



Utiliser ou ne pas utiliser un gène de détermination du sexe : évolution des systèmes de détermination du sexe chez les Esociformes

Qiaowei Pan

► To cite this version:

Qiaowei Pan. Utiliser ou ne pas utiliser un gène de détermination du sexe : évolution des systèmes de détermination du sexe chez les Esociformes. Biologie animale. Agrocampus Ouest, 2017. Français. NNT : 2017NSARB305 . tel-01814706

HAL Id: tel-01814706

<https://theses.hal.science/tel-01814706>

Submitted on 13 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AGRO CAMPUS OUEST

Qiaowei PAN • 12 décembre 2017

Thèse AGROCAMPUS OUEST
sous le label de l'Université Bretagne Loire
pour obtenir le grade de
DOCTEUR D'AGROCAMPUS OUEST
Spécialité Écologie et Évolution

ÉCOLE DOCTORALE • Écologie, Géosciences, Agronomie et
Alimentation (EGAAL)

LABORATOIRE D'ACCUEIL • INRA Laboratoire de Physiologie
et Génomique des Poissons (LPGP)

**Utiliser ou ne pas utiliser un
gène de détermination du sexe :
évolution des systèmes de
détermination du sexe
chez les *Esociformes***

Maria MANZANARES-DAULEUX
Professeur, AGROCAMPUS OUEST, UMR INRA-UR1-AO IGEPP /
présidente

Susana COELHO
Directrice de recherche, CNRS / *rapporteuse*

Eric PAILHOX
Directeur de recherche INRA / *rapporteur*

Nicolas PERRIN
Professeur, Université de Lausanne / *examineur*

Yann GUIGUEN
Directeur de recherche, INRA / *directeur de thèse*

Amaury HERPIN
Chargé de recherche, INRA / *co-encadrant de thèse*



THESE AGROCAMPUS OUEST

Sous le label de l'Université Européenne de Bretagne pour obtenir le diplôme de :

DOCTEUR DE L'INSTITUT SUPERIEUR DES SCIENCES AGRONOMIQUES, AGRO-ALIMENTAIRES, HORTICOLES ET DU PAYSAGE

Spécialité : Biologie et agronomie

Ecole Doctorale : Vie-Agro-Santé

Présentée par :

Qiaowei Pan

**To use or not use a Master sex determining gene:
evolution of sex determination system in the *Esociformes***

Composition du Jury :

Susana Coelho, Directrice de Recherches, Station Biologique de Roscoff : Rapporteuse

Eric Pailhoux, Directeur de Recherches INRA, Jouy en Josas : Rapporteur

Nicolas Perrin, Professeur, Université de Lausanne : Examineur

Maria Manzanares-Dauleux, Professeure, Agrocampus Ouest : Examinatrice

Yann Guiguen, Directeur de Recherches INRA, Rennes : Directeur de thèse

Amaury Herpin, Chargé de Recherches INRA, Rennes : Co-directeur de thèse

Acknowledgements

First of all, I would like to thank the jury members for accepting to review my thesis!

I would like to take this chance to thank my supervisor Dr. Yann Guiguen for his guidance and support during my PhD. I could not have grown so much without his help, from a person who ‘knew nothing’ three years ago to where I am now, proudly knowing something about sex determination in the *Esociformes*.

I would also like to thank my co-supervisor Dr. Amaury Herpin for teaching and helping me, especially with molecular techniques, and sharing the office and the wisdom of French (and German) idioms with me.

I am grateful for the help with experiments I received in the lab, especially from Elodie Jouanno, as well as from Ming Wen, Brigitte Mourot, Alizée Frezel and Amélie Patinote. I am also grateful to Dr. Julien Bobe and Guylene Roudaut for helping me getting a one-month extension of my thesis to finish the work on the bench. I want to thank the other members of INRA-LPGP for their help and support during the past three years.

I could not have done this project without the fishes! I would like to thank all the kind collaborators, especially Professor Manfred Scharl and Professor John Postlethwait, for sending me precious *Esociformes* samples from Asia, Europe and North America.

I want to thank the PhyloSex project and the funding agencies of my PhD for the opportunity to work on such an interesting thesis project.

I would like to express my gratitude to Professor R. Haven Wiley for inspiring my interest in natural history, evolutionary biology and birding, and encouraging me to take my very first research project.

I am grateful to my friends, especially friends here at Rennes and friends I made during MEME, for their encouragement and support.

I want to thank Romain, for his support and help during each and every day of my PhD, for encouraging me, cheering me up, being proud of me and enduring my occasional outburst of stress and frustration.

Lastly, I want to thank my dad for buying, dissecting and carrying Northern pike samples from China to France for me and my mum for always showing interest and finding values in my research topic. I thank them for their love, care and support!

Table of Contents

Glossary and common abbreviations	5
Objectives and structure of this thesis	7
Preface	9
References	14
Section I : Introduction	17
Vertebrate sex-determining genes play musical chairs	19
Introduction	21
A glimpse into sex determination and SD genes evolution in vertebrates with a focus on fish species.....	21
The RAD-Sex approach: a Genome Wide Association Study (GWAS) screening for exploring the diversity of sex determination systems in fish.....	25
Conclusion	29
Acknowledgements	31
References	31
Evolution of Sex determining genes in fish	33
Abstract.....	33
Sex determination and sex chromosomes.....	35
Fish sex determining genes.....	37
DMRT1: a multi-context SD gene.....	39
Sox3, a conserved role in vertebrate sex determination	41
The TGF- β family: A SD genes jungle	43
<i>sdY</i> : an unusual and conserved SD gene in Salmonids.	47
References	53

Section II : Results	55
Identification of a duplicated and translocated master sex determination gene provides insight into homomorphic sex chromosome differentiation in Northern pike (<i>Esox lucius</i>)	59
Authors contributions	61
Abstract.....	63
Introduction	65
Material and Methods	69
Results	89
Discussion.....	103
Conclusion	113
References	113
Supplementary: Identification of a duplicated and translocated master sex determination gene provides insight into homomorphic sex chromosome differentiation in Northern pike (<i>Esox lucius</i>)	121
Rise and fall of an ancient master sex determining gene in the <i>Esociformes</i>	133
Authors contributions	135
Abstract.....	137
Introduction	139
Material and Methods	145
Results	155
Discussion.....	161
Conclusion	171
References	172
Supplementary: Rise and fall of an ancient master sex determining gene in the <i>Esociformes</i> ..	177
Sex determination and karyotype evolution in the Eastern mudminnow, <i>Umbra pygmaea</i>	185
Abstract.....	185
Introduction	187
Material and Methods	189

Results	201
Discussion.....	209
Conclusions and perspectives	213
References	215
Supplementary: Sex determination and karyotype evolution in the Eastern mudminnow, <i>Umbra pygmaea</i>	219
Section IV: Perspectives	223
The importance of AMH in teleost sex determination	225
Evolution of sex determination (SD) systems	229
References	232
Appendix.....	235
Induced spawning of Eastern mudminnow, <i>Umbra pygmaea</i>	235
Abstract.....	235
Introduction	237
Material and Method	237
Results	243
Discussion.....	245
Limitations	247
References	247
Supplementary: Induced spawning of Eastern mudminnow, <i>Umbra pygmaea</i>	249

Glossary and common abbreviations

Amh

Anti-Müllerian hormone

Amhr2

Anti-Müllerian hormone receptor type 2

Allelic diversification

A process that starts from a step-wise diversification of two alleles of a sex chromosome pair. One allele then becomes the sex determiner, while the other allele either retains a non-SD function or favors the development of the opposite sex.

DMRT1

Doublesex and mab-3 related transcription factors

Environmental sex determination (ESD)

When the sex of an individual is triggered by various external factors (i.e. temperature, PH, social interaction).

Gdf6

Growth differentiation factor 6.

Genetic sex determination (GSD)

When the sex of an individual is triggered by its genotype only, including both mono or polygenic sex determination.

Gonochorism

A state of having just one of at least two distinct sexes in any one individual organism.

Gsdf

Gonadal soma derived growth factor

Haploinsufficiency

When a diploid organism has lost one copy of a gene and is left with a single functional copy of that gene and this single copy does not produce enough product to display the wild type's phenotypic characteristics.

Homomorphic/ Heteromorphic sex chromosomes

If the two sex chromosomes in a pair can be distinguished cytologically by their differences in morphology, they are termed heteromorphic sex chromosomes. If the two sex chromosomes are morphologically identical, they are called homomorphic sex chromosomes. The degree of differentiation is usually correlated with the age of the sex chromosome pair.

Irf9

Interferon related factor 9

Master sex determination gene

A gene responsible for triggering the embryo towards either male or female development. It is not necessarily that this gene codes for a protein.

Neofunctionalization

A gene changes its function or acquires a new function by altering the structure of its gene product and/or its expression pattern.

PGC

Primordial germ cells, which are the precursors of the stem cells that will give rise to the germ cell lineage.

Recombination

The process of forming new allelic combination in offspring by exchanges between parental genetic materials.

Sex chromosomes

Chromosomes that usually harbor a master sex determination gene and are involved in the primary genetic sex determination

Sex determination (SD)

Primary mechanism leading to the expression of the phenotypic sex

SdY

Sexual dimorphic on the Y chromosome (Salmonids)

SNP

Single nucleotide polymorphism

SOX

Sry-related high mobility group box

SRY

Sex determining region Y

Therian mammal

Non-egg-laying mammals, including placental mammals and marsupials

TGF- β

Transforming growth factor beta

TE

Transposable element

Objectives and structure of this thesis

The first section of this thesis is a preface describing my focal group, the *Esociformes*, and explaining the reasons for studying sex determination in this group. This preface provides context on the origin of this PhD project.

The introduction section consists of two review papers published during my PhD. The first one focuses on the need for unbiased approaches such as RADSeq (Restriction Associated DNA Sequencing) to explore the diversity of SD systems in the teleosts. The second paper focuses on the evolution of sex determining genes in the teleosts. Together, these two papers provide background on several important elements of my thesis.

The results section is organized around the two main objectives of my thesis. The first objective of this project was to identify and characterize the MSD gene in *Esox lucius*, for which genomic resources were already available and transgenic animals had been successfully obtained in our laboratory. The second objective was to survey the other *Esociformes* species to investigate whether and to what extent this MSD gene is conserved in this lineage, and potentially identify transitions of SD systems in this monophyletic group of teleosts. Both chapters were written in a publication-oriented format, but because we are still waiting for additional data that is being generated, these papers have not yet been submitted to a journal, and they may be subject to small changes prior to submission. Therefore, I took the liberty to write a more extensive discussion of our findings in these two chapters. In addition, I included the author list that will appear in each publication. To help clarifying my own contribution to these papers, I also included a detailed list of the contribution of each author before each of these chapters. An additional chapter is included in this results section, which is titled “*Sex determination and karyotype evolution in the Eastern mudminnow, Umbra pygmaea*”. This chapter is also written in publication format, but we are still missing a key element, namely, a good quality whole genome sequence with enough contiguity to tie the different parts together and as such, to conclude this story. Consequently, this chapter is shorter and less mature than the previous two chapters, and it will be subject to extensive changes before submission to a journal.

The main content of this thesis ends with a conclusion section which contextualizes our results and discusses perspectives on future directions from this thesis project. Lastly, a small report on the reproduction of *U. pygmaea* that has been submitted for review to BMC Scientific reports and is currently under revision is included as an appendix.

Preface

Why study sex determination (SD) in a small non-model fish family like the Esociformes?

The field of vertebrate sex determination has been advancing rapidly in recent years, especially because of advances in sequencing technologies and in functional genomics (gene inactivation using CRISPR) that can be readily applied to non-model species. While initial findings revealed highly conserved genetic sex determination systems in mammals and avians (Fridolfsson *et al.*, 1998; Mank and Ellegren, 2007; Graves, 2008; Potrzebowski *et al.*, 2008), a dynamic evolution involving repeated transitions among different genetic sex determination systems or between genetic and environmental sex determination was found in other vertebrates groups including reptiles (Ezaz *et al.*, 2009; Gamble and Zarkower, 2014), amphibians (Yoshimoto *et al.*, 2008; Miura *et al.*, 2012; Scharl, 2015; Rodrigues *et al.*, 2017), and fishes (Kikuchi and Hamaguchi, 2013; Heule *et al.*, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017). Studies in reptiles, amphibians and fishes also brought to the spotlight many novel genes acting as the master switch for sex determination in vertebrates. The majority of these recently identified master sex determining (MSD) genes come from studies on the teleosts, and they have brought new perspectives on both the evolution of SD systems and sex chromosomes and on the molecular regulation of MSD genes.

Two mechanisms underlying the birth of new MSD genes were identified in the teleosts: allelic diversification, and gene duplication followed by neofunctionalization (Kikuchi and Hamaguchi 2013; Heule *et al.*, 2014; Herpin and Scharl 2015; Pan *et al.*, 2017). Among the currently known or suspected MSD genes in the teleosts, both mechanisms appear to be equally frequent in creating novel MSD genes, while in other taxa, only allelic diversification has been identified. The appearance of new MSD genes through either mechanism can drive the turnover of sex chromosomes; however, whether differences exist in the sex chromosome formation process results from two mechanisms needs to be explored empirically.

In contrast with the sole conserved MSD genes *Sry* on the Y chromosome in therian mammals and *Dmrt1* on the Z chromosome in the avians, a diversity of MSD genes is employed in teleost fishes. However, upon closer examination, all but one of these genes belong to three protein families that are known to be implicated in the sex differentiation pathways: SOX, DMRT and TGF- β (Graves and Peichel, 2010; Kikuchi and Hamaguchi, 2013; Heule, Salzburger, and Böhne, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017).

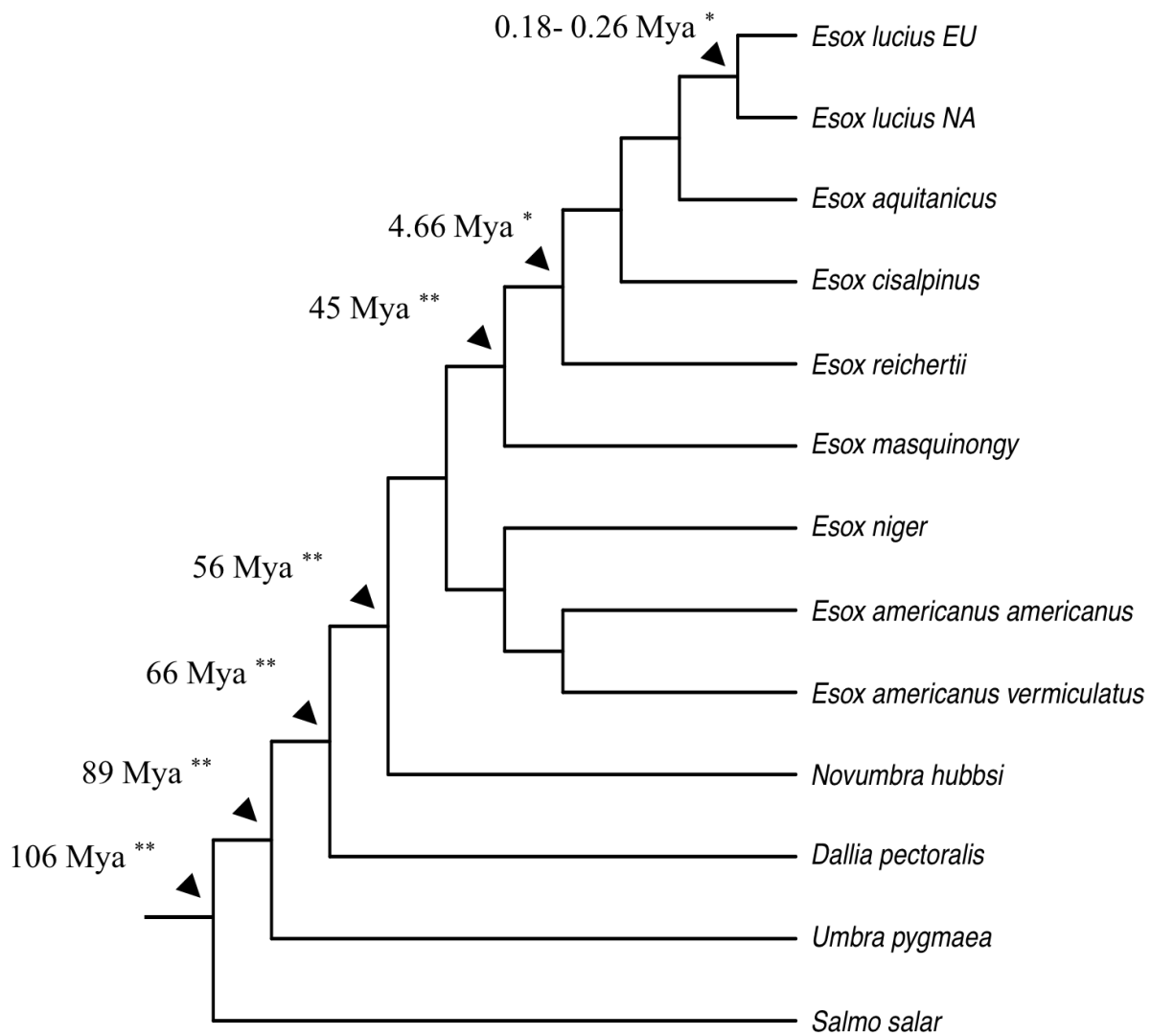


Figure 1: Taxonomy and divergence time estimation of the *Esociformes*, with the Latin name, common name, and an illustration of each species. Taxonomical relationships among the *Esociformes* were retrieved from literature and presented as a cladogram. Estimated divergence times are listed on the left side of the node.

*: Skog *et al.*, 2014

** : Campbell *et al.*, 2013.

The *Sox* and *Dmrt* genes have been identified as MSD genes in other vertebrate groups, but barring the exception of the Platypus, where *Amh* is also suspected to be the MSD gene (Cortez *et al.*, 2014), members of the TGF- β family have only been found to trigger sex determination in teleosts, which suggests a particular importance of this protein family in teleost sex determination and differentiation. In addition, the distribution of MSD genes from these three families does not show a clear phylogenetic pattern among the teleosts and is thus more likely to be the result of repeated and independent recruitments than the result of shared ancestry. This finding prompted the use of the term ‘usual suspects’ (Herpin and Scharl, 2015) to describe the members of these three families in regards to identifying MSD genes. Consequently, the identification of the salmonid MSD gene, named *sdY*, came as a surprise as it originated from the duplication of an immune related gene, previously unsuspected to be involved in sex determination or sex differentiation. The conservation of *sdY* has been confirmed in all salmonids, and to further understand the evolution of such a ‘unusual’ MSD gene and to identify the ancestral MSD gene before the birth of *sdY*, exploring SD in the *Esociformes* was then a logical follow up as this fish order is the sister clade of the salmonids.

The *Esociformes* constitute a small, monophyletic order of teleost fishes, which diverged from its sister clade *Salmoniformes* about 113 million years ago (Mya) (Campbell *et al.*, 2013). The order *Esociformes* contains two families: the *Esocidae*, which regroups the *Esox*, *Dallia*, and *Novumbra* genera, and the *Umbridae* which only contains the *Umbra* genus (**Figure 1**). These two families diverged from a common ancestor about 89 Mya (Campbell *et al.*, 2013).

The most studied species in the *Esociformes* is *Esox lucius*, the Northern pike. This species is a large and long-lived keystone predator found in freshwater and brackish coastal water systems in Europe, North America, and Asia (Forsman *et al.*, 2015). Due to its important role as a top predator in shaping the structure of local fish communities, and as valuable food and sport fish (Raatt, 1988), Northern pike has emerged as an important model species for ecology and fishery. Furthermore, because it is a key species for studying the additional whole genome duplication of the sister group *salmonids* (Leong *et al.*, 2010; Rondeau *et al.*, 2014; Lien *et al.*, 2016), genomic resources including a whole genome assembly anchored on chromosomes (Rondeau *et al.*, 2014) and tissue-specific transcriptomes (Pasquier *et al.*, 2016) are available for this species. Genetic sex determination with a male heterogametic system (XX/XY) was already identified in *E. lucius* sampled in Europe (Luczynski *et al.*, 1997); however, little more is known about the sex determination for this species.

Another outstanding feature of *E. lucius* is that it is the only species of the *Esociformes* with a natural circumpolar distribution. The sister species of *E. lucius* are either restricted to Eurasia (*Esox aquitanicus*, *Esox cisalpinus* and *Esox reichertii*) or to North America (*Esox masquinongy*, *Esox americanus* and *Esox niger*). The current distribution of the extant populations of *E. lucius* is the result of an expansion from three different glacial refugia dated around 0.18 to 0.26 Mya, with the North American populations belonging to the monophyletic circumpolar lineage with a Eurasian origin (Skog *et al.*, 2014). The assembled genome was generated from a male sample from the North American population (Rondeau *et al.*, 2014), but no sex locus could be identified in that study.

In comparison to *Esox sp.*, little genomic resource is available for the other genus of the *Esociformes*. The most closely related genus to *Esox* is *Novumbra*, and the divergence time between these two genera is estimated to be around 56 Mya. There is only one extant species belonging to *Novumbra*, the Olympia mudminnow (*Novumbra hubbsi*). This species is endemic to the West of the Washington state of the United States, and its distribution is limited to the ponds and creeks in this area with geological barriers often found between different populations. The third genus of the *Esocidae* is *Dallia*, which diverged from the rest of the *Esociformes* around 66 Mya. The taxonomy is still debated for this genus, but it is usually considered to include between one and three nominal species of small, cold- and hypoxia-tolerant freshwater fishes restricted to Beringia (Campbell *et al.*, 2015). The *Umbridae* is the second family within the *Esociformes*, and it contains three species within a single extant genus *Umbra*: the European mudminnow (*Umbra krameri*), the Central mudminnow (*Umbra limi*), and the Eastern mudminnow (*Umbra pygmaea*). While the natural habitat of both *U. limi* and *U. pygmaea* is restricted to North America, *U. pygmaea* was introduced in Europe about one century ago (Kuehne and Olden, 2014).

Karyotype evolution is another interesting feature of the *Esociformes*. While limited genomic information was available for these *Esociformes* species prior to our studies, classic cytogenetic studies already revealed surprisingly dynamic karyotype evolution in this group. All *Esox* species have 25 pairs of chromosomes (2N= 50), and *Novumbra hubbsi* has 24 pairs of chromosomes (2N= 48), similar to most teleost species (Beamish *et al.*, 1971, Amores *et al.*, 2017). In contrast, *Dallia pectoralis* has 39 pairs of chromosomes (2N= 78), while *Umbra pygmaea* and *Umbra limi* have only 11 pairs of chromosomes (2N= 22) (Beamish *et al.*, 1971). The mechanisms underlying these drastic chromosomes fusions and fissions are not clear, and whether karyotype evolution could impact the sex chromosome is still an open question. Hence, this feature of the *Esociformes* provides additional incentives to study their genomes and sex determination systems.

References:

- Amores, A., Wilson, C. A., Allard, C. A. H., Detrich, H. W., & Postlethwait, J. H. (2017). Cold Fusion: Massive Karyotype Evolution in the Antarctic Bullhead Notothen Notothenia coriiceps. *G3 (Bethesda, Md.)*, 7(7), 2195–2207. <https://doi.org/10.1534/g3.117.040063>
- Beamish, R. J., Merrilees, M. J., & Crossman, E. J. (1971). Karyotypes and DNA values for members of the suborder Esocoidei (Osteichthyes: Salmoniformes). *Chromosoma*, 34(4), 436–447.
- Campbell, M. A., López, J. A., Sado, T., & Miya, M. (2013). Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. *Gene*, 530(1), 57–65. <https://doi.org/10.1016/j.gene.2013.07.068>
- Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., ... Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature*, 508(7497), 488–493. <https://doi.org/10.1038/nature13151>
- Ezaz, T., Sarre, S. D., O'Meally, D., Graves, J. A. M., & Georges, A. (2009). Sex chromosome evolution in lizards: independent origins and rapid transitions. *Cytogenetic and Genome Research*, 127(2–4), 249–260. <https://doi.org/10.1159/000300507>
- Forsman, A., Tibblin, P., Berggren, H., Nordahl, O., Koch-Schmidt, P., & Larsson, P. (2015). Pike *Esox lucius* as an emerging model organism for studies in ecology and evolutionary biology: a review. *Journal of Fish Biology*, 87(2), 472–479. <https://doi.org/10.1111/jfb.12712>
- Fridolfsson, A.-K., Cheng, H., Copeland, N. G., Jenkins, N. A., Liu, H.-C., Raudsepp, T., ... Ellegren, H. (1998). Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proceedings of the National Academy of Sciences*, 95(14), 8147–8152.
- Gamble, T., & Zarkower, D. (2014). Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Molecular Ecology Resources*, n/a-n/a. <https://doi.org/10.1111/1755-0998.12237>
- Graves, J. A. M. (2008). Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes. *Annual Review of Genetics*, 42(1), 565–586. <https://doi.org/10.1146/annurev.genet.42.110807.091714>
- Graves, J. A. M., & Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biology*, 11(4), 205.
- Herpin, A., & Scharf, M. (2015). Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Reports*, 16(10), 1260–1274. <https://doi.org/10.15252/embr.201540667>
- Heule, C., Salzburger, W., & Böhne, A. (2014). Genetics of Sexual Development: An Evolutionary Playground for Fish. *Genetics*, 196(3), 579–591. <https://doi.org/10.1534/genetics.114.161158>
- Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution: Novel Sex-Determining Genes in Fish. *Developmental Dynamics*, 242(4), 339–353. <https://doi.org/10.1002/dvdy.23927>
- Leong, J. S., Jantzen, S. G., von Schalburg, K. R., Cooper, G. A., Messmer, A. M., Liao, N. Y., ... Koop, B. F. (2010). *Salmo salar* and *Esox lucius* full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. *BMC Genomics*, 11, 279. <https://doi.org/10.1186/1471-2164-11-279>
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., ... Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533(7602), 200–205. <https://doi.org/10.1038/nature17164>
- Luczynski, M. J., Dabrowski, K., Kucharczyk, D., Glogowski, J., Luczynski, M., Szczerbowski, A., & Babiak, I. (2007). Gynogenesis in northern pike: UV-inactivation of spermatozoa and the

- heat shock inducing meiotic diploidization. *Environmental Biotechnology*, Vol. 3, No. 2, 39–43.
- Mank, J. E., & Ellegren, H. (2007). Parallel divergence and degradation of the avian W sex chromosome. *Trends in Ecology & Evolution*, 22(8), 389–391.
<https://doi.org/10.1016/j.tree.2007.05.003>
- Miura, I., Ohtani, H., & Ogata, M. (2012). Independent degeneration of W and Y sex chromosomes in frog *Rana rugosa*. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 20(1), 47–55.
<https://doi.org/10.1007/s10577-011-9258-8>
- Pan, Q., Guiguen, Y., & Herpin, A. (2017). Evolution of sex determining genes in Fish. In *Encyclopedia of Reproduction, Second Edition*.
- Pasquier, J., Cabau, C., Nguyen, T., Jouanno, E., Severac, D., Braasch, I., ... Bobe, J. (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics*, 17, 368. <https://doi.org/10.1186/s12864-016-2709-z>
- Potrzebowski, L., Vinckenbosch, N., Marques, A. C., Chalmel, F., Jégou, B., & Kaessmann, H. (2008). Chromosomal Gene Movements Reflect the Recent Origin and Biology of Therian Sex Chromosomes. *PLOS Biology*, 6(4), e80. <https://doi.org/10.1371/journal.pbio.0060080>
- Rodrigues, N., Studer, T., Dufresnes, C., Ma, W.-J., Veltsos, P., & Perrin, N. (2017). Dmrt1 polymorphism and sex-chromosome differentiation in *Rana temporaria*. *Molecular Ecology*.
<https://doi.org/10.1111/mec.14222>
- Rondeau, E. B., Minkley, D. R., Leong, J. S., Messmer, A. M., Jantzen, J. R., von Schalburg, K. R., ... Koop, B. F. (2014). The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PloS One*, 9(7), e102089. <https://doi.org/10.1371/journal.pone.0102089>
- Schartl, M. (2015). Sex determination by multiple sex chromosomes in *Xenopus tropicalis*. *Proceedings of the National Academy of Sciences*, 112(34), 10575–10576.
<https://doi.org/10.1073/pnas.1513518112>
- Skog, A., Vøllestad, L. A., Stenseth, N. C., Kasumyan, A., & Jakobsen, K. S. (2014). Circumpolar phylogeography of the northern pike (*Esox lucius*) and its relationship to the Amur pike (*E. reichertii*). *Frontiers in Zoology*, 11(1), 67.
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., ... Ito, M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences*, 105(7), 2469–2474.
<https://doi.org/10.1073/pnas.0712244105>

Section I :

Introduction

Vertebrate sex-determining genes play musical chairs

Qiaowei Pan¹, Jennifer Anderson^{1&}, Sylvain Bertho^{1,2}, Amaury Herpin¹, Catherine Wilson³, John H. Postlethwait³, Manfred Scharl^{2,4,5}, Yann Guiguen^{1*}

¹ INRA, Fish Physiology and Genomics laboratory, 35042 Rennes, France

² University of Wuerzburg, Physiological Chemistry, Biocenter, 97074 Wuerzburg, Germany

³ University of Oregon, Institute of Neuroscience, Eugene OR 97403, USA

⁴ Comprehensive Cancer Center Mainfranken, University Hospital, 97080 Wuerzburg, Germany

⁵ Hagler Institute for Advanced Study and Department of Biology, Texas A&M University, College Station, Texas 77843, USA

* Corresponding author: Yann Guiguen, Tel: 33 (0) 2 23 48 50 09, Fax: 33 (0) 2 23 48 50 20, Email: yann.guiguen@rennes.inra.fr

& Present address: Department of Systematic Biology, EBC, Uppsala University, Norbyägen 18D 752 36 Uppsala, Sweden

Abstract: Sexual reproduction is one of the most highly conserved processes in evolution. However, the genetic and cellular mechanisms making the decision of whether the undifferentiated gonad of animal embryos develops either towards male or female are manifold and quite diverse. In vertebrates, sex-determining mechanisms range from environmental to simple or complex genetic mechanisms and different mechanisms have evolved repeatedly and independently. In species with simple genetic sex determination, master sex-determining genes lying on sex chromosomes drive the gonadal differentiation process by switching on a developmental program, which ultimately leads to testicular or ovarian differentiation. So far very few sex-determining genes have been identified in vertebrates and apart from mammals and birds, these genes are apparently not conserved over a larger number of related orders, families, genera or even species. To fill this knowledge gap and to better explore genetic sex determination we propose a strategy (RAD-Sex) that makes use of next generation sequencing technology to identify genetic markers that define sex specific segments of the male or female genome.

Introduction

Sex has been termed the “queen of problems in evolutionary biology”, and how sex is determined in the animal kingdom has always been a major question with “debates on the origin of the sexes that took place long before the beginning of the scientific era” [1]. Indeed, since the beginning of this “scientific era”, important knowledge has been acquired on this topic with, for instance within the quite recent “genetics era”, the important discovery of sex chromosomes followed by the identification of the master gene governing the acquisition of the male sex in human and mouse [2]. Despite huge efforts in the last decades, however, our knowledge of master genes controlling genetic sex determination has remained limited [3]. The relative scarcity of information on sex determination (SD) genes is mainly due to major scientific and technical barriers that hinder the precise identification of these SD genes in many species. Classical approaches for the characterization of SD genes have always relied on the intimate knowledge of a group of specialists working on very few species. Their success has often been guided either by educated guess and candidate gene approaches [4,5], or by tedious searches for sex-linked markers with various time-consuming methods [6–9]. The relatively few species investigated now with regards to SD genes have been selected based on practical interests (for instance species easy to maintain in captivity [8–12]) or economical (for instance important aquaculture species [5,13–15]) and these species always had pre-existing genetic and/or genomic information available. There is then an urgent need to address the question of the evolution of SD genes using a large scale, unbiased approach that would not rely on or need such previous knowledge.

A glimpse into sex determination and SD genes evolution in vertebrates with a focus on fish species

The field of evolution of sex determination and SD genes in vertebrates has long been shaped by knowledge gathered in mammals, because previously this was the only group for which the master SD gene, *SRY*, was known [2]. The generally accepted hypothesis for the mechanism by which an SD gene can evolve, namely by allelic diversification of an autosomal gene towards a male or female sex promoting function, and subsequent maintenance of the male-specifying allele as a male dominant SD gene on the proto-Y chromosome, was satisfied by *SRY* and its supposed progenitor *SOX3* [3]. In 1999, *dmrt1* was found as a candidate gene for SD in chicken, which proposed a different mechanism, namely dosage sensitivity [16]. In 2002, the first fish SD gene was identified in the medaka, *Oryzias latipes*.

