PLOS GENETICS

RESEARCH ARTICLE

Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation

Qiaowei Pan[1,13], Romain Feron[1,13], Ayaka Yano[1], René Guyomard[2], Elodie Jouanno[1], Estelle Vigouroux[1], Ming Wen[1], Jean-Mickaël Busnel[3], Julien Bobe[1], Jean-Paul Concordet[4], Hugues Parrinello[5], Laurent Journot[5], Christophe Klopp[6,7], Jérôme Lluch[8], Céline Roques[8], John Postlethwait[9], Manfred Schartl[10,11,12], Amaury Herpin[1], Yann Guiguen[1]*

1 INRA, UR1037 LPGP, Campus de Beaulieu, Rennes, France, **2** GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, **3** Fédération d'Ille-et-Vilaine pour la pêche et la protection du milieu aquatique (FDPPMA35), CS 26713, Rennes, France, **4** INSERM U1154, CNRS UMR7196, MNHN, Muséum National d'Histoire Naturelle, France, **5** Institut de Génomique Fonctionnelle, IGF, CNRS, INSERM, Université de Montpellier, Montpellier, France, **6** Plate-forme bio-informatique Genotoul, Mathématiques et Informatique Appliquées de Toulouse, INRA, Castanet Tolosan, France, **7** SIGENAE, GenPhySE, Université de Toulouse, INRA, ENVT, Castanet Tolosan, France, **8** INRA, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France, **9** Institute of Neuroscience, University of Oregon, Eugene, Oregon, United States of America, **10** University of Wuerzburg, Physiological Chemistry, Biocenter, Würzburg, Germany, **11** Comprehensive Cancer Center Mainfranken, University Hospital, Würzburg, Germany, **12** Hagler Institute for Advanced Study and Department of Biology, Texas A&M University, College Station, Texas, United States of America, **13** Department of Ecology and Evolution, University of Lausanne,1015, Lausanne, Switzerland

* yann.guiguen@inra.fr

Citation: Pan Q, Feron R, Yano A, Guyomard R, Jouanno E, Vigouroux E, et al. (2019) Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. PLoS Genet 15(8): e1008013. https://doi.org/10.1371/journal.pgen.1008013

Editor: Catherine L. Peichel, Universitat Bern, SWITZERLAND

Received: February 11, 2019

Accepted: July 26, 2019

Published: August 22, 2019

Data Availability Statement: Sequencing data and assembly for the Esox lucius genome can be found under NCBI Bioproject PRJNA514887. Sequencing data for identifying and charactering the sex locus, including RAD-seq and pool-seq reads, can be found under the NCBI Bioproject PRJNA514888.

Funding: This project was supported by funds from the "Agence Nationale de la Recherche" and the "Deutsche Forschungsgemeinschaft" (ANR/DFG, PhyloSex project, 2014-2016) to YG and MS.

Abstract

Teleost fishes, thanks to their rapid evolution of sex determination mechanisms, provide remarkable opportunities to study the formation of sex chromosomes and the mechanisms driving the birth of new master sex determining (MSD) genes. However, the evolutionary interplay between the sex chromosomes and the MSD genes they harbor is rather unexplored. We characterized a male-specific duplicate of the anti-Müllerian hormone (*amh)* as the MSD gene in Northern Pike (*Esox lucius*), using genomic and expression evidence as well as by loss-of-function and gain-of-function experiments. Using RAD-Sequencing from a family panel, we identified Linkage Group (LG) 24 as the sex chromosome and positioned the sex locus in its sub-telomeric region. Furthermore, we demonstrated that this MSD originated from an ancient duplication of the autosomal *amh* gene, which was subsequently translocated to LG24. Using sex-specific pooled genome sequencing and a new male genome sequence assembled using Nanopore long reads, we also characterized the differentiation of the X and Y chromosomes, revealing a small male-specific insertion containing the MSD gene and a limited region with reduced recombination. Our study reveals an unexpectedly low level of differentiation between a pair of sex chromosomes harboring an old MSD gene in a wild teleost fish population, and highlights both the pivotal role of genes from the *amh* pathway in sex determination, as well as the importance of gene duplication as a mechanism driving the turnover of sex chromosomes in this clade.

The GeT core facility (JL, CR), Toulouse, France and the MGX facility, Montpellier, France (LJ, HP) were supported by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

In stark contrast to mammals and birds, a high proportion of teleosts have homomorphic sex chromosomes and display a high diversity of sex determining genes. Yet, population level knowledge of both the sex chromosome and the master sex determining gene is only available for the Japanese medaka, a model species. Here we identified and provided functional proofs of an old duplicate of anti-Müllerian hormone (Amh), a member of the Tgf-β family, as the male master sex determining gene in the Northern pike, *Esox lucius*. We found that this duplicate, named *amhby* (Y-chromosome-specific anti-Müllerian hormone paralog b), was translocated to the sub-telomeric region of the new sex chromosome, and now *amhby* shows strong sequence divergence as well as substantial expression pattern differences from its autosomal paralog, *amha*. We assembled a male genome sequence using Nanopore long reads and identified a restricted region of differentiation within the sex chromosome pair in a wild population. Our results provide insight on the conserved players in sex determination pathways, the mechanisms of sex chromosome turnover, and the diversity of levels of differentiation between homomorphic sex chromosomes in teleosts.

Introduction

The evolution of sex determination (SD) systems and sex chromosomes has sparked the interest of evolutionary biologists for decades. While initial insights on sex chromosome evolution came from detailed studies in *Drosophila* and in mammals [1–5], recent research on other vertebrate groups, such as avian [6,7], non-avian reptiles [8,9], amphibians [10–13], and teleost fishes [14–17], has provided new information that helps us understand the evolution of SD systems and sex chromosomes.

Teleosts display the highest diversity of genetic sex determination systems in vertebrates, including several types of monofactorial and polygenic systems [14,18]. In addition, in some species, genetic factors can interact with environmental factors, most commonly temperature, *i.e.* in *Odontesthes bonariensis* [19], generating intricate sex determination mechanisms. Moreover, sex determination systems in fish can differ between very closely related species, as illustrated by the group of Asian ricefish (genus *Oryzias*) [20–25], and sometimes even among different populations of one species, as in the Southern platyfish, *Xiphophorus maculatus* [26]. Beside this remarkable dynamic of sex determination systems, the rapid turnover of sex chromosomes in teleosts provides many opportunities to examine sex chromosome pairs at different stages of differentiation. Finally, recent studies on fish sex determination have revealed a dozen new master sex determining (MSD) genes [14,16,27], providing additional insight to the forces driving the turnover of SD systems and the formation of sex chromosomes.

The birth of new MSD genes drives the formation of sex chromosomes and transitions of SD systems. The origin of new MSDs falls into two categories: either gene duplication followed by sub- or neo-functionalization, or allelic diversification [14]. To date, teleosts are the only group where examples of both gene duplication and allelic diversification mechanisms have been found [14–17]. Yet, among teleost species with a known MSD gene, sex chromosomes have only been sequenced and characterized in two species: the Japanese medaka (*Oryzias latipes*), whose MSD gene originated from gene duplication / insertion of the duplicated copy [28–31], and the Chinese tongue sole (*Cynoglossus semilaevis*), whose MSD gene originated from allelic diversification [32]. The duplication / insertion mechanism could suppress

recombination more readily than the allelic diversification mechanism, as the newly inserted genomic segment immediately lacks homologous regions for recombination. Furthermore, the duplicated MSD gene could potentially be inserted into different chromosomes and thus provide further flexibility in sex chromosome turnover. To understand how the mechanism of MSD origin could impact the evolution of sex chromosomes, additional empirical studies identifying the origin of MSDs and sex chromosomes are urgently needed. Such studies will form a rich knowledge base allowing advances of theories of sex chromosome evolution.

Among identified teleost MSD genes, the *salmonid* MSD gene, named *sdY*, is the most intriguing because it revealed a previously unexpected flexibility in SD pathways in teleosts. While all other currently identified MSDs belong to one of three protein families (SOX, DMRT and TGF-β and its signaling pathway) that were known to be implicated in the SD pathways, the *salmonid* MSD gene *sdY* arose from duplication of an immune-related gene [33]. Despite *sdY* being conserved in the majority of salmonid species [34], it was not found in *Esox lucius*, the most studied member of the salmonid's sister order the Esociformes [34]. The restriction of *sdY* to the salmonids raised the question of what was the ancestral MSD before the emergence of the "unusual" *sdY*. The first step to answer this question was to identify the genetic component responsible for sex determination in *E. lucius*.

E. lucius, commonly known as the Northern pike, is a large and long-lived keystone predatory teleost species found in freshwater and brackish coastal water systems in Europe, North America, and Asia [35]. It has emerged as an important model species for ecology and conservation because of its pivotal role as a top predator that shapes the structure of local fish communities, and also as a valuable food and sport fish [36]. Consequently, genomic resources have recently been generated for *E. lucius*, including a whole genome assembly anchored on chromosomes [37] and a tissue-specific transcriptome [38]. Yet, little is known about genetic sex determination in *E. lucius* beyond males being the heterogametic sex [39], and its sex locus and MSD gene remain elusive.

In this study, we identified a duplicate of anti-Müllerian hormone (*amh*) with a testis- specific expression pattern as a candidate male MSD gene for *E. lucius*. Using pooled sequencing (pool-seq) reads from a wild population and a new draft genome sequenced with Nanopore long reads, we found limited differentiation between the homomorphic sex chromosomes and that this male-specific duplicate of *amh*, which we call *amhby*, is located within the Y-specific sequence. Using RAD-sequencing of a family panel, we identified Linkage Group 24 as the sex chromosome and positioned the sex locus in its sub-telomeric region. In addition, we showed that *amhby* has an expression profile characteristic of a male MSD gene and is functionally both sufficient and necessary to trigger testis development, providing robust support for *amhby* as the MSD gene in this species. Finally, through phylogenetic and synteny analyses, we showed that this *amh* duplication occurred around 40 million years ago and that *amhby* was translocated after its duplication, which likely initiated the formation of the proto-Y chromosome.

Taking advantage of recent advances in functional genomics and sequencing technologies, our study combines the location and characterization of the sex locus and the identification of a master sex determining (MSD) gene with substantial functional validation in a non-model species. Our results expand the knowledge of sex determination genes and provide insight on the evolution of sex chromosomes in teleosts.

Results

Identification of a male-specific duplicate of *amh* with testis-specific expression in *E. lucius*

As many currently characterized MSD genes in teleosts belong to a few 'usual suspect' protein families, sex-specific allelic variants or sex-specific duplicate members of genes from these

'usual suspect' families are strong candidates for being potential MSD. By searching tissue-specific transcriptomes of *E. lucius* ([38], phylofish.sigenae.org), we identified duplication of such a 'usual suspect' gene; an anti-Mullerian hormone (*amh*) gene. The two *amh* transcripts share 78.9% nucleotide identity with one transcript being predominantly expressed in adult testis and expressed at a low level in adult ovary and adult muscle, and the other transcript being exclusively expressed in adult testis (**S1 Fig**). PCR amplification on genomic DNA from 221 wild-caught individuals, whose phenotypic sex was determined by gonadal inspection, showed that the genomic sequence of one *amh* copy was present in all phenotypic males and females, while the genomic sequence of the testis-specific *amh* was present in 98% of phenotypic males (157/161) and 0% of phenotypic females (0/60) (**Fig 1A**). The significant association between this testis-specific copy of *amh* and male phenotype (Chi-squared test, $p < 2.2e\text{-}16$) indicates that the genomic sequence of this testis-specific copy of *amh* transcript is Y chromosome specific. This male-specific *amh* was named *amhby* (Y-chromosome-specific anti-Müllerian hormone paralog b) and the autosomal gene was named *amha* (*amh* paralog a).

To compare the genomic regions containing *amha* and *amhby*, clones were isolated from a phenotypic male genomic fosmid library and sequenced. The *amha*-containing fosmid included the entire 5' intergenic region of *amha* up to the closest gene, *dot1l*, and the *amhby*-containing fosmid included a 22 kb region upstream of *amhby* which contained no coding sequences for other proteins (blastx search against Teleostei, taxid:32443). Nucleotide identity between *amha* and *amhby* exon sequences ranges from 74% to 95% (**Fig 1B and 1C**) and the only gross structural difference between the two genes is a 396 bp specific insertion of a repeated region in *amhby* intron 1. Little sequence similarity was found between the proximal sequences of the two genes (**Fig 1D**), except for a 1020 bp repetitive element (transposase with conserved domain HTH_Tnp_Tc3_2), which was present in many copies in the genome of *E. lucius* and was not specifically enriched around the genomic regions containing either *amh* or *amhby* (**S2 Fig**). Predicted proteins contain 580 amino-acids (AA) for *amha* and 560 AA for *amhby*, sharing 68.7% identity and 78.4% similarity. Both proteins have a complete 95 AA C-terminal TGF-β domain with seven canonical cysteines (**S3 Fig**, **S4 Fig**), sharing 62.5% identity and 74.0% similarity.

The sex locus of *E. lucius* is located in the sub-telomeric region of LG24

To identify the sex chromosome in the genome of *E. lucius*, we generated RAD-Seq data from a single full-sib family including 37 phenotypic male offspring, 41 phenotypic female offspring and the two parents. In total, 6,922 polymorphic markers were aligned to the 25 Northern pike linkage groups and 40 polymorphic markers were aligned to unplaced scaffolds (GenBank assembly accession: GCA_004634155.1). Genome-wide average F_{ST} between male and female offspring was 0.00085 and the linkage group with the highest average F_{ST} was LG24 with an F_{ST} of 0.052 while other chromosomes had an average F_{ST} of 0.0067. Furthermore, only markers mapped to LG24 showed genome-wide association with sex phenotype (**Fig 2A**), indicating that LG24 is the sex chromosome of *E. lucius*.

