

# The rise and fall of the ancient northern pike master sex determining gene

Qiaowei Pan<sup>1,2</sup>, Romain Feron<sup>1,2,3</sup>, Elodie Jouanno<sup>1</sup>, Hugo Darras<sup>2</sup>, Amaury Herpin<sup>1</sup>, Ben Koop<sup>4</sup>, Eric Rondeau<sup>4</sup>, Frederick W. Goetz<sup>5</sup>, Wesley A. Larson<sup>6</sup>, Louis Bernatchez<sup>7</sup>, Mike Tringali<sup>8</sup>, Stephen S. Curran<sup>9</sup>, Eric Saillant<sup>10</sup>, Gael P.J. Denys<sup>11,12</sup>, Frank A. von Hippel<sup>13</sup>, Songlin Chen<sup>14</sup>, J. Andrés López<sup>15</sup>, Hugo Verreycken<sup>16</sup>, Konrad Ocalewicz<sup>17</sup>, Rene Guyomard<sup>18</sup>, Camille Eche<sup>19</sup>, Jerome Lluch<sup>19</sup>, Celine Roques<sup>19</sup>, Hongxia Hu<sup>20</sup>, Roger Tabor<sup>21</sup>, Patrick DeHaan<sup>21</sup>, Krista M. Nichols<sup>22</sup>, Laurent Journot<sup>23</sup>, Hugues Parrinello<sup>23</sup>, Christophe Klopp<sup>24</sup>, Elena A. Interesova<sup>25</sup>, Vladimir Trifonov<sup>26</sup>, Manfred Scharl<sup>27</sup>, John Postlethwait<sup>28</sup>, Yann Guiguen<sup>1&</sup>.

&: Corresponding author.

1. INRAE, LPGP, 35000, Rennes, France.
2. Department of Ecology and Evolution, University of Lausanne, 1015, Lausanne, Switzerland.
3. Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.
4. Department of Biology, Centre for Biomedical Research, University of Victoria, Victoria, BC, V8W 3N5, Canada.
5. Environmental and Fisheries Sciences Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA, United States of America.
6. Fisheries Aquatic Science and Technology Laboratory at Alaska Pacific University. 4101 University Dr, Anchorage, AK 99508.
7. Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada, G1V 0A6.
8. Fish and Wildlife Conservation Commission, Florida Marine Research Institute, St. Petersburg, Florida 33701, USA.
9. School of Fisheries and Aquatic Sciences, Auburn University.
10. Gulf Coast Research Laboratory, School of Ocean Science and Technology, The University of Southern Mississippi, Ocean Springs, MS, USA.
11. Laboratoire de Biologie des organismes et écosystèmes aquatiques (BOREA), MNHN, CNRS, IR D, SU, UCN, UA, 57 rue Cuvier CP26, 75005 Paris, France.
12. Unité Mixte de Service Patrimoine Naturelle – Centre d'expertise et de données (UMS 2006 AFB, CNRS, MNHN), Muséum national d'Histoire naturelle, 36 rue Geoffroy Saint-Hilaire CP 41, 75005 Paris, France.
13. Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA.
14. Yellow Sea Fisheries Research Institute, CAFS, Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266071, China.
15. College of Fisheries and Ocean Sciences Fisheries, Fairbanks, AK 99775, USA.
16. Research Institute for Nature and Forest (INBO), Havenlaan 88 box 73, 1000, Brussels, Belgium.
17. Department of Marine Biology and Ecology, Institute of Oceanography, University of Gdansk, Poland.
18. INRAE, GABI, 78350, Jouy-en-Josas, France.
19. US 1426, GeT-PlaGe, INRAE, Genotoul, Castanet-Tolosan, France.
20. Beijing Fisheries Research Institute & Beijing Key Laboratory of Fishery Biotechnology, Beijing, China.
21. U.S. Fish and Wildlife Service, Lacey, Washington, USA.
22. Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA.

23. Institut de Génomique Fonctionnelle, IGF, CNRS, INSERM, Univ. Montpellier, F-34094 Montpellier, France.
24. INRAE, Sigeneae, Genotoul Bioinfo, Toulouse, France.
25. Tomsk State University, Tomsk, Russia.
26. Institute of Molecular and Cellular Biology, Siberian Branch of the Russian Academy of Sciences, Novosibirsk State University, Novosibirsk, Russia.
27. University of Würzburg, Developmental Biochemistry, Biocenter, 97074 Würzburg, Germany; and The *Xiphophorus* Genetic Stock Center, Texas State University, San Marcos, TX 78666 USA.
28. Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA.

## Abstract

Sexual reproduction is a ubiquitous basic feature of life and genetic sex determination is thus widespread, at least among eukaryotes. Understanding the remarkable diversity of sex determination mechanisms, however, is limited by the paucity of empirical studies. Here, we traced back the evolution of sex determination in an entire clade of vertebrates and uncovered that the northern pike (*Esox lucius*) master sex-determining gene initiated from a 65 to 90 million-year-old gene duplication and remained sex-linked on undifferentiated sex chromosomes for at least 56 million years. Contrasting with its ancient origin, we identified several independent species- or population-specific transitions of sex determination mechanisms in this lineage, including an unexpected complete and recent Y-chromosome loss in some North American northern pike populations. These findings highlight the diversity of the evolutionary fates of master sex-determining genes and raise the importance of careful considerations of population demographic history in sex determination studies. Our study also puts forward the hypothesis that occasional sex reversals and genetic bottlenecks provide a non-adaptive explanation for sex determination transitions.

## Introduction

Genetic sex determination (GSD) evolved independently and repeatedly in diverse taxa, including animals, plants and fungi<sup>1,2</sup>, but the stability of such systems varies drastically among groups. In mammals and birds, the same sex determination (SD) system has been maintained over a long evolutionary time with conserved master sex determining (MSD) genes on sex chromosomes<sup>3–6</sup>. In stark contrast, teleost fish display both GSD and environmental sex determination (ESD)<sup>7,8</sup>, and a remarkable evolutionary lability, driven by dynamic and rapid

sex chromosome and MSD gene turnovers<sup>9–12</sup>. These characteristics have made teleosts an attractive model group to study the process of the evolution of sex chromosomes and the turnover of SD systems.

In the past two decades, fueled by advances in genomics tools, a large diversity of MSD genes have been discovered in teleosts, advancing new hypotheses on how the birth of MSD genes, either by allelic diversification or duplication / translocation<sup>9–12</sup>, can drive sex chromosome turnover in vertebrates. Teleost MSD genes also provided empirical support for the “limited option” hypothesis which states that certain genes known to be implicated in the sex differentiation pathways are more likely to be recruited as new MSD genes<sup>9–11,13</sup>. The majority of these recently discovered MSD genes, however, were found in many diverse teleost clades, making it challenging to infer evolutionary patterns and conserved themes in sex chromosomes turnover. Comparative studies on closely related species with SD transitions can thus provide valuable insight on the causes and processes of this turnover<sup>9</sup>. These comparative studies have been done in medakas (*Oryzias sp.*), poeciliids, tilapiine cichlids, salmonids, and sticklebacks, in which different SD systems, sex chromosomes and MSD genes were found even in closely related species<sup>14–18</sup>. Apart from the *Oryzias* family<sup>19–29</sup> and the salmonids<sup>18,30–32</sup>, however, relatively few studies have explored the evolution of SD systems and the fate of MSD genes within an entire group of closely related species.

Esociformes are a relatively small but old monophyletic teleost order (**Figure 1**) that diverged from a common ancestor around 90 million years ago (Mya) and from their sister clade Salmoniformes about 110 Mya<sup>33,34</sup>. With two families, i.e., Esocidae and Umbridae, and 13 well-recognized species<sup>35</sup>, Esociformes are an ecologically important group of freshwater species from the northern hemisphere<sup>33,34</sup>. Following the identification of a male-specific duplication of the anti-Müllerian hormone, *amhby*, as the MSD gene in northern pike (*Esox lucius*)<sup>36</sup>, we took advantage of the small number of Esociformes species and their relatively long evolutionary history to explore the evolution of sex determination in this entire lineage.

Based on whole genome sequencing and population genomic data (**Figure 1**) we traced back the evolutionary trajectory of this MSD gene from its birth in a gene duplication event 65 to 90 Mya to a few independent species- or population-specific SD transitions. Among these SD transitions, we strikingly uncovered even within one species, northern pike, that some populations have experienced a complete and recent loss of their Y-chromosome. Our results highlight the diversity of the evolutionary fates that a single MSD gene can experience in a complete and old monophyletic group of vertebrates. We also show how drift, exacerbated by

bottleneck effect, facilitated the loss of a sex chromosome along with its MSD gene in wild populations, providing a first mechanism for such an intriguing entire sex chromosome loss in vertebrates.