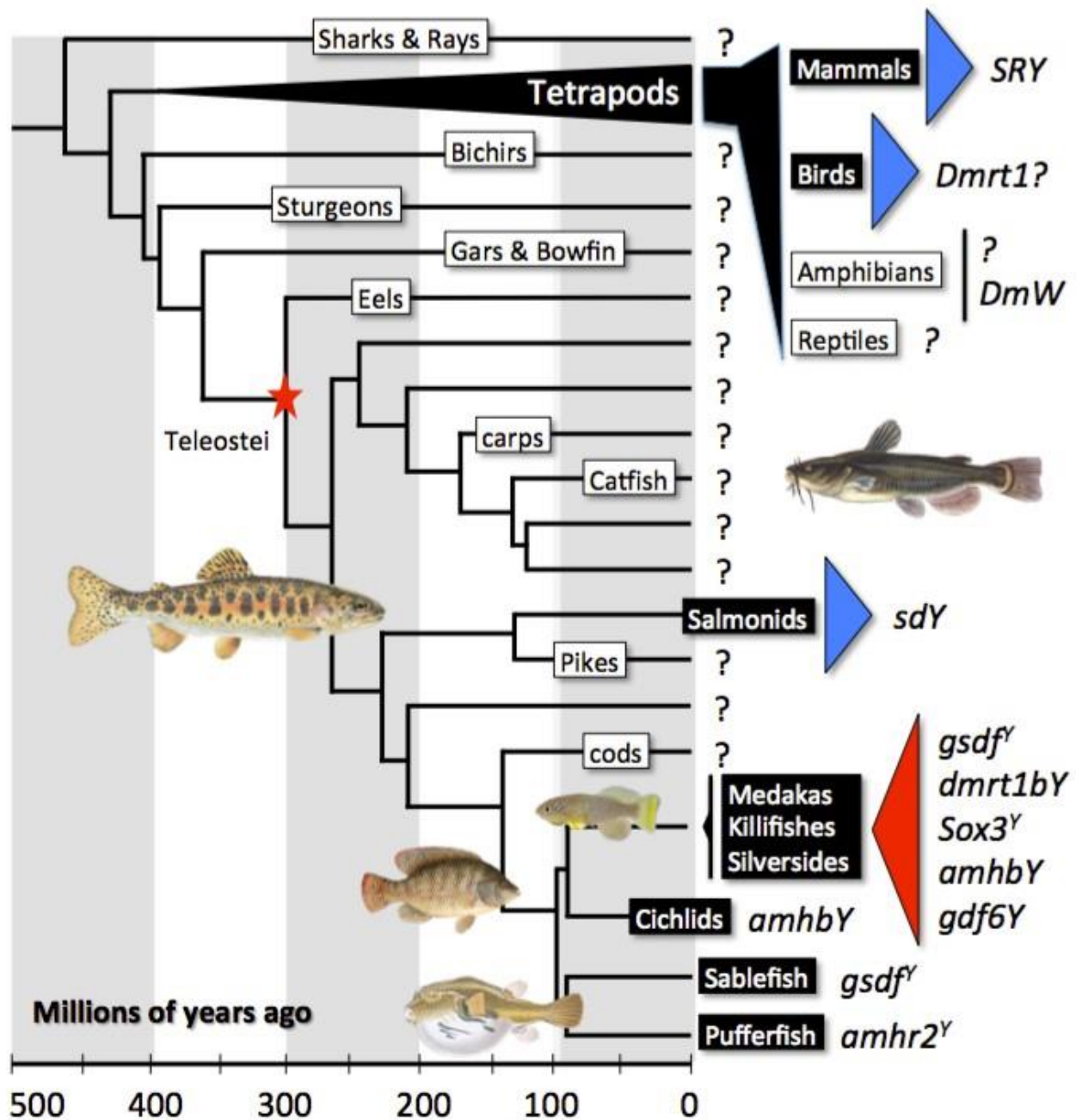


Figure 1: Vertebrate sex determining genes play musical chairs. Evolution and diversity of sex-determining genes in vertebrates (see text for details). Species or group of species with an already known or highly suspected sex-determining gene are shown in black boxes. Blue triangles represent master sex-determining genes with a certain degree of evolutionary conservation. Red triangles represent group of species with a high turnover of sex-determining genes. The red star shows the position of the teleost-specific whole genome duplication.

Unexpectedly, this gene, called *dmrt1bY* or *dmy* [7,10], arose by a gene duplication event from the autosomal *dmrt1a* gene being transposed to the proto-Y, adding another evolutionary mechanism for the emergence of an SD gene. Full sequencing of the male-specific region of the Y-chromosome (MSY), the first to be known after the human Y, revealed that several additional predictions of the general theory of Y-chromosome evolution were also not fulfilled, like stepwise diversification on an autosome, recombination suppression, degeneration of Y-linked genes, and accumulation of male-favoring genes [17]. This finding raised doubts about a unifying concept for the evolution of SD genes and heterogametic sex chromosomes.

Fish are uniquely suited to study the evolution of sex determination and SD genes (**Fig. 1**). Comprising about half of the 60 000 species of vertebrates, fish show also the greatest variety of sex determination. Unisexuality, simultaneous and consecutive hermaphroditism, environmental, and genetic sex determination are found in different, often closely related species and the distribution of various mechanisms follows no obvious phylogenetic pattern. With respect to genetic SD, it became quickly clear after its first discovery that *dmrt1bY* of medaka is not the master SD gene of fish in general [14]. Despite huge efforts over more than a decade, SD genes of only a few more species have been added to this list, and all those species have strong monogenic sex determination. In a sister species to medaka, *O. luzonensis*, and also in sablefish, *Anoplopoma fimbria*, allelic variations of a TGF- β member gene named *gsdf^x* determines male development [8,13]. In *Takifugu rubripes* (pufferfish) as well, allelic variation at a single nucleotide position of the *amh-receptor 2* gene, again a downstream component of the SD cascade in other vertebrates, controls phenotypic sex [6]. In the Patagonian pejerrey, *Odontesthes hatcheri*, a Y-linked duplicate of the *amh* gene drives male sex determination [5], a situation that mirrors the evolutionary scenario of the medaka SD gene. The same SD gene is also found in the tilapia, *Oreochromis niloticus* but in that species it is a Y chromosome missense SNP on a tandem duplication of *amh* that drives sex determination [15]. Interestingly, another TGF- β member, namely *gsdf6Y* has also been found as a potential SD in the killifish *Nothobranchius furzeri* [11], suggesting that members from the TGF- β pathway are often recruited as SD genes in vertebrates. In rainbow trout, *Oncorhynchus mykiss*, gene duplication and neofunctionalization led to the emergence of the SD gene [14], but totally unexpected and contrary to the situation in all other vertebrates for which information on proven or candidate SD genes exists, the duplicated gene *sdY* is not derived from a previously known component of the SD cascade but from the *interferon regulatory factor 9*, which normally functions in the context of immunity.

The current state of knowledge about the evolution of SD in vertebrates - last but not least influenced by the recent findings in fish - has resulted in a very unsatisfying situation. With the examples of male-biasing SD genes identified to date, it is rather impossible to derive a comprehensive understanding of processes and mechanisms that lead to the evolution of new SD genes and to evaluate the prevalence and general relevance of different systems. We do not know whether dosage sensitivity of master SD genes is an enigmatic case that is restricted to the avian lineage or is used more generally. Gene duplication and neofunctionalization is appearing now as a more frequent process to create novel SD genes, but the three fish species in which it has been found might be special in some way and bias our view by chance due to the low number of known SD genes in general. It can also be asked whether teleost fish that experienced in their evolutionary history a whole genome duplication, and certain lineage-specific additional ones, profit from this genomic situation with respect to the evolution of new SD genes. Regarding female heterogamety, there is so far no example for a predicted dominant female determining SD gene in fish that could add to the single case of *Xenopus laevis*, where dominant-negative action of a W-linked truncated *dmrt1* gene is proposed to suppress a postulated default male development in WZ frogs [4]. A totally unresolved complication comes from several reports where even in cases that appear to be clear monogenic SD, polymorphic autosomal modifier genes or environmental influences can modify or override the sex-chromosomal mechanisms. It can only be vaguely hypothesized that these findings indicate transition states between SD switches, and again the prevalence of such situations is absolutely unclear. It is obvious that the first step towards a better understanding of SD mechanisms and how new SD genes evolve is to characterize SD genes in more species, study their inheritance, and eventually illuminate their mode of action

The RAD-Sex approach: a Genome Wide Association Study (GWAS) screening for exploring the diversity of sex determination systems in fish.

RAD-tags (Restriction Associated DNA tags) are sequences adjacent to restriction enzyme cutting sites that can be assayed by a next generation sequencing technique known as RAD sequencing method (RAD-seq) [18]. RAD-seq has been shown to be an extremely powerful method for sampling the genomes of multiple individuals in a population. Because it does not require preliminary information on the genetics or the genome of the studied species, this approach is especially interesting for developing genetic analyses in the context of non-model organisms [19].

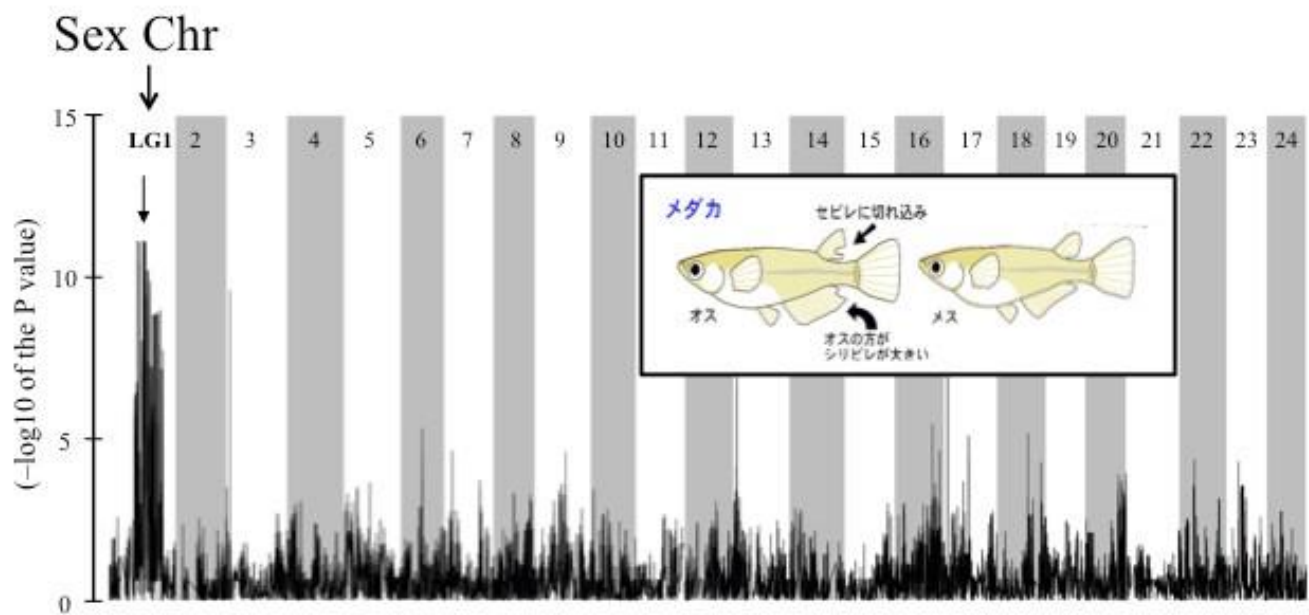


Figure 2: Distribution of RAD-tag markers along the 24 chromosomes of the medaka sequenced genome. The significantly sex-biased markers identified mid LG1 (chromosome Y) as the location of the sex-determining locus. Markers retained in this analysis have been scored in at least 25 individuals within the population.

This technique was recently applied with success to search for sex specific sequences in Gecko lizards [20] and for a sex determining locus in zebrafish, in which strain-specific sex determination systems were characterized [21,22]. To serve as an additional proof of principle, we applied this approach to the medaka, *Oryzias latipes*, which has a strong well-characterized genetic SD system [7,10,17]. The main idea was to investigate if RAD-seq was equally efficient in identifying sex-specific sequences close to *dmrt1bY*, the known SD gene in that species [7,10], using as starting biological material individuals from an outbred population of fish and not a genetic mapping family panel as was used in the case of the zebrafish [21]. To that end, fin clips were sampled from 31 phenotypic males and 32 phenotypic females of the Carbio strain of medaka. Genomic DNA was extracted from these samples and RAD library construction was carried out with each individual given a distinct barcode. Samples were sequenced in two lanes of Illumina HiSeq2000 using single end 100 bp reads. This dataset analysis allowed us to characterize more than 600 sex-polymorphic RAD-tag sequences among an overall total of 120 000 genetic markers. Most of these sex-polymorphic sequences map to the central portion of medaka linkage group LG1, the region that distinguishes the Y from the X chromosome (see **Fig. 2**). Among these sex-polymorphic RAD-tag sequences, 58 were sex-specific, i.e. the sequences were only present in males and totally absent in all the females. When mapped onto the medaka genome, 22 of these sex-specific RAD-tags were found to be on the Y chromosome (LG1) and 30 on unordered scaffolds probably due to an incomplete assembly of the medaka Y chromosome. Among these sex-specific RAD-tag sequences, 9 mapped to the Y non-recombinant sex-specific region (GenBank ID: AP006153) including two sequences that were less than 1 kb downstream of the 5' end of the *dmrt1bY* medaka sex determining gene (GenBank ID: AY129241) [10]. These results clearly demonstrate that the RAD-Sex approach is highly efficient because it enabled the identification of sex-specific sequences located very close to the previously known SD gene even in a GWAS that does not allow a simultaneous genetic mapping of the RAD-tags. It should be noted, that our analysis, although conducted on a species with a sequenced genome, did not use the genome sequence for the analysis of sex-associated RAD-tags: the genome sequence only provided validation of the proof of principle. This comparatively inexpensive approach can thus be applied to multiple species in a relatively high throughput project, without the need for obtaining biological materials from controlled genetic crosses that could be difficult to obtain in some species. This approach is also possible in species with no genetic or genomic information available, thus providing a special opportunity to develop an evolutionary-based project that can investigate the sex-determining system in a vast array of species.

Applying this approach to a large panel of species should provide a first evolutionary picture of SD genes answering simple but important questions like: how many evolutionary innovations can be characterized i.e., novel SD genes, how many “usual suspects” i.e., duplications of genes already known for their implication in the sex differentiation cascade, and how prevalent are gene duplications as primary events generating novel SD genes and thus how relevant is this evolutionary scenario beside the classical hypothesis of allelic diversification and the so far enigmatic situation of SD by gene dosage to explain the diversity of SD mechanisms. These questions on the evolution of SD genes have currently only been addressed for a narrow phylogenetic range, but never for a broad spectrum of taxa that represent a considerable fraction of the diversity and adaptations of a whole class of the chordate phylum. Such information will inform an essential question in evolutionary biology, foster genetics and developmental biology, and last but not least provide immediately useful information for better management of economically important fish species in fisheries and aquaculture.

Conclusion

Sex determination is a very basal and ubiquitous developmental process and the fact that it is so variable even between closely related organism poses many fascinating questions. A number of scenarios and hypothesis have been put forward to explain which evolutionary forces could favor such transitions and turnovers (for review see [23]). They range from fitness advantages of a new sex determination or its linkage to genes that are favorable to one sex or antagonistic to the other sex, over mutational load on the old, disappearing sex determination gene to scenarios that consider that neutral or non-adaptive processes of genetic drift, mutation and recombination can be instrumental in creating new sex determination systems. For all these hypotheses, which appear to a certain extent opposing or even contradictory, examples to support them can be found. A single one obviously cannot explain all the different cases of sex determination systems and the many instances of turnovers and transitions. Rather than being alternatives they may be complementary. To further our understanding of the trajectories that lead to the evolution of diverse sex determination mechanisms and sex chromosomes we need more detailed molecular knowledge about the molecular biology of these various mechanisms and about the ecology and population genetics under which they occur.

Acknowledgements

This project is supported by funds from the “Agence Nationale de la Recherche” and the Deutsche Forschungsgemeinschaft (ANR/DFG, PhyloSex project, 2014-2016).
The authors declare no conflict of interest.

References

- Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., & Postlethwait, J. H. (2011). Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. *Genetics*, 188(4), 799–808.
<https://doi.org/10.1534/genetics.111.127324>
- Anderson, J. L., Rodríguez Marí, A., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., & Postlethwait, J. H. (2012). Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PloS One*, 7(7), e40701.
<https://doi.org/10.1371/journal.pone.0040701>
- Gamble, T., Geneva, A. J., Glor, R. E., & Zarkower, D. (2014). Anolis Sex Chromosomes Are Derived from a Single Ancestral Pair. *Evolution*, 68(4), 1027–1041.
<https://doi.org/10.1111/evo.12328>
- Graves, J. A. M. (2013). How to evolve new vertebrate sex determining genes. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 242(4), 354–359. <https://doi.org/10.1002/dvdy.23887>
- Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., ... Strussmann, C. A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences*, 109(8), 2955–2959.
<https://doi.org/10.1073/pnas.1018392109>
- Herpin, A., & Schartl, M. (2009). Molecular mechanisms of sex determination and evolution of the Y-chromosome: Insights from the medakafish (*Oryzias latipes*). *Molecular and Cellular Endocrinology*, 306(1–2), 51–58. <https://doi.org/10.1016/j.mce.2009.02.004>
- Herpin, A., & Schartl, M. (2015). Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Reports*, 16(10), 1260–1274. <https://doi.org/10.15252/embr.201540667>
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., ... Kikuchi, K. (2012). A Trans-Species Missense SNP in *Amhr2* Is Associated with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*, 8(7), e1002798.
<https://doi.org/10.1371/journal.pgen.1002798>
- Koopman, P. (1995). The molecular biology of SRY and its role in sex determination in mammals. *Reproduction, Fertility, and Development*, 7(4), 713–722.
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., ... Wang, D. (2015). 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(11), e1005678.
<https://doi.org/10.1371/journal.pgen.1005678>
- Matsuda, M., Nagahama, Y., Kobayashi, T., Matsuda, C., Hamaguchi, S., & Sakaizumi, M. (2003). The sex determining gene of medaka: a Y-specific DM domain gene (DMY) is required for male development. *Fish Physiology and Biochemistry*, 28(1–4), 135–139.
<https://doi.org/10.1023/B:FISH.0000030500.29914.7a>

- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., ... Sakaizumi, M. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417(6888), 559–563. <https://doi.org/10.1038/nature751>
- Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17(2), 240–248. <https://doi.org/10.1101/gr.5681207>
- Mittwoch, U. (2005). Sex determination in mythology and history. *Arquivos Brasileiros De Endocrinologia E Metabologia*, 49(1), 7–13. <https://doi.org/S0004-27302005000100003>
- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., ... Sakaizumi, M. (2012). Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, *Oryzias luzonensis*. *Genetics*, 191(1), 163–170. <https://doi.org/10.1534/genetics.111.137497>
- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., ... Scharl, M. (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proceedings of the National Academy of Sciences*, 99(18), 11778–11783. <https://doi.org/10.1073/pnas.182314699>
- Nanda, I., Shan, Z., Scharl, M., Burt, D. W., Koehler, M., Nothwang, H., ... Schmid, M. (1999). 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nature Genetics*, 21(3), 258–259. <https://doi.org/10.1038/6769>
- Reichwald, K., Petzold, A., Koch, P., Downie, B. R., Hartmann, N., Pietsch, S., ... Platzer, M. (2015). Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell*, 163(6), 1527–1538. <https://doi.org/10.1016/j.cell.2015.10.071>
- Rondeau, E. B., Minkley, D. R., Leong, J. S., Messmer, A. M., Jantzen, J. R., von Schalburg, K. R., ... Koop, B. F. (2014). The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PloS One*, 9(7), e102089. <https://doi.org/10.1371/journal.pone.0102089>
- Takehana, Y., Matsuda, M., Myosho, T., Suster, M. L., Kawakami, K., Shin-I, T., ... Naruse, K. (2014). Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nature Communications*, 5. <https://doi.org/10.1038/ncomms5157>
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., ... Guiguen, Y. (2012). An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Current Biology*, 22(15), 1423–1428. <https://doi.org/10.1016/j.cub.2012.05.045>
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., ... Ito, M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences*, 105(7), 2469–2474. <https://doi.org/10.1073/pnas.0712244105>

Evolution of Sex determining genes in fish

Qiaowei PAN, Yann GUIGUEN* & Amaury HERPIN

INRA, Fish Physiology and Genomics laboratory, 35042 Rennes, France.

* Corresponding author: Yann Guiguen, Tel: 33 (0) 2 23 48 50 09, Fax: 33 (0) 2 23 48 50 20, Email: yann.guiguen@inra.fr

Abstract:

Fish have an exceptionally high diversity of sex determination systems including environmental- and genetic-based determinism. Master sex determining (MSD) genes that control genetic sex-determination also evolve extremely rapidly in fish, providing a window to understand the evolutionary mechanisms and processes of sex chromosome formation. Despite this high turnover, most of the currently known MSD genes are classical “usual suspects”, i.e., genes that evolved from downstream sex differentiation gene network. Intriguingly, fish also provide the only known exception to this “usual suspects” rule: a duplicate of an immune-related gene is the conserved salmonid MSD gene, unraveling further flexibility.

Sex determination and sex chromosomes

Sexual reproduction is a widespread phenomenon, which requires two sexes in a population. There are multiple ways to achieve this state, but the most common in vertebrates is gonochorism, where individuals acquire their sex during development and maintain their sexual fate throughout life. This sexual identity is triggered by a process called sex determination (SD), in which a “master switch” controls a downstream regulatory cascade leading to either male or female gonadal development. In contrast with other more conserved developmental programs, SD master switches can be highly variable with primary signals coming either from the genome (genetic sex determination or GSD), or from the environment (environmental sex determination or ESD). Genetic sex determination involves sex chromosomes that harbor inherited master SD genes (not necessarily coding for a protein), responsible for the initiation of the SD process (Herpin and Schartl, 2015). One species can have one master SD gene (monofactorial system) or multiple SD genes (polygenic system). In a monofactorial system, one sex is heterogametic with two different sex chromosomes; the other sex is homogametic with two of the same type of sex chromosome. A male heterogametic system is termed XX-XY system (i.e., like in most mammals), where the Y chromosome carries the male master SD gene; a female heterogametic system is denominated ZZ-ZW system (i.e., like in birds), where the W chromosome carries the female SD gene (Heule *et al.*, 2014). Contrarily to monofactorial systems, polygenic systems (Moore and Roberts, 2013) involve more than two alleles: either one locus with multiple independently segregating alleles, or multiple loci interacting through epistasis. However, such systems are more complex and less well documented than monofactorial systems. Unlike most mammals and birds where one type of monofactorial system has been conserved for over 100 million years, fish display a high diversity of sex determination systems including various types of monofactorial and polygenic systems. Moreover, sex determination systems in fish can differ between very closely related species, sometimes even among different populations of the same species. Furthermore, in some species, genetic factors can interact with environmental cues (most commonly temperature), generating intricate sex determination mechanisms. Such evolutionary dynamics makes fish an attractive group to study the evolution of sex determination systems. It is worth noting that studies from other taxa, such as amphibians and reptiles, have also revealed a comparable degree of diversity in sex determination systems (Kikuchi and Hamaguchi, 2013); however, to date, fish has provided the most abundant knowledge on sex determination and its evolution in vertebrates. While fish is a more generic term that can include several taxonomic groups, so far all the identified master SD genes were found in teleost fish. For this reason, we will be focusing on the evolution of sex determination genes in this subgroup.

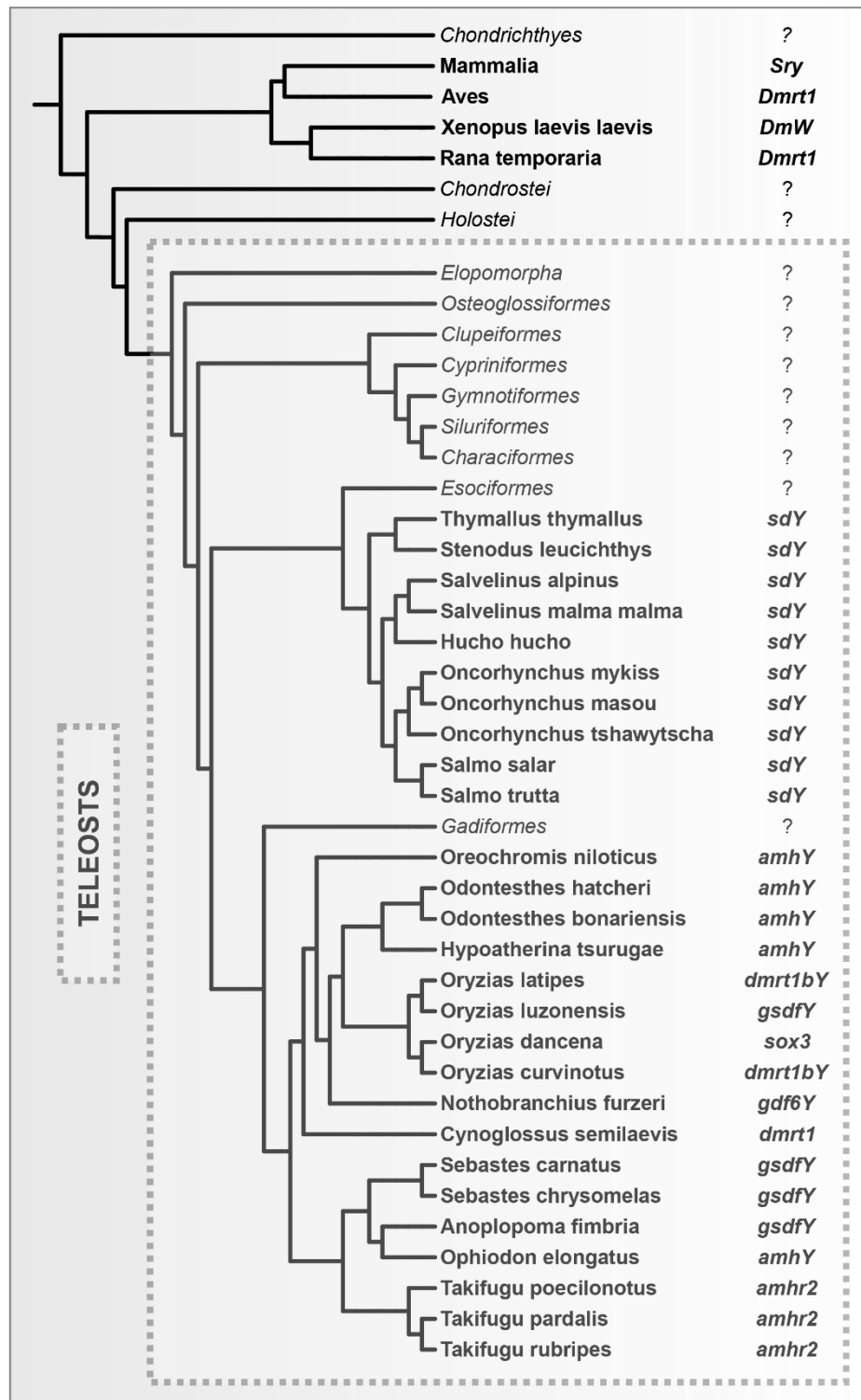


Figure 1: Evolution and diversity of sex-determining genes in vertebrates with a focus on teleost fish. Species with a known or highly suspected SD gene are mentioned in bold type at each tip of the tree leaves along with the name of the SD gene. A few other taxa in which no SD gene is currently reported are also represented in normal type with a question mark.

Fish sex determining genes

The high diversity of sex determination systems seen in teleost fish is linked to the high turnover rate of their sex chromosomes. As a result, most fish sex chromosomes are considered to be relatively young and are often described as being homomorphics, i.e., with little differentiation and no cytological difference. Sex chromosomes turnover is usually connected with the emergence of new master SD genes. Indeed, in therian mammals and birds, master SD genes are highly conserved with the quasi-exclusive usage of *SRY* for mammals and *DMRT1* for birds. But, a high diversity of SD genes (**Figure 1**) has been found in the past two decades in many teleost fish and this has greatly expanded our understanding of the SD mechanisms and SD evolution. Despite this great diversity, most of these newly discovered fish master SD genes are classical “usual suspects” (Herpin and Scharl, 2015) that fall into three independent protein families implicated in the “canonical” vertebrate sex differentiation gene network, namely the Sry-related HMG box (SOX), the Doublesex and mab-3 related transcription factors (DMRT) and the transforming growth factors (TGF- β) family. Their recruitments as master SD genes do not follow any obvious phylogenetic relationship and are more likely to be independent and repeated events during teleost evolution (**Figure 1**). Only one known exception to this “usual suspects” rule is the salmonid SD gene, *sdY* (sex determining region on the Y chromosome), which was derived from the duplication of an immune-related gene unsuspected to be involved in the sex differentiation pathway.

The birth of a new master SD genes can start with a duplication of an autosomal gene that will acquire a new SD function leading to new sex chromosomes. On the other hand, it can begin with allelic diversification from a step-wise diversification of two alleles of the future sex chromosome pair. One allele then becomes the sex determiner, while the other allele either retains a non-SD function or favors the development of the opposite sex. (Kikuchi and Hamaguchi, 2013). The gene duplication and allelic diversification mechanisms have both been found in teleost fish to generate new master SD genes and drive the turnover of sex chromosomes. (**Table 1**).

In the following part of this article, we will review the cases of master sex determination genes in teleost fish and the key publications on their discoveries can all be found in **Table 1**.

Species	SD System	Master SD Gene	SD Gene Ancestor	Gene Group	Evolution Mechanism	Validation	Key Publications
Japanese medaka <i>Oryzias latipes</i>	XY	<i>Dmy / dmrt1bY</i>	<i>Dmrt1</i>	Dmrt	Gene duplication	Expression Function Sex-linkage	Matsuda <i>et al.</i> , 2002 (doi:10.1038/nature751) Nanda <i>et al.</i> , 2002 (doi:10.1073/pnas.182314)
Malabar medaka <i>Oryzias curvinotus</i>	XY	<i>Dmy / dmrt1bY</i>	<i>Dmrt1</i>	Dmrt	Gene duplication	Expression Function Sex-linkage	Matsuda <i>et al.</i> , 2003 (doi:10.2108/zsj.20.159)
Chinese tongue sole <i>Cynoglossus semilaevis</i>	ZW	<i>Dmrt1</i>	<i>Dmrt1</i>	Dmrt	Allelic diversification	Expression Sex-linkage	Chen <i>et al.</i> , 2014 (doi:10.1038/ng.2890)
Indian medaka <i>Oryzias dancena</i>	XY	<i>Sox3Y</i>	<i>Sox3</i>	Sox	Allelic diversification	Expression Function Sex-linkage	Takehana <i>et al.</i> , 2014 (doi:10.1038/ncomms5157)
Luzon ricefish <i>Oryzias luzonensis</i>	XY	<i>GsdfY</i>	<i>Gsdf</i>	TGF- β	Allelic diversification	Expression Function Sex-linkage	Myosho <i>et al.</i> , 2012 (doi:10.1534/genetics.111.127497)
Sable fish <i>Anoplopoma fimbria</i>	XY	<i>Gsdf^f</i>	<i>Gsdf</i>	TGF- β	Allelic diversification	Sex-linkage	Rondeau <i>et al.</i> , 2013 (doi:10.1186/1471-2164-14-452)
Rockfishes <i>Sebastes carnatus</i> & <i>Sebastes chrysomelas</i>	XY	<i>Gsdf</i>	<i>Gsdf</i>	TGF- β	Allelic diversification	Sex-linkage	Flower & Buonaccorsi, 2016 (doi:10.1111/mec.13594)
Patagonia pejerrey <i>Odontesthes hatcheri</i>	XY	<i>amhy</i>	<i>Amh</i>	TGF- β	Gene duplication	Expression Function Sex-linkage	Hattori <i>et al.</i> , 2012 (doi: 10.1016/j.ygcen.2013.03.019)
Nile tilapia <i>Oreochromis niloticus</i>	XY	<i>amhy</i>	<i>Amh</i>	TGF- β	Gene duplication	Expression Function Sex-linkage	Li <i>et al.</i> , 2015a (doi:10.1371/journal.pgen.1005678)
Lingcod <i>Ophiodon elongates</i>	XY	<i>Amh</i>	<i>Amh</i>	TGF- β	Gene duplication	Sex-linkage	Rondeau <i>et al.</i> , 2016 (doi:10.1186/s13104-016-2030-6)
Fugu <i>Takifugu rubripes</i>	XY	<i>Amhr2Y</i>	<i>Amhr2</i>	TGF- β	Allelic diversification	Expression Sex-linkage	Kamiya <i>et al.</i> , 2012 (doi:10.1371/journal.pgen.1002798)
Killyfish <i>Nothobranchius furzeri</i>	XY	<i>Gdf6Y</i>	<i>Gdf6</i>	TGF- β	Allelic diversification	Expression Sex-linkage	Reichwald <i>et al.</i> , 2015 (doi:10.1016/j.cell.2015.10.071)
Rainbow trout <i>Oncorhynchus mykiss</i>	XY	<i>sdY</i>	<i>irf9</i>	Interferon	Gene duplication	Expression Function Sex-linkage	Yano <i>et al.</i> , 2012 (doi: 10.1016/j.cub.2012.05.045) Yano <i>et al.</i> , 2013 (doi:10.1111/eva/12032)

Table 1: Currently known or suspected master sex determining genes in fish. Not all master SD genes found in teleost species have been shown to fulfill all classical SD requirements. Confidence in their sex determining role and function is given in validation column. References to species-specific publications are given in the publication column with a link to the DOI of the publication.