To locate the sex locus and to investigate the pattern of recombination between the X and Y chromosomes in the paternal genome, we compared LG24 in the male and female RAD-Seq genetic maps. Overall, the order of markers in both male and female genetic maps agreed with their order when aligned to the LG24 assembly ($R^2 = 0.94$ for the female map, and $R^2 = 0.89$ for the male map) with little difference between the male and female total map lengths (64.1 cM in female and 61.7 cM in male). In the female map, recombination was observed along the entire length of LG24. In the male map, while recombination with the sex locus (0 cM in the genetic map) was observed for the majority of the LG24 markers, 11 markers aligned to a small

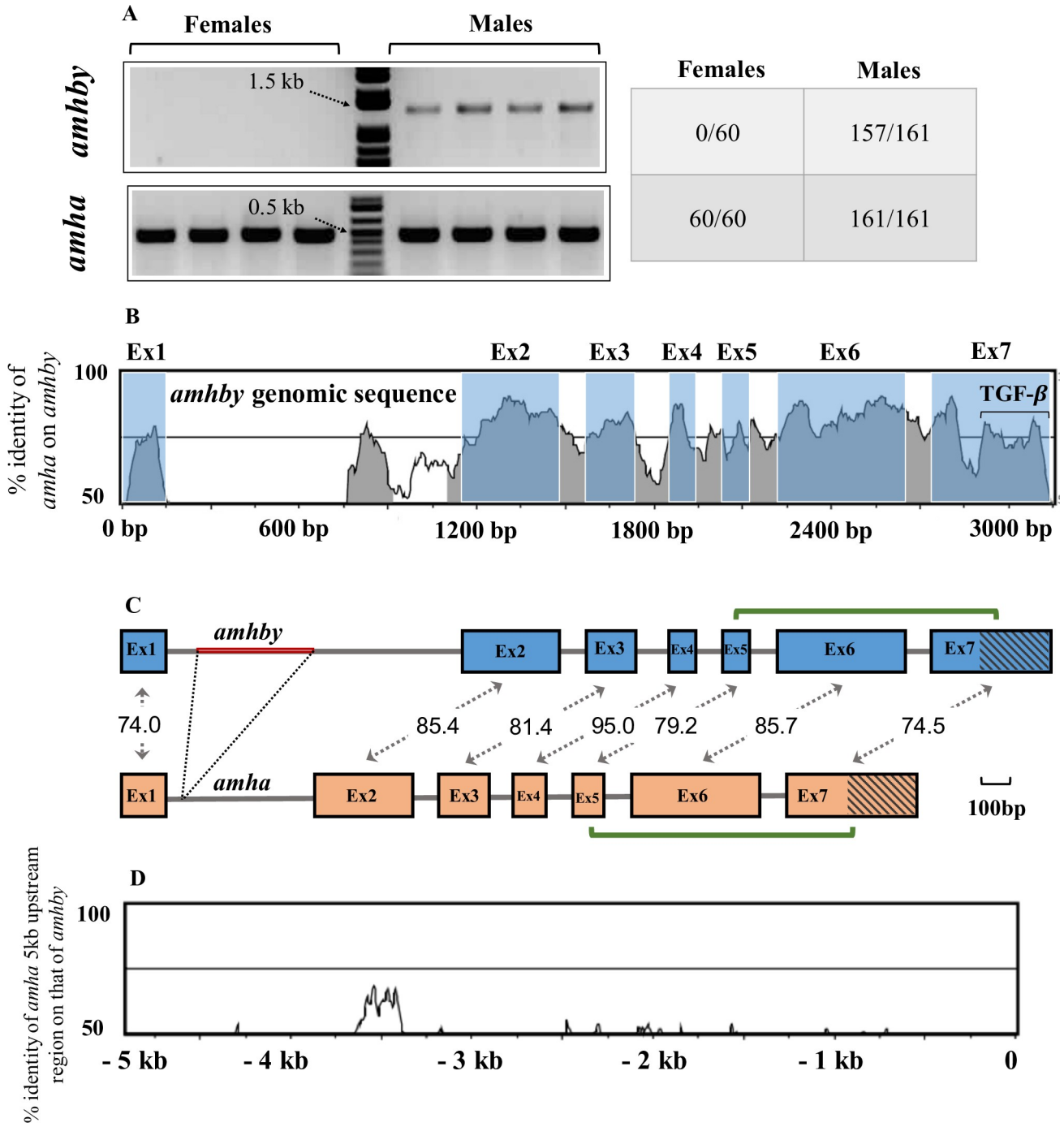

Fig 1. Sequence identity between *amha* and *amhby* and amplification in males and females. A) PCR amplification of *amha* (500 bp band) and *amhby* (1500 bp band) in male (n = 4) and female (n = 4) genomic DNA samples from *E. lucius*. Each lane corresponds to one individual. The number of tested animals of each sex positive for *amha* and *amhby* is indicated in the table on the right. B) Sequence identity of global pairwise alignment between *amha* and *amhby* genomic sequences with the *amhby* sequence as reference. C) Schematic representation of *amha* and *amhby* gene structure in *E. lucius* from the start to the stop codon. Exon1-Exon7 are represented by green boxes with shared percentage identity indicated and introns are represented by white segments. The red segment in intron 1 represents the *amhby* specific insertion. TGF-β domains are indicated with diagonal lines. D) Global pairwise alignment between 5 kb upstream region of *amha* and *amhby* with *amhby* as the reference. The start codon of each gene is positioned at 0 bp and the number on the y-axis indicates distance from the start codon upstream to the coding sequence.

https://doi.org/10.1371/journal.pgen.1008013.g001

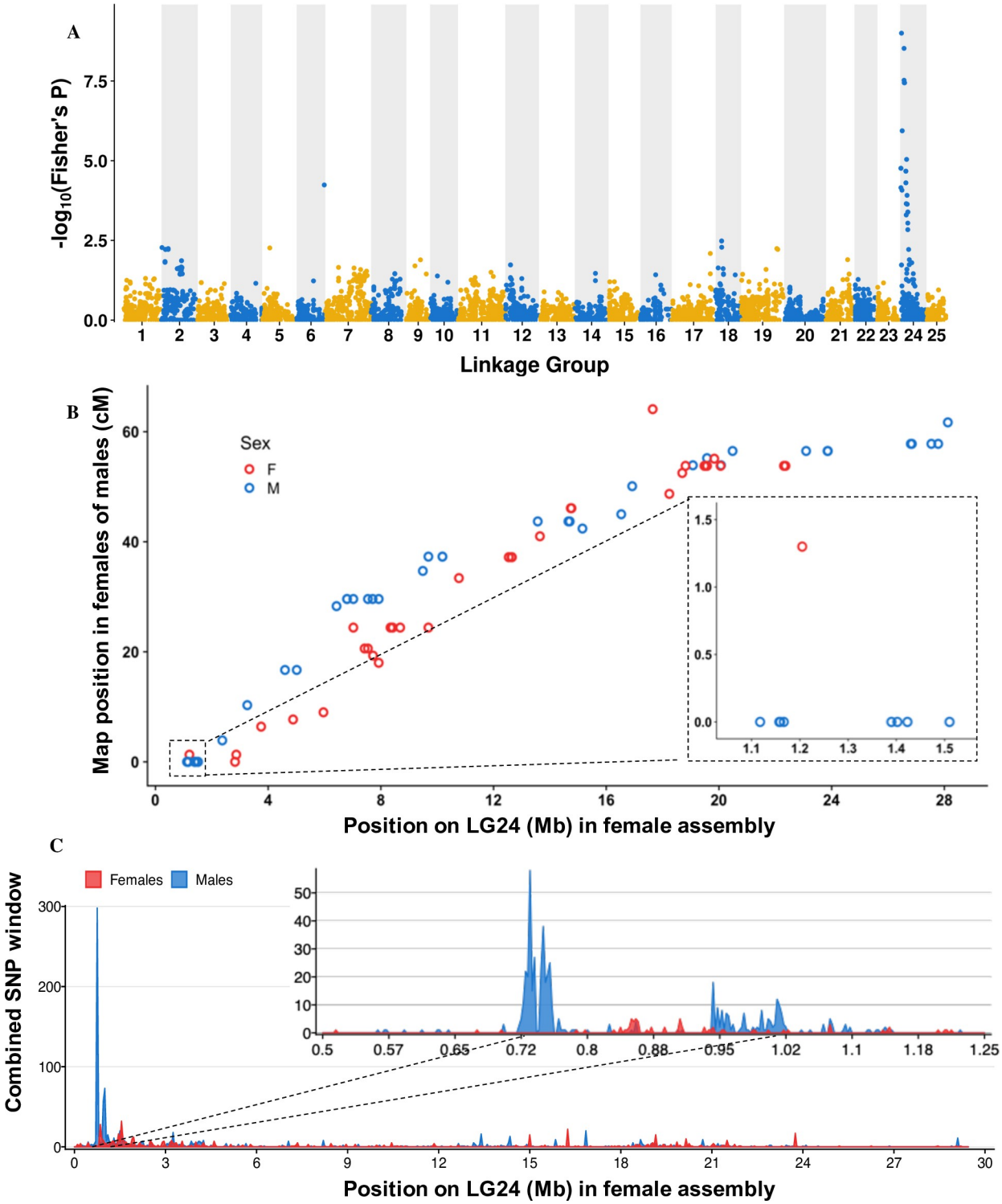

Fig 2. RAD-Seq and pool-seq analysis on a female reference genome of *E. lucius* (GenBank assembly accession: GCA_004634155.1) locates a small sex locus to the sub-telomeric region of Linkage group (LG)24. A: The $\log_{10}$ of p-values from the Fisher's exact test between each marker and sex phenotype is mapped against the linkage group (LG) to which they belong. Linkage group 24 showed a concentration of markers significantly associated with sex phenotype. **B:** Genetic map distance of RAD-Seq markers on LG 24 plotted against their physical position in LG24 for males and females. Recombination is observed along the entire length of LG24 in females and for most male markers except a group aligned to a region between 1.1 mb to 1.5 Mb on the physical map that shows no recombination with the sex locus mapped to 0 cM on the genetic map. Black dotted-line box highlights the group of male markers that shows no recombination with the sex locus with a zoomed-in view on the side. **C:**

Number of male and female-specific SNPs from pool-seq in 50 kb non-overlapping windows is plotted along LG 24 with zoomed-in view on the first 1Mb. The male data are represented in blue and female in red. The ~ 300 kb region between ~ 720 kb and ~ 1.02 Mb containing 442 MSS, showed the strongest differentiation between males and females.

https://doi.org/10.1371/journal.pgen.1008013.g002

400 kb region between 1.1 and 1.5 Mb showed no recombination with the sex locus, suggesting that the region with restricted recombination is small and located only in the sub-telomeric region of the LG24 sex chromosome. Additionally, 20 markers showed no recombination with the sex locus and could not be aligned to the female reference genome, suggesting that they are Y-specific sequences. Two of these 20 markers aligned to the *amhby* sequence, indicating that *amhby* is passed down strictly from father to sons (S3 Table). Furthermore, in a separate analysis, we identified the parental origin of all markers and plotted the distribution of these markers between male and female offspring along LG24 (**S5 Fig**)**.** While maternal markers showed no sex bias in their distribution among offspring, paternal markers displayed an important sex-bias at the proximal end of LG24, again locating the sex locus to this region.

This recombination pattern of RAD-Seq markers aligned to LG24 inferred from a single family panel represents the X and Y recombination pattern of the sire. In order to characterize the differentiation between the X and Y chromosomes at the population level, we sequenced a pool of 30 phenotypic males and a pool of 30 phenotypic females. Reads were aligned to the female Northern pike reference assembly (GCA_004634155.1), and we computed the number of male-specific SNPs (MSS) in a 50 kb non-overlapping window across the whole genome. The genome average was 0.55 MSS per 50 kb window (2.6 MSS per window when excluding windows without MSS). A single window located around 750,000 bp on LG24 contained 298 MSS (**Fig 2C**), close to three times the number of MSS in the next highest window (111 MSS on LG07), indicating that the sex locus is located near the proximal end of LG24. A further analysis with a higher 2.5 kb window resolution revealed that this enrichment of MSS on LG24 is restricted to a 300 kb region located between ~0.72 Mb and ~1.02 Mb containing 442 MSS (**Fig 2C**). However, as no marker from the family panel aligned to the region between 0 Mb to 1 Mb on LG24, we do not have information regarding the suppression of X/Y recombination in the sire of the family panel for this 300 kb. These results confirm that the sex locus is located in the proximal of LG24, and indicate that at the population level, there is strong differentiation between the X and the Y sequences restricted to a small 300 kb sub-telomeric region of the LG24 sex chromosome.

The Y chromosome harbors a small sex locus containing *amhby*

The fosmid clone sequence containing *amhby* was not found in the female reference assembly (GCA_004634155.1), suggesting that the Northern pike Y chromosome contains unique sequences that lack homology to the X chromosome. Therefore, to better characterize the Northern pike sex locus, we sequenced and assembled the genome of a genetic male with Nanopore long reads. Results from BUSCO show that this new assembly has comparable completeness to that of the female reference assembly (**S1 Table**). In this Nanopore assembly, the entire sequence of the *amhby*-containing fosmid was included in a 99 kb scaffold (tig00003316), from 24,050 bp to 60,989 bp, with *amhby* located from 27,062 bp to 30,224 bp.

To identify these potential Y-specific sequences absent in XX females, we aligned the pool-seq reads to the XY Nanopore assembly and searched for non-overlapping 1 kb regions covered only by male reads (MR1k). In total, we found 94 MR1k-containing regions located on only three Nanopore contigs (tig00003316 = 53 MR1k, tig00003988 = 22 MR1k, tig00009868 = 19 MR1k). In contrast, only four non-overlapping 1 kb regions were covered only by female reads, each on different contigs, and we found no MR1k from the same analysis on the female reference assembly (GCA_004634155.1), indicating a low false discovery rate for

sex-specific regions with our method. Blasting results on the three MR1k-containing Nanopore contigs showed that *amhby* is the only protein coding gene apart from transposable elements (TEs) associated proteins in the *E. lucius* sex locus (S2 Table).

To quantify the length of Y-specific sequences, we then looked for regions with few or no female reads mapped and male reads mapped at a depth close to half of the genome average on the three MR1k-containing contigs from the Nanopore assembly (S6 Fig). In total, we identified ~ 180 kb of Y-specific sequences in these three contigs. One of these contigs showed strong homology to a region on LG24 spanning from ~0.72 Mb to ~0.80 Mb (megablast, e-value = 0, identity = 95%), which indicates that the male specific region is located adjacent to the ~300 kb region enriched with male specific SNPs from 0.72 Mb to 1.02 Mb on LG24. Together, these results indicate that the size of the sex locus of *E. lucius* is ~480 kb and that *amhby* is the only non-TE, protein-coding gene in this locus.

amhby is expressed prior to molecular gonadal differentiation in male *E. lucius*

To characterize the temporal and spatial expression of *amhby* in relation to the molecular and morphological differentiation between male and female gonads, both quantitative PCR (qPCR) and *in-situ* hybridization (ISH) were performed.

Expression of *amhby*, *amha*, three other genes (*drmt1*, *cyp19a1a*, and *gsdf*) known for their role in gonadal sex differentiation, and *amhrII*, the putative receptor for the canonical *amh*, was measured by qPCR at four time points from 54 days post-fertilization (dpf) to 125 dpf, prior to the onset of gametogenesis. The entire trunks were used for the first three time points when the gonads were too small to be isolated, and only gonads were used at 125 dpf for both males and females. Expression of *amha* was detected in both males and females starting from 75 days post-fertilization (dpf), with a significantly higher expression in males than in females at 100 dpf (Wilcoxon signed-rank test, $p = 0.043$) (Fig 3A). In contrast, expression of *amhby* was detected only in males starting from 54 dpf and increasing exponentially thereafter till 125 dpf ($R^2 = 0.79$) (Fig 3B). Expression of *drmt1*, *cyp19a1a*, and *gsdf* was only detected from 100 dpf onwards, *i.e.* much later than the first detected expression of *amhby* (S7 Fig). Moreover, among these three genes, only *cyp19a1a* showed significantly different expression between sexes with a higher expression in females at 100 dpf (Wilcoxon signed-rank test, $p = 0.014$). Expression of *amhrII* was not detected until 100 dpf and did not differ significantly between sexes at any stage (S7 Fig).