## Results

### The *Esox lucius amhby* gene originated in an ancient duplication event

Previously, we identified *amhby*, a male-specific duplicate of the anti-Müllerian hormone gene (*amh*) as the MSD gene in northern Pike (*Esox lucius*). To explore the evolution of *amha* (the autosomal copy of *amh*) and *amhby* in the genus *Esox*, we collected phenotypically sexed samples of most species of Esociformes (**Figure 1 & Table 1**). Homologs of *amh* were searched by a combination of homology cloning and whole genome sequencing (**Supplementary note 1**) and two *amh* genes were found in all surveyed species with the exceptions of the basally diverging species *Dallia pectoralis* and *Umbra pygmaea* in which only one *amh* gene was found in their whole genome assemblies (**Supplementary note 2**) and in tissue-specific transcriptome databases<sup>1</sup>. To clarify relationships among these *amh* homologs, phylogenetic trees were inferred from these sequences and they provided a clear and consistent topology (**Figure 2A, Figure S1, Supplementary note 3**) indicating an *amh* duplication (*amha* and *amhby* paralogs) that occurred in the common ancestor of the Esocidae<sup>2</sup>. In agreement with that hypothesis, the single *amh* ortholog of *U. pygmaea*, which is the closest sister species to the Esocidae, roots at the base of the *amha* / *amhby* clade. In contrast, the single *amh* ortholog of *D. pectoralis* was systematically placed within the *amha* cluster, strongly suggesting that this species experienced a secondary loss of its *amhby* gene after the duplication of *amh* genes in the Esocidae lineage (**Figure 2B & S1, Supplementary note 3**).

### Sex-linkage of *amhby* in the Esocidae lineage

To explore whether *amhby* is the MSD gene in all these Esocidae species, we first checked if *amhby* is genetically strongly linked to male phenotype, because male sex-linkage would be a strong indication that this gene is located in the non-recombining region of a Y chromosome (**Supplementary note 4**). Results showed complete sex-linkage in five of the six *Esox* species with the only exception being *E. masquinongy* (**Table 1**), in which significant but incomplete sex-linkage was found in two North American populations from Iowa and Wisconsin, but not

in a population from Quebec, Canada (**Table 1**). In addition, heterozygous state of the *amhby* gene was detected in two males and one female of *E. masquinongy* in the population from Quebec (**Figure S2**), conflicting with the expected hemizygous status of a single *amhby* gene on a single Y sex locus. In *Novumbra hubbsi*, the only member of Esocidae not belonging to the *Esox* genus but with an *amhby* gene, we also found a significant but incomplete sex-linkage (**Table 1**).

### Evolution of the structure of the *amhby* gene in the Esocidae

The typical *amh* gene in teleosts comprises seven exons encoding a protein that contains 500 to 571 amino acids depending on the species, with a C-terminal TGF- $\beta$  domain that is crucial for canonical Amh function (di Clemente et al. 2010; Pfennig et al., 2015). The predicted structures of most of the *amha* and *amhby* genes in Esociformes are consistent with this canonical structure, ruling out an origin by retrotransposition. Furthermore, most *amhby* and *amha* genes do not show any signature of relaxation from purifying selection (**Supplementary note 5**). In both *E. niger* and *N. hubbsi*, however, the predicted Amhby protein is truncated in its N- or C-terminal part. In *E. niger*, this truncated *amhby* gene is flanked by repeated elements and encompasses only three of the seven conserved Amh exons, i.e., exons 5, 6 and an exon 7 with a truncated TGF- $\beta$  domain (**Supplementary note 6**). In *N. hubbsi*, the truncated *amhby* gene contains the seven conserved Amh exons but with a N-terminal truncation of exon 1 encoding only 8 amino acids with no homology to the conserved 50 amino acid sequences of the first Amh exon of other Esocidae (**Supplementary note 6**). Together, these results show that even in some species having a tight *amhby* sex-linkage, the Amhby deduced protein sequence is strongly modified in its N- and C-terminal conserved domains encoding respectively the TGF- $\beta$  signal peptide needed for a proper protein secretion and the conserved cysteines of the TGF- $\beta$  domain required to provide the correct conformation for proper interaction between Amh and its receptor<sup>3</sup> (**Figure S3**).

### Whole genome analyses of the evolution of sex-determination systems in Esociformes

To complement *amhby* sex-linkage analyses, we also searched for whole-genome sex-specific signatures in species with either incomplete *amhby* sex linkage (*E. masquinongy* and *N. hubbsi*) or species without *amhby* (*D. pectoralis* and *U. pygmaea*). For species with incomplete linkage of sex to *amhby*, we produced gDNA Pool-sequencing (Pool-Seq) of *amhby* positive males

versus *amhby* negative females in the Iowa population of *E. masquinongy* and gDNA RAD-sequencing (RAD-Seq) of phenotypic males versus phenotypic females in *N. hubbsi*. Results from these analyses (**Supplementary notes 7 and 8**) provided for these two species a low level of male-specific signal that indicates a small sex locus region with a male heterogametic SD system (XX/XY) (**Figure 3A & Figure S5**). For species without *amhby*, RAD-Seq of phenotypic males versus phenotypic females was carried out in both *D. pectoralis* and *U. pygmaea* (**Supplementary note 7**). In *D. pectoralis*, only three female-biased RAD markers were identified, suggesting that this species has a small sex locus region under a female heterogametic SD system (ZZ/ZW) (**Figure 3B**). This ZZ/ZW SD system was further supported by a substantially higher number of female-specific sequences in Pool-Seq analysis (**Supplementary note 7**). In contrast, a large number ( $N = 137$ ) of male-biased RAD markers were identified in *U. pygmaea*, that does not have the *amhby* gene, supporting the hypothesis that this species has a large sex locus region under a male heterogametic SD system (XX/XY) (**Figure 3C**).

## Some populations of *Esox lucius* lost their Y chromosome and ancestral master sex determining gene

Although *amhby* was demonstrated to be the MSD gene in European populations of northern pike<sup>4</sup>, this gene was surprisingly absent from the high quality genome assembly of a male specimen from a Canadian population (GCA\_000721915.3)<sup>5</sup>. To better explore this discrepancy, we characterized the *amhby* sex linkage in several geographically isolated populations of *E. lucius* and found significant male sex linkage in all investigated populations from China, Europe and Alaska (**Table 1, Supplementary note 9**). In contrast, *amhby* was found to be completely absent, both in males and females, from all Canadian and mainland USA populations (**Table 1**).

To investigate whether the loss of *amhby* coincides with large genomic changes, we first compared phenotypic males and phenotypic females from a European population carrying the *amhby* gene (Ille-et-Vilaine, France) with a North American population that lost its *amhby* gene (Quebec, Canada) using a Pool-Seq approach. These Pool-Seq results were aligned to an improved European *E. lucius* genome assembly (NCBI accession number SDAW000000000) in which all previously identified Y-specific contigs<sup>4</sup> were scaffolded into a single contiguous sex locus on the Y (LG24) sex chromosome (**Table S1**). In the European population of *E. lucius*,



Pool-Seq analysis (**Supplementary note 9**) confirmed, but with a much better resolution, previous results<sup>4</sup>, showing that this European *E. lucius* population harbors a rather small (140 kb) non-recombining sex locus at the proximal end of LG24 (**Figure 4A.1 & 4B.1**). In the Canadian population, virtually no reads from either the male or female pools mapped to this 140 kb Y-specific region identified in the European population (**Figure 4A.2 & 4B.2**) and no sign of differentiation between males and females could be observed along the rest of LG24. In the Quebec population, not only LG24 but the entire rest of the genome had no sex-specific heterozygosity for either male or female reads with either the Pool-Seq analysis or a reference-free RAD-Seq approach (**Figure S5, Supplementary note 9**). Together, these results suggest that this Canadian population, and likely other mainland USA populations lack the male-specific Y region found in European populations, including the MSD gene *amhby*. Moreover, these North American populations did not acquire a sufficiently large new sex locus that could be detected by the RAD-Seq and Pool-Seq approaches, or their current SD relies on a non-genetic mechanism.