DMRT1: a multi-context SD gene

The *dmrt1* gene is a transcriptional factor belonging to the evolutionarily well-conserved DM (Doublesex and Mad-3) domain DMRT gene family, which is involved in sexual development in many phylogenetically distant groups from worms to mammals. In teleost fish, sexually dimorphic expression of *dmrt1* has been correlated with gonadal identity in both gonochoristic and hermaphroditic fish species (Herpin and Scharl, 2011). *dmy/dmrt1bY*, was the first non-mammalian SD gene found in vertebrates and it is a duplicated copy of autosomal *dmrt1/dmrt1a* gene found in two ricefish species of the genus *Oryzias*: the Japanese medaka (*O. latipes*) and the Malabar ricefish (*O. curvinotus*), both having an XX/XY sex determination system. *dmy/dmrt1bY* arose about 5–10 million years ago from a local duplication of an autosomal fragment encompassing the *dmrt1* gene. Consequently, this pair of X and Y chromosomes are very young in comparison with the chromosome pair found in therian mammals, and remain homomorphic.

The expression pattern of the medaka *dmrt1bY* is typical of a master SD gene with an early and transient expression during gonadal primordium differentiation, and low persistence in the adult gonads. Around the time of hatching *dmy/dmrt1bY* is first expressed in the nuclei of the somatic undifferentiated cells surrounding the primordial germ cells (PGC) and later on exclusively in the sertoli cells in males. In comparison, the autosomal *dmrt1/dmrt1a* is expressed much later at 20 days after hatching in males and shows a testis-specific expression pattern in adult males with 50 times higher expression than that of *dmy/dmrt1bY*. Concomitantly to the acquisition of a dominant position within the sex-determining network, *dmy/dmrt1bY* was subjected to a profound rearrangement of its regulatory landscape brought about by exaptation of a “ready-to-use” cis-regulatory element from a transposable element (TE). This element, called Izanagi, acts as a silencer to turn-off the *dmy/dmrt1bY* gene after it has fulfilled its male SD function (Herpin *et al.*, 2010).

In the medaka, the duplication/insertion of *dmrt1* led to a dominant *dmy/dmrt1bY* master SD on the Y chromosome, but interestingly *dmrt1* has also been found as the master SD gene in a fish species with ZZ/ZW sex determination system: the half-smooth tongue sole (*Cynoglossus semilaevis*). In this species *dmrt1* is only found on the Z chromosome while its homolog on the W chromosome is highly degenerated, probably a result of an old allelic diversification process. Even though homogametic sex is male in this species with two copies of Z chromosomes, SD operates through a Z-encoded mechanism that determines male development: the presence of two copies of *dmrt1* in ZZ animals leads to male development, while having only one functional copy (haploinsufficiency) in ZW individuals results in female development.

Dmrt1 is expressed specifically in male germ cells and pre-somatic cells of the undifferentiated gonad at the sex determination stage, fulfilling a key feature of male SD gene. Furthermore, functional knock-out of this gene resulted in male to female sex reversal, indicating its essential role of male SD gene in this species. The sex chromosome pair in tongue sole is thought to be around 30 million years old, but whether closely related species of the tongue sole conserved dmrt1 as a master SD gene is currently unknown. Interestingly, the sex chromosomes of the half-smooth tongue sole share the same ancestral vertebrate protochromosome as the W and Z chromosome in avians, which also have a Z dosage-dependent male sex determination system based on the Dmrt1 gene, highlighting an evolutionary convergence around Dmrt1.

These two teleost cases show that dmrt1 can acquire its position as a master SD gene through both allelic diversification and gene duplication mechanisms and its function as a SD gene can be triggered either by expression dosage effect or rewired transcription. Combined with its known SD roles in avians and some frogs (Yoshimoto *et al.*, 2008), dmrt1 thus appears to be a highly versatile and flexible master SD gene.

Sox3, a conserved role in vertebrate sex determination

SRY, the mammalian master SD gene on the Y chromosome is thought to be the evolution of an allelic diversification of the X-linked *SOX3* gene. This is reminiscent of what has been found the Indian ricefish (*Oryzias dancena*), in which a novel sex chromosome recently evolved by co-option of this *Sox3* gene. In this species, the Y chromosome allele of *Sox3* (*Sox3^Y*) encodes the exact same amino-acid sequence as its X chromosome counterpart. However, a distant and specific *cis*-regulatory element of *Sox3* on the Y chromosome induces a gonadal specific expression of *Sox3* only in XY embryos providing a SD signal in differentiating gonads. XX transgenic fish with the *Sox3^Y* regulatory segment developed as male and that the loss of function mutation of *Sox3^Y* in XY fish developed ovary-type gonads, supporting the SD role of *Sox3^Y* in that species. It was also demonstrated that the early testis specific expression of *Sox3^Y* initiates the expression of *gsdf*, a gene conserved in teleost testicular differentiation pathways. However, the overexpression of *Sox3^Y* in the sister species *O. latipes* which has *dmy/dmrt1bY* as the master SD gene, could not induce testicular differentiation. This result indicates that co-evolved components in the downstream sex differentiation pathways is required for *Sox3^Y* to act as the master switch for male development (Herpin and Scharf, 2015). The evolutionary parallel between *Sry* in therian mammals and *Sox3^Y* in Indian ricefish underlines the conserved importance of *sox3* in vertebrate sex determination.

The TGF- β family: A SD genes jungle

A long-standing hypothesis, following the discovery of *SRY* in mammals, was that vertebrate SD genes should be transcription factors. The identification of *dmy/dmrt1bY* as the SD gene in *O. latipes* also supported this view. However, more recent evidences in teleost have substantiated the idea that SD genes does not necessarily need to be transcriptional factors and members of the TGF- β family have emerged as master SD switches in teleost fish.

The first TGF- β related SD gene was identified in the Patagonian pejerrey (*Odontesthes hatcheri*). In this XX/XY species, a duplication of the anti-müllerian hormone (*amh*) was only found in males. Amh is a TGF- β hormone mainly known in mammals, birds and reptiles for its role in the regression of Müllerian ducts during male fetal development. Although teleost fishes lack the Müllerian ducts, they possess both *amh* and its type II receptor *amhr2* and this pathway has been implicated in the regulation of germ cell proliferation, spermatogenesis and regulation of gonadal sex-steroid synthesis. In the Patagonian pejerrey, this male-specific duplicated *amh*, called *amhy*, shares a high identity with its autosomal copy *amha* but has a Y-specific insertion in its third intron. Being very similar within the coding sequence to the autosomal *amha*, *amhy* is thought to fulfil its SD role through an earlier expression in the differentiating male gonads. This male-specific duplication of *amh* is also conserved in a closely sister species *O. bonariensis*, whose SD is known to be also modulated by temperature (Yamamoto *et al.*, 2014). In addition, in a more distant *Atheriniformes* species, a truncated duplicated copy of *amh* is suspected to be involved in male sex determination. However, due to greater sequence divergence, it has not been resolved whether this truncated duplicate of *amh* shares ancestry with the previous two cases or is an independent recruitment of the same SD gene. In the Nile tilapia (*Oreochromis niloticus*), a tandem duplication of *amh* has been found on the Y chromosome. This tandem duplication consists of a gene encoding a truncated protein lacking the TGF- β domain (*amh- Δ y*), and immediately downstream an *amhy* gene that is identical to the X-linked *amh* except for a single missense single nucleotide polymorphism (SNP). Both *amh- Δ y* and *amhy* are exclusively expressed in XY gonads during the onset of sex determination period in Nile tilapia. Similar expression profile of the X-linked *amh* and their receptor *amhr2* was found in both XX and XY gonads. However, only the knock-out of *amhy* lead to male to female sex reversal in XY individuals, while the knock-out of *amh- Δ y* alone had no effect. Additionally, overexpression of *amhy* caused female to male sex reversal in XX fish, but not the overexpression of the X-linked *amh* or the truncated *amh- Δ y*. Collectively, these results show that *amhy* is the master SD gene in Nile tilapia.

More recently, another fish species, the lingcod (*Ophiodon elongatus*), has been characterized with a male-specific duplicate of *amh*. Although no expression nor functional data were available to consolidate its SD function, it potentially provides another independent case of recruitment of *amh* as a master SD gene in teleost.

Interestingly, the Amh type II specific receptor gene, *amhr2*, has also been identified as a SD gene in fish. In the pufferfish (*Takifugu rubripes*), a species with an XX/XY SD system, a missense SNP in the kinase domain of *amhr2* was found to be the only polymorphism strictly associated with phenotypic sex with males being heterozygous and females homozygous. Sex is then determined by the combination of these two X and Y *amhr2* alleles. This SNP leads to a single amino acid substitution and the X allele confers a decreased AMH signal transduction in XX females, while XY males with the wild-type *amhr2* Y allele would have a better Amh signaling, leading to testicular differentiation. This result is consistent with the mekade *hotei* mutant in which the impaired *amhr2* activity is linked with the feminization of the gonads. Surprisingly, this single missense SNP associated with male development is conserved over 10 million years of evolution, as it is present in two other Takifugu species (*T. pardalis* and *T. peocilonotus*).

The gonadal soma derived growth factor (Gsdg) is another vertebrate TGF- β ligand closely related to Bone Morphogenetic Proteins (BMPs), which has been lost in tetrapods. This *gsdf* gene was first characterized in rainbow trout as an inducer of germ cells proliferation, and has since been associated with testicular development due to its exclusive expression in the early differentiating testis of all teleost examined so far (Herpin and Scharl, 2015). In *Oryzias luzonensis*, an XX-XY ricefish species closely related to the Japanese medaka *O. latipes*, X and Y alleles of *gsdf* have been identified, and *gsdf^Y*, the Y chromosome allelic form of *gsdf*, has been further characterized as a master SD gene. *gsdf^Y* and *gsdf^X* slightly differ in their sequence, but encode identical proteins. The testicular differentiation inducing role of *gsdf^Y* is triggered by a cis-regulatory region affecting a potential binding site for steroidogenic factor 1 (*sf1* or *nr5a1*) in the *gsdf^Y* promoter, conferring an elevated and earlier expression level of *gsdf^Y* during sexual differentiation than that of *gsdf^X* allele. Overexpression of the *gsdf^Y* locus in XX females is able to induce masculinization, corroborating its role as a male master SD gene in *O. luzonensis*. In addition, this overexpression of *gsdf^Y* is also able to induce masculinization of XX females in the sister species *O. latipes* which has *dmy/dmrt1bY* as the master SD gene, suggesting that *gsdf^Y* could functionally replace *dmy/dmrt1bY*. A few other fish species have been also described to rely on *gsdf* as a putative master SD gene, i.e., the sablefish, *Anaplopoma fimbria* and two rockfish species, *Sebastes carnatus* and *S. chrysomelas*.

Growth differentiation factor 6 (*Gdf6*) is a member of the TGF- β family known to be involved in vertebral segmentation and cell differentiation. While no gonadal function was reported previously for this gene, *gdf6Y* emerged recently as a putative SD gene in the turquoise killifish, *Nothobranchius furzeri*. In this species, the *gdf6* locus is sex-linked with specific X and Y chromosomal allelic variations that affect the Gdf6 protein function. So far, functional evidence of its role as a SD gene is not available, yet the Y chromosome allele, *gdf6Y*, exhibits a clear sex-linked expression during testicular differentiation, supporting its SD role in this species.

***sdY*: an unusual and conserved SD gene in Salmonids.**

All of the above-described SD genes belong to the so-called ‘usual suspects’ category but the recent discovery of a new type of SD gene in rainbow trout (*Oncorhynchus mykiss*) demonstrated an unexpected flexibility in vertebrate SD. This gene, called *sdY*, came to the spotlight as it was found to be expressed only in embryonic testes and was also tightly linked to the sex-locus on the Y chromosome. It was further confirmed to be the trout SD gene based on functional proofs showing that *sdY* is by itself necessary and sufficient to induce testicular differentiation. One of the multiple unusual characteristics of *sdY* is that it evolved from the duplication of an immune related gene, *irf9* (*interferon related factor 9*) and compared to its ancestral protein, *sdY* lost the N-terminal DNA-binding domain, but preserved its C-terminal protein-protein interaction domain. Another interesting feature is that *sdY* is sex-linked in most salmonid species, indicating that it is a conserved SD gene in this teleost group. Yet, the SEX locus containing *sdY* is not located in the same chromosome in different salmonids species (Woram *et al.*, 2003), suggesting a rapid turnover of sex chromosome while preserving what seems to be a jumping master SD gene (Faber-Hammond *et al.*, 2015). How this unusual SD gene was recruited and the mechanism through which it regulates the downstream sex differentiation gene networks remains to be clarified. However, it presents a paradigmatic case that un-related genes outside the “usual suspects” are able to acquire *de novo* sex determining functions in teleost.

The evolution of genetic sex determination mechanisms is closely linked to the evolution of sex chromosomes. The discovery of multiple new master SD genes in teleosts provided invaluable insights on the evolution of SD and the processes of the establishment of new sex chromosomes. Before the identification of the first fish master SD gene, models of sex chromosome evolution predicted that SD genes should evolve from a stepwise allelic diversification process starting from a single gene on an autosome.

After fixation of a sex-specific allele in that newly formed SD gene, recombination reduction around that new sex locus in the proto-sex chromosome pair would give rise to classical heteromorphic sex chromosomes. Some newly described fish SD genes have evolved by this allelic diversification mechanism, for instance *sox3^Y* and *gsdf^Y* in some medaka species and *amhr2* in Takifugu. But this mechanism did not lead to strongly differentiated heteromorphic sex chromosomes in these teleost species. This phenomenon can probably be explained if we consider that these SD genes are very young in term of evolution. But the conservation of a single missense SNP in *amhr2* acting as the sole signal to initiate sex determination in at least three Takifugu species that diverged over 10 million years ago, shows that recombination reduction around a new sex locus in the proto-sex chromosome is not necessary to fix a new SD gene. The characterization of new teleost SD genes also provided a novel mechanism for sex chromosome and SD gene evolution as many fish SD genes are the result of a duplication/insertion event. In these cases, a genomic area containing an ancestral gene is duplicated into a different chromosome that will initiate the formation of a new sex chromosome after subsequent functional and/or regulatory changes of the proto SD gene. This mechanism of gene duplication leading to sex chromosome formation first came in light for *dmy/dmrt1b^Y* in Japanese medaka, and also been found to generate other vertebrate master SD genes, such as the female SD gene *DM-W* in the African clawed frog (*Xenopus laevis*) (Yoshimoto *et al.*, 2008).

In addition, thanks to their high turnover rate, teleost sex chromosomes also provide many different models to study the process of sex chromosome formation. One such case is from the study on half-smooth tongue sole in which a pair of relatively young Z and W chromosomes, comparing to those found in birds, have been sequenced and analyzed in details. This analysis showed that suppression of recombination, accumulation of transposable elements, fixation of sex specific genes, and gene dosage compensation are early events of sex chromosomes evolution. Furthermore, this half-smooth tongue sole W chromosome has already lost about two-thirds of its original protein coding genes and thus reached a roughly similar stage as much older mammalian and avian sex chromosomes within 30 million years. Combined with the finding that the gene-poor primate Y chromosome has been stable over the last 25 million years, they provide supports for the hypothesis that the decay of sex chromosome can happen rapidly in the early phase of sex chromosome formation then slow down and stabilize in its final stage, reaching a gene content equilibrium.

The ‘limited options’ hypothesis states that particular sex chromosomes or genes, especially those known to be involved in the SD gene network, might be preferentially recruited repeatedly in vertebrate lineages as new sex chromosomes or new master SD genes as they are ‘good at doing the job’ (Graves and Peichel, 2010). The multiple master SD genes recently found in teleost strongly endorse this idea as all of them with on single exception originate from three “usual suspect” gene family, namely *dmrt*, *sox* and *Tgf-β*. *Dmrt* and *sox* family members have also been also found in other vertebrate groups but several *Tgf-β* members have first emerged in teleosts as new ‘usual suspects’ and since then *Amh* has also been identified as the potential platypus master SD gene (Cortez *et al.*, 2014). Interestingly new knowledge on fish SD genes also provide an exception to this ‘limited options’ hypothesis as *sdY*, the salmonid master SD gene, is seemingly not related to any sex differentiation genes. However, it could not be excluded that this might be representing previously neglected factors in the SD gene networks (Herpin and Scharl, 2015), or that these unusual SD genes are underrepresented because they are just much more difficult to find.

Recently, a proposal with a developmental perspective has been put forward to view sex determination not as one hierarchical cascade with the master SD gene on the top, but as a complex network with different types of genetic and environmental factors influencing cell proliferation and hormone levels to push sexual development over either a male or female threshold (Heule *et al.*, 2014; Uller and Helanterä, 2011). Studies on the teleost SD provided supports to this view in which each node of the SD gene network has the potential of becoming the major-effect loci, not necessarily being on the top of the cascade. A prominent example of this hypothesis would be *amhr2* in *Takifugu* that modulate transduction efficiency of an upstream ligand signal, i.e., AMH acting on germ cell proliferation or regulating steroid synthesis activity (Kikuchi and Hamaguchi, 2013), to direct male or female development.

Many open questions still remain concerning teleost SD and probably the most important one is why the teleost fish use so many SD genes and SD systems with such a high sex chromosomes turnover rate? Rapid turnover of sex chromosomes is thought as a way to escape the decaying fate of sex chromosomes, assuming that the lack of recombination around the sex locus will lead to the accumulation of deleterious mutation. Alternatively, sexual antagonistic selection has also been put forward to explain the expansion of the sex locus and the accumulation of sex specific genes near it. Sex ratio selection provides a scenario that a new master SD gene could rise in frequency when it confers a more balanced sex ratio in the population. Other mechanisms such as random genetic drift and pleiotropic selection have also been used to explain the rise of new SD systems.

Another hypothesis, to explain the lack of sex chromosome turnover in mammals and birds, states that these highly heteromorphic sex chromosomes might be acting as evolutionary traps that would stabilize SD genes for long span of evolutionary time (Bachtrog *et al.*, 2014). Interestingly SD in many teleost species with a genetic SD system can still be modulated by environmental cues like for instance in the Nile tilapia or the half-smooth tongue sole. Transition stages between GSD and TSD are also well illustrated in teleost like in the case of pejerrey species that retains a SD gene (*amhY*) that is used alone in *O. hatcheri* and used in combination with temperature in the closely related species *O. bonariensis* (Yamamoto *et al.*, 2014). These potentially unstable stages could also be seen as transition stages facilitating a rapid turnover leading to the emergence of new SD genes.

References

- Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J.C., The Tree of Sex Consortium, 2014. Sex Determination: Why So Many Ways of Doing It? PLoS Biol. 12, e1001899. doi: 10.1371/journal.pbio.1001899
- Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P.D., Grützner, F., Kaessmann, H., 2014. Origins and functional evolution of Y chromosomes across mammals. Nature 508, 488–493. doi:10.1038/nature13151
- Faber-Hammond, J.J., Phillips, R.B., Brown, K.H., 2015. Comparative Analysis of the Shared Sex-Determination Region (SDR) among Salmonid Fishes. Genome Biol. Evol. 7, 1972–1987. doi:10.1093/gbe/evv123
- Graves, J.A.M., Peichel, C.L., 2010. Are homologies in vertebrate sex determination due to shared ancestry or to limited options? Genome Biol. 11, 205.
- Herpin, A., Braasch, I., Kraeussling, M., Schmidt, C., Thoma, E.C., Nakamura, S., Tanaka, M., Scharl, M., 2010. Transcriptional Rewiring of the Sex Determining dmrt1 Gene Duplicate by Transposable Elements. PLoS Genet. 6, e1000844. doi: 10.1371/journal.pgen.1000844
- Herpin, A., Scharl, M., 2015. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. EMBO Rep. 16, 1260–1274. doi:10.15252/embr.201540667
- Herpin, A., Scharl, M., 2011. Dmrt1 genes at the crossroads: a widespread and central class of sexual development factors in fish. FEBS J. 278, 1010–1019. doi:10.1111/j.1742-4658.2011.08030.x
- Heule, C., Salzburger, W., Böhne, A., 2014. Genetics of Sexual Development: An Evolutionary Playground for Fish. Genetics 196, 579–591. doi:10.1534/genetics.114.161158
- Kikuchi, K., Hamaguchi, S., 2013. Novel sex-determining genes in fish and sex chromosome evolution: Novel Sex-Determining Genes in Fish. Dev. Dyn. 242, 339–353. doi:10.1002/dvdy.23927
- Moore, E.C., Roberts, R.B., 2013. Polygenic sex determination. Curr. Biol. 23, R510–R512. doi: 10.1016/j.cub.2013.04.004
- Smith, C.A., Roeszler, K.N., Ohnesorg, T., Cummins, D.M., Farlie, P.G., Doran, T.J., Sinclair, A.H., 2009. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature 461, 267–271. doi:10.1038/nature08298
- Uller, T., Helanterä, H., 2011. From the Origin of Sex-Determining Factors to the Evolution of Sex-Determining Systems. Q. Rev. Biol. 86, 163–180. doi:10.1086/661118

- Woram, R.A., Gharbi, K., Sakamoto, T., Hoyheim, B., Holm, L.-E., Naish, K., McGowan, C., Ferguson, M.M., Phillips, R.B., Stein, J., Guyomard, R., Cairney, M., Taggart, J.B., Powell, R., Davidson, W., Danzmann, R.G., 2003. Comparative Genome Analysis of the Primary Sex-Determining Locus in Salmonid Fishes. *Genome Res.* 13, 272–280. doi:10.1101/gr.578503
- Yamamoto, Y., Zhang, Y., Sarida, M., Hattori, R.S., Strüssmann, C.A., 2014. Coexistence of Genotypic and Temperature-Dependent Sex Determination in Pejerrey *Odontesthes bonariensis*. *PLoS ONE* 9, e102574. doi: 10.1371/journal.pone.0102574
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T., Ito, M., 2008. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl. Acad. Sci.* 105, 2469–2474. doi:10.1073/pnas.0712244105

Section II :

Results

Notes on the chapter “Identification of a duplicated and translocated master sex determination gene provides insight into homomorphic sex chromosome differentiation in Northern pike (*Esox lucius*)”

This research project to identify the MSD in *Esox lucius* started with an educated guess. Tissue specific transcriptomes, including one library sequenced from testis samples, were available for *E. lucius* on the Phylofish online database. During the investigation of the expression pattern of the ‘usual suspects’ of vertebrate MSD genes, two homologous copies of Anti-Müllerian hormone (*amh*) were identified and one copy showed an exclusive presence in the testis library. Before then, a male-specific duplicate of *amh* had been determined as the male MSD in another teleost species, the Patagonian pejerrey (*Odontesthes hatcheri*), so the *amh* copy with testis-specific expression found in *E. lucius*, named *amhby*, became a very good candidate for being the male MSD in this species.

In our experiments, we raised the Northern pikes in the indoor fish facility in our laboratory. Raising and performing experiments on top predator fishes with big body size and long generation time like the Northern pike presents numerous difficulties, some of them were not foreseen before our study. Pikes exhibit a high level of cannibalism as early as three weeks post fertilization, and they can grow up to 40 cm during the first year. For these reasons, they need to be kept in individual tanks early and for the entire period of the experiment. These characteristics of Northern pikes increased tremendously the difficulty in performing experiments and the cost of keeping large numbers of individuals for each experiment. As a result, we could only perform limited experiments each reproductive season, which occurs annually within a one-month window.

Identification of a duplicated and translocated master sex determination gene provides insight into homomorphic sex chromosome differentiation in Northern pike (*Esox lucius*)

Qiaowei Pan¹, Ayaka Yano¹, René Guyomard², Elodie Jouanno¹, Romain Feron¹, Min Wen¹, Brigitte Mourot¹, Estelle Vigouroux¹, Jean-Mickaël Busnel³, Alizée Frezel¹, Amélie Patinote¹, Frédéric Borel¹, Konrad Ocalewicz⁴, Wojciech WitarSKI⁵, Monika Bugno-Poniewierska⁵, Julien Bobe¹, Jean-Paul Concorde⁶, Hugues Parinello⁷, Laurent Journot⁷, Christophe Klopp⁸, John Postlethwait⁹, Manfred Scharl¹⁰, Amaury Herpin^{1#}, Yann Guiguen^{1#&}.

#: Equal contributors.

&: Corresponding author.

1. INRA, UR1037 LPGP, Campus de Beaulieu, Rennes, France.
2. GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.
3. Fédération d'Ille-et-Vilaine pour la pêche et la protection du milieu aquatique (FDPPMA35), 9 rue Kérautret Botmel – CS 26713 – 35067 RENNES
4. Department of Marine Biology and Ecology, Institute of Oceanography, Faculty of Oceanography and Geography, University of Gdansk, 81-378 Gdynia, Al. M. Piłsudskiego 46, Poland.
5. Laboratory of Genomics, Department of Animal Genetics and Molecular Biology, National Research Institute of Animal Production, 32-083 Balice, Krakowska 1, Poland.
6. INSERM U1154, CNRS UMR7196, MNHN, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris Cedex 05, France.
7. Institut de Génomique Fonctionnelle, IGF, CNRS, INSERM, Univ. Montpellier, F-34094 Montpellier, France.
8. INRA, Sigene, Toulouse, France.
9. Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA.
10. University of Würzburg, Physiological Chemistry, Biocenter, 97074 Würzburg, Germany; Comprehensive Cancer Center Mainfranken, University Hospital, 97080 Würzburg, Germany; Hagler Institute for Advanced Study and Department of Biology, Texas A&M University, College Station, Texas 77843, USA.

Authors contributions

Qiaowei Pan

- Screening of the fosmid library for *amha* and *amhby* fosmids, subcloning, sequencing, assembly of the fosmids, and *in situ* hybridization
- Knock-out of the *amhrII* with Crispr-cas9
- Subcloning of the cDNA of *amhby*, *amha*, and *amhrII*
- RNA extraction, qPCR and analysis of the results
- Phylogenetic reconstruction and synteny analysis of *Amh* in the teleosts
- *In situ* hybridization of *amha* and *amhby*
- Analysis of the RAD-seq data: GWAS and F_{ST} analysis, reanalysis with the RADSex pipeline
- Primer design, with the exception of the primers used in fosmid screening and in the *amhby* KO experiment
- DNA extraction and genotyping of all the animals used in the study, with the exception of *amhby* KO
- Genotyping and histological analysis of the gonads in the animal used in the temperature manipulation experiment.

Other experiments

- *amhby* KO: Jean-Paul Concordet synthesized the TALENs. Ayaka Yano, Elodie Jouanno and Estelle Vigouroux performed the *amhby* KO experiment.
- RAD-Seq based genetic map: Elodie Jouanno and Estelle Vigouroux extracted the DNA and determined the gonadal sex of the animal. René Guyomard made the genetic map
- Fluorescent in-situ hybridization: Konrad Ocalewicz, Wojciech Witariski, and Monika Bugno-Poniewierska

Contribution of other authors

- Romain Feron: development and implementation of the RADSex pipeline, also assisted with data analysis, visualization and revision of the manuscript
- Elodie Jouanno, Ming Wen and Brigitte Mourot: assisted with molecular experiments
- Alizée Frezel, Amélie Patinote, and Frédéric Borel: in-door fish facility at INRA-LPGP
- Jean-Mickael Busnel: provided Northern pike embryos annually at Pisciculture du Boulet
- Hugues Parinello and Laurent Journot: library preparation and sequencing at the MGX platform
- Christophe Klopp: genome sequencing and assembly at the Genotoul platform
- Julien Bobe: provided access and assistance with RNA-seq data from Phylofish database
- John Postlethwait, Manfred Scharl and Yann Guiguen: designed the Phylosex project
- Yann Guiguen and Amaury Herpin: supervised my thesis

Abstract

Teleost fishes, thanks to their rapid turnover of sex chromosomes and their diversity of master sex determining (MSD) genes, provide remarkable opportunities to study sex chromosome evolution relating to the MSD genes they harbor. In this study, we present genomic, cytogenetic, functional, and expression evidence for a male-specific duplicate of *amh* as the MSD gene in Northern Pike (*Esox lucius*). In addition, we show that environmental factors can interfere with the function of this MSD. Through phylogenetic and synteny reconstruction, we demonstrate that this MSD originated from a lineage-specific duplication and was subsequently translocated to the telomere of the future Y chromosome. These results highlight further substantiate the pivotal role of the *amh* pathway in teleost sex determination and provide the first empirical example of the effect of an inserted MSD gene on a pair of morphologically homomorphic sex chromosome.

Introduction

Sex chromosomes contain a species genetic sex determinants and present a unique evolutionary puzzle. On one hand, sex chromosomes have evolved independently in many taxa, and unrelated sex chromosomes often show converging properties. On the other hand, the age of sex chromosomes, the size of their non-recombining region, and their turnover rate can differ drastically across taxa (Beukeboom and Perrin, 2014; Bachtrog *et al.*, 2014). Initial insights on sex chromosome evolution came from detailed studies in *Drosophila* and in mammals (Lahn and Page, 1999; Bachtrog, 2003; Skaletsky *et al.*, 2003; Vicoso and Charlesworth, 2006; Soh *et al.*, 2014), but recent research on other vertebrate groups, such as avians (Smith *et al.*, 2009; Smeds *et al.*, 2015), reptiles (Ezaz *et al.*, 2009; Gamble and Zarkower, 2014), amphibians (Yoshimoto *et al.*, 2008; Miura, Ohtani, and Ogata, 2012; Scharl, 2015; Rodrigues *et al.*, 2017) and teleost fishes (Kikuchi and Hamaguchi, 2013; Heule *et al.*, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017), has provided new information that helps us understand the evolution of sex chromosomes. The teleosts display the highest diversity of sex determination systems in vertebrates, including several types of monofactorial and polygenic systems (Mank and Avise, 2009; Kikuchi and Hamaguchi, 2013). Moreover, sex determination systems in fish can differ even between very closely related species, as illustrated in the group of Asian ricefish (genus *Oryzias*) (Takehana *et al.*, 2007; Takehana *et al.*, 2007; Takehana, 2008; Tanaka *et al.*, 2007; Nagai *et al.*, 2008; Myosho *et al.*, 2015), and sometimes even among different populations of a same species, like in the Southern platyfish, *Xiphophorus maculatus* (Volf and Scharl, 2002). In addition, in some species, genetic factors can interact with environmental factors, most commonly temperature (*i.e.* *Odontesthes bonariensis*, Yamamoto *et al.*, 2014), generating intricate sex determination mechanisms. Beside this remarkable dynamics of sex determination systems, the rapid turnover of sex chromosomes in the teleosts provides many opportunities to examine sex chromosomes pairs at different stages of differentiation corresponding to the variety of master sex determining (MSD) genes that have been adopted during transition of sex chromosomes. Indeed, studies on teleosts have unraveled many cases of rapid transition of sex determination systems and neo-sex chromosomes. In addition, recent studies on fish sex determination have revealed a dozen new master sex determining (MSD) genes, which added an additional layer of understanding of sex chromosome evolution and transition.

The evolutionary trajectories of sex chromosomes are driven by the acquisition of a sex determining role of new MSD genes. The origin of new MSD genes falls into two categories: either gene duplication followed by sub- or neo-functionalization, or allelic diversification.

The first mechanism starts with duplication of an autosomal gene, often as part of a larger chromosome fragment, followed by the insertion of the duplicated fragment into a different chromosome, the acquisition of a new sex determining function by a gene in the duplicated fragment, and finally the origin of a new sex chromosome. The second mechanism starts with the diversification of two alleles of a gene in the future sex chromosome pair; one allele then drives sex determination, while the other allele either retains a non-sex-determining function or favors the development of the opposite sex (Kikuchi and Hamaguchi, 2013). To date, the teleosts is the only group where examples of both gene duplication and allelic diversification mechanisms have been found (Kikuchi and Hamaguchi, 2013; Heule *et al.*, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017).

Another important insight coming from studies on teleost sex determination is that how the MSD genes regulate the downstream pathway could impact profoundly the evolution of the sex chromosome. A remarkable example of allelic diversification is found in the teleost *Takifugu*, where the sex chromosomes are almost identical except for a single missense SNP (single nucleotide polymorphism) in the AMH receptor (Kamiya *et al.*, 2012). Without knowing the crucial and bifurcating role AMH signaling plays in the sex determination gene network in teleosts, it would seem bewildering that this pair of extremely homomorphic sex chromosomes with no suppression of recombination showed strong evolutionary conservation among three species of *Takifugu* over 10 million years. This finding runs contrary to the vast genetic diversification of both sex chromosomes resulting from suppression of recombination that is predicted by theories of sex chromosome evolution (Bergero and Charlesworth, 2009). The *Takifugu* example illustrates the need to integrate molecular studies of MSD gene function with turnovers of SD systems to better identify patterns and processes of sex chromosome evolution in different taxa.

Northern pike (*Esox lucius*) is a large and long-lived keystone predatory teleost species found in freshwater and brackish coastal water systems in Europe, North America, and Asia (Forsman *et al.*, 2015). It has emerged as a significant model species for ecology because of its important role as a top predator that shapes the structure of local fish communities, and is also a valuable food and sport fish (Raaf, 1988). Northern pike exhibits a sexually dimorphic growth rate: females grow faster and can reach a larger size than males, making females more economical than males for aquaculture (Raaf, 1988). To obtain female monosex stocks, fisheries have developed several artificial gynogenesis methods (Ihssen *et al.*, 1990; Purdom, 1983; Luczynski *et al.*, 2007). Despite this interest, little is known about genetic sex determination in *E. lucius* beyond the fact that males are the heterogametic sex (Luczynski *et al.*, 1997).

