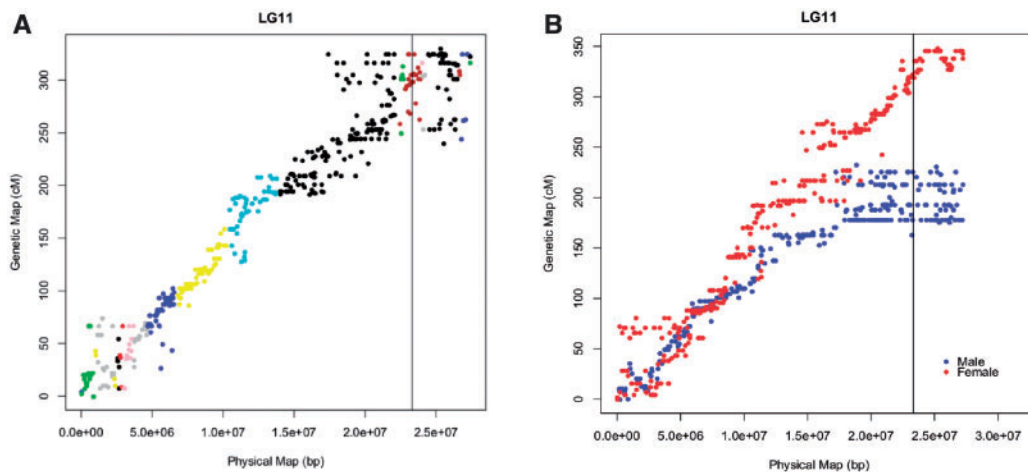
Turbot genetic, physical and cytogenetic information obtained in this work and from previous studies was integrated as far as possible, and accordingly, a unified nomenclature is proposed in this study (Table 1). Four chromosomes (C1, C2, C3 and C22)<sup>34</sup> were unequivocally assigned to specific LGs: the two metacentric C1 (LG02) and C2 (LG16), the submetacentric nucleolus organiser region (NOR)-bearing chromosome C3 (LG15) and the smallest acrocentric chromosome C22 (LG22), constituting the four marker chromosomes identifiable

in the turbot karyotype. The remaining megascaffolds, which correspond to acrocentric chromosomes of decreasing size, were termed according to their physical length from C4 to C21 (Table 1).

The consensus genetic and physical maps showed a linear correspondence along most chromosome length in all LGs, excluding those regions around centromeres and telomeres (Fig. 3A and Supplementary Fig. S3). A cloud of dots indicative of poor mapping accuracy were observed around most mapped centromeres<sup>40</sup> (i.e. LG01, LG10, LG11, LG19 and LG20). Dot clouds were located in intermediate regions of the metacentric chromosomes (LG02 and LG16) or at the terminal regions in the acrocentrics in accordance with their cytogenetic structure. LG15, the NOR-bearing chromosome, showed a moderate dispersion of dots throughout the proximal centromere chromosome half likely related to the tandem rDNA clusters and their associated heterochromatin in that region.<sup>52</sup> Also, close to telomeres, although not so markedly, dots were either

**Table 1.** Average length and constitution (SNP and scaffold number) of LGs in the species, female and male *S. maximus* consensus maps

Karyotype		Physical map				Species consensus map			Female consensus map			Male consensus map		
Chromosome ID	LG	Megasccaffold length (bp)	No SNPs	No scaffolds	Average contig length (bp)	Mapped markers	Average length	Mb/cM	Mapped markers	Average length	Mb/cM	Mapped markers	Average length	Mb/cM
C7	01	26,874,605	1,031	32	839,831.4	671	178.3	0.151	487	176.6	0.152	409	162.9	0.165
C1	02	31,897,078	1,144	46	708,824.0	761	255.8	0.125	513	261.6	0.122	546	215.0	0.148
C17	03	21,400,935	785	23	930,475.4	506	180.7	0.118	317	179.1	0.119	395	156.3	0.137
C5	04	29,497,696	907	31	951,538.6	610	220.8	0.134	412	251.0	0.118	439	171.7	0.172
C12	05	24,809,483	707	20	1,240,474.2	443	125.0	0.198	300	129.1	0.192	275	107.2	0.231
C10	06	25,205,370	851	26	969,437.3	587	186.1	0.135	395	205.3	0.123	455	153.3	0.164
C13	07	24,510,356	699	20	1,225,517.7	412	168.3	0.146	317	179.4	0.137	301	143.3	0.171
C4	08	30,930,439	1,066	38	813,958.9	640	266.3	0.116	434	303.3	0.102	427	196.0	0.158
C8	09	25,788,341	929	20	1,289,417.1	593	200.0	0.129	415	209.6	0.123	406	165.1	0.156
C11	10	25,098,921	922	25	1,003,956.8	650	169.1	0.148	419	166.8	0.150	497	156.5	0.160
C6	11	27,408,095	808	33	830,548.3	494	150.9	0.182	406	174.1	0.157	323	108.9	0.252
C9	12	25,240,469	796	21	1,201,927.1	536	194.9	0.130	395	204.2	0.124	376	162.2	0.156
C19	13	20,256,549	726	20	1,012,827.5	472	154.0	0.132	349	176.3	0.115	346	119.2	0.170
C16	14	21,450,603	778	20	1,072,530.2	514	145.1	0.148	374	158.2	0.136	402	114.3	0.188
C3	15	24,127,324	887	33	731,131.0	593	146.4	0.165	413	156.7	0.154	383	113.0	0.214
C2	16	23,833,596	813	32	768,825.7	531	197.8	0.120	375	201.1	0.119	371	171.1	0.139
C21	17	15,881,475	680	16	992,591.8	467	127.6	0.124	335	141.3	0.112	332	103.9	0.153
C15	19	21,781,057	755	22	990,048.0	492	128.7	0.169	372	123.0	0.177	345	118.0	0.185
C14	20	22,751,219	776	23	989,183.4	456	145.5	0.156	304	163.1	0.139	322	116.4	0.195
C18	21	21,354,662	760	21	1,016,888.7	501	136.2	0.157	356	138.6	0.154	386	123.8	0.172
C22	22	14,917,408	578	14	1,065,529.1	408	121.0	0.123	291	128.6	0.116	259	102.4	0.146
C20	23	19,910,129	796	22	905,005.9	508	155.4	0.128	378	151.4	0.132	378	129.4	0.154
Sub-total		524,925,810	18,194	558		11,845	3,753.9	0.140	8,357	3,978.4	0.132	8,373	3,109.9	0.169
Not anchored		19,312,992	16	15,936	1,211.9	6,369			9,857			9,841		
Total		544,238,802	18,214	16,494		18,214			18,214			18,214		