## **Evolving sex determination systems in the muskellunge, *Esox masquinongy***

As shown above, variable *amhby* sex-linkages were observed among different populations of *E. masquinongy* (**Table 1**). We then searched genome-wide for sex-specific signals using RAD-Seq in a population from Iowa (US) in which *amhby* was incompletely but significantly associated with male phenotype and in a population from Quebec (Canada) where *amhby* association was not significant. In the Iowa population, five RAD markers were significantly associated with male phenotype, while no marker was associated with female phenotype (**Figure 3D**); this result indicates a male heterogametic sex determination system (XX/XY) with a low differentiation between the X and Y chromosomes as observed in the northern pike<sup>4</sup> and *N. hubbsi* (this study). In contrast, no sex-specific marker was identified in some Canadian populations (**Figure 3E, Figure S7, Supplementary note 7**).

# Discussion

Previous results on the *amhby* MSD gene of northern pike<sup>4</sup> showed that its sequence was substantially divergent from its autosomal counterpart *amha*, in contrast to other known cases of *amh* duplication in fishes<sup>6,7</sup>, suggesting that this MSD gene may have an ancient origin and / or may have undergone an accelerated evolution. Here, we show that the duplication that produced *amhby* occurred before the last common ancestor of Esocidae in all species of *Esox* and *Novumbra*, but was subsequently lost in the *Dallia* lineage. Esocidae diverged from the Umbridae between 65 to 90 Mya<sup>8</sup>, hence the high level of differentiation observed between *amha* and *amhby* is likely due to the ancient origin of the duplication event. We also demonstrate that *amhby* is completely sex-linked in the majority of Esocidae species in which this gene was retained with low levels of sex chromosome differentiation at least in some populations of *E. masquinongy* and *N. hubbsi*. Taken together, these results suggest that *amhby* likely acquired an MSD function, either 56 Mya in the last common ancestor of *Esox* and *Novumbra*, or 66 Mya in the common ancestor of all Esocidae followed by a subsequent loss in *D. pectoralis*.

This SD system has been maintained for more than 56 My without driving the evolution of highly differentiated sex chromosomes, similar to what was previously described in *E. lucius*<sup>4</sup>. The identification of an old MSD gene located on a homomorphic sex chromosome with limited differentiation over a long evolutionary time is thus in stark contrast with classic models of sex chromosome evolution that postulate the gradual decay of sex chromosomes<sup>9</sup>. Previous work on cichlid fishes and true frogs suggested that frequent turnovers could keep sex chromosomes undifferentiated<sup>10,11</sup>. It was proposed for these organisms that autosomes that are already hosting genes involved in sex differentiation pathways could be recruited as new sex chromosomes, facilitating turn overs of sex chromosomes and MSD genes<sup>12</sup>. Alternative evolutionary pathways could also involve conserved MSD genes that translocate onto different autosomes as has been found for the salmonid *sdY* gene<sup>13,14</sup>. In addition, frequent sex reversal of the heterogametic sex has been proposed to ‘rejuvenate’ sex chromosomes and keep them homomorphic<sup>15</sup>. Our results, based on whole genome analysis in *E. masquinongy*, suggest that the same small Y chromosome sex locus has been conserved in most Esocidae. Such a stability is unusual in teleosts, where frequent turnover of SD systems is considered the norm<sup>16–19</sup>. In line with our present results, however, substantially undifferentiated sex chromosomes have also been found to be maintained over relatively long evolutionary periods in some Takifugu



fish species<sup>20</sup>, supporting the idea that theoretical models cannot be generalized and that additional and as yet unknown evolutionary forces can prevent important sex chromosome decay.

Even though *amhby* is an ancient and conserved MSD gene, this old MSD gene experienced different fates across the Esocidae lineage. One of these unexpected fates is the evolution of the Amhby protein in *N. hubbsi*, and *E. niger*, which lack parts of the conserved C- and /or N-terminal Amh regions<sup>21</sup>. In these two species, *amhby* was significantly and almost completely sex-linked, suggesting that this gene could still function as a MSD gene. Because the conserved Amh domains that are altered in the *amhby* genes in these two species are needed for proper protein secretion and interaction between Amh and its receptor<sup>3</sup>, it is unclear how these truncated Amhby proteins could still serve as functional MSD proteins. A similar duplication and truncation of an *amh* MSD gene has already been reported in the cobaltcap silverside (*Hypoatherina tsurugae*), another teleost<sup>22</sup>. In addition, other cases of MSD genes having evolved through duplication / truncation of their ancestral copy have also been described, such as the *sdY* MSD gene in Salmonids and the putative *amhr2by* MSD gene in yellow perch, *Perca flavescens*, demonstrating that preservation of all ancestral domains is not always necessary for a duplicated protein to assume an MSD role<sup>13,23,24</sup>. Because domain gains and losses can both contribute to new protein functions<sup>25,26</sup>, the evolution of a new MSD protein structure can even be seen as a strong evolutionary driver required for the emergence and fixation of some new MSD genes.

Another striking evolutionary fate of this ancient *amhby* SD system in Esocidae relies on the several independent SD transitions that we characterized in this lineage. The first is a complete transition in *D. pectoralis* from an XX/XY male heterogametic SD system, which is shared by Esocidae and *U. pygmaea*, to a ZZ/ZW female heterogametic SD system. Interestingly, this SD transition was also accompanied by, or shortly followed by, a secondary loss of *amhby* because our phylogenetic analyses consistently support the timing of the *amhby* duplication in the common ancestor of the Esocidae i.e., before the divergence of the *D. pectoralis* lineage. While all other Esociformes uniformly display a clear XX/XY SD system, this ZZ/ZW SD system in *D. pectoralis*, could suggest that the shift from male to female heterogamety was potentially driven by the loss of the *amhby* MSD gene, because this kind of SD transition has been documented in other species, for example, in the housefly<sup>27</sup>. The other SD transitions that we characterized, those in *E. lucius* and *E. masquinongy*, are probably more recent because they appear within populations of the same species and not between different

species. In *E. masquinongy*, evidence from whole genome analysis and *amhby* sex-linkage both support its conservation as a MSD gene along with a clear XX/XY SD system in two USA populations. In contrast, in a Canadian population from Quebec, our whole genome analysis did not support a male heterogametic SD system. In addition, despite being present in this population, *amhby* was no longer sex-linked and a few males and females even displayed allelic variations in the *amhby* gene, which is not compatible with the expected hemizygous status of a Y sex locus gene. This finding could suggest that the *amhby* locus in this population is not in the non-recombining sex region and thus is no longer an MSD gene. Because we did not find *amhby* sequence differences among populations of *E. masquinongy*, this result would suggest that this SD transition is linked to the recent emergence of a new population-specific SD mechanism. The fact that SD was suggested to be female heterogametic in some populations of *E. masquinongy*<sup>28</sup>, and that the great lakes population and the Quebec population belong to different lineages<sup>29</sup>, further support the hypothesis of population-specific variation in *E. masquinongy* SD. But an interesting alternative hypothesis would be that this SD system relies on a simple monofactorial XX/XY system in some *E. masquinongy* populations and on a polygenic system with multiple sex chromosomes in other populations of the same species. This hypothesis would equally fit with our current results because it would explain both the absence of complete sex-linkage and the heterozygosity of *amhby* in some populations. This explanation would also fit with the suggested female heterogamety of some populations, which was inferred from the presence of males in the offspring of gynogenetic females<sup>28</sup>, a result that could equally be the indication of a ZZ/ZW monofactorial system or of a polygenic system with multiple sex chromosomes. Such polygenic SD systems, including systems with multiple sex chromosomes<sup>30,31</sup>, are complicated to characterize but they have been well described in a limited number of species<sup>30</sup> and even in some populations of the same species such as the Poeciliids *Xiphophorus nigrensis* and *X. multilineatus*<sup>32</sup>.