As the most studied species in the ancient lineage of *Esociformes*, Northern pike is a key species for studying the additional whole genome duplication of the sister group *salmonids* (Leong *et al.*, 2010; Rondeau *et al.*, 2014; Lien *et al.*, 2016). Consequently, important genomic resources such as a whole genome assembly anchored on chromosomes (Rondeau *et al.*, 2014), and tissue-specific transcriptomes (Pasquier *et al.*, 2016) have been recently published for *E. lucius*. In salmonids, an unusual master sex determination gene arising from the duplication of an immune-related gene has been identified (Yano *et al.*, 2012). This master SD gene, named *sdY*, was found to be conserved in the majority of inspected salmonid species, but was not found in *E. lucius* (Yano *et al.*, 2013). The restriction of *sdY* to the salmonids raises the question of what was the ancestral MSD before the emergence of the “unusual” *sdY*. The first step to answer this question is to identify the genetic component responsible for sex determination in *E. lucius*.

In this study, we provide support from gene expression, functional studies, and genomic data for a male-specific duplicate of *amh* as the MSD gene in *E. lucius*. We show that this male-specific paralog of *amh* originated from a lineage-specific duplication and translocation event, and acquired its MSD function through regulatory rewiring. We demonstrate that the MSD gene is located in the telomere of the Y chromosome. Finally, we show that sex determination in *E. lucius* can be influenced by environmental factors, illustrating the flexibility of the genetic sex determination system in this 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 and European regulations on animal welfare. 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 aquaria, 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 fish size. After reaching a length of 4-5 cm, fish were fed with rainbow trout fry.

For genotyping, small fin clips were taken from caudal fins after administering 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). Temperature treatment experiments were performed on 60 Northern pike from the same parental cross. All animals were kept at 12 °C until hatching and then separated into two groups of 30 animals each. One group was raised at a temperature of 12 °C while the other group was raised at a temperature of 18°C.

Animals raised at 18 °C had a higher growth rate and were sampled at 163 days post fertilization (dpf), while animals raised at 12 °C were sampled at 199 dpf to compensate for their slower growth rate. Fin clips from each animal were kept for genotyping, and gonadal sex was first visually assessed and further confirmed by histological analysis when necessary.

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 European *E. lucius* 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.

Pike fosmid clones were screened by several rounds of purification by PCR to identify individual fosmids containing *amhby* and *amha*. To sequence the two 40 kb fosmids, purified fosmid DNA was first shattered into approximately 1.5 kb fragments using a Nebulizer kit supplied in the TOPO shotgun Subcloning kit (Invitrogen), 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 (Solovyev, 2007), and the resulting cDNA sequences were compared to the corresponding transcripts publicly available in the Phylofish database (Pasquier *et al.*, 2016). Shared identity between cDNA sequences was calculated after alignment using Clustal Omega (Sievers *et al.*, 2011) implemented on EMBL-EBI (McWilliam *et al.*, 2013; W. Li *et al.*, 2015) with default settings.

Global pairwise alignments of the two transcripts were performed using the mVISTA LAGAN program with default settings (Brudno *et al.*, 2003). Shared identity and similarity between protein sequences were calculated using EMBOSS Water (Rice *et al.*, 2000). For each *amh* paralog, the putative promoter was arbitrarily defined as the 5kb genomic sequence upstream of the start codon. The two putative promoters were compared using PromoterWise (Rice *et al.*, 2000) and mVISTA LAGAN program (Brudno *et al.*, 2003).

Phylogenetic reconstruction of Amh proteins was performed using the maximum likelihood method implemented in PhyML 3.0 (Guindon *et al.*, 2010), and the proper substitution model was determined with PhyML-SMS (Lefort *et al.*, 2017). 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 **Table S1**.

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 (*Lepisosteus oculatus*), zebrafish (*Danio rerio*), Atlantic cod (*Gadus morhua*), three-spined stickleback (*Gasterosteus aculeatus*), fugu (*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 (www.genomicus.biologie.ens.fr). For Northern pike, genes located upstream and downstream of *amha* on LG08 were deduced from the genome assembly (Eluc_V3, GenBank assembly accession: GCA_000721915.3).

Fluorescent *in situ* hybridization (FISH) on chromosome

For cytogenetic analyses, twelve 8-month old males and nine females *E. lucius* were obtained from a local fish farm in Ruś (Poland). Somatic metaphase plates were prepared from cephalic kidney cells by the conventional air-drying technique with slight modification (Ocalewicz *et al.*, 2013). To accumulate metaphases, fish were injected with 0.3% colchicine (Sigma-Aldrich, St. Louis, MO, USA) solution (0.25 ml/100 g body weight) for 60 minutes before being sacrificed. Portions of cephalic kidneys were removed, placed in 5 ml of a hypotonic solution (0.075 M KCl), and homogenized using glass homogenizers. Cell suspensions were then transferred to 10 ml glass tubes, hypotonized for 40 minutes at room temperature, and fixed with freshly prepared ice cold fixative (methanol: acetic acid, 3: 1). Chromosomes were stained with 4',6-diamidino-2-phenylindole DAPI (Vector, Burlingame, CA, USA) for visualization. Constitutive heterochromatic regions were visualized using the C-banding technique described by Sumner (1972).

Fluorescent *in situ* Hybridization was performed according to the method described by Bugno *et al* (2010) with some modifications. Fosmids containing *amha* and *amhby* sequences were labelled with the biotin-nick translation mix (Roche, Mannheim, Germany) according to the manufacturer's instructions. Metaphase spreads were digested with RNase and pepsin, dehydrated in an ethanol series (70%, 80% and 95%), and dried at room temperature. 10µl of the labelled BAC probes suspended in the hybridization solution were dropped onto chromosome slides, covered with coverslips and sealed with rubber cement. Co-denaturation of labelled DNA and chromosomal DNA was performed on a hot plate at 75 °C for 4 min. Samples were hybridized overnight in a moist chamber at 37 °C. Post-hybridization washes were performed at 45 °C in the following order: twice with a solution of 50% formamide and 2X SSC for 5 min., and twice with 2X SSC. Then, slides were incubated in Detergent Wash Solution for 4 min. Hybridization signals were detected using Star*FISH© detection kit (FITC) for the biotin labelled probes (Cambio, Cambridge, UK) according to the manufacturer's instruction. Chromosomal spreads were then stained with DAPI and analyzed under a fluorescence microscope (Axiophot, Zeiss, Germany) equipped with a digital charge coupled device camera, driven by Lucia software (Laboratory Imaging LTD, Prague, Czech Republic).

DNA extraction and genotyping

All fin clips were collected and stored at 4°C in 75-100% ethanol until DNA extraction. To obtain DNA for genotyping, cells were lysed with 5% Chelex and 25 mg of Proteinase K at 55°C for two 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 (Gharbi *et al.*, 2006). To obtain DNA for sequencing, extractions were performed using NucleoSpin Kits for Tissue (Macherey-Nagel, Duren, Germany) following the producer's protocol. When using this protocol, DNA concentration was quantified with a NanoDrop ND2000 spectrophotometer (Thermo scientific, Wilmington, Delaware). Primers for genotyping were designed using Perl-Primer (version 1.1.21, Marshall, 2004). Primer sequences and corresponding experiments can be found in **Table S2**.

RNA extraction, cDNA synthesis and qPCR

Ten animals were sampled at 54, 75, 100, and 125 days post fertilization (dpf). At the first three time points, the whole trunk was collected for RNA extraction because gonads were too small to isolate routinely at these stages. At 125 dpf, gonads were large enough to be isolated and they were collected for RNA extraction. The genotypic sex of each animal was determined based on *amhby* presence and results are listed in **Table S3**.

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, Delaware). Reverse transcription (RT) was performed by denaturing a mix of 1µg of RNA and 5 µL of 10mM 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.

qPCR primers were designed using Perl-Primer (version 1.1.21, Marshall, 2004) on intron-exon junctions to avoid genomic DNA amplification (primers listed in **Table S2**). Primer pairs were diluted to 6 µg/µL for qPCR. qPCR 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 1µL 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. Samples 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*, *efla*, *gapdh*, *eftud2*, *ubr2*, and *ccdc6b*) following classical qPCR normalization procedures (Vandesompele *et al.*, 2002).

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 (**Table S2**). 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 standard 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. Hybridization was performed using an *in situ* Pro intavis AG robotic station according to the following procedure: male and female samples were dehydrated, 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 1hr 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 hr. 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 2hr, and incubated for 12h with addition of alkaline phosphatase coupled with anti-digoxigenin antibody for *amha* and anti-Fluorescein antibody for *amhby* (1:2000, Roche Diagnostics Corp). Samples were then washed with PBS solution, and color reactions were performed with NBT/BCIP (Roche). After visual inspection of coloration, samples were dehydrated and embedded in plastic molds with paraffin with a HistoEmbedder (TBS88, Medite, Burgdorf, Germany) instrument. Embedded samples were sectioned into 5 mm 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 and CRISPR-cas9 knock-out

TALENs were assembled following a method derived from Huang *et al.*, 2011. For each subunit, the target-specific TALE DNA binding domain consisting of 16 RVD repeats was 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 Δ 152 Nter TALE, +63 Cter TALE, and FokI cDNA sequences with the appropriate half-TALE repeat (derived from the original pCS2 vector (Huang *et al.*, 2011)). 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).

Three pairs of TALENs (Reyon *et al.*, 2012; Cade *et al.*, 2012), called T1, T2 and T3, were designed on exon I and exon II of *amhby* (**Figure S3A**) and diluted to 50 ng/L and 100 ng/L. Groups of embryos were microinjected with one pair of TALENs at the one-cell stage. 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 for genotyping with primers flanking the TALEN targets. Only males with a disrupted *amhby* sequence were kept until the following reproductive season. The sperm of three one-year old males carrying an *amhby* disrupted sequence was collected and then used for *in vitro* fertilization with wild-type eggs.

F1 individuals were genotyped for *amhby* TALEN targeting sites using primers flanking the TALEN targets, and mutants were kept until 153 dpf. Animals carrying a disrupted *amhby* sequence were then dissected and their gonads were subjected to histological analysis to inspect the phenotypes resulting from *amhby* knockout.

CRISPR-Cas9 guide RNAs (gRNAs) targeting three sites located in exon 9 and in intronic regions immediately upstream and downstream of exon 9 of *amhrII* were designed with the ZiFiT software (Sander *et al.*, 2007, 2010) (**Figure S3B**). Annealed oligonucleotide pairs were cloned into a pDR274 vector and linearized with *DraI*, and pCS2-Ncas9n plasmids were linearized with *NotI* (Hwang *et al.*, 2013; Mali *et al.*, 2013). *In vitro* transcription of the gRNA and Cas9 RNA from the T7 initiation site was performed using Maxiscript T7 kit (Applied Biosystems). Synthesized gRNAs were then mixed and adjusted to 25 ng/L, and Cas9 RNA was adjusted to 100 ng/L. Embryos were microinjected with both Cas9 RNA and a mixture of the three gRNAs at the one-cell stage. Fin clips were collected from these embryos two months post fertilization for genotyping with primers flanking the deletion site on *amhrII* (**Figure S3**) and with primers for *amhby*. Chromosomally XY individuals with *amhby* and carrying a mutation in *amhrII* were maintained to the following reproductive season, when spawning was induced with Ovaprim (Syndel, Ferndale, Washington) following the protocol by Szabó (2003). *In vitro* fertilization was performed with eggs and sperm produced by XY *amhrII* mutants. F1 animals were genotyped with the same primers as used in the genotyping of G0 animals at 4 weeks and 9 weeks post fertilization to detect *amhrII* mutations *see note at the end of material and methods section

Genetic map construction

A Restriction Associated DNA Sequencing (RAD-Seq) library was constructed from genomic DNA extracted from the fin clips of two parents and 78 of their offspring according to standard protocols (Amores *et al.*, 2011), and sequenced in one lane of Illumina HiSeq 2500. Raw reads were analyzed with the Stacks (Catchen *et al.*, 2011) 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 using homemade scripts before genetic map construction to 1) remove loci found in fewer than half of the offspring; 2) remove loci with more than two triploid (or higher polyploid level) genotypes (in conserved loci, these unexpected genotypes were replaced by missing data); 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 CARTHAGENE (de Givry *et al.*, 2005), 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 and a distance of 0.3 centiMorgans 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}` (for more details on these settings, see the CARTHAGENE user guide). The resulting map was visualized with MAPCHART version 2.1 (Voorrips, 2002).

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

RAD-Seq reads from the genetic map were analyzed with the Stacks program version 1.44. Raw reads were filtered and demultiplexed using the *process_radtags* wrapper script with all settings set to default. Demultiplexed reads were mapped to the genome assembly of *E. lucius* (Elu_V3, GenBank assembly accession: GCA_000721915.3) using BWA (version 0.7.15-r1140, H. Li and Durbin, 2009) with all settings set to default. The resulting BAM-files were run through the *ref_map.pl* pipeline, with a minimum stack depth of 10 ($m=10$) and all other settings set to default. Results of *ref_map.pl* were analyzed with the *populations* program, 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, Rentería *et al.*, 2013) to estimate association between variants and phenotypic sex. A Manhattan plot was constructed in R with homemade scripts showing $-\log_{10}(\text{Fisher’s test } p\text{-value})$ for all 25 LGs of the *E. lucius* genome.

An additional analysis was performed to analyze both polymorphic and non-polymorphic markers, including potentially hemizygous markers. In this analysis, demultiplexed reads were grouped into stacks using the *denovo_map.pl* wrapper script with minimum number of reads to form a stack set to 10 ($m=10$), maximum number of mismatches allowed when forming stacks in one individual set to 0 ($M=0$), maximum number of mismatches allowed when merging stacks in the catalog set to 0 ($N=0$), gapped assembly deactivated (`--gapped` not set), calling haplotypes from secondary reads disabled (`H` not set), *ustacks* max locus per stack set to 1 (`--max_locus_stacks=1`), and *ustacks* lumberjack step deactivated (`--keep_high_cov` set). These parameter values ensured that each stack represented a single allele (or haplotype) from a single locus. The distribution of haplotypes between males and females was calculated using a custom pipeline, and results were presented in a heatmap with number of males in which a haplotype is present as abscissa (x-axis), number of females in which a haplotype is present as ordinate (y-axis), and number of haplotypes present in x males and y females as color of the tile at coordinates (x , y). For each combination of number of males and number of females in which a haplotype is present, the probability of association with sex was estimated with a chi-square test, applying Bonferroni correction. Sequences from haplotypes with male-biased or female-biased distribution were extracted and mapped to the genome assembly using BWA (version 0.7.12) with all settings to default. The ratio $(M - F) / (M + F)$, where M is the number of males and F the number of females in which the haplotype is present, was plotted against the mapping position of each haplotype to visualize haplotype distributions between males and females along each LG.

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). Each embedded sample was 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).

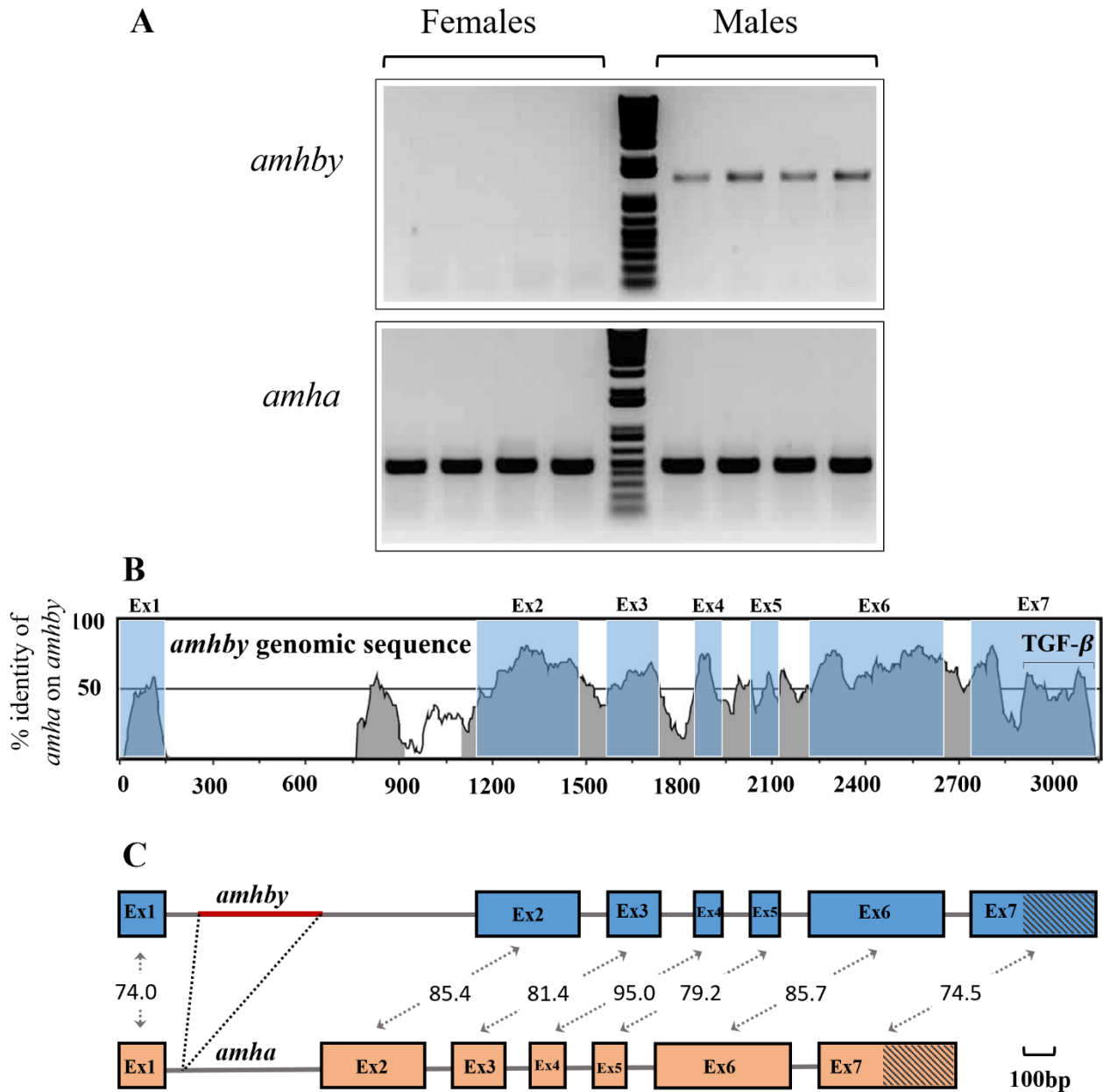


Figure 1: Sequence identify comparison and amplification of *amha* and *amhby* in male and female genomes. **A:** PCR amplification of *amha* and *amhby* in male (n=4) and female (n=4) genomic DNA samples from *E. lucius*. **B:** Global pairwise alignment between *amha* and *amhby* genomic sequences is presented with *amhby* as the reference. The shared identity between the two sequences is plotted along the *amhby* genomic sequence and each exon of *amhby* is highlighted in blue with the TGF- β domain in the Exon 7. **C:** *amhby* and *amha* gene structure are presented in a schematic fashion from the start to the stop codon. Exon1-Exon7 are represented by blue boxes with shared percentage identity indicated in between *amha* and *amhby*, and introns are represented by solid grey lines. The red segment in intron 1 represents the *amhby* specific insertion. The TGF- β domains are indicated with diagonal lines.

Statistical analyses

All statistical analyses, including Wilcoxon signed rank test and Chi-squared test, were performed with R (version 3.3.2).

* Notes

The characterization of the F1 *amhrII* mutants is not included in this version of the thesis as at the time of the writing, the pikes had not yet grown to sufficient size. However, the mutants have already been sampled on the 17th of Oct, 2017.

Results

Characterizing a male-specific duplication of *amh* in *E. lucius*

BLAST searches using *amh* sequence of *Salmo salar*, two *amh* transcripts in the tissue-specific transcriptome libraries of *Esox lucius* from the PhyloFish online database (phylofish.signae.org). While one *amh* transcript was expressed in all tissues investigated, including testis and ovary, the other transcript was found to be exclusively expressed in the testis (**Figure S1**). The two *amh* transcripts share 78.9% nucleotide identity.

Amplification of genomic DNAs using primers specific to each transcript showed that one copy of *amh* was present in both male and female genomes, while the testis-specific copy of *amh* was only present in the XY male genome (**Figure 1A**). The association of the testis-specific copy with male phenotype was confirmed in 140 individuals among three populations. We called the male-specific *amh* paralog *amhby* (Y-chromosome-specific *amh* paralog b) and the other paralog *amha* (*amh* paralog a).

Fluorescent *in situ* hybridization (FISH) of *amha* and *amhby* was performed on interphase chromosomes in both male and female Northern pike using fosmids containing either gene as a probe. Fifty uniarmed chromosomes were observed (FN=50) with no clear sex-specific pattern of heterochromatin distribution.

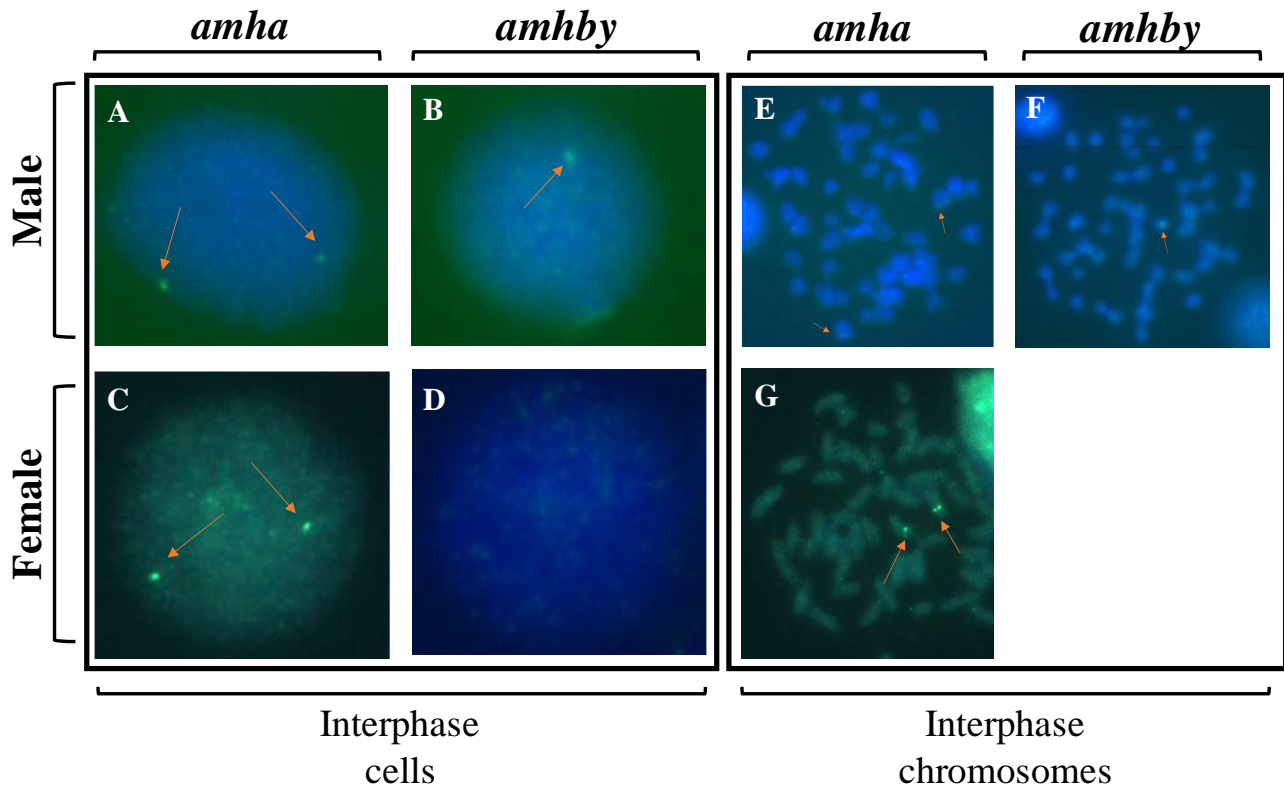


Figure 2: Fluorescent *in situ* hybridization of *amha* and *amhby* in *E. lucius*.

A-D: Two spots of *amha* signal are detected on interphase cells in both male (A) and female (C) samples, while only one spot of *amhby* signal is detected in males (B) but not in females (D). **E-G:** *amha* signal is detected on two chromosomes in both male (E) and female (F) samples, while *amhby* signal is only detected in on chromosome in male samples (G). Orange arrows point to the fluorescent signal of *amha* or *amhby*.

For *amha*, fluorescent signals after FISH were observed on two chromosomes in both male and female cells, and two spots were also observed in the male and female interphase nuclei (**Figure 2**). In contrast, for *amhby*, fluorescent signal was detected in a single medium-sized chromosome in males but not females and only one spot was observed in male interphase nuclei and none in female interphase nuclei (**Figure 2**).

To compare the local genomic sequences as well as putative promoter regions upstream of the coding sequence of each paralog, one fosmid containing *amha* and one containing *amhby* were obtained and sequenced. The nucleotide identity between corresponding exons of *amha* and *amhby* ranged from 74% to 95% (**Figure 1B and 1C**). A 396 bp insertion specific to intron 1 of *amhby* was identified as the only structural difference between the two genes. Hardly any sequence similarity was found between their putative promoters (5 kb region upstream of each start codon), except in a 1020 bp repetitive element located 1309-2321 bp upstream of *amhby* and 1184-2212 bp upstream of *amha* sharing 74% sequence identity between the two promoter regions. For *amha*, this 5 kb putative promoter included the entire intergenic sequence up to the *dot11* gene, located 4.3 kb from the start codon of *amha*. No additional similarity was identified when comparing this *amha* putative promoter to the additional 22 kb region upstream of *amhby* contained in the *amhby* fosmid, and apart from the Amhby protein, no other protein was found in the *amhby* fosmid sequence.

Protein sequences predicted from *amha* and *amhby* transcript sequences contain 580 amino acids for Amha and 560 amino acids for Amhby. These two protein sequences share 68.7% identity and 78.4% amino-acids similarities. Both proteins had a complete TGF- β domain, comprising the last 95 amino acids, with seven canonical cysteine residues within their C-terminal sequence.

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

Prior to the discovery of *amhby* in *E. lucius*, male-specific duplications of *amh* had been identified in Patagonian pejerrey (*Odontesthes hatcheri*) (Hatorri *et al.*, 2012) and in Nile tilapia (*Oreochromis aureus*) (Li *et al.*, 2015), but whether these *amh* duplicates are lineage-specific or results of shared ancestry remained unclear. To clarify, we constructed a phylogeny of Amh from nine species across the teleosts - including all three species with male-specific *amh* duplications - and included spotted gar Amh as outgroup (**Figure 3A**).

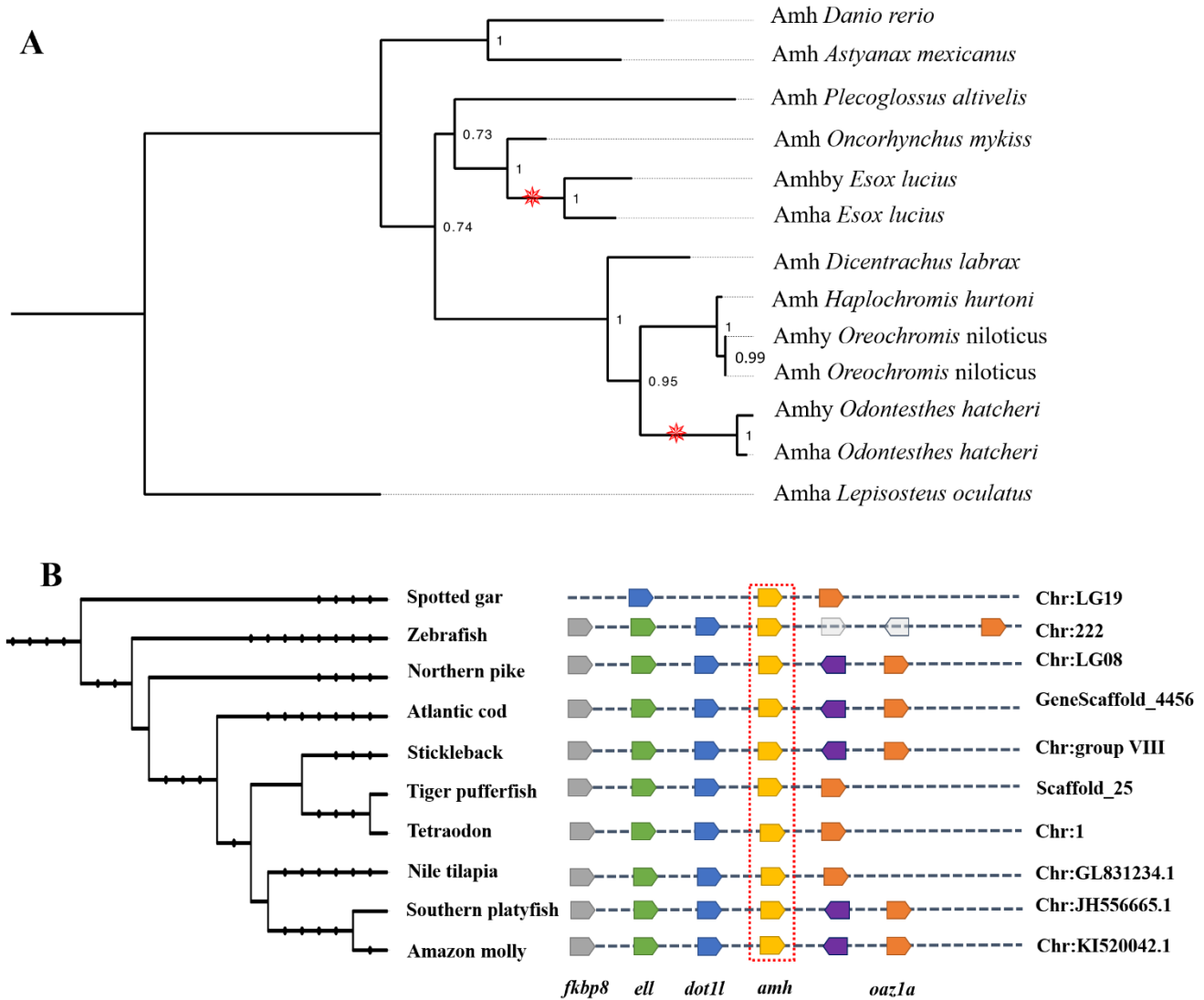


Figure 3: Evolution of Amh in the teleosts. **A:** Phylogenetic reconstruction of the teleosts Amh protein orthologs. Phylogenetic tree was constructed with the maximum likelihood method (bootstrap=1000) and the numbers at the tree nodes are the bootstrap values. Spotted gar (*Lepisosteus oculatus*) Amh was used as an outgroup. Branches with Amh duplication are indicated by the red pinwheel stars. **B:** Synteny map of the genomic regions around *amh* genes (highlighted by the red dotted-line box) in the 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 *Esox lucius*, *amha*, which is located on LG08, is used in this analysis.

In this protein phylogeny, the ‘normal’ and sex-specific Amh paralogs from *Esox lucius*, *Odontesthes hatcheri* and *Oreochromis aureus* cluster together within each species with significant bootstrap values, suggesting that these three pairs of *amh* paralogs were not derived from the same ancestral duplication event, but rather derived from three independent and lineage-specific duplication events.

To determine the origin and genomics of the two pike *amh* paralogs, we generated a map of conserved syntenies for *amh* in several teleost species including the spotted gar (*Lepisosteus oculatus*) as outgroup (**Figure 3B**). The genomic region around *amha* in *E. lucius* was inferred from the genome assembly (GenBank assembly accession: GCA_000721915.3) and verified by the sequence of the fosmid containing *amha* that also contained *dot1l*. The *amhby* sequence was not found in the Northern pike reference genome, even though this genome is recorded as a male genome. Because no other known gene was found in the sequence of the fosmid containing *amhby*, this gene could not be included in the synteny map. Genes located upstream (*i.e.* *dot1l*, *ell*, and *fkbp8*) and downstream (*i.e.* *oazla*) of *amha* showed conserved synteny in all teleost species included in the analysis, which suggests that *amha* was the ancestral *amh* copy in *E. lucius* and that *amhby* evolved from a duplication of *amha*.

amhby* is expressed prior to sexual dimorphic molecular differentiation of the gonads in male *E. lucius

The expression kinetics of *amhby* and *amha* were measured by qPCR at 54 dpf (days post-fertilization), 75 dpf, 100 dpf and 125 dpf (**Figure 4A-B**). Expression of *amha* was significantly higher in males than in females at 100 dpf (Wilcoxon signed-rank test, $p=0.043$) but no significant sex dimorphic pattern was observed at 75 dpf ($p=0.915$) and 125 dpf ($p=0.915$), and no expression was detected at 54 dpf. In contrast, the expression of *amhby* was detected only in males, where it had already begun before the first sampling stage (54 dpf), and it increased exponentially over time ($R^2=0.79$). To better define the timing of *amhby* expression with respect to other sex differentiation events in the gonads, we investigated the expression of *dmrt1*, *cyp19a1a*, and *gsdf*, three genes known for their role in gonadal sex differentiation. The expression of these sex differentiation genes was not detected until 100 dpf, much later than the beginning of the expression of *amhby* (**Figure S2**). Moreover, only *cyp19a1a* was significantly expressed at that time point of 100 dpf (Wilcoxon signed-rank test, $p=0.11$, $p=0.014$, and $p=0.11$ for *dmrt1*, *cyp19a1a*, and *gsdf* respectively), and none of these genes had a significant sex-dimorphic expression at 125 dpf (Wilcoxon signed-rank test, $p=0.75$, 0.11, and 0.11 for *dmrt1*, *cyp19a1a*, and *gsdf* respectively).