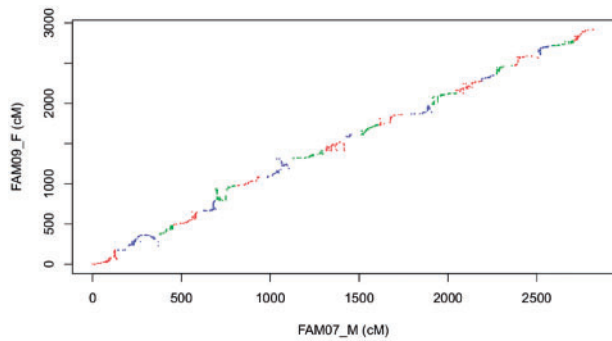
**Figure 3.** Correspondence between physical (abscissae) and genetic (ordinates) maps showing the dispersion of dots around the centromere at LG11 (A) and the difference between male and female maps at LG11 (B). The vertical bar in both figures show the position of the centromere according to Martínez et al. (2008)<sup>42</sup>. Genetic markers corresponding to different scaffolds are shown in different colors.

dispersed (i.e. LG03, LG08, LG11 and LG14) or showing particular drawings (LG02, LG06 and LG15) suggestive of wrong physical assembly or genetic mapping. Finally, the pattern of the acrocentric chromosome C5 (LG04) was remarkable because of the presence of an internal cloud of dots in addition to that associated around the centromere at the chromosome end. This LG very likely correspond to the chromosome 5 of turbot karyotype, which showed an interstitial heterochromatic block with C-banding.<sup>52</sup> When the same comparison between genetic and physical maps was performed for the female and male consensus maps, dot clouds around centromeres were very prominent in the male map and subtle, and sometimes

inexistent, in the female maps (Fig. 3B and Supplementary Fig. S4; i.e. LG01, LG05, LG09, LG15, LG19 and LG22). So, it seems that the length difference between male and female maps is mainly related to the lower recombination rates around centromeres characterising male maps.

### 3.5. Intraspecific chromosome reorganisations

Marker positions showed a good correspondence between the 62 parental genetic maps (including ‘replicates’ from half-sib families) although particular drawings were suggestive of different marker



**Figure 4.** Particular drawings denoting the lack of collinearity between two individual genetic maps of *S. maximus*: Female Fam09 vs Male Fam07 maps; consecutive LGs are represented in different colours from LG01 (left bottom) to LG23 (right top).

order between individual maps (Fig. 4 and Supplementary Fig. S5). The negative correlations between physical and genetic maps at these regions supported this observation, and the presence of putative chromosome rearrangements that involved either single blocks (50 SNPs,  $\sim 1.6$  Mb) or longer chromosome tracts was identified (Supplementary Fig. S6). Inverted single blocks were very frequent at LG02 (blocks 7 and 23), LG05 (block 12), LG08 (block 2), LG13 (blocks 14 and 15) and LG22 (block 12), while larger inverted chromosome tracts were observed in some individuals at LG06 (blocks 5–9) and LG09 (blocks 13–16) (Supplementary Figs S5 and S6). However, these putative inversions, particularly those related to single blocks, were found close to telomeres or centromeres, where mapping accuracy is poorer both due to assembling problems and to the lower ratio of genetic/physical distance. This suggests that they are more likely related to technical devices than to true reorganisations. Otherwise, the longer tracts observed at LG06 and LG09 were located at interstitial regions, and thus, signatures of inversions should be more consistent.