Expression of *amha* and *amhby* was also characterized by *in-situ* hybridization (ISH) performed on histological sections of the entire trunk of male and female *E. lucius* sampled at 80 dpf. Expression of *amha* was detected in the gonads of both female (Fig 3C) and male (Fig 3D) samples, but the signal was much stronger in male gonads. In contrast, expression of *amhby* was strong in male gonads (Fig 3F) but not detected in female gonads (Fig 3E), confirming the specificity of the probe for *amhby*. In addition, no morphological differences were observed between male and female gonads at 80 dpf, even though expression of *amhby* was already detected by qPCR and by ISH at this stage. Our ISH results show that the expression of both *amha* and *amhby* is high in male gonads before the first signs of histological differentiation between male and female gonads.

Collectively, these results show that *amhby* is expressed in the male gonads prior to both molecular and morphological sexual-dimorphic differentiation of gonads in *E. lucius*.

The *amhby* gene is both necessary and sufficient to trigger testicular differentiation

To further investigate the functional role of *amhby* in initiating testicular development, we performed both loss-of-function and gain-of-function experiments.

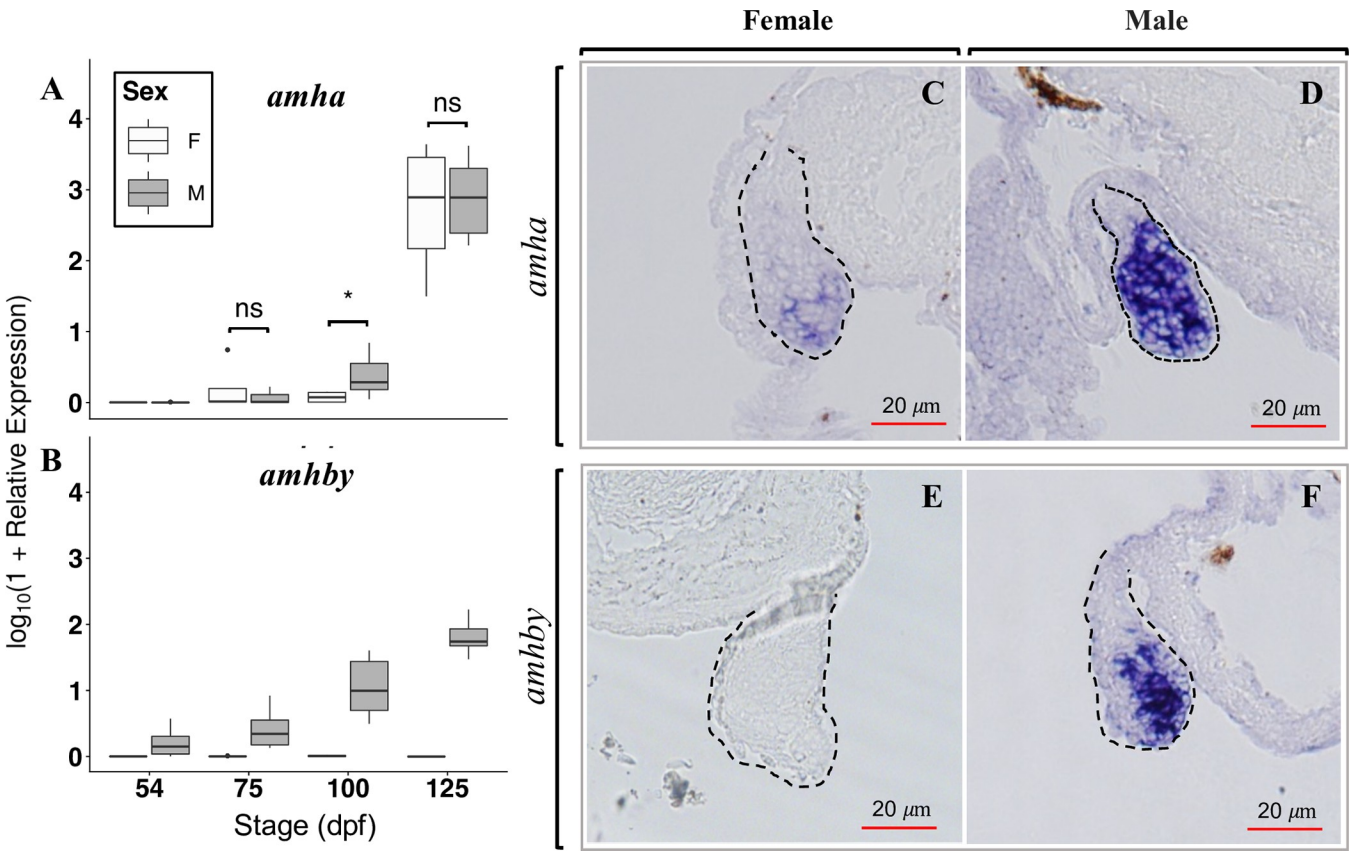

Fig 3. Temporal and spatial expression of *amha* and *amhby* mRNA in male and female developing gonads. A-B: Boxplots showing the first quantile, median, and the third quantile of the temporal expression of (A) *amha* and (B) *amhby* during early development of *E. lucius* measured by qPCR. Outliers are displayed as dots. The mRNA expression of *amha* and *amhby* was measured at 54, 75, 100, and 125 days post fertilization (dpf) in male and female samples of *E. lucius* and the $\log_{10}$ of their relative expression is presented on the graph. Significant P-values (<0.05) for Wilcoxon signed rank test between male and female expression at each time point are indicated by * and 'ns' indicates non-significant P-values. Statistical tests were not performed on *amhby* expression between sexes because of the complete absence of *amhby* expression in females. **C-F**: *In situ* hybridization on histological sections revealed the localization of *amha* in both 80 dpf (C) female and (D) male gonads, with a stronger expression in male gonads. A high *amhby* mRNA expression is detected in the male gonad (F), with no signal detected in female gonad (E). The red scale bars denote 20 μm and the dashed lines outlines the gonadal sections.

https://doi.org/10.1371/journal.pgen.1008013.g003

We knocked out *amhby* using three pairs of TALENs targeting exon 1 and exon 2 of *amhby* (S8A Fig). Only the T2 TALEN pair targeting exon 1 was effective in inducing deletions in the *amhby* sequence. Overall, 12 of 36 (33.3%) surviving G0 males possessed a disrupted *amhby* that resulted in truncated proteins (S8B Fig). G1 XY offspring obtained from three *amhby* mosaic G0 males crossed with wild-type females were maintained until the beginning of testicular gametogenesis at 153 dpf and then processed for histology. Gonads from 23 G1 *amhby* positive XY mutants were compared with those of wild-type control XY males (N = 4) and control XX females (N = 4) of the same age. Control animals developed normal ovaries and testes (Fig 4A and 4B), but all 23 XY F1 *amhby* mutants failed to develop a normal testes. Among these 23 G1 XY mutants, 20 (87%) showed complete gonadal sex reversal, characterized by the formation of an ovarian cavity and the appearance of previtellogenetic oocytes (Fig 4C); the three (13%) remaining mutants developed potentially sterile gonads with no clear ovarian nor testicular structure. None of these 23 G1 XY mutants display a potential off-target mutation in the corresponding region targeted by the T2 TALEN pair in the *amha* gene (S9 Fig).

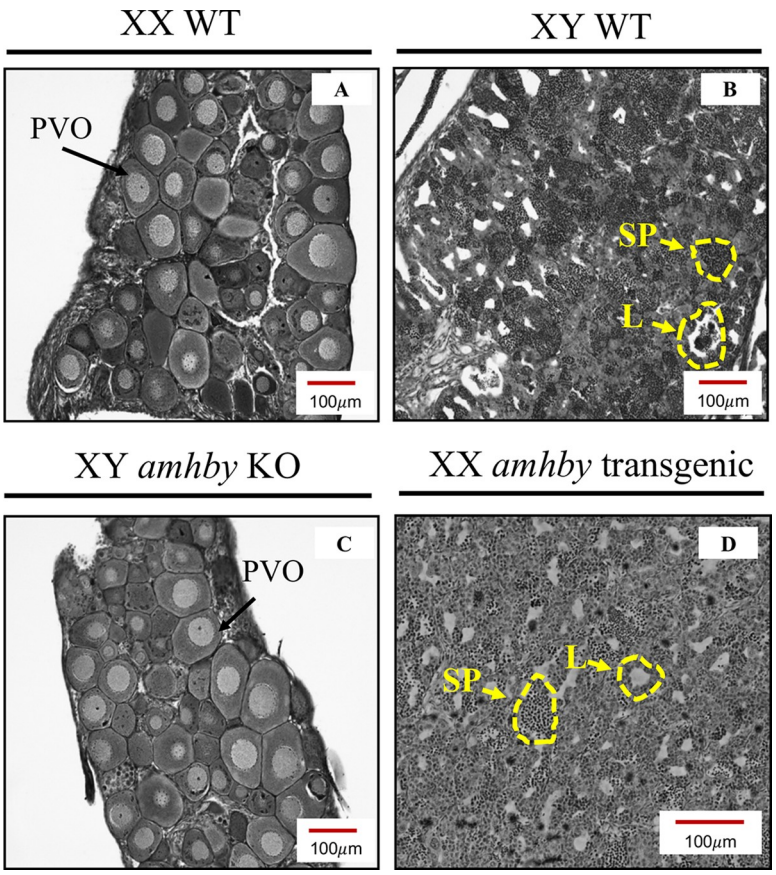

Fig 4. Gonadal phenotypes of *E. lucius* in *amhby* knockout (KO) and additive transgenesis experiments. Gonadal histology of a representative control XX female (A), control XY male (B), an *amhby* KO XY individual (C), and an *amhby* transgenic XX mutant (D). The *amhby* KO XY individual (C) developed ovaries with oocytes and an ovary cavity, indistinguishable from the ovary of the control females (A). The *amhby* transgenic XX mutant individual (D) developed testis with clusters of spermatozoids and testicular lobules identical to that of the control males (B). **PVO**: previtellogenic oocytes; **SP**: spermatozoids; **L**: testicular lobules; **T**: testis; and **O**: ovary.

https://doi.org/10.1371/journal.pgen.1008013.g004

To investigate whether *amhby* alone is sufficient to trigger testicular development, we overexpressed *amhby* in XX genetic females. Two G0 XY mosaic transgenic males possessing the *amhby* fosmid were crossed with wild-type females, and 10 G1 XX offspring carrying the *amhby* fosmid were maintained along with control wild-type siblings until the beginning of testicular gametogenesis at 155 dpf. Upon histological analysis of the gonads, all ten (100%) XX transgenics carrying *amhby* fosmid developed testis with testicular lumen and clusters of spermatozoids (**Fig 4D**), while all 12 control genetic males developed testis and 19 of 24 (79%) control genetic females developed ovaries. The other five control genetic females (21%) developed testes. Such sex reversal was also observed in the natural population at a rate of 2%, and this might have been exacerbated by culture conditions, a phenomenon previously documented in other teleosts [40]. Despite this effect, the XX transgenics with *amhby* fosmid had a significantly higher rate of sex reversal than their control female siblings raised in identical conditions (Chi-squared test, $p = 0.0001148$).

Taken together, these results show that *amhby* is both necessary and sufficient to trigger testicular development in *E. lucius*, and further support the functional role of *amhby* as the MSD gene in this species.

A chromosomal translocation involved *amhby* after a lineage-specific duplication of *amh*

To determine the origin of the two *E. lucius amh* paralogs, we generated a map of conserved syntenies for *amh* in several teleost species, including the spotted gar (*Lepisosteus oculatus*) as outgroup (Fig 5A). Genes located upstream (*i.e. dot1l*, *ell*, and *fkbp8*) and downstream (*i.e. oazla*) of *amha* on *E. lucius* LG08 showed conserved synteny in all teleost species included in the analysis, indicating that LG08 is the conserved location of the *E. lucius* ancestral *amh*, now called *amha*, and that *amhby* evolved from a duplication of *amha* that was later translocated to the sub-telomeric region of the future sex chromosome, LG24. We estimated the duplication event occurred ~ 38 and ~50 million years ago (S1 File, S7 Table), and found no homology between the ~ 180 Kb Y-specific sequence identified in the Nanopore assembly and the sequence of LG08 from the reference assembly, besides the two *amh* genes, suggesting that the translocation is also likely to be ancient.