In *E. lucius*, we found that *amhby* is present and completely sex-linked in all populations we surveyed from Europe, Alaska and Asia, suggesting that this gene functions as an MSD gene in all of these populations. The entire Y sex locus including *amhby*, however, was absent in some North American populations, suggesting that these populations have lost their ancestral Y sex chromosome and that their remaining X chromosomes reverted to autosomes. The current geographic distribution of *E. lucius* results from a post-glacial expansion from three glacial refugia ~ 0.18 to 0.26 Mya, and the North American populations without *amhby* belong to a monophyletic lineage with a circumpolar distribution originating from the same Eurasian

origin<sup>33</sup>. Fossil records from Alaska support the idea of an “out of Alaska” North-American expansion of *E. lucius* within the last 100,000 years<sup>34</sup>, suggesting that the *amhby* sex locus could have been lost during this dispersal period (**Figure 4C**). The complete loss of this sex locus in such a short evolutionary time is unlikely to result from the slow pseudogenization of the MSD gene as was found for instance in the Luzon medaka, *Oryzias luzonensis*<sup>35</sup>. Rather, it is probably the result of the loss of the entire Y chromosome in founder individuals. This founder hypothesis is supported by the fact that North American populations of *E. lucius* likely experienced a strong bottleneck because they display a much lower genetic diversity compared to other circumpolar lineage populations<sup>33</sup>. The loss of the Y chromosome could then stem from founder males being genetic XX females that experienced sex reversal; spontaneously masculinized XX fish, likely due to environmental effects, are occasionally observed in captive *E. lucius*<sup>4</sup>. Environmental influence on GSD is a well documented phenomenon found in many fish species, such as Nile tilapia<sup>36</sup>, Half-smooth tongue sole<sup>37</sup> and Japanese rice fish<sup>38</sup>. These environmental effects on SD may result in the persistence of populations which would otherwise go extinct following the loss of the ancient MSD gene. Notably, this Y chromosome loss in natural populations of *E. lucius* mirrors the loss of the W chromosome in lab strains of zebrafish after just a few decades of selective breeding<sup>39</sup>, with the current SD mechanism in laboratory stocks hypothesized to be polygenic<sup>40</sup>, environmental<sup>41</sup> or even ‘random’<sup>42</sup>, having evolved from the wild, natural ZW strains. Whether ESD, which is common among poikilothermic animals including teleosts<sup>43</sup>, could be the only SD system in these North American populations or whether a new MSD gene has taken the lead of a new GSD system is still unresolved, but our whole-genome analyses failed to identify any sex-associated markers in these North American populations, meaning that if such a new sex locus exists, it would lack detectable signatures of molecular differentiation like what has been found for instance in some *Takifugu* species<sup>20</sup>.

A few evolutionary models have been formulated to capture the dynamics of sex chromosome turnover<sup>12</sup>. The replacement of the ancestral SD locus by a new one could occur via neutral processes driven by drift alone, or via positive selection when the new MSD gene confers a fitness advantage<sup>44</sup>, or when this new MSD gene is associated with strong sexually antagonistic alleles<sup>45</sup>. Moreover, the accumulation of deleterious mutations on the non-recombining sex locus of the sex chromosome could facilitate sex chromosome turnover<sup>46</sup>. In all these models, the ancestral sex chromosomes would revert to autosomes only when the new SD locus is fixed in the population. In North American populations of *E. lucius*, the ancestral

sex chromosome was likely lost due to sudden drift in bottlenecked populations. Interestingly, this process would not necessarily need the simultaneous emergence of a new GSD system, given the flexibility of SD mechanisms in teleosts, in which environmental cues can compensate at least temporarily for the absence of a strict GSD system. ESD, which is known to be easily invaded by new MSD<sup>47,48</sup>, could then serve as a transitional state in between sex chromosome turnovers.

Collectively, our results depict the evolutionary trajectories of a conserved MSD gene in an entire clade of vertebrates, highlighting both the potential stability of MSD genes as well as non-differentiated sex chromosomes in some lineages, and also the dynamics of species- or population-specific independent SD evolution events in teleost fish. Our results also point to the importance of careful consideration of the population demographic history of SD systems, and of taking into account the potential buffering role of ESD during transitions of SD systems in models of sex chromosome evolution.

# Methods

## Sample collection

Information on the different Esociform species, including species collectors, sexing method and experiments performed is given in **Table S2**.

## Genomic DNA extraction

Fin clips were collected and stored at 4°C in 75-100% ethanol until genomic DNA (gDNA) extraction. For genotyping, samples 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<sup>49</sup>. For Illumina sequencing, gDNA was obtained using NucleoSpin Kits for Tissue (Macherey-Nagel, Duren, Germany) following the producer's protocol. DNA concentration was quantified using Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA) and a Qubit3 fluorometer (Invitrogen, Carlsbad, CA). For Pool-Seq, DNA from different samples was normalized to the same quantity before pooling for

male and female libraries separately. High molecular weight gDNA for long read sequencing was extracted as described by Pan et al. (2019)<sup>4</sup>.

## Genome and population genomics sequencing

Draft genomes of northern pike (*Esox lucius*), muskellunge (*Esox masquinongy*), chain pickerel (*Esox niger*), Olympic mudminnow (*Novumbra hubbsi*), and Alaska blackfish (*Dallia pectoralis*) were sequenced using a whole genome shotgun strategy with 2x250 bp Illumina reads. Libraries were built using the Truseq nano kit (Illumina, ref. FC-121-4001) following instructions from the manufacturer. 200 ng of gDNA was briefly sonicated using a Bioruptor (Diagenode). The gDNA was end-repaired and size-selected on beads to retain fragments of size around 550 bp, and these fragments were a-tailed and ligated to Illumina's adapter. The ligated gDNA was then subjected to 8 PCR cycles. Libraries were checked on a Fragment Analyzer (AATI) and quantified by qPCR using a library quantification kit from KAPA. Libraries were sequenced on a HiSeq2500 using a paired end 2x250 bp v2 rapid mode according to the manufacturer's instructions. Image analysis was performed by HiSeq Control Software and base calling by the RTA software provided by Illumina. The output of each run can be found in **Table S3**. For improving our European male genome of *E. lucius*, we generated an extra coverage of Oxford Nanopore long reads using a higher fragment size (50 kb) library made from gDNA extracted from a different male from the same European population as the sample used for previous genome assembly. Library construction and genome sequencing was carried out as previously described<sup>4</sup> and 12.7 Gbp of new data were generated from one PromethION flowcell.

Pool-Seq was performed on the North American population of *E. lucius*, on *E. masquinongy*, and on *D. pectoralis*. Pooled libraries were constructed using a Truseq nano kit (Illumina, ref. FC-121-4001) following the manufacturer's instructions. Two sex-specific DNA Pool-Seq libraries were prepared for each species using the Illumina TruSeq Nano DNA HT Library Prep Kit (Illumina, San Diego, CA) with the same protocol as for the draft genome sequencing. The libraries were then sequenced on a NovaSeq S4 lane (Illumina, San Diego, CA) using paired-end 2x150 bp mode with Illumina NovaSeq Reagent Kits following the manufacturer's instructions. The output of each run for each sex can be found in **Table S3**.

RAD-Seq was performed on the North American population of *E. lucius*, on *E. masquinongy*, on *N. hubbsi*, on *D. pectoralis* and *U. pygmaea*. RAD libraries were constructed from gDNA extracted from fin clips for each species using a single *SbfI* restriction enzyme as

previously described<sup>50</sup>. Each library was sequenced on one lane of Illumina HiSeq 2500. The summary of the output of each dataset can be found in **Table S4**.

## Analysis of population genomic data for sex-specific signals

**RAD-seq:** Raw reads were demultiplexed with the *process\_radtags.pl* script from *stacks* version 1.44<sup>51,52</sup> with all parameters set to default. Demultiplexed reads for each species were analyzed with the RADSex computational workflow using the *radsex* software version 1.1.2 (10.5281/zenodo.3775206). After generating a table of markers depth with *process*, the distribution of markers between sexes was computed with *distrib* and markers significantly associated with phenotypic sex were extracted with *signif* using a minimum depth of 10 (-d 10) for both commands and all other settings to default. All figures for RAD-Seq results were generated with *sgtr* (10.5281/zenodo.3773063)<sup>78</sup>.

**Pool-Seq:** For each dataset, Pool-Seq reads were aligned to the reference genome using BWA mem (version 0.7.17)<sup>53</sup>. Alignment results were sorted by genomic coordinates using samtools sort (version 1.10)<sup>54</sup> and PCR duplicates were removed using samtools rmdup. A file containing nucleotide counts for each genomic position was generated with the *pileup* command from PSASS (version 3.0.1b, 10.5281/zenodo.3702337). This file was used as input to compute  $F_{ST}$  between males and females, number of male- and female-specific SNPs, and male and female depth in a sliding window along the entire genome using the *analyze* command from PSASS with parameters --window-size 50000, --output-resolution: 1000, --freq-het 0.5, --range-het 0.15, --freq-hom 1, --range-hom 0.05, --min-depth 1, and --group-snps. The entire analysis was performed with a snakemake workflow available at <https://github.com/SexGenomicsToolkit/PSASS-workflow>. All figures for Pool-Seq results were generated with *sgtr* version 1.1.1 (10.5281/zenodo.3773063).