### 3.6. Interspecific chromosome reorganisations

The whole turbot megascaffold sequences were aligned against the genomes of Japanese flounder and tongue sole using the LASTZ program and a 1:1 macrosyntentic pattern was observed (Supplementary Fig. S7A and B), as reported previously.<sup>9</sup> However, some Robertsonian translocations (fusions or fissions) and other minor inter-chromosomal rearrangements were observed. The most notable cases when comparing turbot ( $n=22$ ) and Japanese flounder ( $n=24$ ) genomes were the correspondence of turbot LG02 with Japanese flounder Chr06 and Chr14 (Fig. 5A) and that of LG16 with Chr09 and Chr16 (Supplementary Fig. S7A). In the same way, turbot LG02 was related to tongue sole Chr13 and Chr14 (Fig. 5B), which suggests that LG02 fusion is a derived condition (apomorphy) in the turbot lineage. The position of turbot centromeres at LG02 was the expected one if a centric fusion between two chromosomes had occurred in the turbot lineage (Fig. 5), as also occurred for LG16 (Supplementary Fig. S7A). These chromosome fusions, starting from the ancestral teleost karyotype ( $2n=48$ ),<sup>57</sup> would explain the karyotype differences between turbot and Japanese flounder, as suggested by Robledo et al.<sup>9</sup> Despite the general macrosyntentic pattern observed, pairwise alignment of chromosomal sequences indicated that extensive intrachromosomal reorganisations (mainly inversions) would have taken place throughout the evolution of these species (Supplementary Fig. S7A and B). Reorganisations were much more

extensive when comparing turbot to half-smooth tongue sole than to Japanese flounder (Fig. 5C and D). The results of the comparison with stickleback are presented in Supplementary Figure S7C. Similar fusion patterns as those identified in the comparison with the other flatfish species are visible, as for example at LG16, that showed macrosyntentic patterns with stickleback's ChrVII and ChrIV.