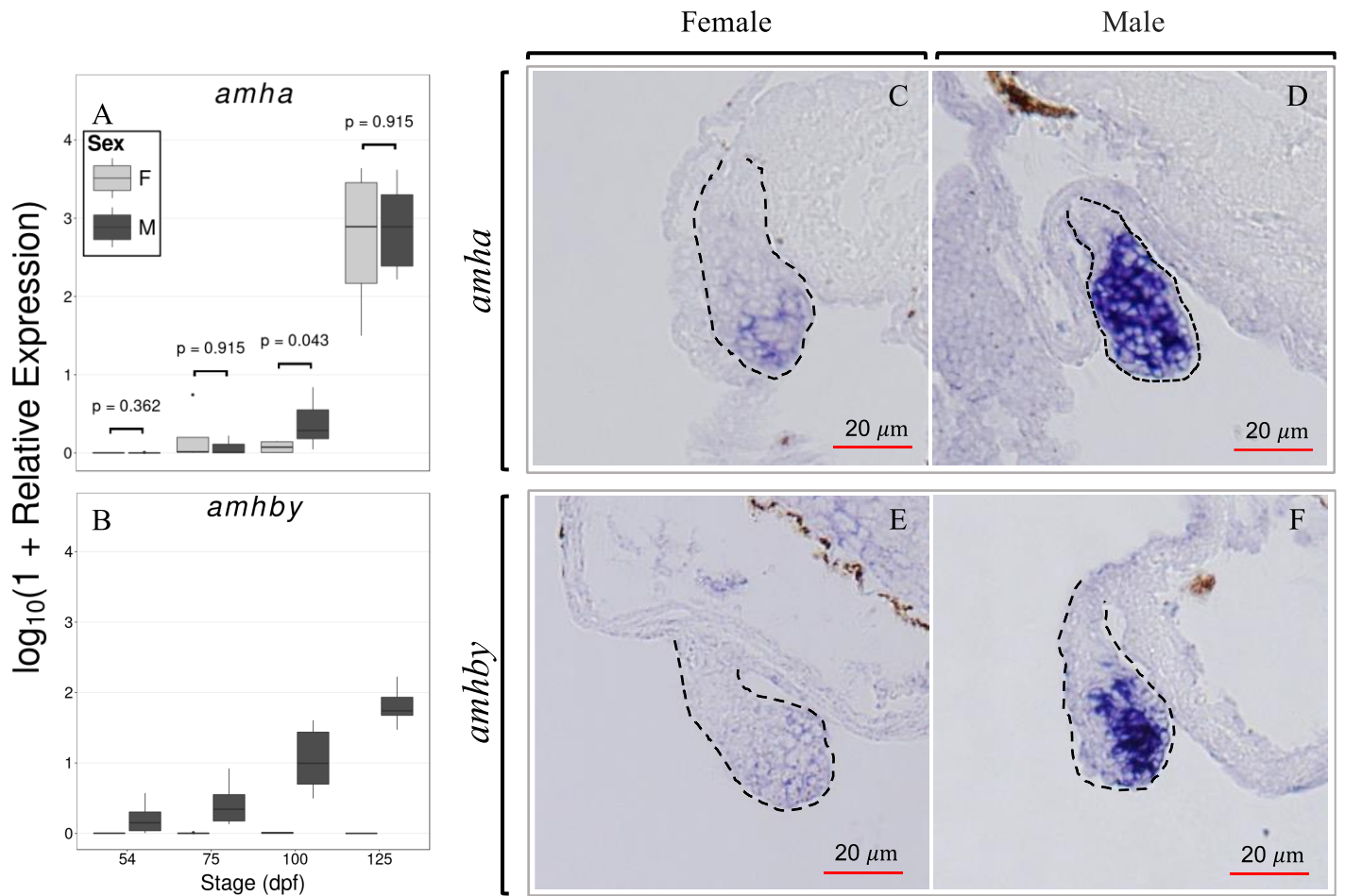


Figure 4: Temporal and spatial expression of *amha* and *amhby* mRNA in male and female developing gonads. **A-B:** Boxplot showing the first quantile, median, and the third quantile of the temporal expression of *amha* (A) and *amhby* (B) measure by qPCR during early development of *E. lucius*. Outliers are displayed as a dot. The expression of mRNA of *amha* and *amhby* were measured at 54, 75, 100, and 125 days post fertilization (dpf) in male and female samples of *E. lucius* and the log₁₀ of their relative expression is presented on the graph. P-values are given for Wilcoxon signed rank test between male and female expression at each time point. **C-F:** *In situ* hybridization revealed the localization of *amha* in both 85dpf female (C) and male (D) gonads, with stronger signal in the male gonads. The signal of *amhby* mRNA expression is only detected in the male gonad with high concentration (F), but not in female gonad (E). The red scale bars denote 20 μ m and the dashed lines outlines the gonadal sections.

The expression of *amhrII*, the putative specific type II receptor for Amh, was also not detected until 100 dpf, which was later than the beginning of the expression of both *amha* and *amhby*. Furthermore, the expression of *amhrII* did not follow a sex-dimorphic pattern at either stage examined (Wilcoxon signed-rank test, $p=0.11$ at 100 dpf and $p=1$ at 125 dpf). However, RNA-Seq data obtained from the Phylofish online database (Pasquier *et al.*, 2016) in adult tissues showed expression of *amhrII* restricted to gonads, with a 35-fold higher expression in testis than in ovary (**Figure S1C**), showing male-biased expression of this gene in adult gonads.

To determine whether *amhby* is expressed in germ cells or support cells of the gonad, we performed *in-situ* hybridization (ISH) on histological sections of the entire trunk of male and female *E. lucius* at 80dpf using RNA probes specifically targeting *amha* and *amhby* (**Figure 4C-4F**). Signal for *amha* was detected in the gonads of both male (**Figure 4D**) and female (**Figure 4C**) samples, but the signal was much stronger in male gonads, in agreement with the higher expression of *amha* in males compared to females measured by qPCR at 100 dpf. In contrast, *amhby* signal was not detected in the female samples (**Figure 4E**), while a strong *amhby* signal was detected in male gonads (**Figure 4F**). Although specific markers would be required to identify the type of cells expressing *amhby* and *amha*, it could be inferred that both genes are likely to be expressed in somatic cells surrounding the germ cells. In addition, no morphological differentiation was observed between male and female gonads at this stage.

amhby* is necessary for testicular differentiation and the sex determination (SD) signal is passed through *amhrII

We knocked out *amhby* using three pairs of TALENs targeting exon 1 of the Amhby protein (**Figure S3A**). Genotyping of G0 animals showed that only one TALEN pair was effective in inducing a deletion in the *amhby* sequence. Overall, 12 *E. lucius* males among the 36 (33.3%) males treated with this pair of TALENs were carrying a disrupted *amhby* that resulted in a truncated and non-functional protein. These G0 mutant males were crossed with wild-type females to generate non-mosaic F1 mutants with disrupted *amhby*. Offspring were obtained from three *amhby* mutant males and only XY offspring were maintained until the beginning of testicular gametogenesis at 153 dpf. Gonads from the F1 XY mutants were subjected to histological analysis and compared with those of wild-type control XY males (N=4) and control XX (N=4) females. Control animals developed normal ovary and testis, but all 23 XY mutants carrying a defective *amhby* failed to develop a normal testis in three independent crosses of different parental pairs (N=7, 8, and 8 for each cross).

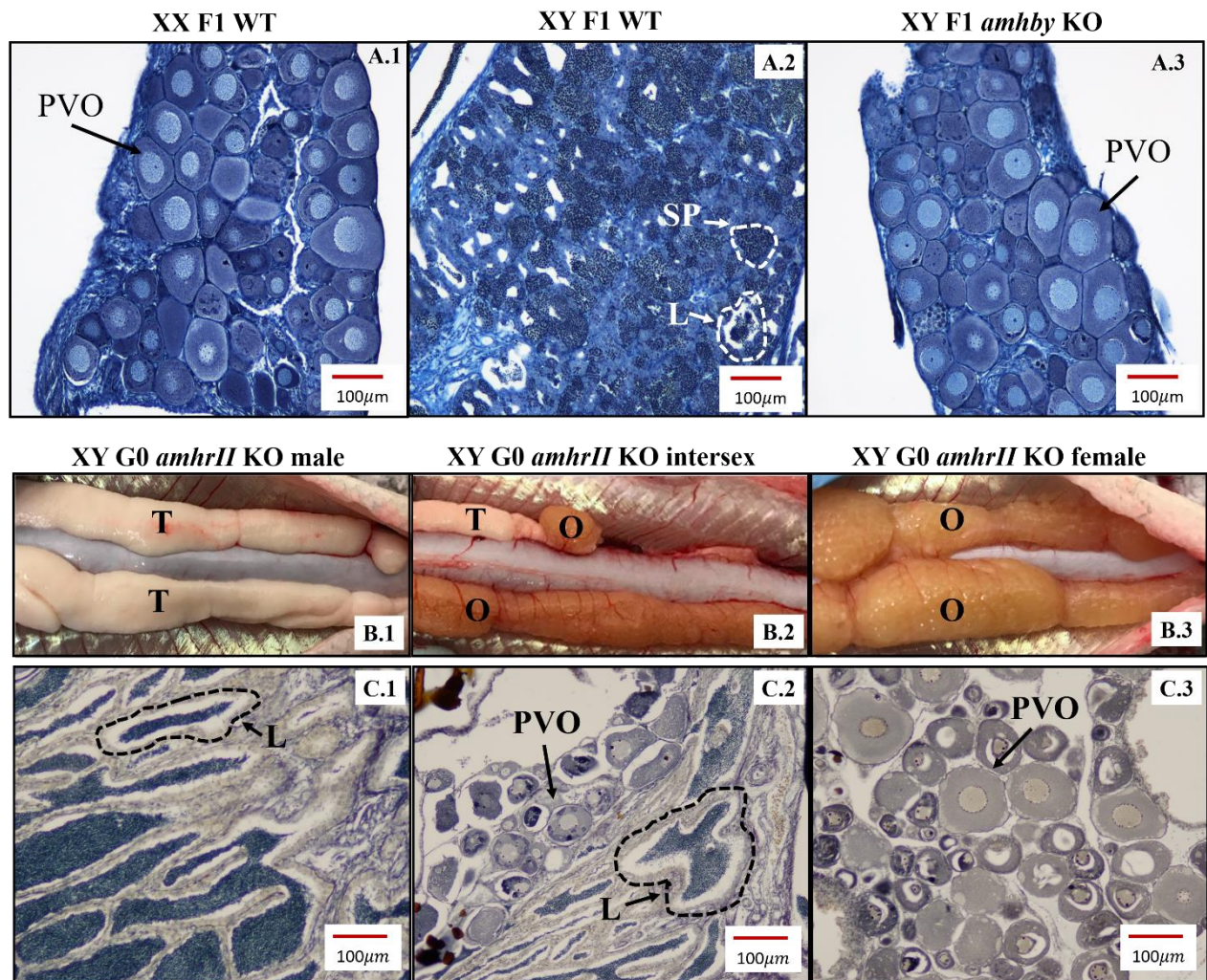


Figure 5: Gonadal phenotypes of *E. lucius* in knockout *amhby* and *amhrII* experiments.

A: F1 Control and mutant phenotypes in *amhby* knockout (KO) experiment. Histology of gonads of control XX female (A1), control XY male (A2) and *amhby* KO XY male (A3) at 153 dpf. The XY mutant with *amhby* KO (A3) showed complete sex reversal that it developed ovary with oocytes and ovary cavity, and is indistinguishable from the ovary of control females (A1). **B&C:** Gonadal phenotypes in *amhrII* knockout G0 mutants. Pictures of the gonads in the first lane (B.1- B.3), with the corresponding histology of the gonad below (C.1-C.3) are presented for three categories of mutants: 1) male gonads of an XY mutant 2) intersex gonads 3) female gonads of an XY mutant. **PVO:** previtellogenic oocytes; **SP:** spermatozooids; **L:** testicular lobules; **T:** testis; and **O:** ovary.

Among these 23 F1 XY mutants, 20 (87%) showed complete sex reversal in gonads, characterized by the formation of an ovarian cavity and the appearance of previtellogenic oocytes; the 3 (13%) remaining mutants developed regressed, potentially sterile gonads (**Figure 5A**).

The *amhrII* gene was knocked out using CRISPR/Cas9 utilizing three gRNAs targeting exon 9 (**Figure S3**). Among the 92 surviving embryos microinjected with the pool of three gRNAs, 63 (68%) carried mutated *amhrII* sequences as well as wild-type sequences. We maintained 24 (38%) mutant males were selected among these 63 animals and 19 (79%) of them survived to the next reproductive season. After spawning induction, 7 (37%) animals produced sperm, and 4 (21%) showed functional sex reversal demonstrated by the fact that they produced viable eggs. Gametes from these 11 individuals were used to perform *in vitro* fertilization to generate an F1 generation of *amhrII* mutants. G0 animals that did not spawn after induction were dissected to determine their gonadal sex. In total, among 19 G0 *amhrII* mutant males, 9 (37%) developed female gonads, 1 (5%) developed an intersex gonad with both ovaries and testes, 9 (47%) developed normal testes, and 2 (10%) were not determined (**Figure 5B**, **Table S4**).

Environmental factors interact with genetic sex determination in captive *E. lucius*

Two temperature regimes were applied to developing wild-type *E. lucius* from after hatching to 5 months post fertilization. One group (n=27) was raised at an elevated temperature of 18°C and another group from the same parental cross (n= 28) was raised at a temperature of 12°C. In the group raised at high temperature, genotyping identified 13 (48%) *amhby* positive animals and 14 (52%) *amhby* negative animals, but 26 (96%) animals developed testis and 1 (4%) animal developed ovaries, resulting in a phenotypic sex ratio significantly skewed towards males (Chi-square test, $p=3.8e-06$). In the group raised at the control temperature, genotyping showed 16 *amhby* positive (57%) and 12 *amhby* negative (43%) animals, and remarkably, all (100%) animals in this group developed testis, also resulting in a sex-ratio significantly skewed towards males (Chi-square test, $p=1.21e-07$).

Because male-skewed sex ratios were found in animals raised at both low and elevated temperature, these results do not allow us to conclude how temperature affects sex determination in *E. lucius*. However, they clearly demonstrate that, at least under laboratory conditions, sex determination in the Northern pike can be affected by the rearing environment.

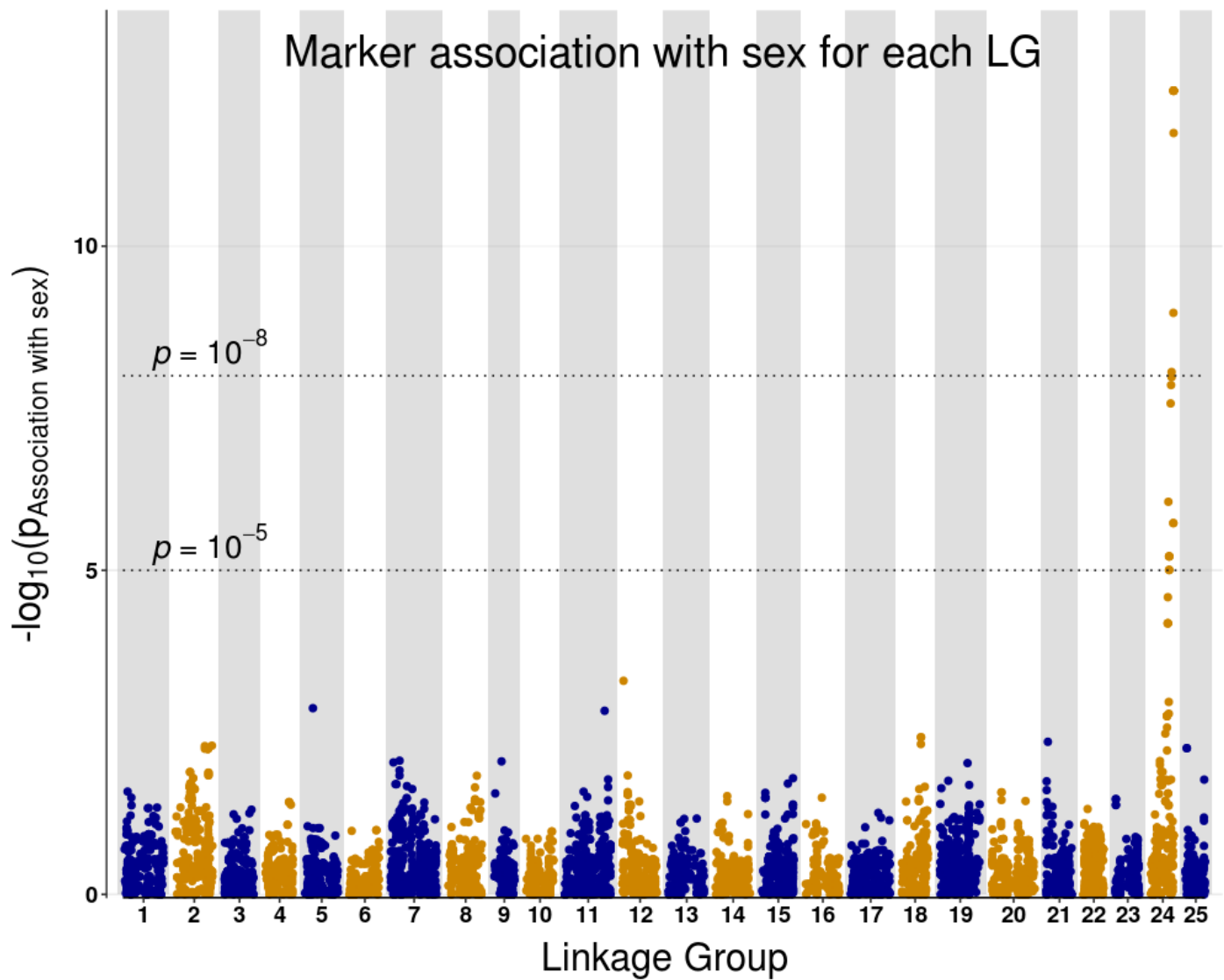


Figure 6: Marker association with sex across 25 linkage groups of *E. lucius* genome. The \log_{10} of p-value from the association test between each marker and sex phenotype is mapped against the linkage group they belong. Linkage group 24 showed a concentration of markers significantly associated with sex phenotype. The colored horizontal lines are indications for genome wide significance level from the association test, with red line indicating p-value of 1×10^{-5} and blue line indicating p-value of 1×10^{-8} .

The sex locus of *E. lucius* is located in the telomere of LG24

A microsatellite based linkage map of *E. lucius* was already available from Rondeau *et al.*, (2014), but the sex locus was not identified, because the phenotypic sex of the samples was not recorded. Therefore, we generated a new genetic map using RAD-Seq data from a single full-sibling family of *E. lucius* in which sex phenotype was recorded for each individual fish. In total, this analysis yielded 255,875 RAD tags of 94 bp each, 8,318 (3.25%) of which contained at least one polymorphic site. Male and female maps were constructed separately and contained 897 and 1210 informative markers respectively, more than three times the number scored in the microsatellite-based map from Rondeau *et al.*, which is likely caused by the differences in polymorphism between populations. The total length of the male map was 1,332.2 cM with an average distance of 1.49 cM between markers, and the total length of the female map was 1,477.1 cM with an average distance of 1.22 cM between markers. The average recombination rate was 1.11 times higher in females than in males, which was similar to the results of Rondeau *et al.*, who obtained a female to male recombination rate ratio of 1.07. The total length and number of markers for each linkage group (LG) are given in **Table S5**. The linkage map for LG24, the sex chromosome, is given in **Figure S5**, and linkage maps of the 25 LGs for males and females can be found in **Appendix 1**.

The position of the sex locus was inferred to be in the telomere region of LG 24 (0 cM) based on phenotypic sex information for all of the offspring. When markers from LG 24 were mapped to the genome assembly of *E. lucius* (Rondeau *et al.*, 2014), a marker close to the sex locus (3.9 cM) was located at 22.0 Mb on the 22.4 Mb assembled LG 24. (**Figure S4**). In addition, 32 male-specific non-polymorphic RAD markers were identified that were found in all of the males but not found in any female, which were considered to be markers coming from the sex-locus. When the sequence of these markers was aligned to the genome assembly, 15 (47%) mapped with a mapping quality > 20. Among these, 5 (33%) RAD-tags aligned to a region located distal to 22.1 Mb on LG 24; 1 (7%) aligned at position 17.3 Mb on LG24; 7 (22%) aligned to unplaced scaffold 0213, and 2 (6%) aligned to other LGs. In addition, 2 of the unmapped male-specific markers aligned perfectly to the sequence of *amhby* obtained from the fosmid but failed to appear in the published genome sequence (**Table S6**). Collectively, these results identify a well-differentiated sex-determining region located at the end of LG 24 that was not assembled well in the available genome assembly. Furthermore, the unplaced scaffold 0213 is likely included in this region, as well as *amhby*, which showed a strict father-to-son inheritance pattern.

In addition to these male-specific non-polymorphic markers, all polymorphic markers were also aligned to the genome assembly of *E. lucius*. In total, 6,512 polymorphic markers aligned to the 25 linkage groups and 741 polymorphic markers aligned to unplaced scaffolds. For each aligned marker, F_{st} between males and females and probability of association with sex were calculated. The genome-wide average F_{st} between males and females was 0.0033. Only two linkage groups had a higher average F_{st} between males and females than the genome average: LG2 had an average F_{st} of 0.006, and LG24 had an average F_{st} of 0.025 – more than eight times the average F_{st} of the genome. In concordance with this result, only markers on LG24 showed genome-wide significant association with the sex phenotype (**Figure 6**).

Discussion

Since the discovery of *dmrt1bY* in the Japanese rice fish (Matsuda *et al.*, 2002; Nanda *et al.*, 2002), the first teleost master sex determination gene, studies in teleosts have unveiled a dozen novel genes as master regulators for sex determination (Kikuchi and Hamaguchi 2013; Heule *et al.*, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017). Interestingly, many of these master sex determiners belong to the TGF- β superfamily; to date, this finding has been mostly restricted to the teleosts. The repeated and independent recruitment of members of the TGF- β family as MSD in the teleosts highlights the crucial role of TGF- β signaling in the sex determination pathway. In the present study, using genomic, transcriptional, and functional evidence, we demonstrated that a male-specific duplicate of *amh* is the MSD gene in an XX/XY SD system in *Esox lucius*. We established that this male MSD gene originated from a lineage specific *amh* duplication event and was later translocated to the telomeric region of the proto-Y chromosome. We found that although the sex chromosomes appear morphologically homomorphic, gradual suppression of recombination occurs along two thirds of their length, reflecting the differentiation between the X and Y chromosome. Finally, we observed that environmental factors can override genetic sex determination and distort sex ratio in captive population of *E. lucius*.

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

Among studies identifying novel MSDs in the teleosts, only a few studies provided combined evidence on the expression, linkage with sex, and functional necessity and sufficiency of the candidate to support its role in triggering sexual development, most of them in the *Oryzias* (Matsuda *et al.*, 2002; Nanda *et al.*, 2002; Matsuda *et al.*, 2003; T. Myosho *et al.*, 2012; Yano *et al.*, 2013; Takehana *et al.*, 2007; Li *et al.*, 2015).

In this study, results from genotyping and from a genetic map demonstrate the association of *amhby* with phenotypic sex and its colocalization with the sex locus. RNA-seq, qPCR and ISH showed that *amhby* is expressed in the somatic tissues surrounding the primordial germ cells in the differentiating testis, thus fulfilling another criterion for being a MSD. Furthermore, knockout of *amhby* lead to complete gonadal sex reversal of XY mutants, demonstrating that *amhby* is necessary for testicular differentiation. All these independent lines of evidence provide strong confidence in the MSD role of *amhby* in *E. lucius*.

AmhrII transduces the sex determination signal in Northern pike

The putative receptor of Amh, *amhrII*, plays a crucial role in the sex-determination pathway in several teleost species, including *Takifugu* (Kamiya *et al.*, 2012), Nile tilapia (Li *et al.*, 2015), and Japanese ricefish (Morinaga *et al.*, 2007). We knocked out the function of this receptor to assess its importance in the sex-determination pathway in *E. lucius*. Among G0 animals, half of the XY mutants showed complete sex reversal and produced viable eggs, while the other half developed testis or intersex gonads. Genotyping results revealed that all G0 mutants carried both the wildtype *AmhrII* sequence and several different truncated *AmhrII* sequences that lead to non-functional protein products. Consequently, this sex-reversal rate could be explained with either of two hypotheses: either sex-reversal could depend on the location of the cells expressing the wild-type *amhrII*, or it could depend on the amount of wild-type product available, thus following a threshold model of *amh* signaling (Morinaga *et al.*, 2007; Kamiya *et al.*, 2012). Overall, this experiment showed that *AmhrII* is necessary for testis differentiation in *E. lucius*, and is thus likely to act early in the sex-determination pathway.

AMH in teleost sex determination

Studies on vertebrate sex-determination have revealed that all but one known master sex determiners belong to three protein families known to be involved in sex-differentiation pathways: the *SOX*, *DMRT*, and *TGF- β* families. In particular, members of the *TGF- β* family - namely Amh, Gsdf and Gdf6 - have emerged as “usual suspects” for master sex determiners according to several recent studies in teleosts (Kikuchi and Hamaguchi, 2013; Heule *et al.*, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017). In the Patagonian pejerrey (*Odontesthes hatcheri*, Hattori *et al.*, 2012) and the Nile tilapia (*Oreochromis niloticus*, Li *et al.*, 2015), male-specific duplicates of *AMH* have been identified and characterized to function as MSD. In *O. bonariensis*, a sister species of the Patagonian pejerrey, a male-specific *amhy* was found to interact with temperature in determining sex (Yamamoto *et al.*, 2014); in the ling cod (*Ophiodon elongatus*), a duplicated male-specific copy of *amh* has been identified (Rondeau *et al.*, 2016). 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 (Bej *et al.*, 2017). Our phylogenetic analysis on *AMH* sequences from several teleost species revealed that these male-specific *AMH* duplications are most likely independent events rather than being the product of shared ancestry. This finding supports the “limited option” hypothesis for master sex determiners (Graves and Peichel, 2010) and makes *AMH* duplicates the most frequently independently recruited master determiners identified so far.

AMH was first identified in tetrapods, where it plays a role in the regression of Müllerian ducts in male embryos, in gonadal development, and in gametogenesis (Josso *et al.*, 2001). Teleost fishes do not possess Müllerian ducts, but AMH ortholog sequences have been identified in all inspected species. The function of AMH in the teleosts has been tightly linked to sex determination networks through the regulation of germ cell maturation and steroidogenesis. In males, AMH has been shown, for instance in zebrafish, to suppress proliferation and differentiation of type A spermatogonia (Skaar *et al.*, 2011). In female Nile tilapia, the knockout of *amh* led to the blockage of oogenesis (M. Li *et al.*, 2015). In addition, expression of *amh* has been found to be positively correlated with the expression of genes conserved in steroidogenesis, such as *cyp17*, *cyp11*, *hsd3-B1*, and *nr5a1* (Baron *et al.*, 2005; Vizziano *et al.*, 2008; Wang and Orban, 2007; Liew *et al.*, 2012). Furthermore, Amh has been shown to suppress the expression of steroidogenic genes such as *star* (steroidogenic acute regulatory proteins) and *cyp17a1* in zebrafish, and knockout of *amh* was linked to an increase in expression of *cyp19a1* in the Patagonian pejerrey and the Nile tilapia (Hattori *et al.*, 2012; Li *et al.*, 2015). Finally, it has been hypothesized that the *amh* pathway could play a more important role in gonadal sex determination in non-mammalian vertebrates than in mammalian vertebrates (Smith and Sinclair, 2004; Kikuchi and Hamaguchi, 2013): in mammals, *amh* is subject to regulation by *sox9* (Arango *et al.*, 1999), but in rainbow trout (*Oncorhynchus mykiss*), the *amh* promoter lacks the putative *sox* binding sites (Jamin *et al.*, 2008), and in chicken (*Gallus gallus*) and the American alligator (*Alligator mississippiensis*) expression of *amh* even precedes expression of *sox9*. The recurrent presence of *amh* in sex-determination pathways, especially in non-mammal vertebrates, could provide one explanation for its prevalence as the MSD gene in the teleosts.

Among the three teleost species with confirmed *amh* duplicates as MSD, *Esox lucius* shows the highest degree of divergence between sequences of the paralogs. In the Nile tilapia, *amhy* is almost identical to the autosomal *amh*, differing by only one SNP (Li *et al.*, 2015), while in the Patagonian pejerrey, the shared identity between paralogs ranges from 89.1% to 100% depending on the exon (Hattori *et al.*, 2012). In *E. lucius*, *amhby* and *Amha* have a higher level of sequence divergence, sharing 74% to 95% identity between corresponding exons and only 68% identity at the protein level. This higher divergence between *amhby* and *amha* could either indicate that the duplication event in *E. lucius* is more ancient than in Patagonian pejerrey and in Nile tilapia, or that different selection regimes acted on these two paralogs in these species. In addition, because of the low divergence between the two paralogs in the Patagonian pejerrey and the Nile tilapia, the authors attributed the new function of *amhy* as the MSD mainly to its novel expression pattern. In *E. lucius*, *amhby* also showed a different expression profile than *amha*, likely due to a completely different promoter region; however, because of the divergence between sequences, it is tempting to hypothesize that the two proteins could have also diverged in their function and in their downstream signaling pathways as well.

Evolution of the sex chromosomes in *E. lucius*

The analysis of the genomic neighborhood of both *amh* duplicates showed that *amha* is located on LG08 in a functional cluster of genes regulating sexual development and cell cycling with conserved synteny in the teleosts (Overbeek *et al.*, 1999; Pfennig *et al.*, 2015;), while *amhby* is located in the telomeric region of LG24 with no other identified genes in its vicinity. These results indicate that *amha* is likely to be the ancestral *amh* copy in *E. lucius*, and that *amhby* was translocated to another chromosome after duplication. This scenario fits the description of a proposed mechanism of sex chromosome emergence and turnover through gene duplication, translocation and neofunctionalization (Herpin and Schartl, 2009; Kondo *et al.*, 2009). Following this model, the translocation of a single copy of *amhby* into another autosome triggered the formation of proto sex chromosomes.

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 (Herpin *et al.*, 2010). 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. The vagabond nature of *sdY* is assumed to be the result of being embedded within a cassette of transposable elements (Yano *et al.*, 2013; Faber-Hammond *et al.*, 2015; Lubieniecki *et al.*, 2015). In both Japanese medaka and *salmonids*, transposable elements facilitated the translocation of the new MSD; it is thus possible that transposable elements also contributed to the translocation of *amhby* in *E. lucius*. We identified some transposable elements in the promoter region and in the intronic region of *amhby*, but a better assembly of the region surrounding *amhby* is needed to conclude on the role of transposable elements in this translocation.

Interactions between Environmental SD and Genetic SD

In many teleost species, an increase in temperature during development has been linked to an increase in the expression of *amh*, eventually resulting in the masculinization of the gonads (Yoshinaga *et al.*, 2004; Kitano *et al.*, 2007; Fernandino *et al.*, 2008; Yamaguchi and Kitano, 2012; Poonlaphdecha *et al.*, 2013; Mankiewicz *et al.*, 2013). In particular, in one species of Patagonian pejerrey (*Odontesthes bonariensis*), the sex ratio can sway as far as 100% females at low temperature (17°C) and 100% males at high temperature (29°C), while at an intermediate temperature, phenotypic sex is strongly associated with the presence of a male-specific *amh* duplicate (Yamamoto *et al.*, 2014). In its sister species *Odontesthes hatcheri*, where *amhy* is also the MSD, the sex-ratio can also be affected by temperature, but to a much lesser extent (Hattori *et al.*, 2012). However, in our experiment testing the potential

effects of temperature on sex determination in *E. lucius*, we observed close to 100% female to male sex reversal in XX animals raised in normal temperature as well as in animals raised at elevated temperature. As a result, we were not able to assess the modulation effect of temperature on sex-determination; however, it is likely that other environmental stressors in our raising conditions may have led to sex-reversal in the animals. In fact, various environmental stresses have been found in other teleost species to stimulate the production of cortisol, which in turn increases the expression of *amh*, eventually leading to masculinization of the gonads (van den Hurk and van Oordt, 1985; Bonga, 1997; Fernandino *et al.*, 2008; Hattori *et al.*, 2009). In our experiment, we observed high levels of cannibalism in captive pikes as early as three weeks post-fertilization; in their natural habitat, early migration of *E. lucius* larvae has been documented as a way to potentially avoid cannibalism (Larsson *et al.*, 2015; Forsman *et al.*, 2015). Therefore, it is likely that the animals experienced constant predation stress throughout development, including sex differentiation. We suspect the predation stress suffered by developing pikes may have been the environmental stimulus that interfered with gonadal differentiation.