## Genome assembly

Raw Illumina sequencing reads for *Dallia pectoralis*, *Esox masquinongy*, *Esox niger*, and *Novumbra hubbsi* were assembled using the DISCOVAR *de novo* software<sup>55</sup> with the following assembly parameters: MAX\_MEM\_GB=256, MEMORY\_CHECK=False, and NUM\_THREADS=16. Assembly metrics were calculated with the *assemblathon\_stats.pl*



script<sup>56</sup> and assembly completeness was assessed with BUSCO (version 3.0.2)<sup>57</sup> using the *Actinopterygii* gene set, and both results are given in **Table S5**.

To facilitate the comparison of sex loci in different populations of *E. lucius*, we improved the continuity of the previously published assembly of an *E. lucius* European male individual (GenBank assembly accession: GCA\_007844535.1). This updated assembly was performed with Flye (version 2.6)<sup>58</sup>, using standard parameters and --genome-size set to 1.1g to match theoretical expectations<sup>59</sup>. Two rounds of polishing were performed with Racon (version 1.4.10)<sup>60</sup> using default settings and the Nanopore reads aligned to the assembly with minimap2 (version 2.17)<sup>61</sup> with the “map-ont” preset. Then, three additional rounds of polishing were performed with Pilon (version 1.23)<sup>62</sup> using the “--fix all” setting and the Illumina reads aligned to the assembly with BWA mem (version 0.7.17)<sup>53</sup>. Metrics for the resulting assembly were calculated with the *assemblathon\_stats.pl* script<sup>56</sup>. The assembly’s completeness was assessed with BUSCO (version 3.0.2)<sup>57</sup> using the *Actinopterygii* gene set (4,584 genes) and the default gene model for Augustus. The resulting assembly was scaffolded with RaGOO (version 1.1)<sup>63</sup> using the published female genome anchored to chromosomes (Genbank accession: GCA\_011004845.1) as reference.

## Sequencing of *amha* and *amhby* genes

For species closely related to *E. lucius* (*E. aquitanicus*, *E. cisalpinus* and *E. reichertii*), the sequence of *amh* homologs was amplified and sequenced with primers designed on the sequences of *E. lucius* (**Table S6**). To search for *amh* homologs in the more divergent species (*E. niger*, *E. masquinongy*, *N. hubbsi*, *D. pectoralis* and *U. pygmaea*), we generated draft genome assemblies from phenotypic males. Lastly, for *E. americanus americanus* and *E. americanus vermiculatus*, ortholog sequences were amplified with primers designed on regions that were found to be conserved in the other *Esox* species. Briefly, we first tried to amplify the complete ortholog sequences of *amha* and *amhby*, if any, from different species of Esociformes using primers designed on the exons of *amha* and *amhby* sequences of *E. lucius*. These two sequences share 78% identity, and PCR primers were chosen in divergent sequences ensuring a specific amplification of each ortholog. When the amplification was successful, the PCR products were purified using a NucleoSpin Gel and a PCR Clean-up Kit (Macherey-Nagel, Duren, Germany) and then sequenced directly by Sanger sequencing using the same PCR primers. A second round of amplification was performed using species-specific primers and primers designed on conserved regions of the *amha* and *amhby* sequences amplified in the first

round. All primers used in this study were designed using Perl-Primer<sup>64</sup> (Marshall 2004) and are listed in **Table S6**. The amplicons for each species were assembled with the CAP3 program<sup>65</sup>. Details of each approach are given in **Supplementary note 1**.

## Gene phylogeny analysis

Phylogenetic reconstructions were performed on all *amh* homolog sequences obtained from Esociformes with the *Salmo salar amh* used as an outgroup. Full-length CDS were predicted with the FGENESH+<sup>66</sup> suite based on the genomic sequence and Amh protein sequence from *Esox lucius*. Sequence alignments were performed with MAFFT (version 7.450)<sup>67</sup>. Both maximum likelihood and Bayesian methods were used for tree construction with IQ-TREE (version 1.6.7)<sup>68</sup> and Phylobayes (version 4.1)<sup>69–71</sup>, respectively. Details for the methods can be found in **Supplementary note 3**. The resulting phylogenetic trees were visualized with Figtree (version 1.4.1).

## Data availability

Whole Genome Shotgun genome assemblies are in the process of being deposited at DDBJ/ENA/GenBank. All gene sequences, genomic, Pool-seq and RAD-Seq reads were deposited under the common project number PRJNA634624.

## Acknowledgments

We are grateful to the fish facility of INRAE LPGP for support in experimental installation and fish maintenance and to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing help, computing and storage resources. This project was supported by funds from the Agence Nationale de la Recherche, the Deutsche Forschungsgemeinschaft (ANR/DFG, PhyloSex project, 2014-2016, SCHA 408/10-1, to Y.G and M.S), and the National Institutes of Health (USA) grant R01GM085318, to J.H.P). The MGX and Get-Plage core sequencing facilities were supported by France Genomique National infrastructure, funded as part of “Investissement d’avenir” program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09). We would like to thank MNHN curator for giving access to the collections of *Esox aquitanicus* and *Esox cisalpinus* as well as for the ONEMA and G.B.

Delmastro who aided the specimen sampling. We would also like to thank Mackenzie Garvey and Penny Swanson for their assistance with dissection of *Novumbra hubbsi*.

## Author contributions

Y.G., J.H.P., M.S. obtained the funding for the project. Y.G., J.H.P., M.S., and Q.P. designed the experiments. R.F. and C.K. assembled the genomes. R.F. and Q.P. conducted analyses on RAD-Seq and Pool-Seq data. Q.P. performed genotyping and obtained all gene sequences and performed the phylogenetic analyses with the input from H.D. Q.P. and E.J. performed DNA extraction. C.R., C.E. and J.L. conducted library preparation and sequencing of Nanopore long reads and Pool-Seq reads. L.J. and H.P. conducted library preparation and sequencing of RAD-Seq. Q.P., R.F., H.D., and Y.G. made the figures. J.H.P., M.S., B.K., E.R., F.W.G., W.A.L., L.B., M.T., S.S.C., E.S., G.P.J.D., F.v.H., S.C., E.A.I., V.T., H.V., J.A.L., K.O., H.H., R.G., R.T., P.D., and K.M.N. performed sample collection and determination of phenotypic sex. Q.P. wrote the manuscript with inputs from Y.G., R.F., H.D., M.S. and J.H.P. All co-authors approved the manuscript.

## Competing interests

The authors declare no competing interests.

## References

1. Pasquier, J. *et al.* Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics* **17**, 368 (2016).
2. Leaché, A. D., Harris, R. B., Rannala, B. & Yang, Z. The Influence of Gene Flow on Species Tree Estimation: A Simulation Study. *Syst Biol* **63**, 17–30 (2014).
3. di Clemente, N. *et al.* Processing of Anti-Müllerian Hormone Regulates Receptor Activation by a Mechanism Distinct from TGF- $\beta$ . *Molecular Endocrinology* **24**, 2193–2206 (2010).
4. Pan, Q. *et al.* Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics* **15**, e1008013 (2019).
5. Rondeau, E. B. *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* **9**, e102089 (2014).
6. Hattori, R. S. *et al.* A Y-linked anti-Mullerian hormone duplication takes over a critical role