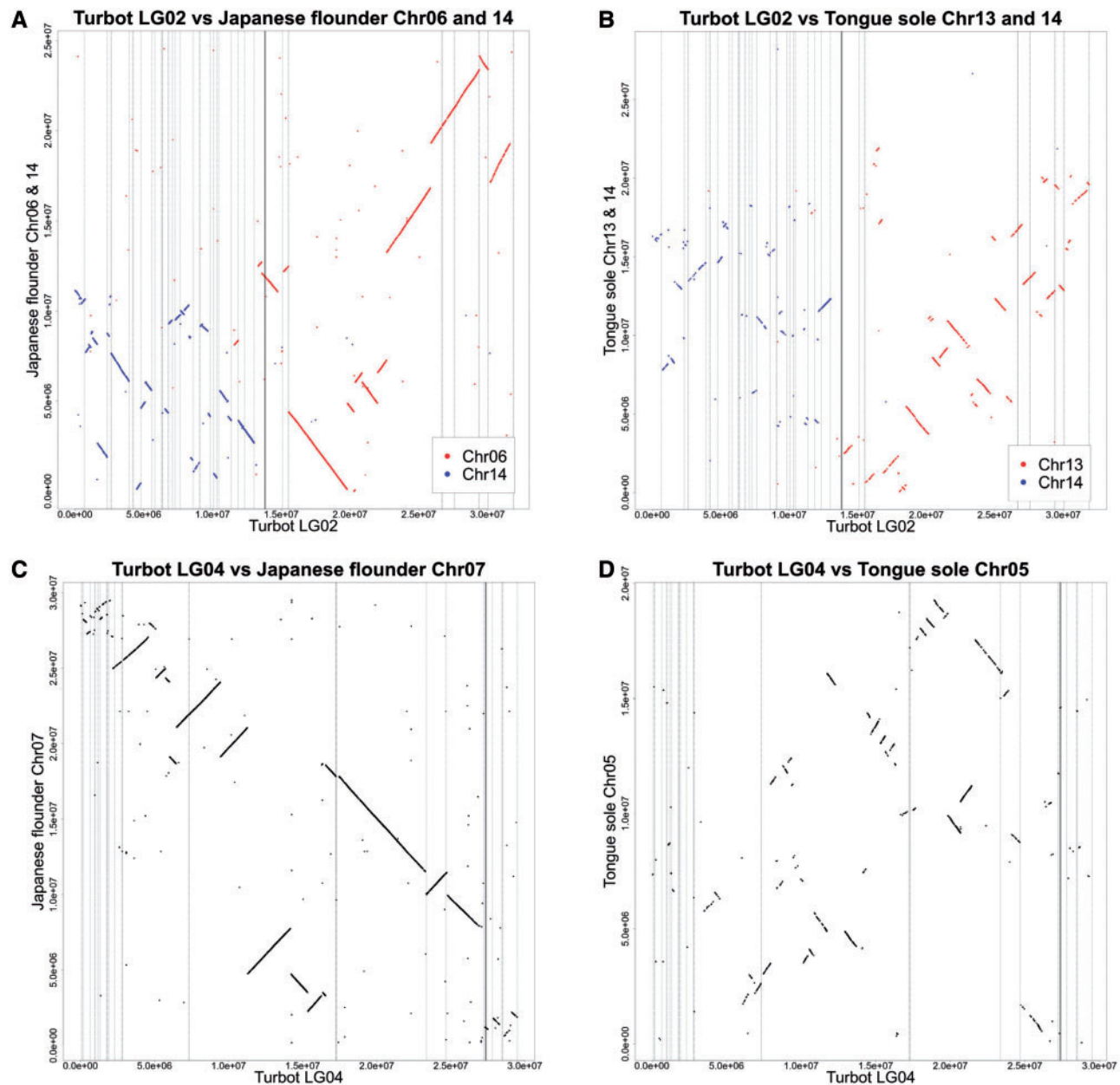
## 4. Discussion

The reduced cost of the new genotyping by sequencing (GBS) techniques has facilitated the development of highly dense genetic maps in aquaculture fish species,<sup>3,17–19</sup> including turbot.<sup>21</sup> Here, we combined for the first time a high number of markers and families for a deep mapping analysis aimed at integrating genetic and physical map resources in turbot to understand its genome organisation and variation, a study not performed to date in fish.<sup>23,58</sup> The average number of SNPs in the 31 full-sib families analysed in our work was in the upper range reported in the most recent linkage mapping studies (5–6k SNPs) and  $\sim 40$  individuals were genotyped at each family. The turbot consensus map (8,532.6 cM) was much longer than any previous linkage map reported for the species (6,647 SNPs, 2,622.1 cM<sup>21</sup>), but this elongation is the consequence of combining a high number of markers with several mapping families and the MergeMaps software, an artefact highlighted previously by different authors.<sup>24,59–61</sup> A direct consequence of this 'inflated' map length is that the average ratio between physical and genetic distance is much lower than usual (0.066 Mb/cM vs  $\sim 0.33$  Mb/cM in teleost). A way to compensate this bias is to normalise for each LG consensus map coordinates to the average genetic distance obtained from the individual maps used to construct the consensus<sup>62</sup> (see Table 1). Accordingly, the average length and average inter-marker distance of the three single family maps used to construct the consensus was 3753.9 and 0.32, respectively, figures much closer to those found by Wang et al.<sup>21</sup> using RADseq. Interestingly, the sex-determining region-bearing chromosome (LG05<sup>30</sup>) showed one of the lowest RF ratio (cM/Mb), which agrees with theoretical expectations on the evolution of the sex-determining chromosome.<sup>32</sup>

The anchorage of the assembled turbot genome to a robust and dense genetic map is essential to transfer marker associations detected by segregation analysis in families to physical positions in the genome for gene mining.<sup>63</sup> In our study,  $\sim 20\%$  of turbot genome assembly<sup>34</sup> was refined, thanks to the mapping information obtained across the 10 most informative families. Furthermore, scaffolds representing 96.5% of turbot genome assembly were anchored to chromosomes, ordered and oriented, and thus, a megascaffold was set up for each LG. This represents a significant improvement from the 80% anchored with the previous medium density turbot map.<sup>34</sup> Also, the available cytogenetic information<sup>32,40,52</sup> was integrated as much as possible in a common framework with genetic and physical maps, and four marker chromosomes of the turbot karyotype were associated with their respective megascaffolds, and the remaining 18 roughly assigned according to their physical length. Using this information, a new nomenclature is here proposed for the 22 chromosomes/LGs of the species, which will facilitate the integration of the different levels of genome organisation for comparative studies within flatfish and teleosts.<sup>9</sup>

The high correlation observed between genetic and physical maps supports the confidence of marker ordering and physical assembling of turbot genome. However, this general trend was lost at specific chromosome structures such as centromeres and telomeres. Around





**Figure 5.** LASTZ plots between *S. maximus* linkage groups (LG) and chromosomes (Chr) of *P. olivaceus* and *C. semilaevis*: (A) LG02 vs Chr06 and 14 of *P. olivaceus*; (B) LG02 vs Chr13 and 14 of *C. semilaevis*; (C) LG04 vs Chr07 of *P. olivaceus*; (D) LG04 vs Chr05 *C. semilaevis*. The gross vertical bar points the position of centromeres according to Martínez et al. (2008)<sup>42</sup> and the light ones the limits between consecutive scaffolds.

centromeres extensive dot clouds were visualised, suggestive of poor mapping accuracy very likely related to the decay of RF in the vicinity of heterochromatic blocks.<sup>64,65</sup> This was observed not only for pericentromeric heterochromatin but also for interstitial heterochromatic blocks such as the most conspicuous one detected at LG04 (C5). A remarkable RF difference was observed between males and females around heterochromatic structures, especially the centromeres, where females hardly showed RF decay. This fact is the main responsible for the length difference between male and female maps of turbot, as reported previously in other fish species.<sup>58,66–68</sup> It is tempting to ascribe the different RF between males and females observed in fish<sup>19,66</sup> to different crossing-over patterns around heterochromatic blocks. However, other explanations have been proposed

for different species, such as the GC content in Asian sea bass<sup>19</sup>; stronger selection pressure in male gametes during the haploid life stage in zebrafish<sup>66</sup> or in some plants<sup>69</sup>; and differences in chromatin distribution between sexes as for example in humans<sup>70</sup>. More information is needed for a comprehensive explanation to these observations, and indeed, NGS technologies can provide suitable data for this task.