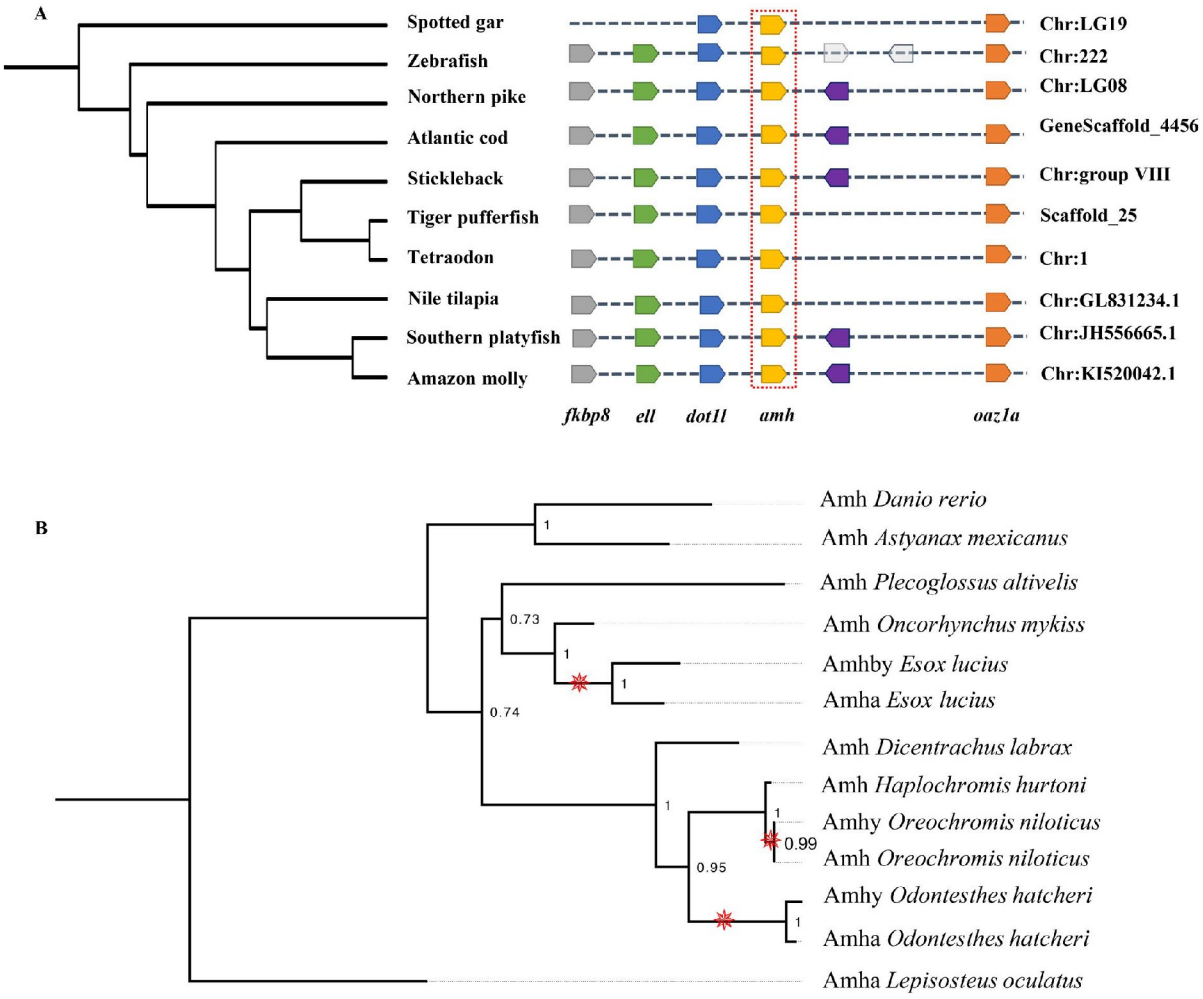

Fig 5. Evolution of Amh in teleosts. A: Synteny map of genomic regions around *amh* genes (highlighted by the red box) in teleosts. Orthologs of each gene are shown in the same color and the direction of the arrow indicates the gene orientation. Ortholog names are listed below and the genomic location of the orthologs are listed on the right side. For Northern pike (*Esox lucius*), *amha*, which is located on LG08, is used in this analysis. **B**: Phylogenetic reconstruction of teleost Amh protein orthologs. The phylogenetic tree was constructed with the maximum likelihood method (bootstrap = 1000). Numbers at tree nodes are bootstrap values. Spotted gar (*Lepisosteus oculatus*) Amh was used as an outgroup. Branches with Amh duplication are indicated by the red pinwheels.

https://doi.org/10.1371/journal.pgen.1008013.g005

Prior to the discovery of *amhby* in *E. lucius*, male-specific duplications of *amh* were identified in Patagonian pejerrey (*Odontesthes hatcheri*) [41] and Nile tilapia (*Oreochromis niloticus*) [42]. To test whether these duplications have a shared origin, we constructed a phylogeny of Amh from nine teleost species, including these three species with male-specific *amh* duplications, and spotted gar Amh as outgroup (**Fig 5B**). In this protein phylogeny, each sex-specific Amh paralog clusters as a sister clade to its own species' 'canonical' Amh with significant bootstrap values, indicating that these three pairs of *amh* paralogs were derived from three independent and lineage-specific duplication events.

Discussion

amhby, the male-specific duplicate of *amh*, is the master sex determination gene in *E. lucius*

Since the discovery of *dmrt1bY* in the Japanese rice fish [28,29], the first identified teleost master sex determination gene, studies in teleosts have unveiled a dozen novel genes as master regulators for sex determination [14–16,27]. Interestingly, many of these master sex determining genes belong to the TGF-β superfamily. To date, this finding has been mostly restricted to teleosts, highlighting the crucial role of TGF-β signaling in the sex determination pathway in this vertebrate group. In the present study, we identified an old duplicate of *amh*, a member of the TGF-β superfamily, as the male MSD gene in *E. lucius*. Results from genotyping demonstrate a strong and significant association of *amhby* with male phenotype. RNA-seq, qPCR and ISH showed that *amhby* is expressed in the male gonadal primordium before histological testis differentiation, thus fulfilling another criterion for being an MSD. Furthermore, knockout of *amhby* leads to complete gonadal sex reversal of XY mutants, while overexpression of *amhby* in XX animals leads to the development of testis, demonstrating that *amhby* is both necessary and sufficient for testicular differentiation. Together, these independent lines of evidence provide strong support that *amhby* is the MSD in *E. lucius*.

This work provides a third functionally validated case of an *amh* duplicate evolving into the MSD gene in a teleost species, along with the Patagonian pejerrey, *Odontesthes hatcheri*, [41], and the Nile tilapia, *Oreochromis niloticus*, [42]. Besides these three examples, association of *amh* duplicates with phenotypic sex was also found in other teleosts: in *O. bonariensis*, the sister species of the Patagonian pejerrey, a male-specific *amhy* was found to interact with temperature in determining sex [19], and in the ling cod, *Ophiodon elongatus*, a male-specific duplicate of *amh* was also identified using molecular marker sequences [43]. More recently, a duplicated copy of *amh* was found in an *Atheriniformes* species, *Hypoatherina tsurugae*, and is suspected to be involved in male sex determination [44]. Besides *amh*, its canonical receptor, *amhrII*, has also been shown to play a pivotal role as a MSD gene in the Tiger Pufferfish, *Takifugu rubripes* [45]. Our phylogenetic analysis on Amh sequences from several teleost species revealed that the three confirmed male-specific Amh duplications are independent, lineage-specific events rather than the product of shared ancestry. This finding supports the "limited option" hypothesis for master sex determining genes [46] and makes Amh pathway members the most frequently and independently recruited master sex determining genes identified in any animal group so far.

Among teleost species with *amh* duplicated MSD genes, *E. lucius* displays the highest degree of sequence divergence between paralogs, with an average of ~ 79.6% genomic sequence identity. In the Nile tilapia, *amhy* is almost identical to the autosomal *amh*, differing by only one SNP [42]. In the Patagonian pejerrey, the shared identity between the two paralogs ranges from 89.1% to 100% depending on the exon [41]. Because of the low divergence between the two paralogous sequences in the Patagonian pejerrey and the Nile tilapia, the new MSD

function of *amhy* was attributed to their novel expression patterns [45]. Yet, low sequence divergence, as little as one amino-acid changing SNP, was shown to be sufficient to impact the signal transduction function of the human AMH protein, leading to persistent Müllerian duct syndrome [47,48]. In *E. lucius*, *amhby* also has an expression profile different from *amha*, likely due to a completely different promoter region. However, because of the relatively high level of divergence between *amha* and *amhby* sequences in *E. lucius*, especially in the C-terminal bioactive domain of the proteins [47], it is tempting to hypothesize that the two proteins could have also diverged in their function. For instance, they may have a different affinity for their canonical AmhrII receptor or even the ability to bind to different receptors, leading to divergence in their downstream signaling pathways. Because we estimated the duplication of *amh* in *E. lucius* to be between ~ 38 and ~ 50 million years old, the long divergence time between the two paralogous *amh* genes potentially provided opportunities for the accumulation of these sequence differences. Further functional studies would be required to unravel the downstream signaling pathways of Amha and Amhby in *E. lucius* to better understand the mechanisms leading to the novel function of *amhby* as the MSD gene in this species [37].

The birth of the MSD and sex chromosomes in *E. lucius*

The analysis of the genomic neighborhood of both *amh* duplicates showed that *amha* is located on LG08 in a cluster of genes regulating sexual development and cell cycling with conserved synteny in teleosts [49,50], while *amhby* is located near the telomeric region of LG24 with no other identified gene besides transposable elements within at least 99 kb in its close vicinity. These results indicate that *amha* is likely to be the ancestral *amh* copy in *E. lucius*, and that *amhby* was translocated near the telomere of the ancestor of LG24 after its duplication. This scenario fits the description of a proposed mechanism of sex chromosome emergence and turnover through gene duplication, translocation and neofunctionalization [31,51]. Following this model, the translocation of a single copy of *amh* into another autosome triggered the formation of proto sex chromosomes, possibly because the newly translocated genomic segment containing the *amh* copy halted recombination with the X chromosome *ab initio* due to a complete lack of homology. This mechanism came to light after the discovery of *dmrt1bY* in the Japanese medaka, which acquired a pre-existing *cis*-regulatory element in its promoter through a transposable element [52]. Besides *dmrt1bY*, the only other well-described case of sex chromosome turnover via gene duplication, translocation and neofunctionalization is the salmonid *sdY* gene, which maps to different linkage groups in different salmonid species. Our study provides a third empirical example of gene duplication and translocation giving rise to new MSD genes and bolsters the importance of this mechanism in the birth and turnover of sex chromosomes.

Theories of sex chromosome evolution predict suppression of recombination around the sex determining locus, eventually leading to its degeneration because of its lack of ability to effectively purge deleterious mutations and repeated elements [53,54]. Here, we found that only a small part (around 480 kb) of the sex chromosome shows differentiation between X and Y chromosome, suggesting that this sex locus encompasses less than 1% of LG24 in *E. lucius*. Sex chromosome differentiation has been characterized in a few other teleost species, including Chinese tongue sole [32], Japanese medaka [30,55,56], stickleback [57], Trinidad guppy [58] and a few cichlid species [59,60]. Compared to these examples, we found that the Northern pike displays a very limited region of suppressed recombination between the sex chromosomes. A usual explanation for a small sex locus is that the rapid transition of sex chromosomes frequently observed in teleosts, facilitated by duplication and translocation, can readily produce neo-sex chromosomes showing little differentiation. This scenario was

demonstrated in the salmonids with the vagabond MSD gene *Sdy* [34,61,62]. However, the ~ 40 million years of divergence time between *amhby* and *amha*, and the lack of homology between the sequences suggest that the sex locus of *E. lucius* is not nascent. Further comparative studies that include sister species in the same clade will be needed to better estimate the age of this MSD gene and the sex chromosome, but a nascent sex locus is likely not the explanation for such a restricted region suppression of recombination between the X and Y chromosome of *E. lucius*. On the other hand, old yet homomorphic sex chromosomes have been observed, for instance in ratite birds [63]. Furthermore, in *Takifugu*, a single SNP conserved for 30 million years determines sex and the rest of the sex chromosomes do not show evidence of suppressed recombination, raising the possibility that decay is not the only possible fate for sex chromosomes [45]. One mechanism for the maintenance of a small sex locus has been proposed in the Japanese rice fish, where long repeats flanking the sex locus on the Y chromosome may recombine with the same repeats on the X chromosomes, thus hindering the spread of suppression of recombination around the MSD gene [51]. We found a slight enrichment of repeated elements on the sequences from the sex locus of *E. lucius*, however a better assembly of the sex locus would be needed to investigate whether a similar mechanism could have contributed to the very limited differentiation between X and Y chromosomes in *E. lucius*.

A widely accepted model of sex chromosome evolution postulates that the presence of sexually antagonistic alleles near the SD locus would favor the repression of recombination along the sex chromosome [64,65]. This process could eventually create 'evolutionary strata' of different ages and different levels of differentiation between the sex chromosome pair [66,67]. While evolutionary strata have been detected in various species [1,68–74], suppression of recombination on sex chromosome can happen at a very different tempo and the intensity of sexual selection has been suggested as a potential reason for this difference observed on the sex chromosomes across avian lineages [74]. As the sex locus in *E. lucius* is old but limited in size without indication of the formation of evolutionary strata, it is unlikely that sexual conflict and sexually antagonistic alleles have impacted sex chromosome evolution in this species.

To date, empirical support for non-decaying sex chromosomes is still rare, possibly due to the difficulties in identifying small differences between sex chromosomes, and because environmental factors affecting the sex determination pathways could weaken the association between genotype and phenotype, particularly in non-endothermic vertebrates. However, our study, as well as other recent efforts in characterizing sex loci in non-model species [75], show that current population genomic approaches can now relatively easily identify and characterize small sex loci. Future studies surveying sex chromosomes at various stages of differentiation and understanding the factors influencing their level of differentiation will form a new empirical basis to update the current models of sex chromosome evolution.

In conclusion, our study identified an old duplication of *amh* in *E. lucius* which generated a Y-chromosome-specific copy that we named *amhby*. We showed that *amhby* is functionally necessary and sufficient to trigger testicular development, and is expressed in the male gonadal primordium, fulfilling key requirements for a classic MSD gene. Furthermore, we located *amhby* on the sub-telomeric region on LG24, a region showing very limited differentiation between the X and Y chromosome. The recurrent identification of *amh* duplicates as MSD genes in teleosts highlights the pivotal role of Amh signaling pathway in teleost sex determination and encourages further analysis on how *amh* MSDs genes initiate testicular differentiation. Moreover, our results provide an intriguing empirical example of an unexpectedly small sex locus with an old MSD gene and highlight the power of exciting new sequencing technologies and population genomics approaches to identify and characterize sex loci in non-model species.

Material and methods

Fish rearing conditions

Research involving animal experimentation conformed to the principles for the use and care of laboratory animals, in compliance with French ("National Council for Animal Experimentation" of the French Ministry of Higher Education and Research and the Ministry of Food, Agriculture, and Forest) and European (European Communities Council Directive 2010/63/UE) guidelines on animal welfare. In agreement with the French legislation the present project was approved by the local "Ethic, Animal Care and Use Committee" [committee N°7 named « Comité Rennais d'Ethique en matière d'Expérimentation Animale (CREEA)»]. This project's agreement number is 01676.02. Fertilized eggs from maturing Northern pike females were obtained from the fish production unit of the fishing federation of Ille-et-Vilaine (Pisciculture du Boulet, Feins, France). Fish were maintained indoors under strictly controlled conditions in individual aquaria to avoid cannibalism, with running dechlorinated local water (pH = 8) filtered with a recirculating system. Rearing temperature and photoperiod followed the trend of the ambient natural environment. Northern pike fry were fed with live prey such as artemia larvae, daphnia, and adult artemia depending on their size. After reaching a length of 4–5 cm, Northern pike juveniles were fed with rainbow trout fry.

Fosmid library screening, sub-cloning and assembling

The Northern pike fosmid genomic DNA library was constructed by Bio S&T (Québec, Canada) from high molecular weight DNA extracted from the liver of a male *E. lucius* from Ille et Vilaine, France, using the CopyControl Fosmid Library Production Kit with pCC1FOS vectors (Epicentre, USA) following the manufacturer's instructions. The resulting fosmid library contained around 500,000 non-amplified clones that were arrayed in pools of 250 individual fosmids in ten 96-well plates.