- in sex determination. *Proceedings of the National Academy of Sciences* **109**, 2955–2959 (2012).
7. Li, M. *et al.* A Tandem Duplicate of Anti-Müllerian Hormone with a Missense SNP on the Y Chromosome Is Essential for Male Sex Determination in Nile Tilapia, *Oreochromis niloticus*. *PLOS Genetics* **11**, e1005678 (2015).
  8. Campbell, M. A., López, J. A., Sado, T. & Miya, M. Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. *Gene* **530**, 57–65 (2013).
  9. Vicoso, B. & Charlesworth, B. Evolution on the X chromosome: unusual patterns and processes. *Nat. Rev. Genet.* **7**, 645–653 (2006).
  10. Gammerdinger, W. J. *et al.* Novel Sex Chromosomes in 3 Cichlid Fishes from Lake Tanganyika. *J. Hered.* **109**, 489–500 (2018).
  11. Jeffries, D. L. *et al.* A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nat Commun* **9**, 4088 (2018).
  12. Vicoso, B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nature Ecology & Evolution* **3**, 1632–1641 (2019).
  13. Yano, A. *et al.* An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Current Biology* **22**, 1423–1428 (2012).
  14. Lubieniecki, K. P. *et al.* Genomic Instability of the Sex-Determining Locus in Atlantic Salmon (*Salmo salar*). *G3: Genes, Genomes, Genetics* **5**, 2513–2522 (2015).
  15. Perrin, N. SEX REVERSAL: A FOUNTAIN OF YOUTH FOR SEX CHROMOSOMES? *Evolution* **63**, 3043–3049 (2009).
  16. Kikuchi, K. & Hamaguchi, S. Novel sex-determining genes in fish and sex chromosome evolution: Novel Sex-Determining Genes in Fish. *Developmental Dynamics* **242**, 339–353 (2013).
  17. Heule, C., Salzburger, W. & Böhne, A. Genetics of Sexual Development: An Evolutionary Playground for Fish. *Genetics* **196**, 579–591 (2014).
  18. Herpin, A. & Scharf, M. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO reports* **16**, 1260–1274 (2015).
  19. Pan, Q., Guiguen, Y. & Herpin, A. Evolution of sex determining genes in Fish. in *Encyclopedia of Reproduction, Second Edition* (2017).
  20. Kamiya, T. *et al.* A Trans-Species Missense SNP in Amhr2 Is Associated with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics* **8**, e1002798 (2012).
  21. Rocha, A., Zanuy, S. & Gómez, A. Conserved Anti-Müllerian Hormone: Anti-Müllerian Hormone Type-2 Receptor Specific Interaction and Intracellular Signaling in Teleosts1. *Biology of Reproduction* **94**, (2016).
  22. Bej, D. K., Miyoshi, K., Hattori, R. S., Strüssmann, C. A. & Yamamoto, Y. A Duplicated, Truncated amh Gene Is Involved in Male Sex Determination in an Old World Silverside. *G3: Genes, Genomes, Genetics* g3.117.042697 (2017) doi:10.1534/g3.117.042697.
  23. Bertho, S. *et al.* The unusual rainbow trout sex determination gene hijacked the canonical vertebrate gonadal differentiation pathway. *PNAS* **115**, 12781–12786 (2018).
  24. Feron, R. *et al.* Characterization of a Y-specific duplication/insertion of the anti-Müllerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Molecular Ecology Resources* **20**, 531–543 (2020).
  25. Moore, A. D. & Bornberg-Bauer, E. The Dynamics and Evolutionary Potential of Domain Loss and Emergence. *Mol Biol Evol* **29**, 787–796 (2012).
  26. Moore, A. D., Grath, S., Schüler, A., Huylmans, A. K. & Bornberg-Bauer, E. Quantification and functional analysis of modular protein evolution in a dense phylogenetic tree. *Biochim.*

- Biophys. Acta* **1834**, 898–907 (2013).
27. Kozielska, M., Feldmeyer, B., Pen, I., Weissing, F. J. & Beukeboom, L. W. Are autosomal sex-determining factors of the housefly (*Musca domestica*) spreading north? *Genet Res (Camb)* **90**, 157–165 (2008).
  28. Dabrowski, K., Rinchar, J., Lin, F., Garcia-Abiado, M. A. & Schmidt, D. Induction of gynogenesis in muskellunge with irradiated sperm of yellow perch proves diploid muskellunge male homogamety. *J. Exp. Zool.* **287**, 96–105 (2000).
  29. Turnquist, K. N. *et al.* Genetic structure of muskellunge in the Great Lakes region and the effects of supplementation on genetic integrity of wild populations. *Journal of Great Lakes Research* **43**, 1141–1152 (2017).
  30. Moore, E. C. & Roberts, R. B. Polygenic sex determination. *Current Biology* **23**, R510–R512 (2013).
  31. Roberts, N. B. *et al.* Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* **17**, (2016).
  32. Kallman, K. D. A New Look at Sex Determination in Poeciliid Fishes. in *Evolutionary Genetics of Fishes* (ed. Turner, B. J.) 95–171 (Springer US, 1984). doi:10.1007/978-1-4684-4652-4\_3.
  33. Skog, A., Vøllestad, L. A., Stenseth, N. C., Kasumyan, A. & Jakobsen, K. S. Circumpolar phylogeography of the northern pike (*Esox lucius*) and its relationship to the Amur pike (*E. reichertii*). *Frontiers in zoology* **11**, 67 (2014).
  34. Wooller, M. J., Gaglioti, B., Fulton, T. L., Lopez, A. & Shapiro, B. Post-glacial dispersal patterns of Northern pike inferred from an 8800 year old pike (*Esox cf. lucius*) skull from interior Alaska. *Quaternary Science Reviews* **120**, 118–125 (2015).
  35. Myosho, T. *et al.* Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, *Oryzias luzonensis*. *Genetics* **191**, 163–170 (2012).
  36. Baroiller, J. F., D’Cotta, H., Bezault, E., Wessels, S. & Hoerstgen-Schwark, G. Tilapia sex determination: Where temperature and genetics meet. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* **153**, 30–38 (2009).
  37. Chen, L., Ding, X. P., Wei, X. & Li, L. X. Investigation of mutations in the SRY, SOX9, and DAX1 genes in sex reversal patients from the Sichuan region of China. *Genet. Mol. Res.* **13**, 1518–1526 (2014).
  38. Sato, T., Endo, T., Yamahira, K., Hamaguchi, S. & Sakaizumi, M. Induction of Female-to-Male Sex Reversal by High Temperature Treatment in Medaka, *Oryzias latipes*. *jzoo* **22**, 985–988 (2005).
  39. Wilson, C. A. *et al.* Wild Sex in Zebrafish: Loss of the Natural Sex Determinant in Domesticated Strains. *Genetics* **198**, 1291–1308 (2014).
  40. Liew, W. C. & Orbán, L. Zebrafish sex: a complicated affair. *Brief Funct Genomics* **13**, 172–187 (2014).
  41. Santos, D., Luzio, A. & Coimbra, A. M. Zebrafish sex differentiation and gonad development: A review on the impact of environmental factors. *Aquat. Toxicol.* **191**, 141–163 (2017).
  42. Perrin, N. Random sex determination: When developmental noise tips the sex balance. *BioEssays* **38**, 1218–1226 (2016).
  43. Navara, K. J. Mechanisms of Environmental Sex Determination in Fish, Amphibians, and Reptiles. in *Choosing Sexes: Mechanisms and Adaptive Patterns of Sex Allocation in Vertebrates* (ed. Navara, K. J.) 213–240 (Springer International Publishing, 2018). doi:10.1007/978-3-319-71271-0\_10.
  44. Bull, J. J. & Charnov, E. L. Changes in the heterogametic mechanism of sex determination. *Heredity (Edinb)* **39**, 1–14 (1977).
  45. van Doorn, G. S. & Kirkpatrick, M. Turnover of sex chromosomes induced by sexual