Conclusion

In this study, we identified a male-specific duplicate of the Anti-Müllerian hormone, that we called *amhby*, as the master sex determination gene in the Northern pike (*Esox lucius*), with evidence from genomic, expression, cytogenetic, and functional experiments. We also revealed the crucial role of *amhrII* in testicular differentiation, highlighting the importance of the *amh* pathway in sex differentiation in the teleosts. In addition, we showed the susceptibility of the genetic sex determination system to environmental perturbations, not just restricted to temperature, in this species. With both phylogenetic and conserved synteny analysis, we revealed that *amhby* originated from a lineage-specific duplication event, and was translocated after duplication to the telomere of LG24. Our results link the mechanism underlying the birth of the MSD to the pattern of differentiation of the sex chromosomes and offers insights into the evolutionary process of sex chromosome formation with a duplicated and inserted MSD.

References

- Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., & Postlethwait, J. H. (2011). Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. *Genetics*, 188(4), 799–808. <https://doi.org/10.1534/genetics.111.127324>
- Arango, N. A., Lovell-Badge, R., & Behringer, R. R. (1999). Targeted mutagenesis of the endogenous mouse *Mis* gene promoter: in vivo definition of genetic pathways of vertebrate sexual development. *Cell*, 99(4), 409–419.
- Bachtrog, D. (2003). Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in *Drosophila*. *Nature Genetics*, 34(2), 215–219. <https://doi.org/10.1038/ng1164>
- Bachtrog, D. (2006). A dynamic view of sex chromosome evolution. *Current Opinion in Genetics & Development*, 16(6), 578–585. <https://doi.org/10.1016/j.gde.2006.10.007>
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., ... Consortium, T. T. of S. (2014). Sex Determination: Why So Many Ways of Doing It? *PLOS Biology*, 12(7), e1001899. <https://doi.org/10.1371/journal.pbio.1001899>
- Baron, D., Houlgatte, R., Fostier, A., & Guiguen, Y. (2005). Large-Scale Temporal Gene Expression Profiling During Gonadal Differentiation and Early Gametogenesis in Rainbow Trout. *Biology of Reproduction*, 73(5), 959–966. <https://doi.org/10.1095/biolreprod.105.041830>
- Bej, D. K., Miyoshi, K., Hattori, R. S., Strüssmann, C. A., & Yamamoto, Y. (2017). A Duplicated, Truncated *amh* Gene Is Involved in Male Sex Determination in an Old World Silverside. *G3: Genes, Genomes, Genetics*, g3.117.042697. <https://doi.org/10.1534/g3.117.042697>
- Bergero, R., & Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends in Ecology & Evolution*, 24(2), 94–102. <https://doi.org/10.1016/j.tree.2008.09.010>

- Bergero, R., Charlesworth, D., Filatov, D. A., & Moore, R. C. (2008). Defining regions and rearrangements of the *Silene latifolia* Y chromosome. *Genetics*, 178(4), 2045–2053. <https://doi.org/10.1534/genetics.107.084566>
- Bergero, R., Forrest, A., Kamau, E., & Charlesworth, D. (2007). Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics*, 175(4), 1945–1954. <https://doi.org/10.1534/genetics.106.070110>
- Beukeboom, L. W., & Perrin, N. (2014). *The Evolution of Sex Determination*. Oxford University Press.
- Brudno, M., Do, C. B., Cooper, G. M., Kim, M. F., Davydov, E., NISC Comparative Sequencing Program, ... Batzoglou, S. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Research*, 13(4), 721–731. <https://doi.org/10.1101/gr.926603>
- Bugno, M., Jablonska, Z., Tischner, M., Klukowska-Rötzler, J., Pienkowska-Schelling, A., Schelling, C., & Slota, E. (2010). Detection of sex chromosome aneuploidy in equine spermatozoa using fluorescence in situ hybridization. *Reproduction in Domestic Animals = Zuchthygiene*, 45(6), 1015–1019. <https://doi.org/10.1111/j.1439-0531.2009.01478.x>
- Cade, L., Reyon, D., Hwang, W. Y., Tsai, S. Q., Patel, S., Khayter, C., ... Yeh, J.-R. J. (2012). Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Research*, 40(16), 8001–8010. <https://doi.org/10.1093/nar/gks518>
- Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, 95(2), 118–128. <https://doi.org/10.1038/sj.hdy.6800697>
- de Givry, S., Bouchez, M., Chabrier, P., Milan, D., & Schiex, T. (2005). CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics (Oxford, England)*, 21(8), 1703–1704. <https://doi.org/10.1093/bioinformatics/bti222>
- Ezaz, T., Sarre, S. D., O'Meally, D., Graves, J. A. M., & Georges, A. (2009). Sex chromosome evolution in lizards: independent origins and rapid transitions. *Cytogenetic and Genome Research*, 127(2–4), 249–260. <https://doi.org/10.1159/000300507>
- Fernandino, J. I., Hattori, R. S., Kimura, H., Strüssmann, C. A., & Somoza, G. M. (2008). Expression profile and estrogenic regulation of anti-Müllerian hormone during gonadal development in pejerrey *Odontesthes bonariensis*, a teleost fish with strong temperature-dependent sex determination. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 237(11), 3192–3199. <https://doi.org/10.1002/dvdy.21731>
- Forsman, A., Tibblin, P., Berggren, H., Nordahl, O., Koch-Schmidt, P., & Larsson, P. (2015). Pike *Esox lucius* as an emerging model organism for studies in ecology and evolutionary biology: a review. *Journal of Fish Biology*, 87(2), 472–479. <https://doi.org/10.1111/jfb.12712>
- Gamble, T., & Zarkower, D. (2014). Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Molecular Ecology Resources*, n/a-n/a. <https://doi.org/10.1111/1755-0998.12237>
- Gharbi, K., Gautier, A., Danzmann, R. G., Gharbi, S., Sakamoto, T., Høyheim, B., ... Guyomard, R. (2006). A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, 172(4), 2405–2419. <https://doi.org/10.1534/genetics.105.048330>
- Graves, J. A. M., & Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biology*, 11(4), 205.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>

- Hattori, R. S., Fernandino, J. I., Kishii, A., Kimura, H., Kinno, T., Oura, M., ... Watanabe, S. (2009). Cortisol-Induced Masculinization: Does Thermal Stress Affect Gonadal Fate in Pejerrey, a Teleost Fish with Temperature-Dependent Sex Determination? *PLOS ONE*, 4(8), e6548. <https://doi.org/10.1371/journal.pone.0006548>
- Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., ... Strussmann, C. A. (2012). A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences*, 109(8), 2955–2959. <https://doi.org/10.1073/pnas.1018392109>
- Herpin, A., Braasch, I., Kraeussling, M., Schmidt, C., Thoma, E. C., Nakamura, S., ... Scharl, M. (2010). Transcriptional Rewiring of the Sex Determining dmrt1 Gene Duplicate by Transposable Elements. *PLoS Genetics*, 6(2), e1000844. <https://doi.org/10.1371/journal.pgen.1000844>
- Herpin, A., & Scharl, M. (2009). Molecular mechanisms of sex determination and evolution of the Y-chromosome: Insights from the medakafish (*Oryzias latipes*). *Molecular and Cellular Endocrinology*, 306(1–2), 51–58. <https://doi.org/10.1016/j.mce.2009.02.004>
- Herpin, A., & Scharl, M. (2015). Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Reports*, 16(10), 1260–1274. <https://doi.org/10.15252/embr.201540667>
- Heule, C., Salzburger, W., & Böhne, A. (2014). Genetics of Sexual Development: An Evolutionary Playground for Fish. *Genetics*, 196(3), 579–591. <https://doi.org/10.1534/genetics.114.161158>
- Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., ... Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*, 31(3), 227–229. <https://doi.org/10.1038/nbt.2501>
- Ihssen, P. E., McKay, L. R., McMillan, I., & Phillips, R. B. (1990). Ploidy Manipulation and Gynogenesis in Fishes: Cytogenetic and Fisheries Applications. *Transactions of the American Fisheries Society*, 119(4), 698–717. [https://doi.org/10.1577/1548-8659\(1990\)119](https://doi.org/10.1577/1548-8659(1990)119)
- Josso, N., di Clemente, N., & Gouédard, L. (2001). Anti-Müllerian hormone and its receptors. *Molecular and Cellular Endocrinology*, 179(1), 25–32.
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., ... Kikuchi, K. (2012). A Trans-Species Missense SNP in Amhr2 Is Associated with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*, 8(7), e1002798. <https://doi.org/10.1371/journal.pgen.1002798>
- Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution: Novel Sex-Determining Genes in Fish. *Developmental Dynamics*, 242(4), 339–353. <https://doi.org/10.1002/dvdy.23927>
- Kitano, T., Yoshinaga, N., Shiraishi, E., Koyanagi, T., & Abe, S.-I. (2007). Tamoxifen induces masculinization of genetic females and regulates P450 aromatase and Müllerian inhibiting substance mRNA expression in Japanese flounder (*Paralichthys olivaceus*). *Molecular Reproduction and Development*, 74(9), 1171–1177. <https://doi.org/10.1002/mrd.20603>
- Kondo, M., Nanda, I., Schmid, M., & Scharl, M. (2009). Sex Determination and Sex Chromosome Evolution: Insights from Medaka. *Sexual Development*, 3(2–3), 88–98. <https://doi.org/10.1159/000223074>
- Lahn, B. T., & Page, D. C. (1999). Four evolutionary strata on the human X chromosome. *Science (New York, N.Y.)*, 286(5441), 964–967.
- Larsson, P., Tibblin, P., Koch-Schmidt, P., Engstedt, O., Nilsson, J., Nordahl, O., & Forsman, A. (2015). Ecology, evolution, and management strategies of northern pike populations in the Baltic Sea. *Ambio*, 44 Suppl 3, 451–461. <https://doi.org/10.1007/s13280-015-0664-6>
- Lefort, V., Longueville, J.-E., & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, 34(9), 2422–2424. <https://doi.org/10.1093/molbev/msx149>

- Leong, J. S., Jantzen, S. G., von Schalburg, K. R., Cooper, G. A., Messmer, A. M., Liao, N. Y., ... Koop, B. F. (2010). *Salmo salar* and *Esox lucius* full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. *BMC Genomics*, 11, 279. <https://doi.org/10.1186/1471-2164-11-279>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., ... Wang, D. (2015). 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(11), e1005678. <https://doi.org/10.1371/journal.pgen.1005678>
- Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., ... Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*, 43(W1), W580-584. <https://doi.org/10.1093/nar/gkv279>
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., ... Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533(7602), 200–205. <https://doi.org/10.1038/nature17164>
- Liew, W. C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K. R., & Orban, L. (2012). Polygenic Sex Determination System in Zebrafish. *PLoS ONE*, 7(4), e34397. <https://doi.org/10.1371/journal.pone.0034397>
- Luczynski, M. J., Dabrowski, K., Kucharczyk, D., Glogowski, J., Luczynski, M., Szczerbowski, A., & Babiak, I. (2007). Gynogenesis in northern pike: UV-inactivation of spermatozoa and the heat shock inducing meiotic diploidization. *Environmental Biotechnology*, Vol. 3, No. 2, 39–43.
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), 823–826. <https://doi.org/10.1126/science.1232033>
- Mank, J. E., & Avise, J. C. (2009). Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, 3(2–3), 60–67. <https://doi.org/10.1159/000223071>
- Mankiewicz, J. L., Godwin, J., Holler, B. L., Turner, P. M., Murashige, R., Shamey, R., ... Borski, R. J. (2013). Masculinizing effect of background color and cortisol in a flatfish with environmental sex-determination. *Integrative and Comparative Biology*, 53(4), 755–765. <https://doi.org/10.1093/icb/ict093>
- Marshall, O. J. (2004). PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics (Oxford, England)*, 20(15), 2471–2472. <https://doi.org/10.1093/bioinformatics/bth254>
- Matsuda, M., Nagahama, Y., Kobayashi, T., Matsuda, C., Hamaguchi, S., & Sakaizumi, M. (2003). The sex determining gene of medaka: a Y-specific DM domain gene (DMY) is required for male development. *Fish Physiology and Biochemistry*, 28(1–4), 135–139. <https://doi.org/10.1023/B:FISH.0000030500.29914.7a>
- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., ... Sakaizumi, M. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417(6888), 559–563. <https://doi.org/10.1038/nature751>
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., ... Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, 41(Web Server issue), W597-600. <https://doi.org/10.1093/nar/gkt376>
- Miura, I., Ohtani, H., & Ogata, M. (2012). Independent degeneration of W and Y sex chromosomes in frog *Rana rugosa*. *Chromosome Research: An International Journal on the Molecular*,

- Supramolecular and Evolutionary Aspects of Chromosome Biology*, 20(1), 47–55.
<https://doi.org/10.1007/s10577-011-9258-8>
- M.j, L., J, G., D, K., M, L., & K, D.-Z. (1997). Gynogenesis in Northern pike [*Esox lucius* L.] induced by heat shock - preliminary data. *Polskie Archiwum Hydrobiologii*, 1–2(44). Retrieved from <https://www.infona.pl/resource/bwmeta1.element.agro-article-5d7c506e-53a2-4005-a9c4-6639a8741d99>
- Morinaga, C., Saito, D., Nakamura, S., Sasaki, T., Asakawa, S., Shimizu, N., ... Kondoh, H. (2007). The hotei mutation of medaka in the anti-Müllerian hormone receptor causes the dysregulation of germ cell and sexual development. *Proceedings of the National Academy of Sciences*, 104(23), 9691–9696.
- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., ... Sakaizumi, M. (2012). Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, *Oryzias latipes*. *Genetics*, 191(1), 163–170. <https://doi.org/10.1534/genetics.111.137497>
- Myosho, T., Takehana, Y., Hamaguchi, S., & Sakaizumi, M. (2015). Turnover of Sex Chromosomes in Celebensis Group Medaka Fishes. *G3: Genes/Genomes/Genetics*, 5(12), 2685–2691. <https://doi.org/10.1534/g3.115.021543>
- Nagai, T., Takehana, Y., Hamaguchi, S., & Sakaizumi, M. (2008). Identification of the sex-determining locus in the Thai medaka, *Oryzias minutillus*. *Cytogenetic and Genome Research*, 121(2), 137–142. <https://doi.org/10.1159/000125839>
- Nam, K., & Ellegren, H. (2008). The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics*, 180(2), 1131–1136. <https://doi.org/10.1534/genetics.108.090324>
- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., ... Scharl, M. (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proceedings of the National Academy of Sciences*, 99(18), 11778–11783. <https://doi.org/10.1073/pnas.182314699>
- Natri, H. M., Shikano, T., & Merilä, J. (2013). Progressive Recombination Suppression and Differentiation in Recently Evolved Neo-sex Chromosomes. *Molecular Biology and Evolution*, 30(5), 1131–1144. <https://doi.org/10.1093/molbev/mst035>
- Ocalewicz, K., Kuzminski, H., Pomianowski, K., & Dobosz, S. (2013). Induction of androgenetic development of the brook charr (*Salvelinus fontinalis*) × Arctic charr (*Salvelinus alpinus*) hybrids in eggs derived from the parental species. *Reproductive Biology*, 13(2), 105–112. <https://doi.org/10.1016/j.repbio.2013.03.002>
- Ohno, S. (1969). Evolution of Sex Chromosomes in Mammals. *Annual Review of Genetics*, 3(1), 495–524. <https://doi.org/10.1146/annurev.ge.03.120169.002431>
- Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D., & Maltsev, N. (1999). The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), 2896–2901.
- Ovulation induction in northern pike *Esox lucius* L. using different GnRH analogues, Ovaprim, Dagin and carp pituitary - Szabó - 2003 - Aquaculture Research - Wiley Online Library. (n.d.). Retrieved October 18, 2017, from <http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2109.2003.00835.x/full>
- Pan, Q., Guiguen, Y., & Herpin, A. (2017). Evolution of sex determining genes in Fish. In *Encyclopedia of Reproduction, Second Edition*.
- Pasquier, J., Cabau, C., Nguyen, T., Jouanno, E., Severac, D., Braasch, I., ... Bobe, J. (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics*, 17, 368. <https://doi.org/10.1186/s12864-016-2709-z>
- Pfennig, F., Standke, A., & Gutzeit, H. O. (2015). The role of Amh signaling in teleost fish – Multiple functions not restricted to the gonads. *General and Comparative Endocrinology*, 223, 87–107. <https://doi.org/10.1016/j.ygcen.2015.09.025>

- Poonlaphdech, S., Pepey, E., Canonne, M., de Verdal, H., Baroiller, J.-F., & D'Cotta, H. (2013). Temperature induced-masculinisation in the Nile tilapia causes rapid up-regulation of both *dmrt1* and *amh* expressions. *General and Comparative Endocrinology*, 193, 234–242. <https://doi.org/10.1016/j.ygcen.2013.06.007>
- Purdom, C. E. (1983). Genetic engineering by the manipulation of chromosomes. *Aquaculture*, 33(1), 287–300. [https://doi.org/10.1016/0044-8486\(83\)90409-X](https://doi.org/10.1016/0044-8486(83)90409-X)
- Raat, A. J. P. (1988). *Synopsis of Biological Data on the Northern Pike: Esox Lucius Linnaeus, 1758*. Food & Agriculture Org.
- Rentería, M. E., Cortes, A., & Medland, S. E. (2013). Using PLINK for Genome-Wide Association Studies (GWAS) and data analysis. *Methods in Molecular Biology (Clifton, N.J.)*, 1019, 193–213. https://doi.org/10.1007/978-1-62703-447-0_8
- Reyon, D., Khayter, C., Regan, M. R., Joung, J. K., & Sander, J. D. (2012). Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. *Current Protocols in Molecular Biology, Chapter 12*, Unit 12.15. <https://doi.org/10.1002/0471142727.mb1215s100>
- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics: TIG*, 16(6), 276–277.
- Rice, W. R. (1996). Evolution of the Y Sex Chromosome in Animals Y chromosomes evolve through the degeneration of autosomes. *BioScience*, 46(5), 331–343. <https://doi.org/10.2307/1312947>
- Rodrigues, N., Studer, T., Dufresnes, C., Ma, W.-J., Veltsos, P., & Perrin, N. (2017). Dmrt1 polymorphism and sex-chromosome differentiation in *Rana temporaria*. *Molecular Ecology*. <https://doi.org/10.1111/mec.14222>
- Rondeau, E. B., Laurie, C. V., Johnson, S. C., & Koop, B. F. (2016). A PCR assay detects a male-specific duplicated copy of Anti-Müllerian hormone (*amh*) in the lingcod (*Ophiodon elongatus*). *BMC Research Notes*, 9(1). <https://doi.org/10.1186/s13104-016-2030-6>
- Rondeau, E. B., Minkley, D. R., Leong, J. S., Messmer, A. M., Jantzen, J. R., von Schalburg, K. R., ... Koop, B. F. (2014). The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PloS One*, 9(7), e102089. <https://doi.org/10.1371/journal.pone.0102089>
- Sander, J. D., Maeder, M. L., Reyon, D., Voytas, D. F., Joung, J. K., & Dobbs, D. (2010). ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Research*, 38(suppl_2), W462–W468. <https://doi.org/10.1093/nar/gkq319>
- Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F., & Dobbs, D. (2007). Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Research*, 35(suppl_2), W599–W605. <https://doi.org/10.1093/nar/gkm349>
- Sandstedt, S. A., & Tucker, P. K. (2004). Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome. *Genome Research*, 14(2), 267–272. <https://doi.org/10.1101/gr.1796204>
- Schartl, M. (2015). Sex determination by multiple sex chromosomes in *Xenopus tropicalis*. *Proceedings of the National Academy of Sciences*, 112(34), 10575–10576. <https://doi.org/10.1073/pnas.1513518112>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539. <https://doi.org/10.1038/msb.2011.75>
- Skaar, K. S., Nübrege, R. H., Magaraki, A., Olsen, L. C., Schulz, R. W., & Male, R. (2011). Proteolytically Activated, Recombinant Anti-Müllerian Hormone Inhibits Androgen Secretion, Proliferation, and Differentiation of Spermatogonia in Adult Zebrafish Testis Organ Cultures. *Endocrinology*, 152(9), 3527–3540. <https://doi.org/10.1210/en.2010-1469>

- Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., ... Page, D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423(6942), 825–837. <https://doi.org/10.1038/nature01722>
- Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., ... Ellegren, H. (2015). Evolutionary analysis of the female-specific avian W chromosome. *Nature Communications*, 6, 7330. <https://doi.org/10.1038/ncomms8330>
- Smith, C. A., Roeszler, K. N., Ohnesorg, T., Cummins, D. M., Farlie, P. G., Doran, T. J., & Sinclair, A. H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature*, 461(7261), 267–271. <https://doi.org/10.1038/nature08298>
- Smith, C. A., & Sinclair, A. H. (2004). Sex determination: insights from the chicken. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 26(2), 120–132. <https://doi.org/10.1002/bies.10400>
- Soh, Y. Q. S., Alföldi, J., Pyntikova, T., Brown, L. G., Graves, T., Minx, P. J., ... Page, D. C. (2014). Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell*, 159(4), 800–813. <https://doi.org/10.1016/j.cell.2014.09.052>
- Sumner, A. T. (1972). A simple technique for demonstrating centromeric heterochromatin. *Experimental Cell Research*, 75(1), 304–306.
- Takehana, Y., Demiyah, D., Naruse, K., Hamaguchi, S., & Sakaizumi, M. (2007). Evolution of different Y chromosomes in two medaka species, *Oryzias dancena* and *O. latipes*. *Genetics*, 175(3), 1335–1340. <https://doi.org/10.1534/genetics.106.068247>
- Takehana, Y., Hamaguchi, S., & Sakaizumi, M. (2008). Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi*. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 16(5), 801–811. <https://doi.org/10.1007/s10577-008-1227-5>
- Takehana, Y., Naruse, K., Hamaguchi, S., & Sakaizumi, M. (2007). Evolution of ZZ/ZW and XX/XY sex-determination systems in the closely related medaka species, *Oryzias hubbsi* and *O. dancena*. *Chromosoma*, 116(5), 463–470. <https://doi.org/10.1007/s00412-007-0110-z>
- Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S., & Sakaizumi, M. (2007). Evidence for Different Origins of Sex Chromosomes in Closely Related *Oryzias* Fishes: Substitution of the Master Sex-Determining Gene. *Genetics*, 177(4), 2075–2081. <https://doi.org/10.1534/genetics.107.075598>
- van den Hurk, R., & van Oordt, P. G. (1985). Effects of natural androgens and corticosteroids on gonad differentiation in the rainbow trout, *Salmo gairdneri*. *General and Comparative Endocrinology*, 57(2), 216–222.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3, research0034. <https://doi.org/10.1186/gb-2002-3-7-research0034>
- Vicoso, B., & Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews. Genetics*, 7(8), 645–653. <https://doi.org/10.1038/nrg1914>
- Vizziano, D., Baron, D., Randuineau, G., Mahè, S., Cauty, C., & Guiguen, Y. (2008). Rainbow Trout Gonadal Masculinization Induced by Inhibition of Estrogen Synthesis Is More Physiological Than Masculinization Induced by Androgen Supplementation. *Biology of Reproduction*, 78(5), 939–946. <https://doi.org/10.1095/biolreprod.107.065961>
- Volff, J.-N., & Schartl, M. (2002). Sex determination and sex chromosome evolution in the medaka, *Oryzias latipes*, and the platyfish, *Xiphophorus maculatus*. *Cytogenetic and Genome Research*, 99(1–4), 170–177. <https://doi.org/10.1159/000071590>
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity*, 93(1), 77–78.

- Wang, X. G., & Orban, L. (2007). Anti-Müllerian hormone and 11 beta-hydroxylase show reciprocal expression to that of aromatase in the transforming gonad of zebrafish males. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 236(5), 1329–1338. <https://doi.org/10.1002/dvdy.21129>
- Wendelaar Bonga, S. E. (1997). The stress response in fish. *Physiological Reviews*, 77(3), 591–625.
- Wright, A. E., Dean, R., Zimmer, F., & Mank, J. E. (2016). How to make a sex chromosome. *Nature Communications*, 7, ncomms12087. <https://doi.org/10.1038/ncomms12087>
- Wright, A. E., Harrison, P. W., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2014). Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the w chromosome. *Evolution; International Journal of Organic Evolution*, 68(11), 3281–3295. <https://doi.org/10.1111/evo.12493>
- Yamaguchi, T., & Kitano, T. (2012). High temperature induces cyp26b1 mRNA expression and delays meiotic initiation of germ cells by increasing cortisol levels during gonadal sex differentiation in Japanese flounder. *Biochemical and Biophysical Research Communications*, 419(2), 287–292. <https://doi.org/10.1016/j.bbrc.2012.02.012>
- Yamamoto, Y., Zhang, Y., Sarida, M., Hattori, R. S., & Strüssmann, C. A. (2014). Coexistence of Genotypic and Temperature-Dependent Sex Determination in Pejerrey *Odontesthes bonariensis*. *PLoS ONE*, 9(7), e102574. <https://doi.org/10.1371/journal.pone.0102574>
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., ... Guiguen, Y. (2012). An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Current Biology*, 22(15), 1423–1428. <https://doi.org/10.1016/j.cub.2012.05.045>
- Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., & Guiguen, Y. (2013). The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications*, 6(3), 486–496. <https://doi.org/10.1111/eva.12032>
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., ... Ito, M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences*, 105(7), 2469–2474. <https://doi.org/10.1073/pnas.0712244105>
- Yoshinaga, N., Shiraishi, E., Yamamoto, T., Iguchi, T., Abe, S., & Kitano, T. (2004). Sexually dimorphic expression of a teleost homologue of Müllerian inhibiting substance during gonadal sex differentiation in Japanese flounder, *Paralichthys olivaceus*. *Biochemical and Biophysical Research Communications*, 322(2), 508–513. <https://doi.org/10.1016/j.bbrc.2004.07.162>

Supplementary: Identification of a duplicated and translocated master sex determination gene provides insight into homomorphic sex chromosome differentiation in Northern pike (*Esox lucius*)

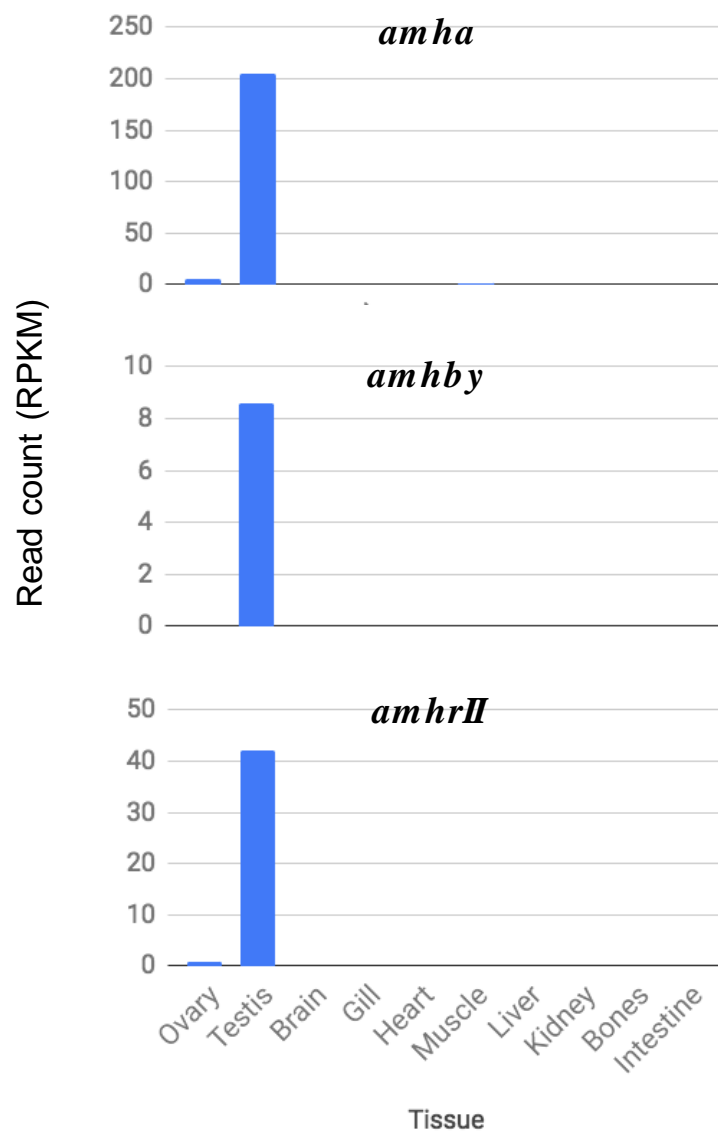


Figure S1: Normalized RNA-Seq read count for *amha*, *amhby*, and *amhrII* in the libraries of different tissues of *Esox lucius*. Data was obtained from the Phylofish database (<http://phylofish.sigenae.org/>) and was normalized by reads per kilobase of sequence per million of reads for each transcript in each library.

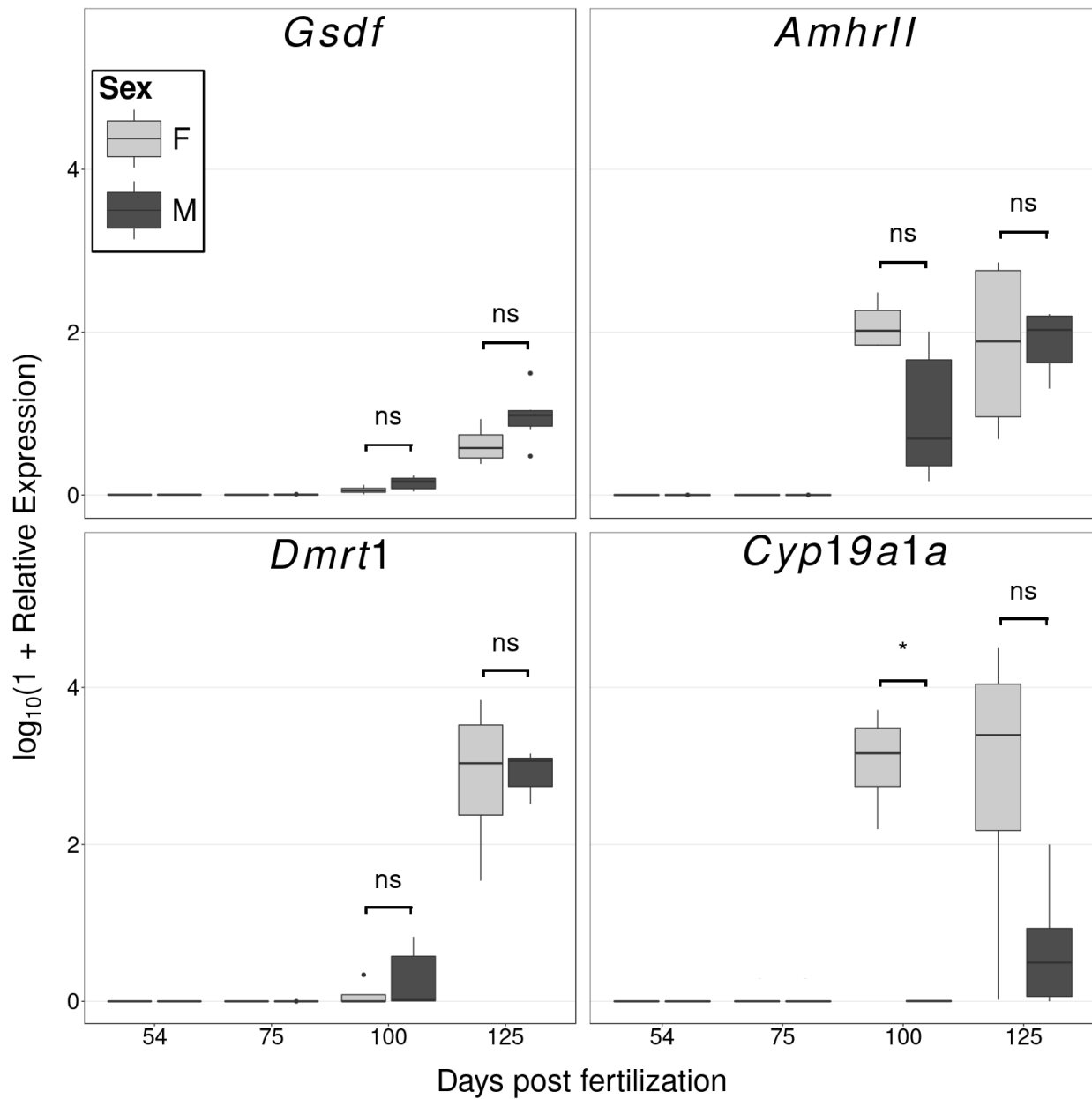
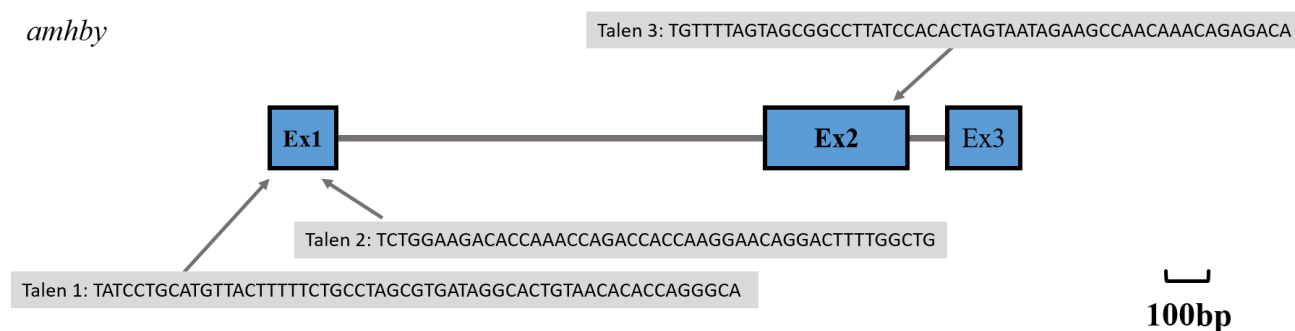


Figure S2 Boxplot showing the first quantile, median, and the third quantile of the temporal expression of *gsdf*, *dmrt1*, *cyp19a1a* and *amhrII* measured by qPCR. Outliers are displayed as a dot. The \log_{10} of relative expression of mRNA of these four genes were measured at 54, 75, 100, and 125 days post fertilization in male and female samples 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 $P > 0.05$.