Northern pike fosmid clones were screened by PCR (**S5 Table**) to identify individual fosmids containing *amhby* and *amha*. To sequence the two ~ 40 kb fosmids, purified fosmid DNA was first fragmented into approximately 1.5 kb fragments using the Nebulizer kit supplied in the TOPO shotgun Subcloning kit (Invitrogen, Carlsbad, CA), and then sub-cloned into pCR4Blunt-TOPO vectors. Individual plasmid DNAs were sequenced from both ends with Sanger sequencing using M13R and M13F primers. The resulting sequences were then assembled using ChromasPro version 2.1.6 (Technelysium Pty Ltd, South Brisbane, Australia).

Phylogenetic and synteny analyses

The protein and cDNA sequences of *amhby* and *amha* were predicted from their genomic sequences using the FGENESH+ suite [76], and the resulting cDNA sequences were compared to the corresponding transcripts publicly available in the Phylofish database [38]. Shared identity between cDNA sequences was calculated after alignment using Clustal Omega [77] implemented on EMBL-EBI [78,79] with default settings.

Global pairwise alignment of the two transcripts was performed using the mVISTA LAGAN program with default settings [80]. Shared identity and similarity between protein sequences were calculated using EMBOSS Water [81]. For each *amh* paralog, we compared the 5 kb genomic sequence upstream of the start codon using PromoterWise [81] and mVISTA LAGAN [80]. For *amha*, this 5 kb upstream region included the entire intergenic sequence up to and including part of the *dot1l* gene, located 4.3 kb from the start codon of *amha*, so the promoter for *amha* is likely located in this region.

Phylogenetic relationship reconstruction of Amh proteins was performed using the maximum likelihood method implemented in PhyML 3.0 [82], and the proper substitution model was determined with PhyML-SMS [83]. Amh protein sequences from eight teleost species and from the spotted gar (*Lepisosteus oculatus)*, which was used as outgroup, were retrieved from the NCBI Protein database. The accession number for each sequence is referenced in **S4 Table**.

A synteny map of the conserved genes in blocks around *amh* was constructed with nine teleost species and spotted gar as outgroup. For spotted gar, zebrafish (*Danio rerio*), Atlantic cod (*Gadus morhua*), threespine stickleback (*Gasterosteus aculeatus*), Japanese pufferfish (*Takifugu rubripes*), tetraodon (*Tetraodon nigroviridis*), Nile tilapia (*Oreochromis niloticus*), Southern platyfish (*Xiphophorus maculatus*) and Amazon molly (*Poecilia formosa*), the synteny map was created with the Genomicus website (http://www.genomicus.biologie.ens.fr, accessed in July 2017 [84,85]). For Northern pike, genes located upstream and downstream of *amha* on LG08 were deduced based on the location of each gene in the genome assembly (Eluc_V4, GenBank assembly accession: GCA_004634155.1).

DNA extraction and genotyping

Small fin clips were taken from caudal fins of Northern pikes after administrating fish anesthetics (3 ml of 2-phenoxyethanol per L). When needed, fish were euthanized with a lethal dose of anesthetics (10 ml of 2-phenoxyethanol per L). All fin clips were collected and stored at 4˚C in 75–100% ethanol until DNA extraction.

To obtain DNA for genotyping, fin clips were lysed with 5% Chelex and 25 mg of Proteinase K at 55˚C for 2 hours, followed by incubation at 99˚C for 10 minutes. After brief centrifugation, the supernatant containing genomic DNA was transferred to clean tubes without the Chelex beads [86]. Primers for genotyping were designed using Perl-Primer (version 1.1.21, [87]). Primer sequences and corresponding experiments can be found in **S5 Table**.

To obtain DNA for Sanger and Illumina short read sequencing, DNA was extracted from fin clips using NucleoSpin Kits for Tissue (Macherey-Nagel, Duren, Germany) following the producer's protocol. To obtain high molecular weight DNA for long read sequencing, DNA was isolated from blood samples lysed in TNES-Urea buffer (TNES-Urea: 4 M urea; 10 mM Tris-HCl, pH 7.5; 125 mM NaCl; 10 mM EDTA; 1% SDS) [88] and extracted with phenol-chloroform [89]. Afterwards, the DNA pellet was washed three times with 80% ethanol before being stored at 4˚C in 80% ethanol.

When preparing DNA for sequencing, DNA concentration was first quantified with a NanoDrop ND2000 spectrophotometer (Thermo scientific, Wilmington, DE) to estimate the range of concentration, and was then measured again with Qubit3 fluorometer (Invitrogen, Carlsbad, CA) to determine the final concentration.

RNA extraction, cDNA synthesis and qPCR

Ten Northern pikes from a local French population were sampled at 54, 75, 100, and 125 days post fertilization (dpf). For the first three time points, the whole trunk, defined as the entire body without head and tail, was collected for RNA extraction because gonads were too small to be dissected consistently at these stages. At 125 dpf, gonads were large enough to be isolated and were thus collected for RNA extraction. The genotypic sex of each animal was determined based on *amhby* amplification and results are listed in **S6 Table**.

Samples (trunks or gonads) were immediately frozen in liquid nitrogen and stored at -70˚C until RNA extraction. RNA was extracted using Tri-Reagent (Molecular Research Center, Cincinnati, OH) following the manufacturer's protocol and RNA concentration was quantified

with a NanoDrop ND2000 spectrophotometer (Thermo scientific, Wilmington, DE). Reverse transcription was performed by denaturing a mix of 1μg of RNA and 5 μL of 10 mM dNTP at 70˚C for 6 minutes, followed by 10 minutes on ice. Random hexamers and M-MLV reverse transcriptase (Promega, Madison, WI) were then added, and the mixture was incubated at 37˚C for 75 minutes followed by 15 minutes of enzyme inactivation at 70˚C, and then chilled at 4˚C. During these last two steps, a negative control without reverse transcriptase was prepared for each sample. The resulting cDNA was diluted 25-fold before qPCR.

Primers for qPCR were designed using Perl-Primer (version 1.1.21, [87]) on intron-exon junctions to avoid genomic DNA amplification (primers are listed in **S5 Table**). Primer pairs were diluted to 6 μg/μL for qPCR, which was performed with SYBER GREEN fluorophore kits (Applied Biosystems, Foster City, CA) on an Applied Biosystems StepOne Plus instrument. For each reaction, 4 μL of diluted cDNA and 1 μL of diluted primers were added to 5 μL of SYBER Green master mix. The following qPCR temperature cycling regime was used: initial denaturation at 50˚C for 2 minutes and then at 95˚C for 2 minutes, followed by 40 PCR cycles at 95˚C for 15s, 60˚C for 30s and 72˚C for 30s, and a final dissociation step at 95˚C for 3s, 60˚C for 30s and 95˚C for 15s. Primer pairs were checked for nonspecific amplification and primer-dimers using the dissociation curve implemented in the StepOne software. Relative abundance of each target cDNA was calculated from a standard curve of serially diluted pooled cDNA from all samples, and then normalized with the geometric mean of the expression of six housekeeping genes (*β-actin*, *ef1α*, *gapdh*, *eftud2*, *ubr2*, and *ccdc6b*) following classical qPCR normalization procedures [90].

Whole mount *in situ* hybridizations

In situ hybridization RNA probes for *amhby* and *amha* were synthesized from cDNA PCR products amplified from 125 dpf testis samples using primers including T7 sequences in the reverse primer (**S5 Table**). These PCRs were performed with the Advantage2 Taq Polymerase (Clontech, Mountain View, CA) for high fidelity. Gel electrophoresis was performed on the PCR products, and products of the expected size were cut out and purified using NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Duren, Germany). 10 ng of purified PCR product was then used as template for RNA probe synthesis. RNA probes were synthesized using T7 RNA polymerase (Promega, Harbonnieres, France) following the manufacturer's protocol, with digoxigenin-11-UTP for *amha* and Fluorescein-12-UTP for *amhby*. Afterwards, probes were purified using NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Duren, Germany), re-suspended in 50 μL of DEPC water, and stored at -70 ˚C.

Samples used for *in situ* hybridization were entire trunks of males and females collected at 85 dpf and fixed overnight at 4˚c in 4% paraformaldehyde solution, then washed and stored at -20˚c in 100% methanol. Hybridization was performed using an *in situ* Pro intavis AG robotic station according to the following procedure: male and female samples were rehydrated, permeabilized with proteinase K (25 mg/ml) for 30 minutes at room temperature, and incubated in post-fix solution (4% paraformaldehyde and 0.2% glutaraldehyde) for 20 minutes. Then, samples were incubated for 1 h at 65 ˚C in a hybridization buffer containing 50% formamide, 5% SCC, 0.1% Tween 20, 0.01% tRNA (0.1 mg/ml), and 0.005% heparin. RNA probes were added and samples were left to hybridize at 65 ˚C for 16 h. Afterwards, samples were washed three times with decreasing percentages of hybridization buffer, incubated in blocking buffer (Triton .1%, Tween 20 0.2%, 2% serum/PBS) for 2 h, and incubated for 12 h with the addition of alkaline phosphatase coupled with anti-digoxigenin antibody for *amha* and anti-Fluorescein antibody for *amhby* (1:2000, Roche Diagnostics Corp, Indianapolis, IN). Samples were then washed with PBS solution, and color reactions were performed with NBT/BCIP (Roche

Diagnostics Corp, Indianapolis, IN). After visual inspection of coloration, samples were dehydrated and embedded in plastic molds containing paraffin with a HistoEmbedder (TBS88, Medite, Burgdorf, Germany) instrument. Embedded samples were sectioned into 5 μm slices using a MICRO HM355 (Thermo Fisher Scientific, Walldorf, Germany) instrument. Imaging of the slides was performed with an automated microscope (Eclipse 90i, Nikon, Tokyo, Japan).

TALENS knock-out

Three pairs of TALENs [91,92], called T1, T2 and T3, were designed to target exon 1 and exon 2 of *amhby* (**S8A Fig**, **S8B Fig**). TALENs were assembled following a method derived from Huang *et al.* [93]. For each subunit, the target-specific TALE DNA binding domain consisted of 16 RVD repeats obtained from single RVD repeat plasmids kindly provided by Bo Zhang (Peking University, China). Assembled TALE repeats were subcloned into a pCS2 vector containing appropriate Δ152 Nter TALE, +63 Cter TALE, and FokI cDNA sequences with the appropriate half-TALE repeat (derived from the original pCS2 vector [93]). TALEN expression vectors were linearized by *NotI* digestion. Capped RNAs were synthesized using the mMESSAGE mMACHINE SP6 Kit (Life Technologies, Carlsbad, CA) and purified using the NucleoSpin RNA II Kit (Macherey-Nagel, Duren, Germany).

Groups of embryos at the one-cell stage were microinjected with one pair of TALENs at the concentration of either 50 ng/L or 100 ng/L. Among surviving embryos, 66 were injected with T1 (51 embryos at 50 ng/L, 15 embryos at 100 ng/L); 101 were injected with T2 (73 embryos at 50 ng/L, 28 embryos at 100 ng/L); and 45 were injected with T3 (all at 50 ng/L). At two months post fertilization, fin clips were collected from 36 surviving animals for genotyping with primers flanking the TALENs targets. Amplification of primers flanking the TALENs targets on *amhby* was also used for sex genotyping. Only genetic males with a disrupted *amhby* sequence were raised until the following reproductive season. The sperm of three one-year old G0 mosaic phenotypic males with a disrupted *amhby* sequence was collected and then used in *in vitro* fertilization with wild-type eggs. G1 individuals were genotyped for *amhby* TALEN targeting sites using primers flanking the TALENs targets (**S8C Fig**), and XY mutants were kept until 153 dpf. The homologous region on *amha* was also amplified and sequenced in all the G1 mutants to verify that there were no off target effects on *amha* (**S9 Fig**). G1 *amhby* mutants were then euthanized and dissected, and their gonads were subjected to histological analysis to inspect the phenotypes resulting from *amhby* knockout.

Additive transgenesis with *amhby* fosmid

Northern pike embryos from wild-type parents were microinjected with the *amhby* fosmid at a concentration of 100 ng/L at the one-cell stage. Two months post fertilization, fin clips were collected and used for genotyping with primer pairs in which one primer was located on the fosmid vector sequence, which does not come from the Northern pike genome, and the other primer was located in the insert sequence from Northern pike genomic DNA to ensure specificity (**S5 Table**). Only G0 males possessing the *amhby* fosmid were kept until the following reproductive season when spawning was induced with Ovaprim (Syndel, Ferndale, Washington) following the protocol from [94]. Sperm from two G0 males possessing the *amhby* fosmid was collected and used for *in vitro* fertilization with wild-type eggs. F1 XX individuals were identified due to the absence of amplification of Y-specific sequences outside of the sequence contained in the *amhby* fosmid. The XX transgenics with *amhby* fosmid were kept until 155 dpf together with wild-type sibling control genetic males and females and then dissected. Gonads from mutant and control fish were subjected to histological analysis.

Histology

Gonads to be processed for histology were fixed immediately after dissection in Bouin's fixative solution for 24 hours. Samples were dehydrated with a Spin Tissue Processor Microm (STP 120, Thermo Fisher Scientific, Walldorf, Germany) and embedded in plastic molds in paraffin with a HistoEmbedder (TBS88; Medite, Burgdorf, Germany). Embedded samples were cut serially into slices of 7 μm using a MICRO HM355 (Thermo Fisher Scientific, Walldorf, Germany) and stained with Hematoxylin. Imaging was performed with an automated microscope (Eclipse 90i, Nikon, Tokyo, Japan).

Statistical analyses

All statistical analyses, including Wilcoxon signed rank test and Chi-squared test, were performed with R (version 3.5.1 [95]).

Population analysis for male and female *E. lucius* RAD-Seq markers

A Restriction Associated DNA Sequencing (RAD-Seq) library was constructed from genomic DNA extracted from the fin clips of two parents, 37 male offspring and 41 female offspring according to standard protocols [96]. The phenotypic sex of the offspring was determined by histological analysis of the gonads at 8-months post fertilization. The library was sequenced in one lane of Illumina HiSeq 2500. Raw reads were analyzed with the Stacks [97] program version 1.44. Quality control and demultiplexing of the 169,717,410 reads were performed using the *process_radtags.pl* script with default settings. In total, 128,342,481 (76%) reads were retained after this filtering step, including ~ 1.6 M. retained sequences from the mother, ~ 0.9 M. retained sequences from the father, and between 1.0 M. and 2.2 M. retained sequences from each offspring.

Demultiplexed reads were mapped to the genome assembly of *E. lucius* (Elu_V4, GenBank assembly accession: GCA_004634155.1) using BWA (version 0.7.15-r1140, [98]) with default settings. The resulting BAM files were run through the *ref_map.pl* pipeline with default settings except a minimum stack depth of 10 (m = 10). Results of *ref_map.pl* were analyzed with *populations* using the—fstats setting to obtain population genetic statistics between sexes. Fisher's exact test was performed on all polymorphic sites using Plink (version 1.90b4.6 64-bit, [99]) to estimate association between variants and phenotypic sex. A Manhattan plot was constructed in R with homemade scripts showing $-\log_{10}$ (Fisher's test p-value) for all 25 LGs of the *E. lucius* genome.