The lack of correspondence between physical and genetic maps observed at telomeres in several families and LGs is likely a consequence of a wrong assembly and/or mapping at chromosome ends. These regions were characterised by the presence of a number of short contigs, suggestive of assembling problems likely due to the presence of repetitive DNA. Similarly, when many markers and a

comparatively small offspring number per family are handled, mapping programs through the least confident markers to LG ends, which lead to the loss of collinearity between physical and genetic maps.<sup>15</sup> Improving genomic assembly at chromosome ends by using long-read sequencing methods and increasing progeny number per family will be needed to solve this problem.

Genetic maps of bony fish reported to date have been based on a few families,<sup>19,38,58</sup> and thus, the presence of inter-individual rearrangements affecting marker order has been scarcely addressed. Previous information showed that, despite a general macrosyntentic 1:1 pattern in *Acanthopterygii*<sup>38</sup> and specifically in *Pleuronectiformes*, a huge intrachromosomal reorganisation (mainly inversions) has taken place along flatfish evolution.<sup>9,34</sup> This observation was confirmed in our study, where the genomic comparison between turbot and tongue sole, species pertaining to the distant families *Scophthalmidae* and *Cynoglossidae*,<sup>71,72</sup> revealed great intrachromosomal reorganisations. Accordingly, polymorphic inversions could occur within flatfish species, and the comparison between individual maps could shed some light on this issue. Using nine families and a much lower amount of markers, Bouza et al.<sup>38</sup> suggested a general collinearity among individual maps, and only minor discrepancies were detected at 11 LGs involving markers at short distances (<3 cM). In our study, two medium-sized chromosome tracts (~7Mb) at LG06 and LG09 showed features that could be compatible with polymorphic inversions although further work will be necessary to confirm this observation. Combining fluorescence in situ hybridization with BAC probes on these chromosomes or checking linkage between markers within and around the potential inversions with a much higher offspring number could be possible strategies. If confirmed, this information should be considered not only for evolutionary studies but also for marker assisted selection in turbot breeding programs.

Flatfish represent one of the most extraordinary examples of anatomical specialisation to a particular lifestyle in vertebrates, and a bitter evolutionary controversy took place in the past about their origin and phylogeny.<sup>73,74</sup> The 1:1 macrosyntentic pattern observed between *Acanthopterygii*<sup>38</sup> was confirmed when comparing the available flatfish genomes: turbot, tongue sole and Japanese flounder.<sup>9,21,42</sup> The integration of our data with previous information, which includes the recently assembled genome of *P. olivaceus*,<sup>43</sup> enabled us to confirm and refine the suggested chromosome fusions along flatfish evolution.<sup>9</sup> Our results highlight that the fusion of ancestral chromosomes in the evolution of turbot genome resulted in the formation of new chromosomes, whose centromeres should correspond to the ancestral centromere of one of the fusing chromosomes, as reported in the origin of human chromosome 2.<sup>75</sup> The precise location of centromere-associated markers in the genetic and physical maps strongly supports two main centric fusions in the origin of turbot karyotype (LG02 and LG16) from the ancestral  $2n=48$  observed in Japanese flounder.<sup>57,76</sup> Flatfish show large variation in chromosome number even across closely related taxa and more than half of flatfish species present a chromosome number lower than the ancestral, suggestive of several fusions, with the minimum diploid number ( $2n=26$ ) reported in *Citharichthys spilopterus*.<sup>76</sup> It would be worth investigating in more detail if the high number of chromosome fusions occurring in the families with lower diploid numbers (e.g. *Soleidae*, *Bothidae* and *Achiridae* and some *Paralichthyidae*) show a preferential fusion pattern, similar to what recently found in notothenioids.<sup>77</sup> In addition to chromosome fusions, other minor reorganizations between non-orthologous chromosomes of turbot, Japanese flounder and half-smooth tongue sole were identified in our study. These chromosome rearrangements are

thought to be important in karyotypic evolution and species differentiation,<sup>77–79</sup> and thus, further comparative mapping studies are encouraged to understand the reasons behind these rearrangements in the evolution of flatfish.

The turbot HD consensus genetic map here constructed using a set of 18,124 SNPs was a suitable reference to anchor the turbot genome and to integrate all previous genomic information. The comparison of the male and female consensus maps enabled to identify sharp RF differences around centromeres that explained their length differences. The 62 individual genetic maps constructed allowed to detect suggestive polymorphic rearrangements in the species putatively related to coadapted gene blocks. Finally, inter- and intrachromosomal reorganisations in *Pleuronectiformes* were identified comparing the chromosome sequences available for three flatfish species taking the turbot map as reference. The consistent HD genetic map reported in turbot represents an invaluable genomic tool for further genome-wide association and evolutionary genomic studies.

## Acknowledgements

This study has been supported by the FISHBOOST project (ref. 613611) from the European Community's Seventh Framework Programme (FP7/2007-2013) and by Consellería de Cultura, Educación e Ordenación Universitaria, Xunta de Galicia local government (ref. GRC2014/010). Computational support for bioinformatic analysis were provided by Centro de Supercomputación de Galicia (CESGA).

## Accession numbers

CP026243-CP026264.

## Conflict of interest

None declared.

## Supplementary data

Supplementary data are available at *DNARES* online.