A

amhby



B

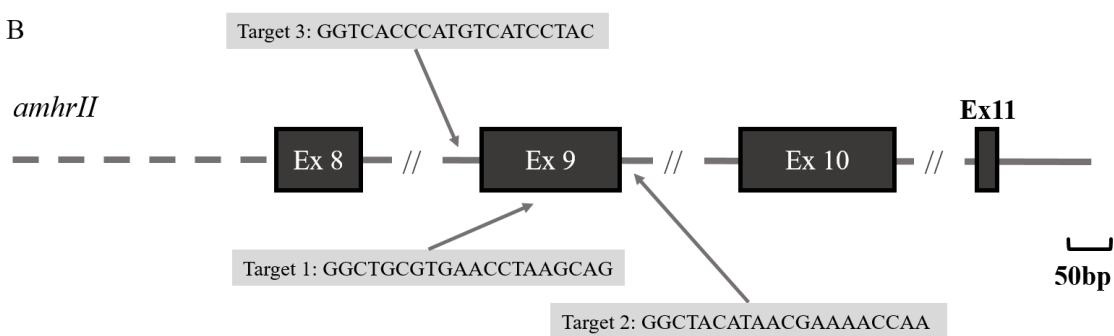


Figure S3: **A:** Position of TALEN targeting *amhby* exon 1 and exon 2. **B:** Position of guide RNAs designed to target *amhrII* exon 9.

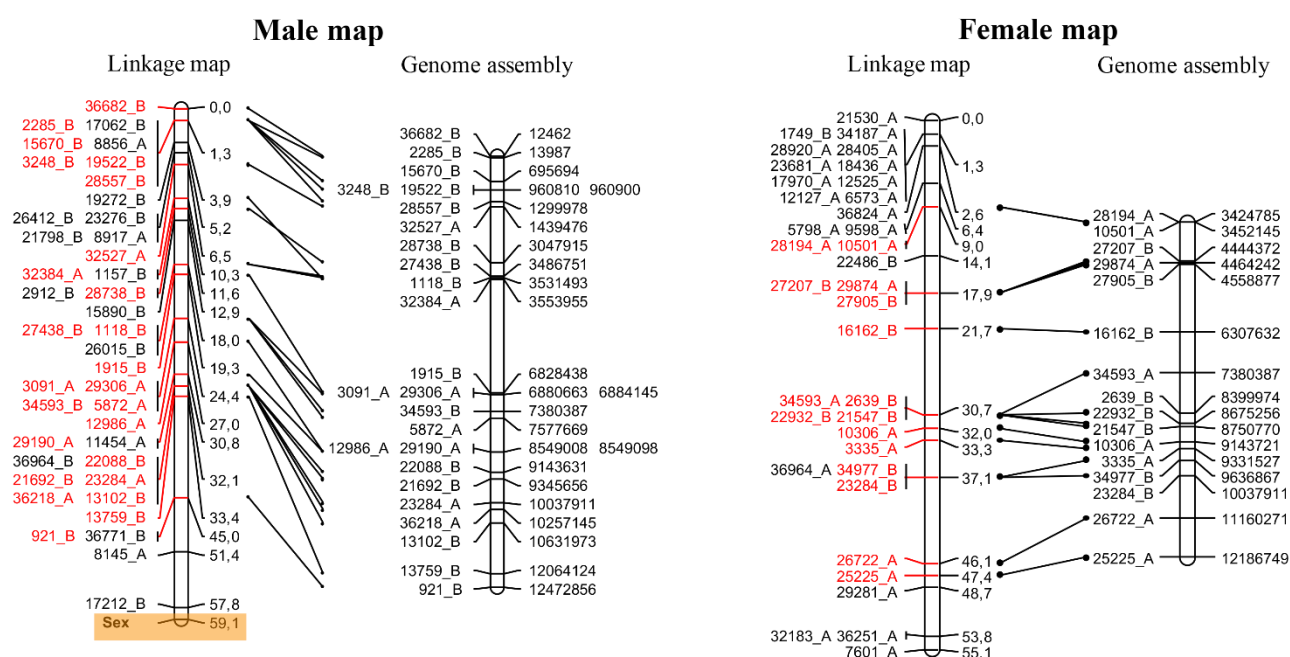


Figure S4: Male and female map of **LG24** of *E. lucius*. For each sex, the linkage map in on the left and the genome assembly on the right, linked by shared markers. Numbers at the left side of the linkage map indicate cumulative distances in cM and numbers at the left side of the genome assembly indicate the position of a given marker. The sex locus on the map linkage map is highlighted with orange shading.

Species name	Protein name	Sequence accession number
<i>Lepisosteus oculatus</i>	Amh	XP_015220826.1
<i>Danio rerio</i>	Amh	NP_001007780.1
<i>Astyanax mexicanus</i>	Amh	XP_007233319.1
<i>Odontestes hatcheri</i>	Amha	ABF47515.2
	Amhy	AHG98063.1
<i>Dicentrarchus labrax</i>	Amh	CAJ 78431.1
<i>Haplochromis burtoni</i>	Amh	XP_005932502.1
<i>Oreochromis aureus</i>	Amh	ABW98500.1
	Amhy	ABB69057.1
<i>Esox lucius</i>	Amha	XP_010870000.1
<i>Oncorhynchus mykiss</i>	Amh	XP_021459508.1

Table S1: Sequence accession number of Amh protein sequences used to construct teleost Amh phylogeny from eight teleost species and Spotted gar (*Lepisosteus oculatus*).

Primer name	Sequence	Target gene	Usage
Con_amha_F1	GTCACATAGGTTGGTTCATC	<i>amha</i>	Genotyping
Con_amha_R1	CGATTTGCCATTTCAAGTG	<i>amha</i>	Genotyping
Con_amhby_F1	GTTACTTTTTCTGCCTAGCGTGA	<i>amhby</i>	Genotyping
Con_amhby_R1	GTGATAGGCACTGTAACACACCA	<i>amhby</i>	Genotyping
Ctrl_AmhrII_F1	GCAACAACACTTGGCAAGTGAAC	<i>amhrII</i>	Genotyping
Ctrl_AmhrII_R1	TATAAGCCATTCTGTGTTCTGGCAC	<i>amhrII</i>	Genotyping
amhby_Xho1_ATG_F1	AACTCGAGATGAGGCTGTCGGGTATATCCT	<i>amhby</i>	Cloning cDNA
amhby_Xba1_Stop_R1	AATCTAGATCAGCGGCACTCCACACTTAGTTGC	<i>amhby</i>	Cloning cDNA
amha_Xho1_ATG_F1	AACTCGAGGCTACCATGAGGCTGTTGCGC	<i>amha</i>	Cloning cDNA
amha_Xba1Stop_R1	AATCTAGATCAGCGGCACTCAGACTTCGTTGC	<i>amha</i>	Cloning cDNA
El_bmpr1a_Xho1_F	CTCGAGATGATGGGCTTCCCTTTTACATG	<i>Bmpr1a</i>	Cloning cDNA
El_bmpr1a_Xba1_R	TCTAGATCAAATTTTGATGTCCTGTGATTCACC	<i>Bmpr1a</i>	Cloning cDNA
amhrII_bamATG_F1	GGATCCCGGCAATGTTATGTTGCACTCTGGTT	<i>amhrII</i>	Cloning cDNA
amhrII_ecoStop_R1	GAATTCTCATACGGAGTGACACAGTCGTAG	<i>amhrII</i>	Cloning cDNA
Probe_amha_F1	ACAACGAAATGTCAAATGCC	<i>amha</i>	ISH probes
Probe_amha_R1_T7	GAAATTAATACGACTCACTATAGGGACGGCAGTCGAAGATATTGG	<i>amha</i>	ISH probe
Probe_amhby_F1	GAAGGAACCAAGTGCATATCCC	<i>amhby</i>	ISH probe
Probe_amhby_R1_T7	GAAATTAATACGACTCACTATAGGGGGATACATCTTGAATTCCTCCA	<i>amhby</i>	ISH probe
El(q)dmrt1_For2	CAATACCTGCAACGAGACCA	<i>dmrt1</i>	qPCR
El(q)dmrt1_Rev2	CTCTACGATGTTCAACAGGGA	<i>dmrt1</i>	qPCR
El(q)bactinS1	CAGGTCATCACCATCGGCAACGAG	<i>β-actin</i>	qPCR
El(q)bactinAS1	GCGTACAGGTCCTTACGGATGTCG	<i>β-actin</i>	qPCR
El(q)ef1aS1	AGTCCTCCACCACCGGCCATC	<i>Ef1-α</i>	qPCR
El(q)ef1aAS1	CATCTCAGCGGCTTCTCTCTCG	<i>Ef1-α</i>	qPCR
El(q)gapdhS1	ATCAACGATCCCTTCATTGACCT	<i>gapdh</i>	qPCR
El(q)gapdhAS1	CATGCTCACCTCACCTTGTAAC	<i>gapdh</i>	qPCR
El(q)gsdfS1	ATCTCTCACAACACCTGGTCAA	<i>gsdf</i>	qPCR
El(q)gsdfAS1	GACCAGCACTAGTCTCAGGTCCA	<i>gsdf</i>	qPCR
El(q)EFTUD1_F	CTGATCCAGGAGACAGGAGAACA	<i>eftud</i>	qPCR
El(q)EFTUD1_R	TGATCTTGGCAAACCTCTCTCTG	<i>eftud</i>	qPCR
El(q)UBR2_F	CGACCGACCTCTACAAGGAAGTT	<i>ubr2</i>	qPCR
El(q)UBR2_R	GTCTTCACTCTGGGGTTCAAGT	<i>ubr2</i>	qPCR
El(q)CCDC6B_F	ATGGAGAACGACACCATCTCAAA	<i>ccdc6b</i>	qPCR
El(q)CCDC6B_R	GCTTCCAGCTTGTCATTCTTTT	<i>ccdc6b</i>	qPCR
El(q)AMH1_For2	TGGATGAAGAGTGACAGAAAGG	<i>amhby</i>	qPCR
El(q)AMH1_Rev2	AGAGAACAACGGGACTGGAG	<i>amhby</i>	qPCR
El(q)AMHR2_For1	GGCTACAGATTGTGGGTACG	<i>amhr2</i>	qPCR
El(q)AMHR2_Rev1	ACTCCCTTATGCTCCCTTCAG	<i>amhr2</i>	qPCR
El(q)amh2S1	CGCATATTCTGCTTGTGTTGTTGCTG	<i>amha</i>	qPCR
El(q)amh2AS1	GGTCCTCCTGATTCTGGCCTTGATAGG	<i>amha</i>	qPCR

Table S2: Primers used in different experiments with their names, sequence, target genes and their usage.

Developmental stage	Total number of samples	Number of females	Number of males
54 dpf	10	3	7
75 dpf	10	4	6
100 dpf	10	4	6
125 dpf	10	4	6

Table S3: Number of total samples, female samples and male samples at different developmental stages used for qPCR. dpf: days post fertilization.

ID	Gonadal phenotype	Evidence
83	Female	Viable eggs used in <i>in vitro</i> fertilization for F1
63	Female	Viable eggs used in <i>in vitro</i> fertilization for F1
45	Female	Viable eggs used in <i>in vitro</i> fertilization for F1
22	Male	Spermiation
49	Female	Dissection: Ovary
27	Male	Spermiation
101	Male	Spermiation
2	Intersex	Spermiation, Dissection: Intersex gonads
41	Undetermined	No spermiation
19	Male	Spermiation
35	Male	Spermiation
59	Female	Viable eggs used in <i>in vitro</i> fertilization for F1
94	Male	Spermiation
92	Male	Spermiation
61	undetermined	No spermiation
55	Female	Dissection: Ovary
104	Male	Spermiation
50	Female	Dissection: Ovary
102	Male	Spermiation

Table S4: Animal ID, gonadal phenotype and evidence for gonadal phenotype of 1-year-old Northern Pike with Crispr-Cas9 *amhrII* knock-out.

LG	Male map		Female map	
	Number of markers	Total cM	Number of markers	Total cM
1	56	84.7	76	68.1
2	38	44.9	79	74.5
3	51	89.4	38	29.5
4	39	56.4	38	72.1
5	39	57.9	55	78.3
6	15	48.9	35	59.1
7	69	75	82	90.3
8	20	9.1	40	52.6
9	19	40	33	54.3
10	42	58	48	73.7
11	38	52.7	64	99.2
12	45	32.1	35	64.4
13	42	38.7	120	72.7
14	32	53.5	43	56.6
15	53	78.6	43	45.3
16	32	42.4	27	19.3
17	52	36.2	36	51.4
18	28	72.2	39	55.3
19	74	56.5	44	65.5
20	64	91.5	54	72.2
21	23	56.6	25	10.2
22	44	57.8	53	59.2
23	10	10.4	27	20.8
24	42	59.1	36	55.1
25	14	29.6	40	77.4
Total	897	1332.2	1210	1477.1

Table S5. The number of maker and the length of the 25 linkage group of *E. lucius* male and female.

ID	LG	Position	Mapping Quality	Sequence
1184	LG08	12863544	41	AGCCTCCTTAGCCTGTGCATTCCCTGCACTACCAACCTCCTCTGCCCTCATCCGACCACA ACCTCCTCCAGCCTCTGCCTGAGCACCTCCTGCA
5188	LG24	22188065	60	TTTAGTGTGTAAGCAACTGGCAAAACTGTAATGGTCAGAGATCCTGCGGTTCCAGATAC CTGACTGGCAAACAGTGTGATGGATCCACCTGCA
6108	un placed scaffold0213	311242	26	TGCAGGCAGAACAGAGCTGAACATCACTGAAACAAAGAGGATTTATTCTTTGAAAGGAGC CAATGCTAAATAATTTACATTTAAATTCCACAAA
7894	un placed scaffold0213	217994	58	TGCAGGTTTAACCACATATGTTGCAAGTTATCATGTTTTCACTACAGGCCTTGACAGCAAA ATAATCTCTCAAGAATAAGTTAATTTAAGTAAT
11743	LG24	1726883	28	CAAGGGAAGTCAAAGCAGCAAGAAGATCTATGCTGAAAACTGGAAACAAATTCACAG CCAATGACCGCTCGTCGGTGTGAAAGGCCTGCA
12976	un placed scaffold0213	323835	60	TGCAGGATGAAAAGTTGATGCGGGGCTAGTGAAGTGGTGGAGAAGAAACGAGGACCGCT ATCTACATTTTCTTTGACGGCCAATATTGCTGTTA
13441	un placed scaffold0213	323745	60	ATTTGAATGGCAAAATGCGCAACATAAATGCTTCATTGTCTGCCATTCTGTCTATGACCAT CAATAAGAAAACACTGACTGCTGCTCACCTGCA
14619	un placed scaffold0213	482281	60	TGCAGGACATCATCAAGGCTGAGGAAGAAATAAATGCCTGTTAGGCAGCGTGATTCTGGA AAGCGTATATTGTCTGTTAATTCCTTTAGGTTA
25769	un placed scaffold0213	482191	60	TTCCATTGAGCAAGGCCCGCTCCACCACATGGGCTCTTGATACAGTGAGAGGAAATTCT TCCTCGTTGTTAACAGCAGCAGAGATGGCCTGCA
30707	LG24	22146752	60	TGCAGGCGCACCTGCGTGAAGTTTGCCCCCCCACCAACGCAAAACGTCGTCAAACATATG CTCGTTTCGTTTGTGCTATGTTTGTGCTGGTTTGTG
26135	un placed scaffold0213	466829	41	TGCAGGTAGAAAAGTGTGGATATCACTAAAACAAAAACAGTAAGCATCTTATTATTAAAC AGTAAGACACATTTTATTTACTGAAATGTACA
29009	LG24	22146662	60	AGTCCTGCTCAGTTTATATATATGAATCATGTGAGACGTGGTGTCTCTCTCCCAGTTGGT GTTTGGAGTGTTAAAGTCCCATGTTCTCCTGCA
6153	LG24	22199821	60	CAAGCAAGGGTTGGTTCCTGCAGCACAATATCTAGACAGCCGTTGTCATAACCCCTTAT GTACCAGGCCCTCTATAAGCGGTTATAACCTGCA

Table S6: Mapping of male specific sequences of *E. lucius* with locus ID, sequence, mapping score and mapped linkage group and position on the genome assembly.

Notes on the chapter “Rise and fall of an ancient master sex determining gene in the Esociformes”

After the identification of the Northern pike MSD *amhby*, which showed considerable sequence divergence compared to its paralog *amha*, we wanted to determine how conserved this MSD was among *Esociformes*. To answer this question, we needed to obtain sexed samples from every species in the *Esociformes* to identify potential *amhby* ortholog sequences and determined the association of these potential sequences with male phenotype. However, obtaining sexed samples for all these species turned out to be one of the biggest challenges of this project. As most of *Esox* species have their distribution restricted to North America or Asia, we relied on collaborators to obtain the samples. As a consequence, we usually did not have control over the number of samples and the method of determination of the sex phenotype of the fishes.

Rise and fall of an ancient master sex determining gene in the *Esociformes*

Pan, Q¹., Feron, R¹., Jouanno, E¹., Koop, B²., Rondeau, E²., Goetz, F. W³., Larson, W. A⁴., Bernatchez, L⁵., Perreault-Payette⁵, A., Tringali, M⁶., Saillant, E⁷., Denys, G. P⁸., van Hippel, F⁹., Chen, S¹⁰., Trifonov, V¹¹., Journot, L¹²., Parrinello, H¹²., Klopp, C¹³., Scharl, M¹⁴., Postlethwait, J.H¹⁵., Herpin, A¹., Guiguen, Y^{1&}.

&: Corresponding author.

1. INRA, UR1037 LPGP, Campus de Beaulieu, Rennes, France.
2. Department of Biology, Centre for Biomedical Research, University of Victoria, Victoria, BC, V8W 3N5, Canada.
3. Environmental and Fisheries Sciences Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA, United States of America.
4. College of Natural Resources, University of Wisconsin-Stevens Point, 800 Reserve Street, Stevens Point, WI 54481
5. Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada, G1V 0A6.
6. Department of Environmental Protection, Florida Marine Research Institute, St. Petersburg, Florida 33701, USA.
7. Gulf Coast Research Laboratory, School of Ocean Science and Technology, The University of Southern Mississippi, Ocean Springs, MS, USA.
8. Muséum national d'histoire naturelle, Département des milieux et peuplements aquatiques, UMR BOREA MNHN - CNRS 7208 - IRD 207 - UPMC, 57, rue Cuvier, CP26, 75005 Paris, France.
9. Department of Biological Sciences and Center for Bioengineering Innovation, Northern Arizona University, Flagstaff, AZ 86011, USA
10. Yellow Sea Fisheries Research Institute, CAFS, Key Laboratory for Sustainable Development of Marine Fisheries, Ministry of Agriculture, Qingdao 266071, China
11. Institute of Molecular and Cellular Biology SB RAS, Lavrentiev ave 8/2, 630090 Novosibirsk, Russia
12. Institut de Génomique Fonctionnelle, IGF, CNRS, INSERM, Univ. Montpellier, F-34094 Montpellier, France.
13. INRA, Sigene, Toulouse, France.
14. University of Wuerzburg, Physiological Chemistry, Biocenter, 97074 Würzburg, Germany; Comprehensive Cancer Center Mainfranken, University Hospital, 97080 Würzburg, Germany; Hagler Institute for Advanced Study and Department of Biology, Texas A&M University, College Station, Texas 77843, USA.
15. Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA.

Authors contributions

Qiaowei Pan

- Performed all experiments and analyses

Contribution of other authors

- Romain Feron: development and implementation of the RADSex pipeline, assisted with bioinformatics analyses, data visualization and revision of the manuscript
- Elodie Jouanno: assisted with molecular experiments
- Ben Koop, Eric Rondeau, Frederik W. Goetz, Wes Larson, Louis Bernatchez, Alysse Perreault-Payette, Mike Tringali, Eric Saillant, Gaël Denys, Frank van Hippel, Songlin Chen, Vladimir Trifonov: provided fish samples
- Hugues Parinello and Laurent Journot: library preparation and sequencing at the MGX platform
- Christophe Klopp: genome sequencing and assembly at the Genotoul platform
- John Postlethwait, Manfred Scharl and Yann Guiguen: designed the Phylosex project and provided fish samples
- Yann Guiguen and Amaury Herpin: supervised my thesis

Abstract

Comparative studies among closely related species have brought important insights on the evolution and transition of sex determination (SD) systems. In the meantime, the identification of master sex determining (MSD) genes added additional layers of understanding of the processes and mechanism involved in the evolution and turnover of SD systems. Here, we surveyed a lineage of teleosts, the *Esociformes* (pikes, pickerels and mudminnows), for the orthologs of *amhby*, a male specific duplicate of the anti-Müllerian hormone that was first identified as the MSD gene in the Northern pike (*Esox lucius*). We estimated that the duplication event that gave birth to *amhby* occurred between 89 and 66 Mya, and that *amhby* was recruited as the MSD gene between 66 and 56 Mya, making it the most ancient MSD gene identified in the teleosts to date. We found varying levels of truncation of this old MSD gene in several lineages within the *Esociformes* even while the gene remained significantly associated with male phenotype, suggesting the emergence of a new mechanism in the function of its protein. In addition, we identified a transition of SD system in the Musky pike (*Esox masquinongy*) and in the North American population of the Northern pike (*Esox lucius*), in which a rapid and complete loss of the sex locus with *amhby* occurred within 0.2 million years. Finally, we investigated the SD system in the Alaska blackfish (*Dallia pectoralis*), where *amhby* was lost after the duplication event, and we found evidence suggesting a potential polygenic SD system in this species. Collectively, our results constitute the first complete evolutionary history of an MSD gene in a group of vertebrate, revealing a branching evolutionary fate for an old MSD gene.

Introduction

Genetic sex determination (GSD) has evolved independently and repeatedly in diverse taxa, including animals, plants and fungi (Bachtrog *et al.*, 2014; Beukeboom and Perrin, 2014), but the stability of such systems varies drastically between groups. In mammals and birds, the same sex determination system has been maintained over a long evolutionary time with conserved master sex determining genes on the Y or the W chromosome respectively (Fridolfsson *et al.*, 1998; Mank and Ellegren, 2007; Graves, 2008; Potrzebowski *et al.*, 2008). In stark contrast, teleost fishes display a wide variety of reproductive strategies and sex determination mechanisms, including unisexuality, hermaphroditism, mono- or polygenic sex determining mechanisms, and environmental influences (Devlin and Nagahama, 2002; Mank and Avise, 2009). Moreover, the GSD systems in teleost fishes are also characterized by a remarkable evolutionary instability, driven by dynamic and rapid sex chromosome and master sex determining gene turnovers (Kikuchi and Hamaguchi, 2013; Heule *et al.*, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017). These properties have made the teleosts an attractive model group to study the process of sex chromosome formation and the turnover of sex determination systems.

In the past two decades, the interest in deciphering sex determination mechanisms in teleosts, facilitated by advances in genomics tools, led to the discovery of a diversity of master sex determination genes (MSD). Incorporating knowledge from these new teleost MSDs, two mechanisms were formulated to describe how the birth of MSD genes can drive sex chromosome turnover in vertebrates: in one mechanism, the birth of a new sex chromosome starts with allelic diversification, while in the other, it starts with gene duplication and translocation of the duplicate to the future sex chromosome, with subsequent neo-functionalization leading to the MSD role (Kikuchi and Hamaguchi, 2013; Heule *et al.*, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017). Teleost MSDs also provided empirical support for the “limited option” hypothesis which states that certain genes, especially those from the SOX, DMRT and TGF- β protein families that are known to be implicated in the sex differentiation pathways, are more likely to be recruited as new sex determiners and eventually cause turnover of sex chromosomes (Graves and Peichel, 2010; Kikuchi and Hamaguchi, 2013; Heule *et al.*, 2014; Herpin and Scharl, 2015; Pan *et al.*, 2017). However, the majority of these recently discovered MSDs were found in diverse teleost clades, making it challenging to infer evolutionary patterns and conserved themes in sex chromosomes turnover.

Comparative studies on closely related species with transition of sex determination systems can thus provide valuable insight on the causes and processes of this turnover (Kikuchi and Hamaguchi, 2013). This approach has been applied to several teleost groups, such as the medakas (*Oryzias sp.*), the Poeciliids, the Tilapiine cichlids, the Salmonids, and the sticklebacks, in which different sex determination systems or sex chromosomes were found even in closely related species (Cnaani *et al.*, 2008; Kondo *et al.*, 2009; Schultheis *et al.*, 2009; Ross *et al.*, 2009; Yano *et al.*, 2013). However, even in these studies, MSDs have only been identified in a few species of the same family in *Oryzias* and salmonids.

Three MSDs have already been identified in the *Oryzias* group: *dmy/dmrt1bY* in *Oryzias latipes* (Matsuda *et al.*, 2002; Nanda *et al.*, 2002), *Sox3^Y* in *Oryzias dancena* (Takehana *et al.*, 2014) and *Gsd^Y* in *O. luzonensis* (Myosho *et al.*, 2012). The identification of these MSDs has brought new insights on the genetic regulation underlying sex chromosomes turnover. In *Oryzias*, it was postulated that a change in the regulation of a gene located downstream from the MSD gene in the sex differentiation cascade could cause this gene to replace the existing MSD gene (Myosho *et al.*, 2012). Studies in this group also revealed concomitant change in the downstream gonadal gene-regulatory network during the transition of MSD gene (Takehana *et al.*, 2014). In salmonids, although only one conserved MSD, *sdY*, was identified (Yano *et al.*, 2012, 2013), the sex locus was mapped to different chromosomes in different species (Davidson *et al.*, 2009), providing support for a jumping MSD relocating itself across the genome with the help of transposable elements (Faber-Hammond *et al.*, 2015; Lubieniecki *et al.*, 2015). Interestingly, the transition of SD systems in *Oryzias* and in *Salmonids* was powered by two very different mechanisms. These examples illustrate that the identification of the MSD gene and its genetic regulation can add a crucial layer to our understanding of the genomic changes associated with SD system turnover mechanisms. Furthermore, they revealed the need for more detailed comparative studies linking MSDs and sex chromosomes to identify conserved patterns in the dynamic evolution of SD systems.

Nascent sex chromosomes have attracted a lot of research interest, as they are excellent models to test theoretical work on sex chromosome evolution, for instance suppression of recombination, accumulation of deleterious elements, and sexual antagonistic selection (Bachtrog and Charlesworth, 2002; Peichel *et al.*, 2004; Hobza *et al.*, 2006; Kondo *et al.*, 2006; Bachtrog *et al.*, 2008; Stöck *et al.*, 2011). The rapid turnover of sex determination systems in the teleosts provides many opportunities to study these nascent sex chromosomes, as many fish species harbor relatively young sex chromosomes (Mank and Avise, 2009; Kikuchi and Hamaguchi, 2013; Heule *et al.*, 2014).

In comparison, other consequences of the rapid turnover of sex determination systems have been neglected, for instance the process of obsolete sex chromosomes reverting to autosome and the potential degeneration of old MSD genes. Among teleost fishes, the case of degenerating an old MSD has only been discussed in *O. luzonensis*, in which *gsdf^y* was identified as the MSD gene and a pseudogene homologous to the ancestral MSD *dmy/dmrt1bY* was also found (Kondo *et al.*, 2006). However, because the precise origin of this pseudogene was not found (Tanaka *et al.*, 2007), the degeneration of the former MSD gene was not further investigated. The paucity of studies on the fate of former MSD genes is likely due to the difficulty in identifying ancestral sex chromosomes and ancestral MSDs; therefore, there is a need for groups of species in which the ancestral state of sex determination can be precisely inferred in order to fill these knowledge gaps on sex chromosome evolution.

The *Esociformes* constitute a small, monophyletic order of teleost fishes, which diverged from its sister clade *Salmoniformes* about 113 million years ago (Mya) (Campbell *et al.*, 2013). The order *Esociformes* contains two families, *Esocidae* (*Esox*, *Dallia* and *Novumbra*) and *Umbridae* (*Umbra*), which diverged from a common ancestor about 89 Mya (Campbell *et al.*, 2013). In one of the most well-known *Esociformes*, the Northern pike (*Esox lucius*) we identified a male-specific copy of the Anti-Müllerian hormone, named *amhby*, as the male MSD gene. In contrast to other known fish MSDs that evolved from an *amh* duplication event (Hattori *et al.*, 2012; M. Li *et al.*, 2015), Northern pike *amhby* differs substantially from its *amh* paralog, which suggests the duplication event occurred a long time ago. As a result, it was interesting to verify whether this *amhby* ortholog was also present in other *Esociformes* species, and if it was, whether it conserved its function as a MSD gene. The relatively small number of species within the *Esociformes* provided a rare opportunity to comprehensively study the evolution of sex determination systems in a monophyletic group of vertebrates.

In this study, we surveyed all 10 species of the *Esocidae* as well as one species of the *Umbridae* to trace the evolutionary trajectory of the master sex determining gene (*amhby*) previously identified in the Northern pike. We estimated that the duplication event of *amh* took place between 89 and 66 Mya in a common ancestor of the *Esocidae*, and that *amhby* was recruited as the MSD gene between 66 and 56 Mya prior to the divergence of the genera *Esox* and *Novumbra*. This old MSD showed different level of truncation in the Chain pickerel (*Esox niger*) and the Olympic mudminnow (*Novumbra hubbsi*), but remained significantly associated with the male phenotype in both species. In contrast, we found a dissociation of *amhby* with the male phenotype in the Musky pike (*Esox masquinongy*), suggesting a potential transition of MSD gene in this species.

Species	Number	Number of females	Number of males	Sample origin	Samples collector	Experiment
<i>Esox lucius</i>	124	62	62	Quebec, Canada	Bernatchez, L.	genotyping RAD-Seq
	20	14	6	British Colombia, Canada	Koop, B. & Rondeau, E.	genotyping
	10	7	3	New Jerseys, USA	Koop, B. & Rondeau, E.	genotyping
	8	5	3	Manitoba, Canada	Koop, B. & Rondeau, E.	genotyping
<i>Esox masquinongy</i>	40	20	20	St. Laurent river, Canada	Bernatchez, L.	genotyping RAD-Seq genome
	49	27	22	Great Lakes	Goetz, R.	genotyping RAD-Seq
	96	48	48	Michigan	Larson, W.A.	genotyping
<i>Esox niger</i>	13	5	8	Quebec, Canada	Bernatchez, L., Perreault-Payette, A.	genotyping
<i>Esox cisalpinus</i>	2	1	1	The Pô Delta, Italy	Denys, G.P., Delmastro, G.	genotyping
<i>Esox aquitanicus</i>	3	1	2	Eyre, Charente and Adour river, France	Denys, G.P.	genotyping
<i>Esox reichertii</i>	4	2	2	China	Chen, S.	genotyping
<i>Esox americanus americanus</i>	10	N.A.	N.A.	Sample used in Ouellet-Cauchon <i>et al.</i> , 2014	Bernatchez, L.	genotyping
<i>Esox americanus vermiculatus</i>	10	N.A.	N.A.	Sample used in Ouellet-Cauchon <i>et al.</i> , 2014	Bernatchez, L.	genotyping
<i>Novumbra hubbsi</i>	50	26	24	Olympic National park, WA,USA	Tabor, R.	genotyping RAD-Seq genome
<i>Dallia pectoralis</i>	60	30	30	Rabbit Slough, AK,USA	Postlethwait, J.H. & von Hippel, F.	genotyping RAD-Seq genome

Table 1: Origins and number of the samples of the Esociformes used in this study.

More surprisingly, we identified a complete loss of the sex locus, including *amhby*, in the North American population of the Northern pike within a short evolutionary time of 0.2 million years. Finally, we investigated the SD system in the Alaska blackfish (*Dallia pectoralis*), where *amhby* was lost in the genome after its duplication, and found evidence of a potential polygenic SD system. Overall, our results allowed us to trace the evolution of sex determination systems in the *Esociformes* starting from one identified MSD. Our results highlight the diversity of evolutionary fates that a single MSD can experience in a monophyletic group of vertebrates.

Material and Methods

Origins of the samples

Information on the animals used in this study can be found in **Table 1**.