Analysis of sex bias in distribution pattern of maternal and paternal-specific markers

To identify RAD markers segregating between male and female offspring, we used the RADSex pipeline [100]. RADSex groups demultiplexed reads sharing the exact same sequence into markers, and thus reads belonging to the same polymorphic locus are split into multiple markers. As a consequence, markers generated by RADSex are non-polymorphic, which allows a comparison of the presence/absence of markers between individuals. First, a table of depth for each marker in each individual from the dataset was generated with RADSex *process*, and all markers were aligned to the reference assembly (NCBI accession number GCA_004634155.1) using RADSex *map* with a minimum depth to consider a marker present in an individual set to 5. This command aligns the markers to a reference assembly using the BWA mem algorithm and estimates sex-bias for each marker, defined as the difference between the proportion of males and the proportion of females in which this marker is present: $S = M / M_{TOTAL} - F /$

F_{TOTAL}, where M and F are the number of males and females in which a marker is present, respectively, and M_{TOTAL} and F_{TOTAL} are the total number of males and females in the population, respectively. With this definition, as our population is a family panel, sex-bias can be used as a measure of segregation between male and female offspring. Then, for each aligned marker present in at least ten individuals, parental origin was assigned as follows: markers present in the dam but not in the sire were considered maternal, markers present in the sire but not in the dam were considered paternal, and markers present in both the sire and the dam were considered non-specific.

Genetic map construction

Rad-Seq reads were also used to construct a genetic map. Raw reads were analyzed with the Stacks [97] program version 1.28. Quality control and demultiplexing of the 169,717,410 reads were performed using the *process_radtags.pl* wrapper script with all settings set to default except the minimum mean quality score on the reading window, which was set to 20 (-s 20). In total, 129,128,677 (76%) reads were retained after this filtering step, including ~2,500,000 retained sequences from the mother, ~3,000,000 retained sequences from the father, and between 1,000,000 and 2,000,000 retained sequences from each offspring. Stacks of reads were built using the *denovo_map.pl* wrapper script with minimal number of reads to form a stack set to 3 (m = 3), high bound set to 0.05 (—bound_high 0.05) and all other settings set to default. A total of 255,813 loci were obtained, among which 8,246 were polymorphic. These polymorphic loci were sorted and filtered to 1) remove loci found in less than half of the offspring; 2) remove loci with more than two triploid (or higher polyploid level) genotypes in retained loci, these unexpected genotypes were replaced with NA; 3) remove loci with heterozygosity higher than 0.66 or lower than 0.33; 4) select markers with two alleles and two genotypes; 5) select markers with two alleles and three genotypes, a heterozygosity between 0.47 and 0.53, and a ratio "number of homozygote 1/ number of homozygote 2" between 0.8 and 1.2 (such markers are informative in both parents and can help to merge male and female maps); and 6) output markers in a format compatible with Carthagène [101], which was used to help sort linkage groups, order markers, and calculate map distances. To be consistent with the cytogenetic observation that *E. lucius* has 25 pairs of chromosomes, a LOD threshold of 7 was chosen. To check for alternative orders with higher likelihood and to test inversions and permutations of groups of markers, the following optimization settings were used: -annealing 30 300.0 0.1 0.8, -flips 5 0 2, -greedy 1–1 5 30 0, -polish, and -squeeze {50 50}.

Genome sequencing and assembly

The genome of one local phenotypic male Northern pike (*Esox lucius*) was assembled using Oxford Nanopore long reads and polished with Illumina reads.

Illumina short reads libraries were built using the Truseq nano kit (Illumina, ref. FC-121-4001) following the manufacturer's instructions. First, 200 ng of gDNA was briefly sonicated using a Bioruptor sonication device (Diagenode, Liege, Belgium), end-repaired and size-selected on beads to retain fragments of around 550 bp, and these fragments were then A-tailed and ligated to Illumina's adapter. Ligated DNA was then subjected to eight PCR cycles. Libraries were checked with a Fragment Analyzer (AATI) and quantified by qPCR using the Kapa Library quantification kit. Libraries were sequenced on one lane of a HiSeq2500 using the paired end 2x250 nt v2 rapid mode according to the manufacturer's instruction. Image analysis was performed with the HiSeq Control Software and base calling with the RTA software provided by Illumina.

Nanopore long read libraries were prepared and sequenced according to the manufacturer's instruction (SQK-LSK108). DNA was quantified at each step using the Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA). DNA purity was checked using a NanoDrop ND2000 spectrophotometer (Thermo scientific, Wilmington, DE) and size distribution and degradation were assessed using a Fragment analyzer (AATI) High Sensitivity DNA Fragment Analysis Kit. Purification steps were performed using AMPure XP beads (Beckman Coulter). In total, five flowcells were sequenced. For each flowcell, approximately 7 μg of DNA was sheared at 20 kb using the megaruptor system (Diagenode, Liege, Belgium). A DNA damage repair step was performed on 5 μg of DNA, followed by an END-repair and dA tail of double stranded DNA fragments, and adapters were then ligated to the library. Libraries were loaded on R9.4.1 flowcells and sequenced on a GridION DNA sequencer (Oxford Nanopore, Oxford, UK) at 0.1 pmol for 48 h.

In total, 1,590,787 Nanopore reads corresponding to 17,576,820,346 nucleotides were used in the assembly. Adapters were removed using Porechop (version 0.2.1,https://github.com/rrwick/Porechop). The assembly was performed with Canu (version 1.[102]) using standard parameters, with genomeSize set to 1.1G to match theoretical expectations [103] and maxMemory set to 240. Two rounds of polishing were performed with racon version 1.3.1 using standard parameters. For this step, Nanopore long reads were mapped to the assembly with Minimap2 (version 2.5 [104]) using the map-ont parameter preset. Afterwards, four additional rounds of polishing were performed with Pilon [105] using default parameters and the Illumina short reads mapped to the assembly with BWA mem (version 0.7.17 [98]). Metrics for the resulting assembly were calculated with the assemblathon_stats.pl script [106]. The assembly's completeness was assessed with BUSCO [107] using the *Actinopterygii* gene set (4,584 genes) and the default gene model for Augustus. The same analysis was performed on the Elu_V4 reference genome (GenBank assembly accession: GCA_004634155.1) for comparison.

Sequencing of male and female pools

DNA from 30 males and 30 females from the fish production unit of the fishing federation of Ille-et-Vilaine (Pisciculture du Boulet, Feins, France) was extracted with a NucleoSpin Kit for Tissue (Macherey-Nagel, Duren, Germany) following the manufacturer's instructions. DNA concentration was quantified using a Qubit dsDNA HS Assay Kits (Invitrogen, Carlsbad, CA) and a Qubit3 fluorometer (Invitrogen, Carlsbad, CA). DNA from different samples was normalized to the same quantity before pooling for male and female libraries separately. Libraries were constructed using a Truseq nano kit (Illumina, ref. FC-121-4001) following the manufacturer's instructions. DNAseq short reads sequencing was performed at the GeT-PlaGe core facility of INRA Toulouse, France (http://get.genotoul.fr/en/). Two DNA poolseq libraries were prepared using the Illumina TruSeq Nano DNA HT Library Prep Kit (Illumina, San Diego, CA) following the manufacturer's protocol. First, 200ng of DNA from each sample (male pool and female pool) was briefly sonicated using a Bioruptor sonication device (Diagenode, Liege, Belgium), and then end-repaired and size-selected on beads to retain fragments of size around 550 bp, and these fragments were then A-tailed and ligated to indexes and Illumina's adapter. Libraries were checked with a Fragment Analyzer (Advanced Analytical Technologies, Inc., Ankeny, IA) and quantified by qPCR using the Kapa Library Quantification Kit (Roche Diagnostics Corp, Indianapolis, IN). Sequencing was performed on a NovaSeq S4 lane (Illumina, San Diego, CA) using paired-end 2x150 nt mode with Illumina NovaSeq Reagent Kits following the manufacturer's instructions. The run produced 129 million read pairs for the male pool library and 136 million read pairs for the female pool library.

Identification of sex-differentiated regions in the reference genome

Reads from the male and female pools were aligned separately to the reference genome (GenBank assembly accession: GCA_004634155.1) using BWA mem version 0.7.17 [98] with default parameters. Each resulting BAM file was sorted and PCR duplicates were removed using Picard tools version 2.18.2 (http://broadinstitute.github.io/picard) with default parameters. Then, a pileup file combining both BAM files was created using samtools mpileup version 1.8 [108] with per-base alignment quality disabled (-B). A sync file containing the nucleotide composition in each pool for each position in the reference was generated from the pileup file using popoolation mpileup2sync version 1.201 [109] with a min quality of 20 (—min-qual 20).

An in-house C++ software was developed to identify sex-specific SNPs, defined as positions heterozygous in one sex while homozygous in the other sex, compute per base between-sex F_{ST}, and coverage for each sex from the output of popoolation mpileup2sync (PSASS v2.0.0, doi: 10.5281/zenodo.2615936). PSASS outputs 1) all positions with sex-specific SNPs 2) the number of such positions in a sliding window over the genome, 3) all positions with high between-sex FST, 4) between-sex FST in a sliding window over the genome, 5) average absolute and relative coverage for each sex in a sliding window over the genome, and 6) number of sex-specific SNPs as well as coverage for each sex for all genes and CDS in a user-supplied GFF file.

We used PSASS to identify non-overlapping 50 kb windows enriched in sex-specific SNPs in *E. lucius*, using the following parameters: minimum depth 10 (—min-depth 10), allele frequency for a heterozygous locus 0.5 ± 0.15 (—freq-het 0.5,—range-het 0.15), allele frequency for a homozygous locus 1 ± 0.05 (—freq-het 1,—range-het 0.05),—window-size 50000, and—output-resolution 50000, and group snps (—group-snps). To precisely delimitate the region enriched in male-specific SNPs on LG24, the number of sex-specific SNPs was similarly computed in 2,500 bp windows, only changing the parameters—window-size to 2500 and—output-resolution to 2500.

Identification of Y specific sequences in the Nanopore assembly

A sync file containing the nucleotide composition in each pool for each position in the reference was generated as described in the previous section, using the Nanopore assembly (NCBI accession number SDAW00000000) to align the reads. This time, because we were comparing read coverage levels, the BAM files were filtered with samtools version 1.8 [108] to only retain reads with a properly mapped pair and a mapping quality higher than 30 to reduce the impact of false positive mapped reads. The resulting sync file was used as input for PSASS to compute coverage for each sex in 1 kb non-overlapping windows along the genome using the following parameters:—min-depth 10,—window-size 1000, and—output-resolution 1000. Results from this analysis were filtered in R (version 3.5.1 [95]) to first identify 1 kb regions with mean absolute depth higher than 10 in males and no aligned reads in females (MR1k), and then regions with mean relative coverage between 0.3 and 0.7 in males, and mean absolute coverage lower than 1 in females.

To identify protein coding sequences, we performed alignments between the Y-specific sequences and the teleostei (taxid:32443) non-redundant protein database using blastx (https://blast.ncbi.nlm.nih.gov/Blast.cgi, version 2.8.1 [110]) with the parameters "Max target sequences" set to 50 and "Max matches in a query range" set to 1. Regions matching potential homologs with an e-value $<$ 1E-50 were considered as protein coding sequences.

To determine repeat content of the Y-specific sequence, RepeatMasker (version open 4.0.3 [111]) was run on these Y-specific sequences with NCBI/RMBLAST (version 2.2.27+) against

the Master RepeatMasker Database (Complete Database: 20130422). The same analysis was also performed on the entire Nanopore assembly.

Supporting information

S1 File. Text: Pool-seq analysis revealed ~180 kb of Y-specific sequences and estimation of divergence time between *amha* and *amhby* suggests an ancient duplication event.
(DOCX)

S1 Fig. Normalized RNA-Seq read count for *amha* and *amhby* in different tissues of *Esox lucius*. Data was obtained from the Phylofish database (http://phylofish.sigenae.org/) and was normalized by reads per kilobase of sequence per million reads for each transcript in each library for *amha* (**A**) and *amhby* (**B**). Non-zero values are indicated on top of the bar.
(TIF)

S2 Fig. Density of transposable element domain HTH_Tnp_Tc3_2 in the genome of *Esox lucius* across 25 linkage groups (LG) and unplaced scaffolds. Total number of transposable element domain HTH_Tnp_Tc3_2 divided by the length of each chromosome was plotted for each of the 25 LGs and unplaced scaffolds in the female reference genome (GenBank assembly accession: GCA_004634155.1). *amha* is located on LG 08 and *amhby* is located on LG 24.
(TIF)

S3 Fig. Sequence alignment of the TGF- β domain of Amha and Amhby. The seven conserved cysteines are highlighted by colored triangles: blue triangles for cysteines involved in forming intra-chain disulfide bonds and a red triangle for the cysteine involved in forming inter chain disulfide bonds.
(TIF)

S4 Fig. Alignment of AMH protein sequence from different vertebrate species, including Amha and Amhby from *Esox lucius*. The AMH domain at the N-terminus is highlighted with green dashed-line boxes, and the TGF-β domain at the C-terminus is highlighted with purple dashed-line boxes with the seven conserved cysteines indicated by the red triangles. EL: *Esox lucius*, XP_010870000.1(Amha) MK355503 (Amhby); OM: *Oncorhynchus mykiss*, XP_021459508.1; DR: *Danio rerio*, NP_001007780.1; HS: *Homo sapiens*, AAH49194.1; MM: *Mus musculus*, NP_031471.2; GG: *Gallus gallus*, NP_990361.1.
(TIF)

S5 Fig. Parental marker distribution bias in offspring along LG24. Sex bias in offspring, defined as the difference between the proportion of male offspring and the proportion of female offspring in which a marker is present, is plotted against genomic position of LG24 for all RADSex markers aligned to this chromosome. Markers found in the dam but not in the sire (i.e. maternal-specific markers) are colored in red, markers found in the sire but not in the dam (i.e. paternal-specific markers) are colored in blue, and markers found in both the sire and the dam are colored in grey. Maternal-specific markers and markers found in both parents were evenly distributed among male and female offspring along the entire chromosome. However, paternal markers aligned to the first one third of the chromosome segregated in either male or female offspring. This sex bias in marker segregation is strongest at the proximal end of the chromosome and decreased gradually along the chromosome, corelating with a marker's physical distance to the sex locus.
(TIF)

S6 Fig. Three contigs with male-specific coverage in *E. lucius* genome based on pool-seq analysis results. A-C: Three contigs containing regions with only male coverage in the Oxford Nanopore assembly. Relative coverage of male and female reads are indicated by blue and red lines, respectively. The location of *amhby* on tig00003316 is indicated by a solid black arrow. The lower dotted line indicates 0.5 genome average coverage and the higher dotted line indicates genome average coverage. The grey shaded regions correspond to repeated elements and the purple shaded regions correspond to regions with strong homology with the reference genome (GenBank assembly accession: GCA_000721915.3) scaffold1067.
(TIF)