## References

- Mardis, E. R. 2008, The impact of next-generation sequencing technology on genetics, *Trends Genet.*, **24**, 133–41.
- Nie, H., Yan, X. and Huo, Z. 2017, Construction of a high-density genetic map and quantitative trait locus mapping in the Manila clam *Ruditapes philippinarum*, *Sci. Rep.*, **7**, 229.
- Nunes, J., de, R., Liu, S. and Pétille, F. 2017, Large-scale SNP discovery and construction of a high-density genetic map of *Colossoma macropomum* through genotyping-by-sequencing, *Sci. Rep.*, **7**, 46112.
- Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. and Mardis, E. R. 2013, The next-generation sequencing revolution and its impact on genomics, *Cell*, **155**, 27–38.
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. 2014, Ten years of next-generation sequencing technology, *Trends Genet.*, **30**, 418–26.
- Moll, K. M., Zhou, P., Ramaraj, T., et al. 2017, Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, *Medicago truncatula*, *BMC Genomics*, **18**, 578.
- Roesti, M., Moser, D. and Berner, D. 2013, Recombination in the threespine stickleback genome—patterns and consequences, *Mol. Ecol.*, **22**, 3014–27.

8. Fierst, J. L. 2015, Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools, *Front. Genet.*, **6**, 220.
9. Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P. and Houston, R. 2017, Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.*, doi: 10.1111/raq.12193.
10. Meyer, A. and Van de Peer, Y. 2005, From 2R to 3R: evidence for a fish-specific genome duplication (FSGD), *Bioessays*, **27**, 937–45.
11. Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A. and Johnson, E. A. 2007, Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers, *Genome Res.*, **17**, 240–8.
12. Baird, N. A., Etter, P. D., Atwood, T. S., et al. 2008, Rapid SNP discovery and genetic mapping using sequenced RAD markers, *PLoS One*, **3**, e3376.
13. Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M. and Blaxter, M. L. 2011, Genome-wide genetic marker discovery and genotyping using next-generation sequencing, *Nat. Rev. Genet.*, **12**, 499–510.
14. O'Quin, K. E., Yoshizawa, M., Doshi, P. and Jeffery, W. R. 2013, Quantitative genetic analysis of retinal degeneration in the blind cavefish *Astyanax mexicanus*, *PLoS One*, **8**, e57281.
15. Gonen, S., Lowe, N. R., Cezard, T., Gharbi, K., Bishop, S. C. and Houston, R. D. 2014, Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing, *BMC Genomics*, **15**, 166.
16. Carlson, B. M., Onusko, S. W. and Gross, J. B. 2015, A high-density linkage map for *Astyanax mexicanus* using genotyping-by-sequencing technology, *G3 Genes Genomes Genet.*, **5**, 241–51.
17. Fu, B., Liu, H., Yu, X. and Tong, J. 2016, A high-density genetic map and growth related QTL mapping in bighead carp (*Hypophthalmichthys nobilis*), *Sci. Rep.*, **6**, 28679.
18. Liu, P., Wang, L., Wong, S.-M. and Yue, G. H. 2016, Fine mapping QTL for resistance to VNN disease using a high-density linkage map in Asian seabass, *Sci. Rep.*, **6**, 32122.
19. Wang, L., Bai, B., Liu, P., et al. 2017, Construction of high-resolution recombination maps in Asian seabass, *BMC Genomics*, **18**, 63.
20. Song, W., Pang, R., Niu, Y., et al. 2012, Construction of high-density genetic linkage maps and mapping of growth-related quantitative trait loci in the Japanese flounder (*Paralichthys olivaceus*), *PLoS One*, **7**, e50404.
21. Wang, W., Hu, Y., Ma, Y., Xu, L., Guan, J. and Kong, J. 2015, High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex- and growth-related regions detection, *PLoS One*, **10**, e0120410.
22. Sun, C., Niu, Y., Ye, X., et al. 2017, Construction of a high-density linkage map and mapping of sex determination and growth-related loci in the mandarin fish (*Siniperca chuatsi*), *BMC Genomics*, **18**, 446.
23. Tsai, H. Y., Robledo, D., Lowe, N. R., et al. 2016, Construction and annotation of a high density SNP linkage map of the Atlantic salmon (*Salmo salar*) genome, *G3 Genes Genomes Genet.*, **6**, 2173–9.
24. Palaiokostas, C., Ferrarresso, S., Franch, R., Houston, R. D. and Bargelloni, L. 2016, Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing, *G3 Genes Genomes Genet.*, **6**, 3693–700.
25. van der Beek, S. and van Arendonk, J. A. 1993, Criteria to optimize designs for detection and estimation of linkage between marker loci from segregating populations containing several families, *Theoret. Appl. Genetics*, **86-86**, 269–80.