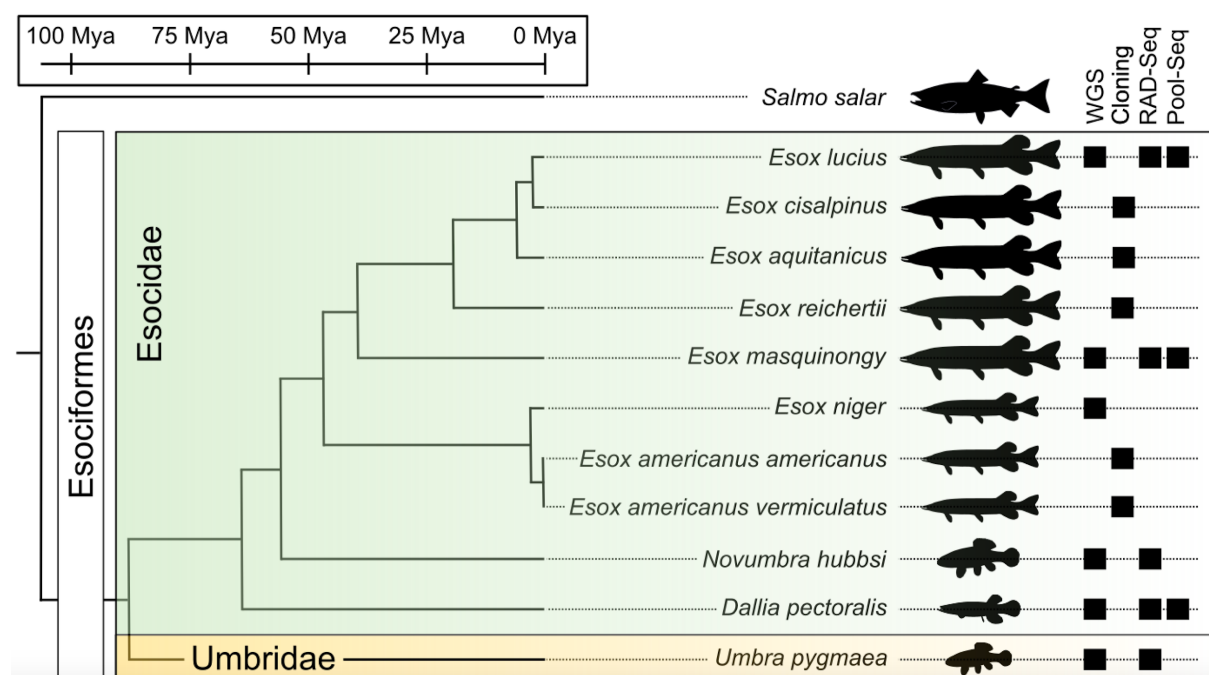


- conflict. *Nature* **449**, 909–912 (2007).
46. SEX-CHROMOSOME TURNOVERS INDUCED BY DELETERIOUS MUTATION LOAD - Blaser - 2013 - Evolution - Wiley Online Library. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1558-5646.2012.01810.x>.
47. van Doorn, G. S. Evolutionary transitions between sex-determining mechanisms: a review of theory. *Sex Dev* **8**, 7–19 (2014).
48. Muralidhar, P. & Veller, C. Sexual antagonism and the instability of environmental sex determination. *Nature Ecology & Evolution* **2**, 343–351 (2018).
49. Gharbi, K. *et al.* A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics* **172**, 2405–2419 (2006).
50. Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. & Postlethwait, J. H. Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. *Genetics* **188**, 799–808 (2011).
51. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3 (Bethesda)* **1**, 171–182 (2011).
52. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124–3140 (2013).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
54. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
55. Weisenfeld, N. I. *et al.* Comprehensive variation discovery in single human genomes. *Nat Genet* **46**, 1350–1355 (2014).
56. Earl, D. *et al.* Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Res* **21**, 2224–2241 (2011).
57. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
58. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540–546 (2019).
59. Hardie, D. C. & Hebert, P. D. Genome-size evolution in fishes. *Can. J. Fish. Aquat. Sci.* **61**, 1636–1646 (2004).
60. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
61. Li, H. & Birol, I. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
62. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963 (2014).
63. Alonge, M. *et al.* RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology* **20**, 224 (2019).
64. Marshall, O. J. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* **20**, 2471–2472 (2004).
65. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).
66. Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* **7**, S10 (2006).
67. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**, 772–780 (2013).

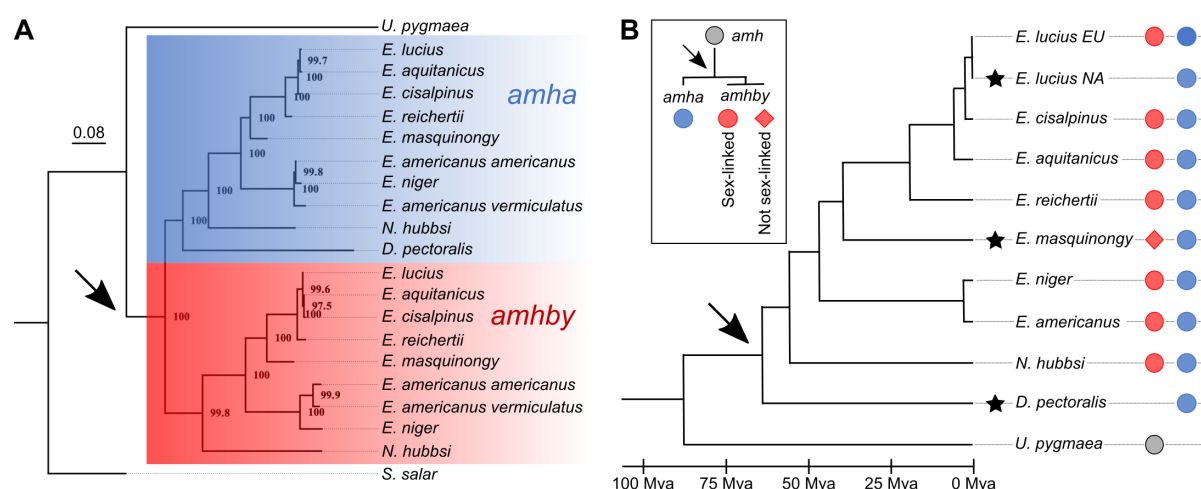


68. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).
69. Lartillot, N. & Philippe, H. A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Mol Biol Evol* **21**, 1095–1109 (2004).
70. Lartillot, N. & Philippe, H. Computing Bayes factors using thermodynamic integration. *Syst. Biol.* **55**, 195–207 (2006).
71. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evolutionary Biology* **7**, S4 (2007).
72. López, J. A., Chen, W.-J. & Ortí, G. Esociform phylogeny. *Copeia* **2004**, 449–464 (2004).
73. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* **34**, 1812–1819 (2017).