S7 Fig. Temporal expression of *gsdf*, *dmrt1*, *cyp19a1a* and *amhrII* mRNA in male and female developing gonads. Boxplots showing the first quantile, median, and the third quantile of *gsdf*, *dmrt1*, *cyp19a1a* and *amhrII* mRNA expression. Outliers are displayed as dots. The log10 of the expression of mRNA of these four genes were measured with qPCR at 54, 75, 100, and 125 days post fertilization in male and female trunks of *E. lucius*. Significance levels of p-value are given for Wilcoxon signed rank test between male and female expression at each time point: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$ and 'ns' indicates $P > 0.05$.
(TIF)

S8 Fig. TALENs targeting regions and G1 mutations on *amhby*. A: Design of TALENs targeting *amhby* exon 1 and exon 2. Grey arrows indicate the position targeted by TALENs with the corresponding sequences in grey boxes. Orange arrows indicate the position of primers used for genotyping of *amhby* knockout mutants. **B:** Alignment of amha and *amhby* sequence around the region targeted by Talen 2 designed specifically to cleave *amhby* sequence. Talen 2 targeting sequence is highlighted by a black dashed line box. **C:** Alignment of *amhby* exon 1 sequences between the wildtype and three different mutants showing different sequences deletion. The number of G1 animals with each type of the mutation is indicated by the red number.
(TIF)

S9 Fig. Alignment of *amha* sequence from all 23 G1 mutants in the amhby Knockout experiment using TALENs. Identical bases to the reference are represented as dots and the regions targets by TALENs on *amhby* are highlighted in yellow in the reference sequence.
(TIF)

S1 Table. Assemblathon and BUSCOs metrics for the genome assembly with Nanopore reads of a genetic male *E.lucius*.
(XLSX)

S2 Table. Blastx result for potential protein coding sequences on the Y-specific sequences from tig00003316, tig00009868 and tig00003988 of the Nanopore genome sequence from a male *E. lucius* against the teleostei (taxid:32443) non-redundant protein database.
(XLSX)

S3 Table. Mapping of male specific sequences of *E. lucius* with RAD tag ID, sequence, mapping score from BWA and mapped linkage group and position on the reference genome assembly (GenBank assembly accession: GCA_000721915.3).
(XLSX)

S4 Table. Sequence accession numbers for Amh protein sequences used to construct the teleost Amh phylogeny from eight teleost species and Spotted gar (*Lepisosteus oculatus*).
(XLSX)

S5 Table. Primers used in different experiments with their names, sequence, target genes and their usage.
(XLSX)

S6 Table. Number of total samples, genetic female samples and genetic male samples at different developmental stages used for qPCR. dpf: days post fertilization.
(XLSX)

S7 Table. Number, sequence length, and percentage of sequence of different classes of repeated elements in the sequence of the sex locus and the entire Nanopore genome assembly of *E. lucius*.
(XLSX)

S1 Numerical_Data. Fig 2. A) Data for the log10 of p-values from the Fisher's exact test between each marker and sex phenotype with their aligned position on each linkage group (LG) in the reference genome of *Esox lucius*. **B)** Genetic map distance of RAD-Seq markers on LG 24 and their physical position in LG24 for males and females and marker sequences. **C**) Number of male and female-specific SNPs from pool-seq in 50 kb and 2.5kb non-overlapping windows is plotted along LG 24 of the reference female genome.
(XLSX)

S2 Numerical_Data. Fig 3. The log10 of the expression of mRNA of *amha* and *amhby* measured with qPCR at 54, 75, 100, and 125 days post fertilization in male and female trunks of *E. lucius*.
(XLSX)

S3 Numerical_Data. S1 Fig. Raw read count and normalized read count for *amha* and *amhby* in 10 adult tissue RNA-Seq libraries from *Esox lucius*. Data was obtained from the Phylofish database (http://phylofish.sigenae.org/) and was normalized by reads per kilobase of sequence per million reads for each transcript in each library for each gene.
(XLSX)

S4 Numerical_Data. S2 Fig. Number and density of transposon TC1-like in each of the linkage group and unplaced scaffolds in the genome of *Esox lucius*.
(XLSX)

S5 Numerical_Data. S5 Fig. Sex bias in offspring, defined as the difference between the proportion of male offspring and the proportion of female offspring in which a marker is present, is plotted against genomic position of LG24 for all RADSex markers aligned to this chromosome.
(XLSX)

S6 Numerical_Data. S6 Fig. Relative and absolute read coverage depth of male and female Pool-Seq reads on the Nanopore genome of a male *Esox lucius* from an European population. This data file was used to search for 1kb windows that only has male coverage in the genome.
(XLSX)

S7 Numerical_Data. S7 Fig. The log10 of the expression of mRNA of *cyp19a1a*, *dmrt1*, *gsdf* and *amhrII* measured with qPCR at 54, 75, 100, and 125 days post fertilization in male and female trunks of *E. lucius*.
(XLSX)

Acknowledgments

We thank three anonymous reviewers for their constructive comments in improving this manuscript. We are grateful to the fish facility of INRA LPGP for support in experimental installation and fish maintenance and to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing help and/or computing and/or storage resources. We want to thank Nicolas Perrin, Susana Coelho and Eric Pailhoux for feedback and helpful discussion on an earlier version of the manuscript. We also want to thank Tomas Kay for improving the writing of the manuscript.

Author Contributions

Conceptualization: Qiaowei Pan, John Postlethwait, Manfred Schartl, Amaury Herpin, Yann Guiguen.

Data curation: Romain Feron.

Formal analysis: Qiaowei Pan, Romain Feron, René Guyomard, Christophe Klopp.

Funding acquisition: John Postlethwait, Manfred Schartl, Yann Guiguen.

Investigation: Qiaowei Pan, Ayaka Yano, Elodie Jouanno, Estelle Vigouroux, Ming Wen, Hugues Parrinello, Laurent Journot, Jérôme Lluch, Céline Roques.

Resources: Jean-Mickaël Busnel, Julien Bobe, Jean-Paul Concordet.

Software: Romain Feron.

Supervision: Amaury Herpin, Yann Guiguen.

Visualization: Qiaowei Pan, Romain Feron.

Writing – original draft: Qiaowei Pan, Romain Feron, Jean-Paul Concordet, Christophe Klopp, Yann Guiguen.

Writing – review & editing: Qiaowei Pan, Romain Feron, John Postlethwait, Manfred Schartl, Amaury Herpin, Yann Guiguen.

References

1. Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. Science. 1999; 286: 964–967. https://doi.org/10.1126/science.286.5441.964 PMID: 10542153
2. Bachtrog D. Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila. Nat Genet. 2003; 34: 215–219. https://doi.org/10.1038/ng1164 PMID: 12754509
3. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 2003; 423: 825–837. https://doi.org/10.1038/nature01722 PMID: 12815422
4. Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet. 2006; 7: 645–653. https://doi.org/10.1038/nrg1914 PMID: 16847464
5. Soh YQS, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, et al. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell. 2014; 159: 800–813. https://doi.org/10.1016/j.cell.2014.09.052 PMID: 25417157
6. Smith CA, Roeszler KN, Ohnesorg T, Cummins DM, Farlie PG, Doran TJ, et al. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature. 2009; 461: 267–271. https://doi.org/10.1038/nature08298 PMID: 19710650
7. Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, et al. Evolutionary analysis of the female-specific avian W chromosome. Nature Communications. 2015; 6: 7330. https://doi.org/10.1038/ncomms8330 PMID: 26040272

8. Ezaz T, Sarre SD, O'Meally D, Graves JAM, Georges A. Sex chromosome evolution in lizards: independent origins and rapid transitions. Cytogenet Genome Res. 2009; 127: 249–260. https://doi.org/10.1159/000300507 PMID: 20332599
9. Gamble T, Zarkower D. Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. Molecular Ecology Resources. 2014; n/a-n/a. https://doi.org/10.1111/1755-0998.12237 PMID: 24506574
10. Yoshimoto S, Okada E, Umemoto H, Tamura K, Uno Y, Nishida-Umehara C, et al. A W-linked DM-domain gene, DM-W, participates in primary ovary development in Xenopus laevis. PNAS. 2008; 105: 2469–2474. https://doi.org/10.1073/pnas.0712244105 PMID: 18268317
11. Miura I, Ohtani H, Ogata M. Independent degeneration of W and Y sex chromosomes in frog Rana rugosa. Chromosome Res. 2012; 20: 47–55. https://doi.org/10.1007/s10577-011-9258-8 PMID: 22143254
12. Schartl M. Sex determination by multiple sex chromosomes in Xenopus tropicalis. PNAS. 2015; 112: 10575–10576. https://doi.org/10.1073/pnas.1513518112 PMID: 26283399
13. Rodrigues N, Studer T, Dufresnes C, Ma W-J, Veltsos P, Perrin N. Dmrt1 polymorphism and sex-chromosome differentiation in Rana temporaria. Mol Ecol. 2017; https://doi.org/10.1111/mec.14222 PMID: 28675502
14. Kikuchi K, Hamaguchi S. Novel sex-determining genes in fish and sex chromosome evolution: Novel Sex-Determining Genes in Fish. Developmental Dynamics. 2013; 242: 339–353. https://doi.org/10.1002/dvdy.23927 PMID: 23335327
15. Heule C, Salzburger W, Bohne A. Genetics of Sexual Development: An Evolutionary Playground for Fish. Genetics. 2014; 196: 579–591. https://doi.org/10.1534/genetics.114.161158 PMID: 24653206
16. Herpin A, Schartl M. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. EMBO reports. 2015; 16: 1260–1274. https://doi.org/10.15252/embr.201540667 PMID: 26358957
17. Pan Q, Anderson J, Bertho S, Herpin A, Wilson C, Postlethwait JH, et al. Vertebrate sex-determining genes play musical chairs. C R Biol. 2016; 339: 258–262. https://doi.org/10.1016/j.crvi.2016.05.010 PMID: 27291506
18. Mank JE, Avise JC. Evolutionary diversity and turn-over of sex determination in teleost fishes. Sex Dev. 2009; 3: 60–67. https://doi.org/10.1159/000223071 PMID: 19684451
19. Yamamoto Y, Zhang Y, Sarida M, Hattori RS, Strüssmann CA. Coexistence of Genotypic and Temperature-Dependent Sex Determination in Pejerrey Odontesthes bonariensis. Orbán L, editor. PLoS ONE. 2014; 9: e102574. https://doi.org/10.1371/journal.pone.0102574 PMID: 25036903
20. Takehana Y, Naruse K, Hamaguchi S, Sakaizumi M. Evolution of ZZ/ZW and XX/XY sex-determination systems in the closely related medaka species, Oryzias hubbsi and O. dancena. Chromosoma. 2007; 116: 463–470. https://doi.org/10.1007/s00412-007-0110-z PMID: 17882464
21. Takehana Y, Demiyah D, Naruse K, Hamaguchi S, Sakaizumi M. Evolution of different Y chromosomes in two medaka species, Oryzias dancena and O. latipes. Genetics. 2007; 175: 1335–1340. https://doi.org/10.1534/genetics.106.068247 PMID: 17194774
22. Takehana Y, Hamaguchi S, Sakaizumi M. Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, Oryzias javanicus and O. hubbsi. Chromosome Res. 2008; 16: 801–811. https://doi.org/10.1007/s10577-008-1227-5 PMID: 18607761
23. Tanaka K, Takehana Y, Naruse K, Hamaguchi S, Sakaizumi M. Evidence for Different Origins of Sex Chromosomes in Closely Related Oryzias Fishes: Substitution of the Master Sex-Determining Gene. Genetics. 2007; 177: 2075–2081. https://doi.org/10.1534/genetics.107.075598 PMID: 17947439
24. Nagai T, Takehana Y, Hamaguchi S, Sakaizumi M. Identification of the sex-determining locus in the Thai medaka, Oryzias minutillus. Cytogenet Genome Res. 2008; 121: 137–142. https://doi.org/10.1159/000125839 PMID: 18544937
25. Myosho T, Takehana Y, Hamaguchi S, Sakaizumi M. Turnover of Sex Chromosomes in Celebensis Group Medaka Fishes. G3 (Bethesda). 2015; 5: 2685–2691. https://doi.org/10.1534/g3.115.021543 PMID: 26497145
26. Volff J-N, Schartl M. Sex determination and sex chromosome evolution in the medaka, Oryzias latipes, and the platyfish, Xiphophorus maculatus. CGR. 2002; 99: 170–177. https://doi.org/10.1159/000071590 PMID: 12900561
27. Pan Q, Guiguen Y, Herpin A. Evolution of sex determining genes in Fish. Encyclopedia of Reproduction, Second Edition. 2017. https://doi.org/10.1530/REP-17-0396
28. Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, et al. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, Oryzias latipes. PNAS. 2002; 99: 11778–11783. https://doi.org/10.1073/pnas.182314699 PMID: 12193652

29. Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, et al. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. Nature. 2002; 417: 559–563. https://doi.org/10.1038/nature751 PMID: 12037570

30. Kondo M, Hornung U, Nanda I, Imai S, Sasaki T, Shimizu A, et al. Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Res. 2006; 16: 815–826. https://doi.org/10.1101/gr.5016106 PMID: 16751340

31. Kondo M, Nanda I, Schmid M, Schartl M. Sex Determination and Sex Chromosome Evolution: Insights from Medaka. SXD. 2009; 3: 88–98. https://doi.org/10.1159/000223074 PMID: 19684454

32. Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, et al. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nature Genetics. 2014; 46: 253–260. https://doi.org/10.1038/ng.2890 PMID: 24487278

33. Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Klopp C, et al. An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, Oncorhynchus mykiss. Current Biology. 2012; 22: 1423–1428. https://doi.org/10.1016/j.cub.2012.05.045 PMID: 22727696

34. Yano A, Nicol B, Jouanno E, Quillet E, Fostier A, Guyomard R, et al. The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. Evol Appl. 2013; 6: 486–496. https://doi.org/10.1111/eva.12032 PMID: 23745140

35. Forsman A, Tibblin P, Berggren H, Nordahl O, Koch-Schmidt P, Larsson P. Pike Esox lucius as an emerging model organism for studies in ecology and evolutionary biology: a review. J Fish Biol. 2015; 87: 472–479. https://doi.org/10.1111/jfb.12712 PMID: 26077107

36. Raat AJP. Synopsis of Biological Data on the Northern Pike: Esox Lucius Linnaeus, 1758. Food & Agriculture Org.; 1988.

37. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, et al. The genome and linkage map of the northern pike (Esox lucius): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS ONE. 2014; 9: e102089. https://doi.org/10.1371/journal.pone.0102089 PMID: 25069045

38. Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, et al. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. BMC Genomics. 2016; 17: 368. https://doi.org/10.1186/s12864-016-2709-z PMID: 27189481

39. Luczynski M, Dabrowski K, Kucharczyk D, Glogowski J, Luczynski M, Szczerbowski A, Babiak I. Gynogenesis in Northern pike [Esox lucius L.] induced by heat shock—preliminary data. Polskie Archiwum Hydrobiologii. 1997;1–2. Available: https://www.infona.pl//resource/bwmeta1.element.agro-article-5d7c506e-53a2-4005-a9c4-6639a8741d99

40. Nanda I, Hornung U, Kondo M, Schmid M, Schartl M. Common spontaneous sex-reversed XX males of the medaka Oryzias latipes. Genetics. 2003; 163: 245–251. PMID: 12586712

41. Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, et al. A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. Proceedings of the National Academy of Sciences. 2012; 109: 2955–2959. https://doi.org/10.1073/pnas.1018392109 PMID: 22323585

42. Li M, Sun Y, Zhao J, Shi H, Zeng S, Ye K, et al. A Tandem Duplicate of Anti-Mullerian Hormone with a Missense SNP on the Y Chromosome Is Essential for Male Sex Determination in Nile Tilapia, Oreochromis niloticus. Naruse K, editor. PLOS Genetics. 2015; 11: e1005678. https://doi.org/10.1371/journal.pgen.1005678 PMID: 26588702

43. Rondeau EB, Laurie CV, Johnson SC, Koop BF. A PCR assay detects a male-specific duplicated copy of Anti-Müllerian hormone (amh) in the lingcod (Ophiodon elongatus). BMC Research Notes. 2016; 9. https://doi.org/10.1186/s13104-016-2030-6 PMID: 27103037

44. Bej DK, Miyoshi K, Hattori RS, Strüssmann CA, Yamamoto Y. A Duplicated, Truncated amh Gene Is Involved in Male Sex Determination in an Old World Silverside. G3: Genes, Genomes, Genetics. 2017; g3.117.042697. https://doi.org/10.1534/g3.117.042697 PMID: 28611256

45. Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, et al. A Trans-Species Missense SNP in Amhr2 Is Associated with Sex Determination in the Tiger Pufferfish, Takifugu rubripes (Fugu). Peichel CL, editor. PLoS Genetics. 2012; 8: e1002798. https://doi.org/10.1371/journal.pgen.1002798 PMID: 22807687

46. Graves JAM, Peichel CL. Are homologies in vertebrate sex determination due to shared ancestry or to limited options? Genome biology. 2010; 11: 205. https://doi.org/10.1186/gb-2010-11-4-205 PMID: 20441602

47. Belville C, Van Vlijmen H, Ehrenfels C, Pepinsky B, Rezaie AR, Picard J-Y, et al. Mutations of the anti-mullerian hormone gene in patients with persistent mullerian duct syndrome: biosynthesis, secretion, and processing of the abnormal proteins and analysis using a three-dimensional model. Mol Endocrinol. 2004; 18: 708–721. https://doi.org/10.1210/me.2003-0358 PMID: 14673134

48. Knebelmann B, Boussin L, Guerrier D, Legeai L, Kahn A, Josso N, et al. Anti-Müllerian hormone Bruxelles: a nonsense mutation associated with the persistent Müllerian duct syndrome. Proc Natl Acad Sci U S A. 1991; 88: 3767–3771. https://doi.org/10.1073/pnas.88.9.3767 PMID: 2023927

49. Pfennig F, Standke A, Gutzeit HO. The role of Amh signaling in teleost fish–Multiple functions not restricted to the gonads. General and Comparative Endocrinology. 2015; 223: 87–107. https://doi.org/10.1016/j.ygcen.2015.09.025 PMID: 26428616

50. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA. 1999; 96: 2896–2901. https://doi.org/10.1073/pnas.96.6.2896 PMID: 10077608

51. Herpin A, Schartl M. Molecular mechanisms of sex determination and evolution of the Y-chromosome: Insights from the medakafish (Oryzias latipes). Molecular and Cellular Endocrinology. 2009; 306: 51–58. https://doi.org/10.1016/j.mce.2009.02.004 PMID: 19481684

52. Herpin A, Braasch I, Kraeussling M, Schmidt C, Thoma EC, Nakamura S, et al. Transcriptional Rewiring of the Sex Determining dmrt1 Gene Duplicate by Transposable Elements. Petrov DA, editor. PLoS Genetics. 2010; 6: e1000844. https://doi.org/10.1371/journal.pgen.1000844 PMID: 20169179

53. Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci. 2000; 355: 1563–1572. https://doi.org/10.1098/rstb.2000.0717 PMID: 11127901

54. Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. Heredity (Edinb). 2005; 95: 118–128. https://doi.org/10.1038/sj.hdy.6800697 PMID: 15931241

55. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, et al. The medaka draft genome and insights into vertebrate genome evolution. Nature. 2007; 447: 714–719. https://doi.org/10.1038/nature05846 PMID: 17554307

56. Wilson CA, High SK, McCluskey BM, Amores A, Yan Y -l., Titus TA, et al. Wild Sex in Zebrafish: Loss of the Natural Sex Determinant in Domesticated Strains. Genetics. 2014; 198: 1291–1308. https://doi.org/10.1534/genetics.114.169284 PMID: 25233988

57. Natri HM, Shikano T, Merilä J. Progressive Recombination Suppression and Differentiation in Recently Evolved Neo-sex Chromosomes. Mol Biol Evol. 2013; 30: 1131–1144. https://doi.org/10.1093/molbev/mst035 PMID: 23436913

58. Charlesworth D. The Guppy Sex Chromosome System and the Sexually Antagonistic Polymorphism Hypothesis for Y Chromosome Recombination Suppression. Genes (Basel). 2018; 9. https://doi.org/10.3390/genes9050264 PMID: 29783761

59. Gammerdinger WJ, Conte MA, Acquah EA, Roberts RB, Kocher TD. Structure and decay of a proto-Y region in Tilapia, Oreochromis niloticus. BMC Genomics. 2014; 15: 975. https://doi.org/10.1186/1471-2164-15-975 PMID: 25404257

60. Gammerdinger WJ, Conte MA, Sandkam BA, Ziegelbecker A, Koblmüller S, Kocher TD. Novel Sex Chromosomes in 3 Cichlid Fishes from Lake Tanganyika. J Hered. 2018; 109: 489–500. https://doi.org/10.1093/jhered/esy003 PMID: 29444291

61. Faber-Hammond JJ, Phillips RB, Brown KH. Comparative Analysis of the Shared Sex-Determination Region (SDR) among Salmonid Fishes. Genome Biol Evol. 2015; 7: 1972–1987. https://doi.org/10.1093/gbe/evv123 PMID: 26112966

62. Lubieniecki KP, Lin S, Cabana EI, Li J, Lai YYY, Davidson WS. Genomic Instability of the Sex-Determining Locus in Atlantic Salmon (*Salmo salar*). G3: Genes, Genomes, Genetics. 2015; 5: 2513–2522. https://doi.org/10.1534/g3.115.020115 PMID: 26401030

63. Vicoso B, Kaiser VB, Bachtrog D. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc Natl Acad Sci USA. 2013; 110: 6453–6458. https://doi.org/10.1073/pnas.1217027110 PMID: 23547111

64. Bull JJ. Evolution of sex determining mechanisms. Evolution of sex determining mechanisms. 1983; Available: https://www.cabdirect.org/cabdirect/abstract/19840180646

65. Rice WR. Evolution of the Y Sex Chromosome in AnimalsY chromosomes evolve through the degeneration of autosomes. BioScience. 1996; 46: 331–343. https://doi.org/10.2307/1312947

66. Bergero R, Charlesworth D. The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol (Amst). 2009; 24: 94–102. https://doi.org/10.1016/j.tree.2008.09.010 PMID: 19100654

67. Wright AE, Dean R, Zimmer F, Mank JE. How to make a sex chromosome. Nature Communications. 2016; 7: ncomms12087. https://doi.org/10.1038/ncomms12087 PMID: 27373494

68. Wang J, Na J-K, Yu Q, Gschwend AR, Han J, Zeng F, et al. Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. PNAS. 2012; 109: 13710–13715. https://doi.org/10.1073/pnas.1207833109 PMID: 22869747

69. Schultheiß R, Viitaniemi HM, Leder EH. Spatial Dynamics of Evolving Dosage Compensation in a Young Sex Chromosome System. Genome Biol Evol. 2015; 7: 581–590. https://doi.org/10.1093/gbe/evv013 PMID: 25618140

70. White MA, Kitano J, Peichel CL. Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. Mol Biol Evol. 2015; 32: 1981–1995. https://doi.org/10.1093/molbev/msv078 PMID: 25818858

71. Bergero R, Forrest A, Kamau E, Charlesworth D. Evolutionary strata on the X chromosomes of the dioecious plant Silene latifolia: evidence from new sex-linked genes. Genetics. 2007; 175: 1945–1954. https://doi.org/10.1534/genetics.106.070110 PMID: 17287532

72. Nam K, Ellegren H. The chicken (Gallus gallus) Z chromosome contains at least three nonlinear evolutionary strata. Genetics. 2008; 180: 1131–1136. https://doi.org/10.1534/genetics.108.090324 PMID: 18791248

73. Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. PLOS Biology. 2013; 11: e1001643. https://doi.org/10.1371/journal.pbio.1001643 PMID: 24015111

74. Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science. 2014; 346: 1246338. https://doi.org/10.1126/science.1246338 PMID: 25504727

75. Keinath MC, Timoshevskaya N, Timoshevskiy VA, Voss SR, Smith JJ. Miniscule differences between sex chromosomes in the giant genome of a salamander. Sci Rep. 2018; 8. https://doi.org/10.1038/s41598-018-36209-2 PMID: 30552368

76. Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 2006; 7: S10. https://doi.org/10.1186/gb-2006-7-s1-s10 PMID: 16925832

77. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7: 539. https://doi.org/10.1038/msb.2011.75 PMID: 21988835

78. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 2013; 41: W597–600. https://doi.org/10.1093/nar/gkt376 PMID: 23671338

79. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 2015; 43: W580–584. https://doi.org/10.1093/nar/gkv279 PMID: 25845596

80. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, NISC Comparative Sequencing Program, et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 2003; 13: 721–731. https://doi.org/10.1101/gr.926603 PMID: 12654723

81. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000; 16: 276–277. PMID: 10827456

82. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59: 307–321. https://doi.org/10.1093/sysbio/syq010 PMID: 20525638

83. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol. 2017; 34: 2422–2424. https://doi.org/10.1093/molbev/msx149 PMID: 28472384

84. Louis A, Muffato M, Roest Crollius H. Genomicus: five genome browsers for comparative genomics in eukaryota. Nucleic Acids Res. 2013; 41: D700–705. https://doi.org/10.1093/nar/gks1156 PMID: 23193262

85. Louis A, Nguyen NTT, Muffato M, Roest Crollius H. Genomicus update 2015: KaryoView and MatrixView provide a genome-wide perspective to multispecies comparative genomics. Nucleic Acids Res. 2015; 43: D682–D689. https://doi.org/10.1093/nar/gku1112 PMID: 25378326

86. Gharbi K, Gautier A, Danzmann RG, Gharbi S, Sakamoto T, Høyheim B, et al. A linkage map for brown trout (Salmo trutta): chromosome homeologies and comparative genome organization with other salmonid fish. Genetics. 2006; 172: 2405–2419. https://doi.org/10.1534/genetics.105.048330 PMID: 16452148

87. Marshall OJ. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics. 2004; 20: 2471–2472. https://doi.org/10.1093/bioinformatics/bth254 PMID: 15073005

88. Asahida T, Kobayashi T, Saitoh K, Nakayama I. Tissue Preservation and Total DNA Extraction form Fish Stored at Ambient Temperature Using Buffers Containing High Concentration of Urea. Fisheries science. 1996; 62: 727–730. https://doi.org/10.2331/fishsci.62.727

89. Barker. Phenol-Chloroform Isoamyl Alcohol (PCI) DNA Extraction. 1998.

90. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology. 2002; 3: research0034. https://doi.org/10.1186/gb-2002-3-7-research0034 PMID: 12184808

91. Reyon D, Khayter C, Regan MR, Joung JK, Sander JD. Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. Curr Protoc Mol Biol. 2012;Chapter 12: Unit 12.15. https://doi.org/10.1002/0471142727.mb1215s100 PMID: 23026907

92. Cade L, Reyon D, Hwang WY, Tsai SQ, Patel S, Khayter C, et al. Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. Nucleic Acids Res. 2012; 40: 8001–8010. https://doi.org/10.1093/nar/gks518 PMID: 22684503

93. Huang P, Xiao A, Zhou M, Zhu Z, Lin S, Zhang B. Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol. 2011; 29: 699–700. https://doi.org/10.1038/nbt.1939 PMID: 21822242

94. Szabó T. Ovulation induction in northern pike Esox lucius L. using different GnRH analogues, Ovaprim, Dagin and carp pituitary. Aquaculture Research. 2003; 34: 479–486. https://doi.org/10.1046/j.1365-2109.2003.00835.x

95. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 2018.

96. Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. Genetics. 2011; 188: 799–808. https://doi.org/10.1534/genetics.111.127324 PMID: 21828280

97. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. G3 (Bethesda). 2011; 1: 171–182. https://doi.org/10.1534/g3.111.000240 PMID: 22384329

98. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

99. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015; 4: 7. https://doi.org/10.1186/s13742-015-0047-8 PMID: 25722852

100. Feron R. RADSex. 2019. https://github.com/RomainFeron/RadSex

101. de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T. CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics. 2005; 21: 1703–1704. https://doi.org/10.1093/bioinformatics/bti222 PMID: 15598829

102. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017; 27: 722–736. https://doi.org/10.1101/gr.215087.116 PMID: 28298431

103. Hardie DC, Hebert PD. Genome-size evolution in fishes. Can J Fish Aquat Sci. 2004; 61: 1636–1646. https://doi.org/10.1139/f04-106

104. Li H, Birol I. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018; 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191 PMID: 29750242

105. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE. 2014; 9: e112963. https://doi.org/10.1371/journal.pone.0112963 PMID: 25409509

106. Earl D, Bradnam K, St. John J, Darling A, Lin D, Fass J, et al. Assemblathon 1: A competitive assessment of de novo short read assembly methods. Genome Res. 2011; 21: 2224–2241. https://doi.org/10.1101/gr.126599.111 PMID: 21926179

107. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351 PMID: 26059717

108. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

109. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011; 27: 3435–3436. https://doi.org/10.1093/bioinformatics/btr589 PMID: 22025480

110. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389 PMID: 9254694

111. Smit AFA, Hubley R & Green P. RepeatMasker Open-4.0 [Internet]. 2010. Available from: www.repeatmasker.org.