26. Li, Y., Liu, S., Wang, R. and Qin, Z. and Liu, Z. 2017, Bioinformatics considerations and approaches for high-density linkage mapping in aquaculture, In: Liu, Z., (ed.) *Bioinformatics in Aquaculture: Principles and Methods*, John Wiley & Sons, Ltd, Chichester, UK, pp. 356–79.
27. Sodeland, M., Jorde, P. E., Lien, S., et al. 2016, 'Islands of divergence' in the Atlantic Cod genome represent polymorphic chromosomal rearrangements, *Genome Biol. Evol.*, **8**, 1012–22.
28. Jones, F. C., Grabherr, M. G., Chan, Y. F., et al. 2012, The genomic basis of adaptive evolution in threespine sticklebacks, *Nature*, **484**, 55–61.
29. Bouza, C., Vandamme, S., Hermida, M., Cabaleiro, S., Volkaert, F., Martínez, P. 2014, Turbot. AquaTrace Species Leaflet. <https://aquatrace.eu/leaflets/turbot>.
30. Martínez, P., Bouza, C., Hermida, M., et al. 2009, Identification of the major sex-determining region of turbot (*Scophthalmus maximus*), *Genetics*, **183**, 1443–52.
31. Rodríguez-Ramilo, S. T., Toro, M. A., Wang, J. and Fernandez, J. 2014, Improving the inference of population genetic structure in the presence of related individuals, *Genet. Res.*, **96**, e003.
32. Taboada, X., Hermida, M., Pardo, B. G., et al. 2014, Fine mapping and evolution of the major sex determining region in turbot (*Scophthalmus maximus*), *G3 Genes Genomes Genet.*, **4**, 1871–80.
33. Robledo, D., Ribas, L., Cal, R., et al. 2015, Gene expression analysis at the onset of sex differentiation in turbot (*Scophthalmus maximus*), *BMC Genomics*, **16**, 973.
34. Figueras, A., Robledo, D., Corvelo, A., et al. 2016, Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life, *DNA Res.*, **23**, 181–92.
35. Robledo, D., Fernández, C., Hermida, M., et al. 2016, Integrative transcriptome, genome and quantitative trait loci resources identify single nucleotide polymorphisms in candidate genes for growth traits in turbot, *IJMS*, **17**, 243.
36. Bouza, C., Hermida, M., Pardo, B. G., et al. 2007, A microsatellite genetic map of the turbot (*Scophthalmus maximus*), *Genetics*, **177**, 2457–67.
37. Ruan, X., Wang, W., Kong, J., Yu, F. and Huang, X. 2010, Genetic linkage mapping of turbot (*Scophthalmus maximus* L.) using microsatellite markers and its application in QTL analysis, *Aquaculture*, **308**, 89–100.
38. Bouza, C., Hermida, M., Pardo, B. G., et al. 2012, An Expressed Sequence Tag (EST)-enriched genetic map of turbot (*Scophthalmus maximus*): a useful framework for comparative genomics across model and farmed teleosts, *BMC Genet.*, **13**, 54.
39. Hermida, M., Bouza, C., Fernández, C., et al. 2013, Compilation of mapping resources in turbot (*Scophthalmus maximus*): a new integrated consensus genetic map, *Aquaculture*, **414-415**, 19–25.
40. Martínez, P., Hermida, M., Pardo, B. G., et al. 2008, Centromere-linkage in the turbot (*Scophthalmus maximus*) through half-tetrad analysis in diploid meiotogonogenesis, *Aquaculture*, **280**, 81–8.
41. Saitoh, K., Hayashizaki, K., Yokoyama, Y., Asahida, T., Toyohara, H. and Yamashita, Y. 2000, Complete nucleotide sequence of Japanese flounder (*Paralichthys olivaceus*) mitochondrial genome: structural properties and cue for resolving teleostean relationship, *J. Hered.*, **91**, 271–8.
42. Chen, S., Zhang, G., Shao, C., et al. 2014, Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle, *Nat. Genet.*, **46**, 253–60.
43. Shao, C., Bao, B., Xie, Z., et al. 2017, The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry, *Nat. Genet.*, **49**, 119–24.
44. Cruz, V. P., Vera, M., Pardo, B. G., et al. 2017, Identification and validation of single nucleotide polymorphisms as tools to detect hybridization and population structure in freshwater stingrays, *Mol. Ecol. Resour.*, **17**, 550–6.
45. Wang, S., Meyer, E., McKay, J. K. and Matz, M. V. 2012, 2b-RAD: a simple and flexible method for genome-wide genotyping, *Nat. Methods*, **9**, 808–10.
46. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. and Postlethwait, J. H. 2011, Stacks: building and genotyping loci de novo from short-read sequences, *G3 (Bethesda)*, **1**, 171–82.
47. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. and Cresko, W. A. 2013, Stacks: an analysis tool set for population genomics, *Mol. Ecol.*, **22**, 3124–40.
48. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.*, **10**, R25.
49. Van Ooijen, J. W. 2006, JoinMap<sup>®</sup>4. *Software for the calculation of genetic linkage maps in experimental populations*, vol. 33. Kyazma BV: Wageningen.
50. Wu, Y., Close, T. J. and Lonardi, S. 2011, Accurate construction of consensus genetic maps via integer linear programming, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **8**, 381–94.