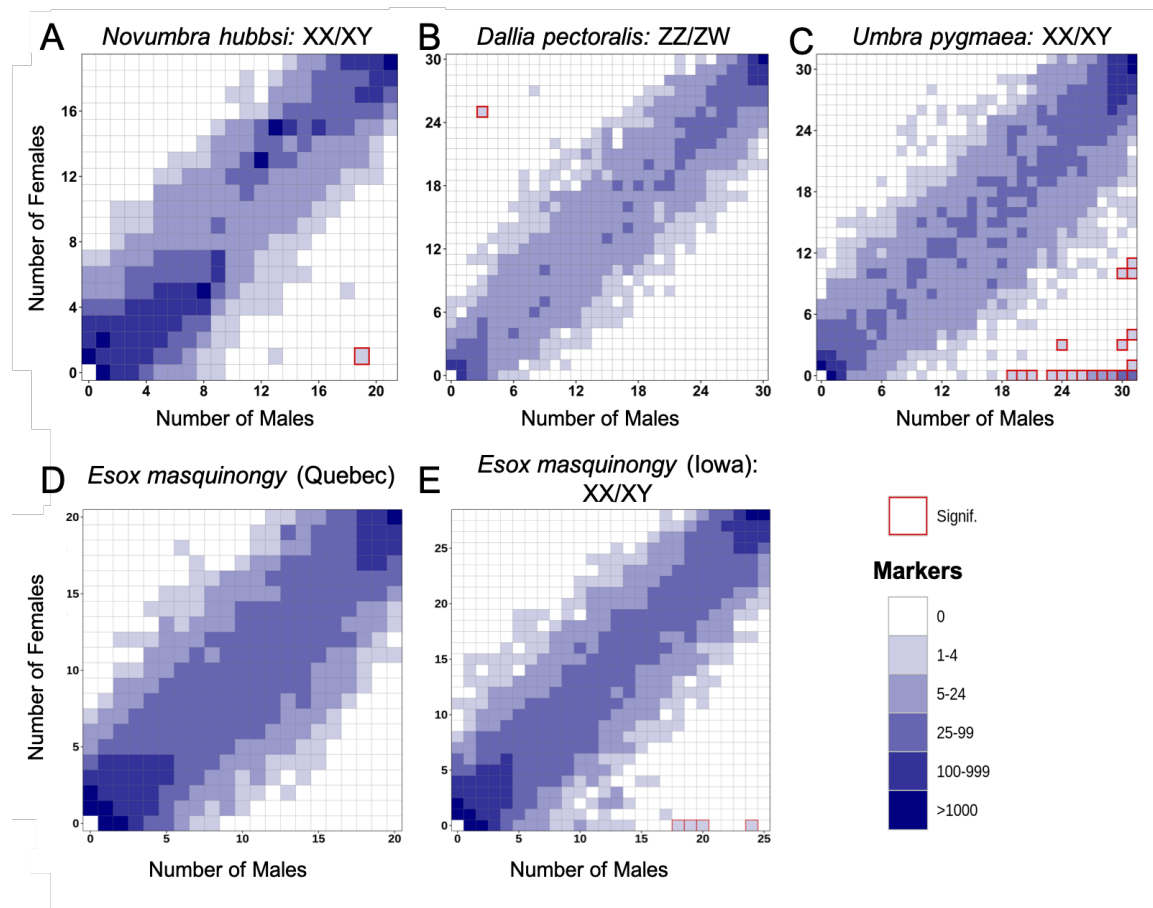
# Figures



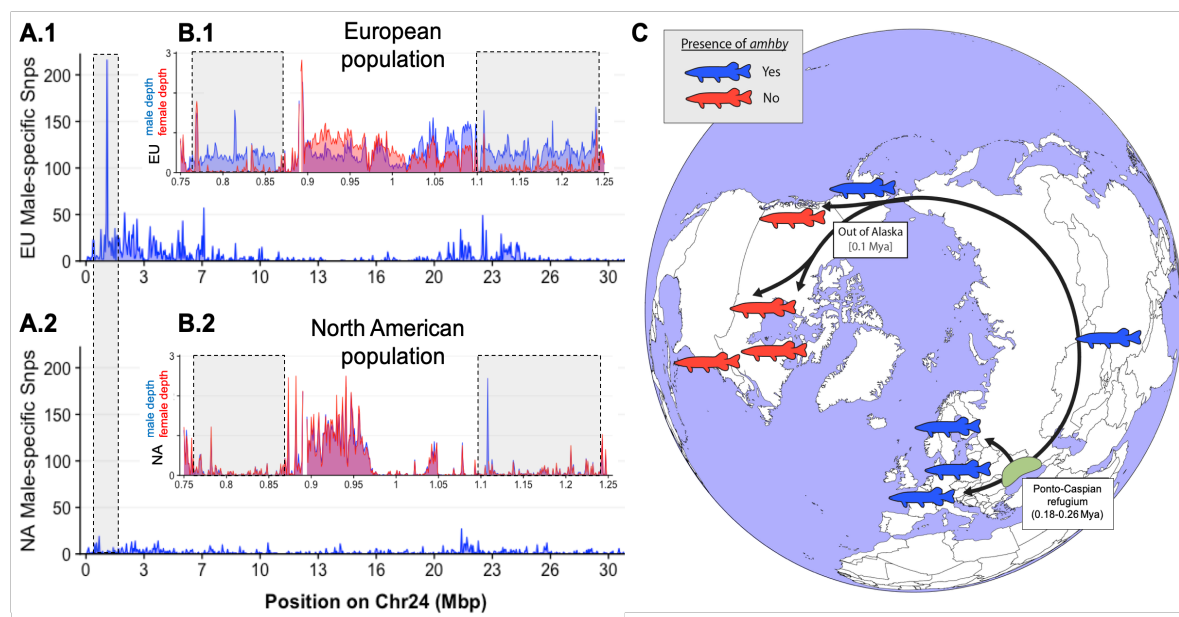
**Figure 1: Species phylogeny and technical approaches for investigating sex determination systems in Esociformes.** Phylogenetic relationships and estimated divergence time<sup>8,72,73</sup>. The two families within Esociformes, the Esocidae and the Umbridae, are highlighted with green and yellow background, respectively. Whole genome sequencing (WGS), homology cloning (Cloning), Restriction-site associated DNA sequencing (RAD-Seq) and Pooled-Sequencing (Pool-Seq) technical approaches used in each species are shown by a black square on the right side of the figure. Fish silhouettes were obtained from [phylopic.org](http://phylopic.org)<sup>100</sup>.



**Figure 2: Phylogenetic analysis of *amh* homologs from the Esociformes revealed an ancient origin of *amhby*.** **A:** Phylogeny tree of *amh* coding sequences built with the maximum likelihood method. Bootstrap values are given on each node of the tree. The *amha* ortholog cluster is highlighted with blue background and the *amhby* ortholog cluster is highlighted with red background. **B:** Species phylogeny of the Esociformes<sup>8</sup>. The three putative sex determination transitions are shown by a black star. The presence of pre-duplication *amh*, *amha*, and *amhby* along with its sex-linkage is represented by colored dots at the end of each branch. The earliest duplication timing of *amh* is denoted by a black arrow at the root of the Esocidae lineage.



**Figure 3: Characterization of sex determination systems in different Esociformes species through RAD-Seq analyses.** Each tile plot shows the distribution of non-polymorphic RAD-Seq markers shared between phenotypic males (horizontal axis) and phenotypic females (vertical axis). The intensity of color for a tile reflects the number of markers present in the respective number of males and females. Tiles that are significantly associated with phenotypic sex (Chi-squared test,  $p < 0.05$  after Bonferroni correction) are highlighted with a red border. **A:** In *N. hubbsi*, two markers (indicated by the shade corresponding to 1-4 markers) were present in 19 males but only one female, indicating significant male association; reciprocally, no marker was present in many females but few males. This result indicates an XX/XY sex determination system with a small sex locus. **B:** In *D. pectoralis*, one marker showed a significant association with females while no marker showed association with males, indicating a ZZ/ZW sex determination system. **C:** In *U. pygmaea*, 140 markers showed a significant association with males, while no marker showed association with females, indicating a XX/XY sex determination system with a large non-recombining sex locus. **D:** In the population of *E. masquinongy* from Quebec (Canada), no marker was found to be associated with either sex. **E:** In a population of *E. masquinongy* from Iowa (USA), analysis identified five markers significantly associated with male phenotype, indicating a XX/XY sex determining system.



**Figure 4: Complete loss of the ancestral sex locus and master sex determination gene in North American Northern pike (*Esox lucius*).** **A:** The numbers of male- and female-specific SNPs in 50 kb non-overlapping windows deduced from Pool-Seq data of a European population (A.1) and North American population (A.2) of *E. lucius* are plotted along the Y chromosome (LG24) of the sequence of *E. lucius* from a European male genome (SDAW00000000). Male data are represented in blue and female in red. While the highest number of male specific SNPs is found in a single window located at the proximal end of LG24 (0.95 Mb to 1.0 Mb) that contained 182 male-specific SNPs (highlighted by the gray box) in the European *E. lucius* population, the same region showed no differentiation between males and females in the North American population. **B:** Relative coverage of male and female Pool-Seq reads is indicated by blue and red lines, respectively, in contigs containing *amhby* in the European population (B.1) and in the North American population (B.2). While in the European population, this region is only covered by male reads at a coverage depth close to 0.5 genome average, it is covered by no male nor female reads in the North American population, indicating the loss of the entire Y-specific region, highlighted by the gray boxes. **C:** Schematic representation of one hypothesized route of post-glacial *E. lucius* expansion from a Eurasian refugia ~ 0.18 to 0.26 Mya and the presence (red silhouettes) / absence (blue silhouettes) of *amhby* in different populations, showing that this master sex determining gene was lost in the North American population during this out-of-Alaska dispersal around 0.1 Mya. The hypothesized refugium in the Ponto Caspian region is indicated with a yellow highlight. An alternative route of post-glacial dispersal is shown in **Figure S6**.

# Tables

Species	Sampling location	Males with <i>amhby</i>	Females with <i>amhby</i>	p-value of association with sex
<i>Esox lucius</i>	France	161/161 (100%)	60/0 (0%)	<b>&lt; 2.2E-16</b>
<i>Esox lucius</i>	Sweden	20/20 (100%)	0/20(0%)	<b>&lt; 2.2E-16</b>
<i>Esox lucius</i>	Poland	20/20 (100%)	0/20(0%)	<b>&lt; 2.2E-16</b>
<i>Esox lucius</i>	Xinjiang, China	5/5 (100%)	0/5 (0%)	<b>0.0114</b>
<i>Esox lucius</i>	Continental USA and Canada	0/88 (0%)	0/74 (0%)	<i>amhby absent</i>
<i>Esox lucius</i>	Alaska, USA	17/19 (89%)	0/19 (0%)	<b>1.79E-07</b>
<i>Esox aquitanicus</i>	France	1/1 (100%)	0/2 (0%)	0.665
<i>Esox cisalpinus</i>	Italy	2/2 (100%)	0/2 (0%)	0.317
<i>Esox reichertii</i>	China	11/11 (100%)	0/10 (0%)	<b>3.40E-05</b>
<i>Esox masquinongy</i>	Iowa, USA	18/27 (66.7%)	4/18 (22%)	<b>2.01E-08</b>
<i>Esox masquinongy</i>	Quebec, Canada	9/20 (45%)	4/20 (20%)	0.18
<i>Esox masquinongy</i>	Wisconsin, USA	57/61 (93%)	22/61 (36%)	<b>1.17E -10</b>
<i>Esox americanus (americanus)</i>	Mississippi, USA	6/6 (100%)	0/4 (0%)	<b>0.0123</b>
<i>Esox niger</i>	Quebec, Canada	5/5 (100%)	0/8 (0%)	<b>0.00253</b>
<i>Novumbra hubbsi</i>	Washington, USA	21/23 (91%)	0/22 (0%)	<b>5.28E-09</b>
<i>Dallia pectoralis</i>	Alaska, USA	0/30 (0%)	0/30 (0%)	<i>amhby absent</i>
<i>Umbra pygmaea</i>	Belgium	0/31 (0%)	0/31 (0%)	<i>amhby absent</i>

**Table 1:** Summary of the identification of *amhby* and their association with sex phenotypes in 11 Esociforme species, including six different populations of *E. lucius*, and three different populations of *E. masquinongy*.