51. Voorrips, R. E. 2002, MapChart: software for the graphical presentation of linkage maps and QTLs, *J. Hered.*, **93**, 77–8.
52. Bouza, C., Sanchez, L. and Martínez, P. 1994, Karotypic characterization of turbot (*Scophthalmus maximus*) with conventional, fluorochrome and restriction endonuclease-banding techniques, *Mar. Biol.*, **120**, 609–13.
53. Team, R. C. 2013, R: a language and environment for statistical computing.
54. Harris, R. S. 2007, Improved pairwise alignment of genomic DNA. PhD Thesis, The Pennsylvania State University, Pennsylvania.
55. Vera, M., Álvarez-Dios, J. A., Millán, A., et al. 2011, Validation of single nucleotide polymorphism (SNP) markers from an immune Expressed Sequence Tag (EST) turbot, *Scophthalmus maximus*, *database, Aquaculture*, **313**, 31–41.
56. Vera, M., Alvarez-Dios, J. A., Fernandez, C., et al. 2013, Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping, *IJMS*, **14**, 5694–711.
57. Ohno, S., Wolf, U. and Atkin, N. B. 1968, Evolution from fish to mammals by gene duplication, *Hereditas*, **59**, 169–87.
58. Li, Y., Liu, S., Qin, Z., et al. 2015, Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish, *DNA Res.*, **22**, 39–52.
59. Palaiokostas, C., Bekaert, M., Taggart, J. B., et al. 2015, A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*), *Genet. Sel. Evol.*, **47**, 68.
60. Wang, J., Lydiate, D. J., Parkin, I. A., et al. 2011, Integration of linkage maps for the amphidiploid *Brassica napus* and comparative mapping with *Arabidopsis* and *Brassica rapa*, *BMC Genomics.*, **12**, 101.
61. Khan, M. A., Han, Y., Zhao, Y. F., Troggo, M. and Korban, S. S. 2012, A multi-population consensus genetic map reveals inconsistent marker order among maps likely attributed to structural variations in the apple genome, *PLoS One*, **7**, e47864.
62. Close, T. J., Bhat, P. R., Lonardi, S., et al. 2009, Development and implementation of high-throughput SNP genotyping in barley, *BMC Genomics*, **10**, 582.
63. Martínez, P., Robledo, D., Rodríguez-Ramilo, S. T., et al. 2016, Turbot (*Scophthalmus maximus*) genomic resources: application for boosting aquaculture production, In: MacKenzie, S., Jentoft, S. (eds) *Genomics in Aquaculture*, Elsevier, London, pp. 131–63.
64. Roberts, P. A. 1965, Difference in the behaviour of eu- and hetero-chromatin: crossing-over, *Nature*, **205**, 725–6.
65. Lou, Q., He, Y., Cheng, C., et al. 2013, Integration of high-resolution physical and genetic map reveals differential recombination frequency between chromosomes and the genome assembling quality in cucumber, *PLoS One*, **8**, e62676.
66. Singer, A., Perlman, H., Yan, Y., et al. 2002, Sex-specific recombination rates in zebrafish (*Danio rerio*), *Genetics*, **160**, 649–57.
67. Moen, T., Hayes, B., Baranski, M., et al. 2008, A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers, *BMC Genomics.*, **9**, 223.
68. Limborg, M. T., Helyar, S. J., de Bruyn, M., et al. 2012, Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*), *Mol. Ecol.*, **21**, 3686–703.
69. Lenormand, T. and Duthel, J. 2005, Recombination difference between sexes: a role for haploid selection, *PLoS Biol.*, **3**, e63.
70. Gruhn, J. R., Rubio, C., Broman, K. W., Hunt, P. A. and Hassold, T. 2013, Cytological studies of human meiosis: sex-specific differences in recombination originate at, or prior to, establishment of double-strand breaks, *PLoS One*, **8**, e85075.
71. Pardo, B. G., Machordom, A., Foresti, F., et al. 2005, Phylogenetic analysis of flatfish (Order Pleuronectiformes) based on mitochondrial 16S rDNA sequences, *Sci. Mar.*, **69**, 531–43.
72. Campbell, M. A., Chen, W.-J. and López, J. A. 2013, Are flatfishes (Pleuronectiformes) monophyletic?, *Mol. Phylogenet. Evol.*, **69**, 664–73.
73. Friedman, M. 2008, The evolutionary origin of flatfish asymmetry, *Nature*, **454**, 209–12.
74. Harrington, R. C., Faircloth, B. C., Eytan, R. I., et al. 2016, Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye, *BMC Evol. Biol.*, **16**, 224.
75. Ijdo, J. W., Baldini, A., Ward, D. C., Reeders, S. T. and Wells, R. A. 1991, Origin of human chromosome 2: an ancestral telomere-telomere fusion, *Proc. Natl. Acad. Sci. USA.*, **88**, 9051–5.
76. Azevedo, M. F. C., Oliveira, C., Pardo, B. G., Martínez, P. and Foresti, F. 2007, Cytogenetic characterization of six species of flatfishes with comments to karyotype differentiation patterns in Pleuronectiformes (Teleostei), *J. Fish Biology*, **70**, 1–15.
77. Amores, A., Wilson, C. A., Allard, C. A., Detrich, H. W. and Postlethwait, J. H. 2017, Cold fusion: massive karyotype evolution in the Antarctic Bullhead Notothen *Notothenia coriiceps*, *G3 Genes Genomes Genet.*, **7**, 2195–207.
78. Livingstone, K. and Rieseberg, L. 2004, Chromosomal evolution and speciation: a recombination-based approach, *New Phytol.*, **161**, 107–12.
79. Navarro, A. and Barton, N. H. 2003, Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes, *Science*, **300**, 321–4.