Genomic DNA (gDNA) extraction

All fin clips were collected and stored at 4°C in 75-100% ethanol until DNA extraction. Two DNA extraction protocols were used depending on the subsequent procedures. 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. After brief centrifugation, the supernatant containing the genomic DNA was transferred to clean tubes without the Chelex beads (Gharbi *et al.*, 2006). For sequencing preparation, DNA extraction was performed using NucleoSpin Kits for Tissue (Macherey-Nagel, Duren, Germany) following the producer's protocol. When using this protocol, DNA concentration was quantified with a NanoDrop ND2000 spectrophotometer (Thermo scientific, Wilmington, Delaware). Primers for genotyping were designed using PerlPrimer (version 1.1.21, (Marshall, 2004)). The primer sequences and corresponding experiments can be found in **Table S1**.

Genome sequencing and assembly

The genomes of the Northern pike (*Esox lucius*), muskellunge (*Esox masquinongy*), chain pickerel (*Esox niger*), the Olympic mudminnow (*Novumbra hubbsi*), and the Alaska blackfish (*Dallia pectoralis*) were sequenced using a whole genome shotgun strategy with short Illumina reads. Libraries were built using the Truseq nano kit (Illumina, ref. FC-121-4001) following the instruction from the manufacturer. 200 ng of gDNA was briefly sonicated using a Bioruptor (Diagneode). The DNA 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 DNA 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. The libraries were sequenced on a HiSeq2500 using a paired end 2x250 nt v2 rapid mode according to the manufacturer's instruction. Image analysis was performed by the HiSeq Control Software and base calling by the RTA software provided by Illumina

Raw sequencing reads were assembled using the DISCOVAR *de novo* software (Weisenfeld *et al.*, 2014) with the following assembly parameters: MAX_MEM_GB=256, MEMORY_CHECK=False, and NUM_THREADS=16. The resulting assembly metrics were calculated with the assemblathon_stats.pl script (Earl *et al.*, 2011) and are given in **Table S4**.

Sequencing of *amha* and *amhby* sequences

The orthologous sequences of *E. lucius amha* and *amhby* were first PCR amplified from the gDNA of different species of *Esociformes* with primers designed on the *amha* and *amhby* sequences of *E. lucius*. 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 (Eurofins, France) using the same PCR primers. Based on these first partial sequences of *amha* and *amhby* obtained from different species, additional conserved primers were designed based on an alignment of *amha* and *amhby* sequences, and species specific primers were also designed. The partial sequences for each species were assembled with the CAP3 program (Huang and Madan, 1999). All primers used in this study were designed using Perl-Primer (Marshall, 2004). The primers used to amplify *amha* and *amhby* sequences in *Esociformes* are listed in **Table S3**.

Gene phylogeny analysis

Phylogenetic reconstructions were performed on all *amha* and *amhby* sequences obtained from *Esociformes*, with the exception of *E. niger amhby* that was excluded from this analysis because no clear consensus sequence could be deduced from the different sequences obtained in this species. The *amh* sequence from *Salmo salar* was used as an outgroup. Full-length cDNA sequences were predicted with the FGENESH+ suite based on the genomic sequences and were aligned using Clustal X (version 1.2.4, Sievers *et al.*, 2011) with all settings set to default. The suitable substitution model, HKY+G, was determined using JModelTest (version 2.1.4, Darriba *et al.*, 2012). Phylogenetic analyses were performed using the Bayesian method implemented in MrBayes with 500,000 MCMC generations (version 3.2, Huelsenbeck and Ronquist, 2001; Ronquist *et al.*, 2012), and using the Maximum likelihood method implemented in phyML with 1,000 bootstrap resampling replicates (version 3.0, Guindon *et al.*, 2010). An additional phylogenetic tree was constructed with the protein sequences of Amha and Amhby using the Maximum likelihood method (**Figure S1**). The resulting phylogenetic trees were rooted with the Amh protein sequence from *Salmo salar* and visualized with Figtree (version 1.4.1).

Positive selection analysis

The ratio of non-synonymous to synonymous substitutions ($dN/dS = \omega$) among *amha* and *amhby* ortholog sequences was estimated using PALM (version 4.7, Yang, 2007) based on aligned full length cDNA sequences, and the phylogeny obtained with the Bayesian method was used in the analysis. *E. niger* and *N. hubbsi* were excluded from this analysis due to the truncation of their *amhby* ortholog sequences. First, a relative rate test on amino-acid substitution (Hughes, 1994) was performed between *amha* and *amhby* pairs with the *amh* of *U. pygmaea* used as an outgroup sequence: for each species with *amh* duplication, ω was calculated between the *amha* ortholog and *amh* of *U. pygmaea*, and was compared to the ω calculated between the *amhby* ortholog and *amh* of *U. pygmaea*. A Wilcoxon test was used to compare the resulting ω values for the *amh* paralogs of these species. Then, several branch and site models (M0, M1a, M2a, M7, M8, free-ratio, and branch-site) implemented in the CODEML package were fitted to the obtained data. The goodness of fit of these models was compared using the likelihood ratio test implemented in PALM.

Restriction Site Associated DNA (RAD) Sequencing applied to Sex determination (RADSex)

RAD-Sequencing (RAD-Seq) was done on the North American population of *E. lucius*, on *E. masquinongy*, on *N. hubbsi*, and on *D. pectoralis*. RAD libraries were constructed from genomic DNA extracted from fin clips for each species according to standard protocols, using a single *SbfI* restriction enzyme (Amores *et al.*, 2011). These libraries were sequenced on one lane of Illumina HiSeq 2500 for each species. Raw reads were analyzed with *stacks* version 1.44 (Catchen *et al.*, 2011; Catchen *et al.*, 2013). Quality control and demultiplexing of the raw reads were performed with the *process_radtags.pl* wrapper script with all settings set to default. For all species, RAD-Seq reads were analyzed with a custom RADSex analysis pipeline (Feron *et al.*, *in prep*) in order to analyze both polymorphic and non-polymorphic markers, including potentially hemizygous markers. In this analysis, demultiplexed reads were grouped into stacks using the *denovo_map.pl* wrapper script with minimum number of reads to form a stack set to 10 ($m=10$), maximum number of mismatches allowed when forming stacks in one individual set to 0 ($M=0$), maximum number of mismatches allowed when merging stacks in the catalog set to 0 ($N=0$), gapped assembly deactivated (`--gapped` not set), calling haplotypes from secondary reads disabled (`H` not set), *ustacks* max locus per stack set to 1 (`--max_locus_stacks=1`), and *ustacks* lumberjack step deactivated (`--keep_high_cov` set). These parameter values ensured that each stack represented a single allele (or haplotype) from a single locus. The distribution of haplotypes between males and females was calculated using a custom pipeline, and results were presented in a heatmap with number of males in which a haplotype is present as abscissa (x-axis), number of females in which a haplotype is present as ordinate (y-axis), and number of haplotypes present in x males and y females as color of the tile at coordinates (x , y). For each combination of number of males and number of females in which a haplotype is present, probability of association with sex was estimated with a chi-square test, applying Bonferroni correction.

For the North American population of *E. lucius*, for which an assembled genome anchored to linkage groups (LGs) was available (Elu_V3, GenBank assembly accession: GCA_000721915.3), demultiplexed reads were mapped to the genome using BWA (version 0.7.12, Li and Durbin, 2015) with all settings set to default. The resulting BAM-files were run through the *ref_map.pl* pipeline, with a minimum stack depth of 10 ($m=10$) and all other settings set to default. The results of *ref_map.pl* were analyzed with the *populations* program, using the `--fstats` setting to obtain population genetics statistics between sexes.

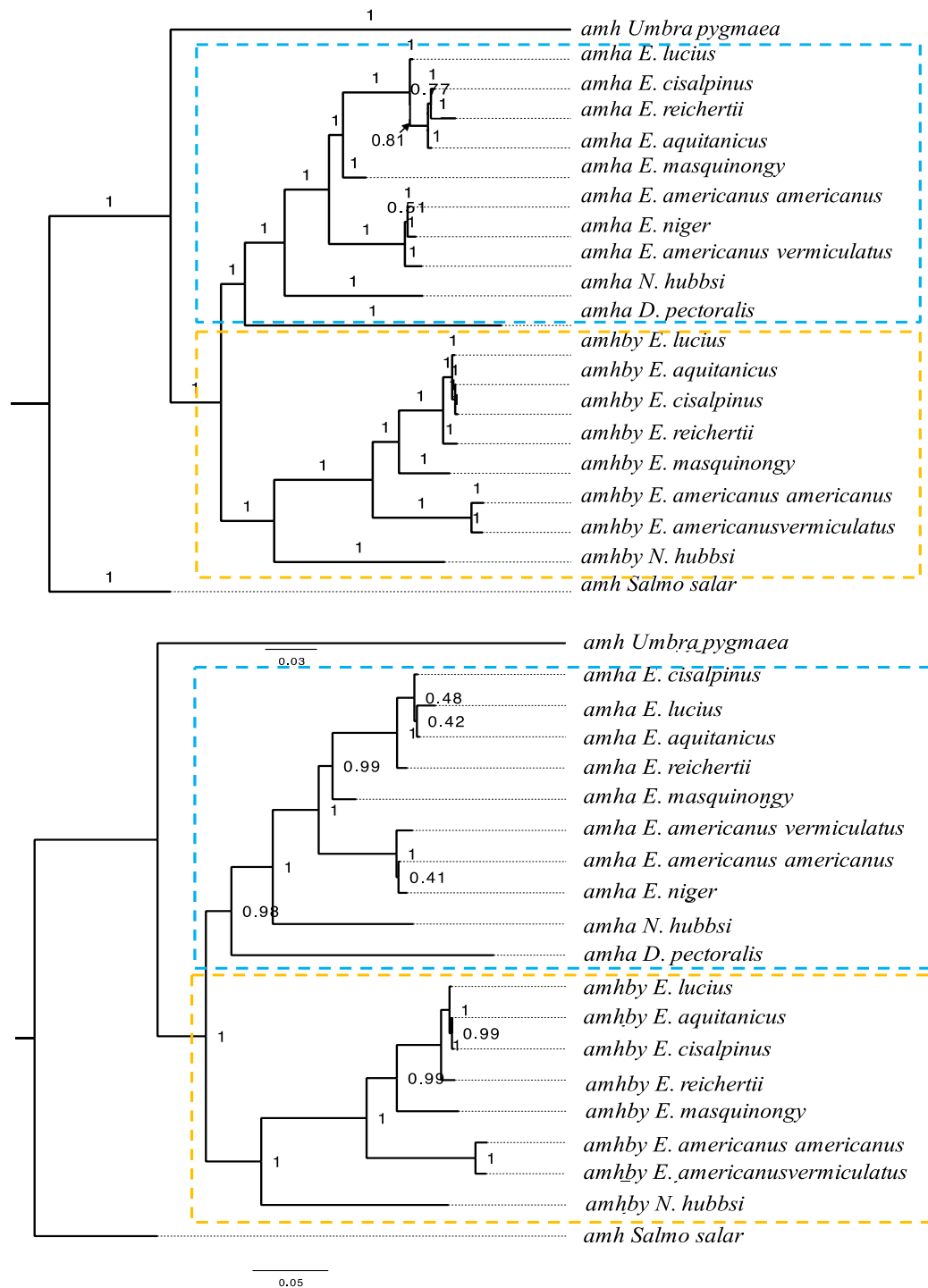


Figure 1: Phylogenetic reconstruction of *amha* and *amhby* orthologs from the *Esociformes* with *amh* sequence from *Salmo salar* as an outgroup. A: Phylogeny tree build with the Bayesian method and B: Phylogeny tree build with the Maximum likelihood method. Bootstrap values are given on each nod of the tree. *amha* ortholog cluster is highlighted with yellow dotted lines and *amhby* ortholog cluster is highlighted with blue dotted

Fisher's exact test was calculated on all polymorphic sites using Plink (version 1.90b4.6 64-bit, Rentería, et al, 2013) to estimate association between variants and phenotypic sex. A Manhattan plot was constructed using a custom R script, showing $-\log_{10}$ (Fisher's test p -value) for all LGs of the *E. lucius* genome.

Pooled-Sequencing (Pool-Seq) to analyze sex determination

In *D. pectoralis*, pooled sequencing (Futschik and Schlötterer, 2010) data was obtained to increase the resolution of detection of sex-specific signal. DNA from 20 males and 20 females was extracted with a NucleoSpin Kit for Tissue (Macherey-Nagel, Duren, Germany) following to the producer's instructions. DNA concentration was quantified using a Qubit dsDNA HS Assay Kits (Invitrogen, Carlsbad, CA) with a Qubit3 fluorometer (Invitrogen, Carlsbad, CA). DNA from different samples were 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. For each sex, the library was then sequenced in a 250 bp paired-end format in an Illumina HiSeq2500.

Pool-Seq reads from *D. pectoralis* were mapped to the *amhby* sequence of *E. lucius* using BWA (version 0.7.12, Li and Durbin, 2009) with all settings set default. Mapped reads were extracted with samtools (version 1.3.1, Li *et al.*, 2009) and remapped to a reference containing both *amha* and *amhby* sequences of *E. lucius*. When reads were mapped to a different position in each analysis, mapping scores were used to determine the origin of the read.

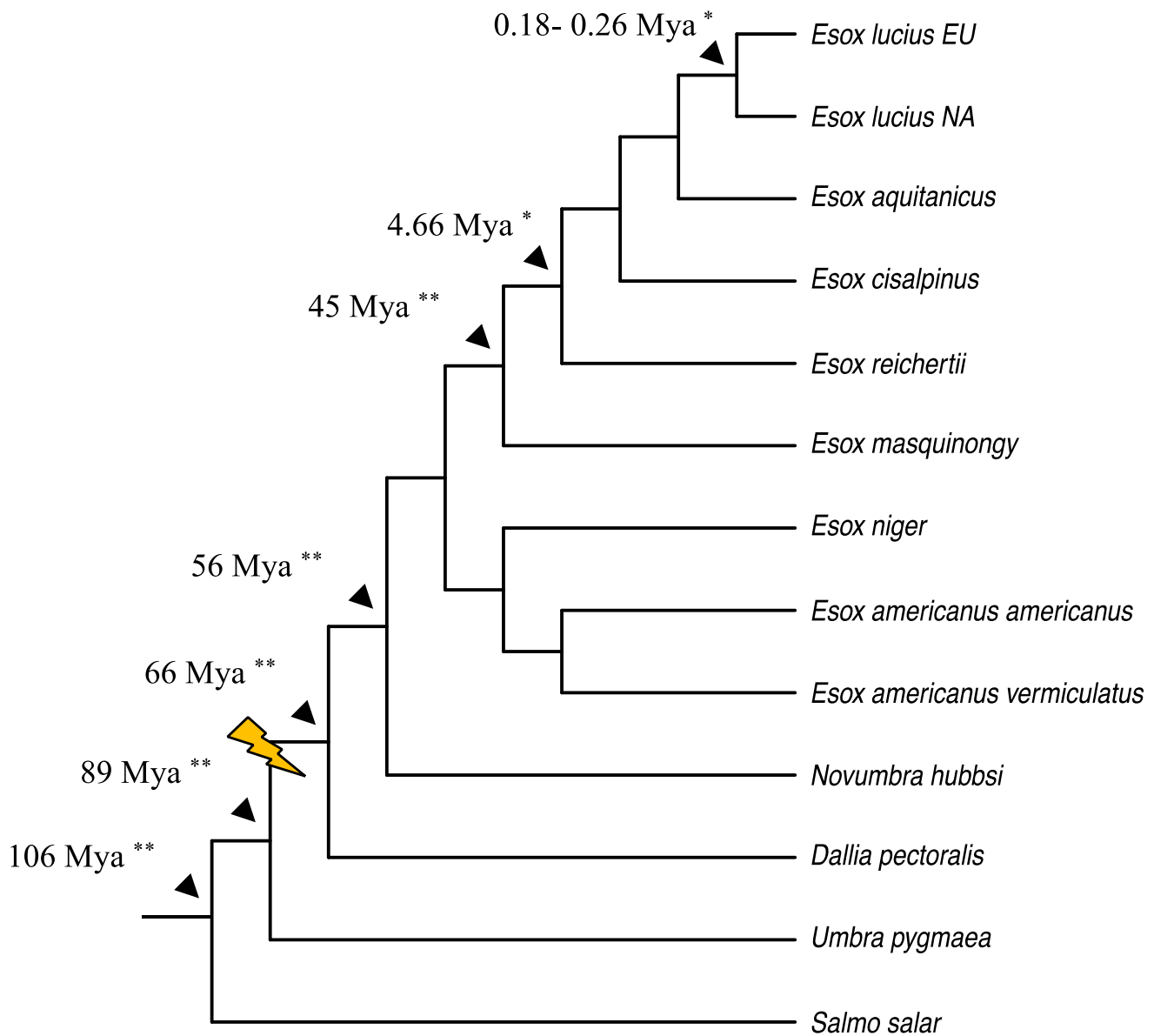


Figure 2: Taxonomy and divergence time estimation of the *Esociformes*. Taxonomical relationship among the *Esociformes* analyzed in this studies was retrieved from literature and presented as a cladogram. Estimated divergence times are listed on the left side of the node.

*: Skog *et al.*, 2014

** : Campbell *et al.*, 2013.

Results

amhby is the result of an old gene duplication

The male specific duplication of *amh*, *amhby*, was first identified in several European populations of *Esox lucius* (Pan *et al.*, *in prep*). To study its evolution within the *Esociformes*, samples were collected in populations of *E. lucius* from North America as well as from the six other species belonging to the genus *Esox*, the monotypic genus *Novumbra*, from one species belonging to the genus *Dallia*, and one species belonging to the genus *Umbra*.

Orthologs of *amha* and *amhby* paralogs were characterized using PCR cloning and sequencing (**Table 2**) and from draft genomes (**Table S2**). An ortholog of *amhby* was identified in all *Esox* species, with the notable exception of the North American population of *E. lucius*, and in *N. hubbsi*, while in *D. pectoralis* and in *U. pygmaea*, *amhby* was not found neither by homology cloning nor by genome analysis. Furthermore, the *amhby* ortholog was found to be truncated in *N. hubbsi*, *E. niger* and *E. masquinongy* (**Figure S2**). In *N. hubbsi*, the exon 1 coding sequence was missing, resulting in a truncated Amhby predicted protein lacking the first 55 amino acids. In *E. niger*, two *amhby* alleles of different lengths were identified in different individuals, and the coding sequence of exon 1 as well as part of the TGF- β domain were missing from both sequences. In *E. masquinongy*, two *amhby* alleles were also identified, sometimes even coexisting within the same individual. The predicted protein sequences from these two *amhby* alleles shared 99.2% identity and 99.6% similarity, and the first conserved cysteine of the TGF- β domain was missing in both proteins.

To better understand the evolution of *amh* duplicates in the *Esociformes*, two phylogenies of *amha* and *amhby* were constructed using either a Bayesian method or a Maximum likelihood method with the HKY+G model (**Figure 1**). These phylogenies show that most *amha* orthologs cluster in a well-supported group containing all *Esocidae* species (*Esox*, *Dallia* and *Novumbra*), and that all *amhby* orthologs cluster in a separate well-supported group. The topology within the cluster of *amhby* orthologs is in complete agreement with the *Esociformes* taxonomy (**Figure 2**). The general topology within the cluster of *amha* orthologs agreed with the *Esociformes* taxonomy, with two poorly supported branches within clusters of closely related species. The *amh* sequence of *U. pygmaea* was always placed outside of the *amha* and *amhby* clusters, indicating that the divergence of the *Umbridae* most likely occurred before the duplication of *amh*. This finding is in agreement with the fact that no *amhby* was found in the genome of *U. pygmaea*.

	Multiple populations	Male samples	Female samples	amhby presence	Number of males with <i>amhby</i>	Number of females with <i>amhby</i>	Association significance level
<i>Esox lucius</i> (North American)	Yes (4)	88	74	No	0	0	N.A.
<i>Esox aquitanicus</i>	No	1	2	Yes	1	0	N.S.
<i>Esox cisalpinus</i>	No	1	1	Yes	1	0	N.S.
<i>Esox reichertii</i>	No	2	2	Yes	2	0	N.S.
<i>Esox masquinongy</i> [#]	Yes (3)	95	90	Yes (truncated)	82	25	**
<i>Esox niger</i>	No	5	8	Yes (truncated)	5	0	*
<i>Esox americanus americanus</i>	No	10 unsexed samples		Yes	Unknown	Unknown	N.A.
<i>Esox americanus vermiculatus</i>	No	10 unsexed samples		Yes	Unknown	Unknown	N.A.
<i>Novumbra hubbsi</i>	Yes	23	22	Yes (truncated)	21	1	***
<i>Dallia pectoralis</i>	No	30	30	No	N.A.	N.A.	N.A.
<i>Umbra pygmaea</i>	No	31	31	No	N.A.	N.A.	N.A.

Table 2: summary of the identification of *amhby* ortholog and their association with sex phenotype in 11 species of the *Esociformes*. Truncation of the obtained sequences was noted. Significance level of p value from chi-squared test on the association between *amhby* ortholog and sex phenotype is also listed in the table:
* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$

[#]: In *E. masquinongy*, *amhby* ortholog showed overall significantly association with male phenotypes, however, it was not tested true for every population surveyed for this species.

In contrast, the *amh* sequence found in *D. pectoralis* was placed within the *amha* orthologs cluster with a significant bootstrap value, strongly indicating that this sequence is orthologous to *amha*. As *amhby* was neither detected in the draft genome nor in the Pool-Seq analysis performed in this species, this result indicates that the duplication of *amh* occurred before the divergence of *Dallia* from the rest of the *Esocidae* and that *amhby* was subsequently lost in this genus. Taken together and combined with a time-calibrated phylogeny of the *Esociformes* (Campbell *et al.*, 2013), these results suggest that the duplication of *amh* which gave rise to *amhby* occurred 65 to 88 Mya, after the divergence of the *Esocidae* from the *Umbridae* (**Figure 2**).

To compare selection pressure between *amha* and *amhby* orthologs, dN/dS ratios were computed between each ortholog sequence and the *amh* sequence of *U. pygmaea*, which was found to have diverged from the other *Esociformes* prior to the duplication of *amh*. Sequences from *E. niger* and *N. hubbsi* were excluded as their *amhby* orthologs were substantially shorter and could introduce bias in the analysis. The resulting (**Table 3**) dN/dS ratios were not significantly different between the two *amh* paralogs, with a mean of 0.378 for *amha* and a mean of 0.376 for *amhby* (Wilcoxon signed-ranked test $p=0.9015$). In order to identify potential positive selection signal on specific sites, several site models allowing different dN/dS ratios among sites were used. However, none of these alternative models fitted the data better than the null model, which allowed a single dN/dS for all sites and all branches in the gene phylogenetic tree of *amh* orthologs (**Table S3**). In summary, no evidence for positive selection was detected on neither *amha* nor *amhby* orthologs, and the paralog pair experienced overall similar selection pressure.

***amhby* was recruited as a MSD gene early during its evolution but experienced some independent declines and losses**

With the exception of the North American population of *E. lucius*, in which *amhby* was not found, and *E. masquinongy*, in which *amhby* association with male phenotype was found to be variable between populations, a clear linkage was found between male phenotype and the presence of *amhby* in all species investigated. This sex linkage was significant in the European population of *E. lucius*, in two populations of *E. masquinongy*, in *E. niger*, and in *N. hubbsi*. It was supported but not significant due to low sample size in *E. reichertii*, in *E. cisalpinus*, and in *E. aquitanicus* (**Table 4 and Supplementary data 1**). These results suggest that *amhby* likely gained its role as a MSD gene quite early after its birth from the duplication of an autosomal *amh* gene.

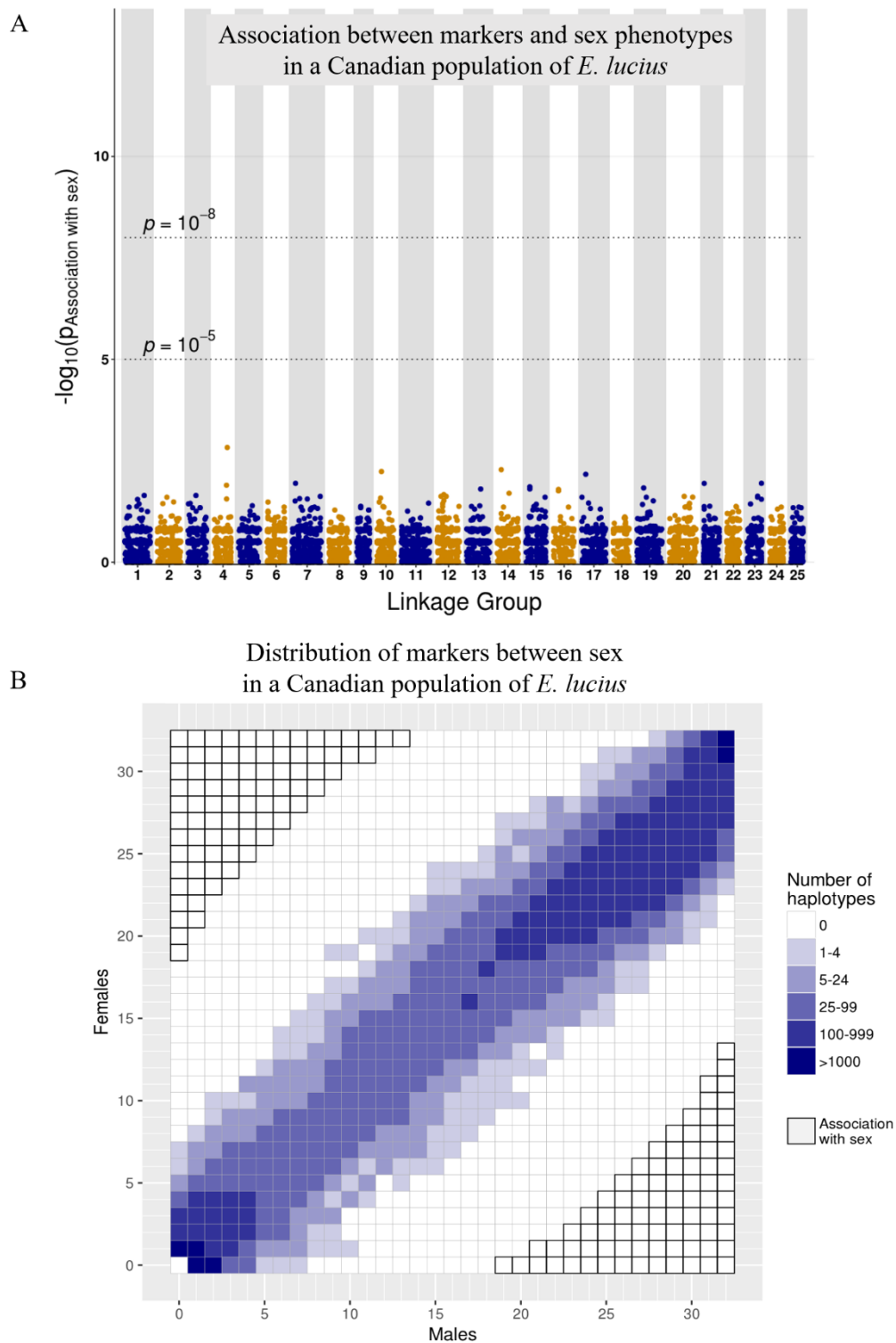


Figure 3: marker association with sex phenotypes in a North American population of *E. lucius*. **A:** Manhattan plot representation of polymorphic markers association with sex phenotypes among the 25 linkage groups of *E. lucius*. **B:** Heat map of the distribution of markers between males and females from the RADSex pipeline. Positions that show significant association with sex are highlighted with the black outline on the lower right corner and upper left corner for male and female respectively. The intensity of the color corresponds to the number of markers

Although a complete sex linkage of *amhby* was previously found in European populations of *E. lucius* (Pan *et al.*, *in prep.*), *amhby* could not be amplified in any sample from several North American populations of *E. lucius*, neither with primers designed specifically on *E. lucius amhby* nor with primers anchored on regions with high conservation among different *amhby* orthologs from several *Esociformes* species. In addition, none of the numerous male specific sequences identified in the European population, including *amhby* (Pan *et al.*, *in prep.*), could be mapped to the North American *E. lucius* genome assembly (Rondeau *et al.*, 2014), which is referenced as a male genome. These results show that all North American *E. lucius* populations lost the entire male specific region on the Y chromosome found in the European populations, including *amhby*.

In order to identify a potential new sex locus and a potential new MSD in these North American *E. lucius* populations, a genome-wide association analysis was performed using RAD-Seq data for 30 males and 30 females. In total, 11,227 polymorphic markers were found and were mapped to the genome assembly. Among these markers, only 35 showed a non-zero F_{ST} between male and female populations and none of them was significantly associated with sex (**Figure 3A**). In order to include both polymorphic and non-polymorphic markers in the analysis, including potentially hemizygous markers, the data were analyzed again using the RADSex pipeline (Feron *et al.*, *in prep.*). The distribution of markers between the two sexes is presented in **Figure 3B**. None of the RAD-Seq markers from this analysis showed significant association with sex, confirming the complete absence of sex-associated signal, at least within the resolution of the RAD-Sequencing approach (around one pair of markers every 52 kb, deduced from the number of *SbfI* restriction sites found in the genome of *E. lucius*).

In *E. masquinongy*, genotyping revealed different frequencies of males and females carrying *amhby* in different populations, and even some animals carrying two copies of *amhby* (**Table 2**). To understand the origin of these differences, RADSeq data was generated for 25 males and 28 females from a Great lakes population and for 20 males and 20 females from a Canadian population, and each population was analyzed separately using the RADSex pipeline. In the Great lakes population, 11 markers were significantly associated with male phenotype, and no marker was significantly associated with female phenotype, indicating an XX/XY sex determination system (**Figure 4A**). In contrast, no marker showed significant association with either sex in the Canadian population (**Figure 4B**). The 11 markers significantly associated with male phenotype were mapped to the genome assembly of *E. lucius*, but none of the markers was uniquely mapped. In addition, when mapped to our draft genomes of *E. masquinongy*, 4 of the 11 markers mapped perfectly on the male genome but mapped with multiple positions and/or with a 2 bp gap on the female genome.

ω (comparing with <i>U. pygmaea amh</i>)		
Species	<i>amha</i> ortholog	<i>amhby</i> ortholog
<i>D. pectoralis</i>	0.37	NA
<i>N. hubbsi</i>	0.4301	truncated
<i>E. niger</i>	0.3732	truncated
<i>E. americanus americanus</i>	0.3826	0.3579
<i>E. americanus vermiculatus</i>	0.3933	0.3561
<i>E. masquinongy</i>	0.3877	0.3808 & 0.3857
<i>E. reichertii</i>	0.3786	0.3797
<i>E. aquitanicus</i>	0.3706	0.3848
<i>E. cisalpinus</i>	0.3629	0.3876
<i>E. lucius</i>	0.3678	0.3804

Table 3: ω value between the *amh* paralogs in different *Esociformes* and *amh* ortholog of *U. pygmaea*.

Population	Males with <i>amhby</i>	Females with <i>amhby</i>	P
Canada (St. Laurent river)	9/20 (45%)	4/20 (20%)	0.177
Great Lakes	27/27 (100%)	4/22 (18%)	2.01E-08
Michigan	46/48 (96%)	20/48 (42%)	2.91E-07

Table 4: Proportion of males and females carrying *amhby* ortholog in three populations of *E. masquinongy*. P value from chi-squared test on the association between *amhby* ortholog and sex phenotype is also listed in the table.

The remaining markers were either not mapped (N=2), or were mapped to multiple positions with mismatches on both male and female genomes (N=5). The size of the four male genomic contigs hit by the perfectly mapped markers ranged from 382 to 2,403 bp, and none of these four contigs could be mapped on the *E. lucius* genome.

What is the sex determination system in *D. pectoralis*, a species that lost *amhby*?

RADSeq data was generated for 30 males and 30 females *D. pectoralis* and analyzed using the RADSex pipeline. Intriguingly, both markers significantly associated with male phenotype and markers significantly associated with female phenotype were found (**Figure 5**). Two of these markers were found in 28 males (93%) and 4 females (13%), and one additional marker was found in 28 males (93%) and 6 females (20%). One marker was found in 28 females (93%) and 11 males (37%), and another marker was found in 25 females (83%) and 3 males (10%). Two of the significantly male-associated markers were mapped on the same contig in the genome assembly of a male individual. However, this contig is only 1,193 bp in length and did not contain any protein coding sequence, nor did it show any homology with sequences from the genome of *E. lucius*. Overall, these results do not support a simple monofactorial sex determining system, and suggest sex-determination in *D. pectoralis* could be polygenic.

Discussion

A male-specific copy of the anti-müllerian hormone, named *amhby*, was recently identified as the male MSD gene in the Northern pike, *E. lucius* (Pan *et al.*, *in prep.*). In contrast to other known fish MSD genes that evolved from a duplication of *amh* (Hatorri *et al.*, 2012, Li *et al.*, 2015), the Northern pike *amhby* diverged substantially from its autosomal counterpart *amha*, suggesting a long divergence time since its duplication. In agreement with this idea, we identified *amhby* orthologs in all species from the genera *Esox* and *Novumbra*. Phylogenetic reconstruction of the evolution of *amha* and *amhby* showed that the duplication of *amh* took place prior to the divergence of the genera *Dallia*, *Esox* and *Novumbra* but after the divergence of the genus *Umbra* from the rest of the *Esociformes*.

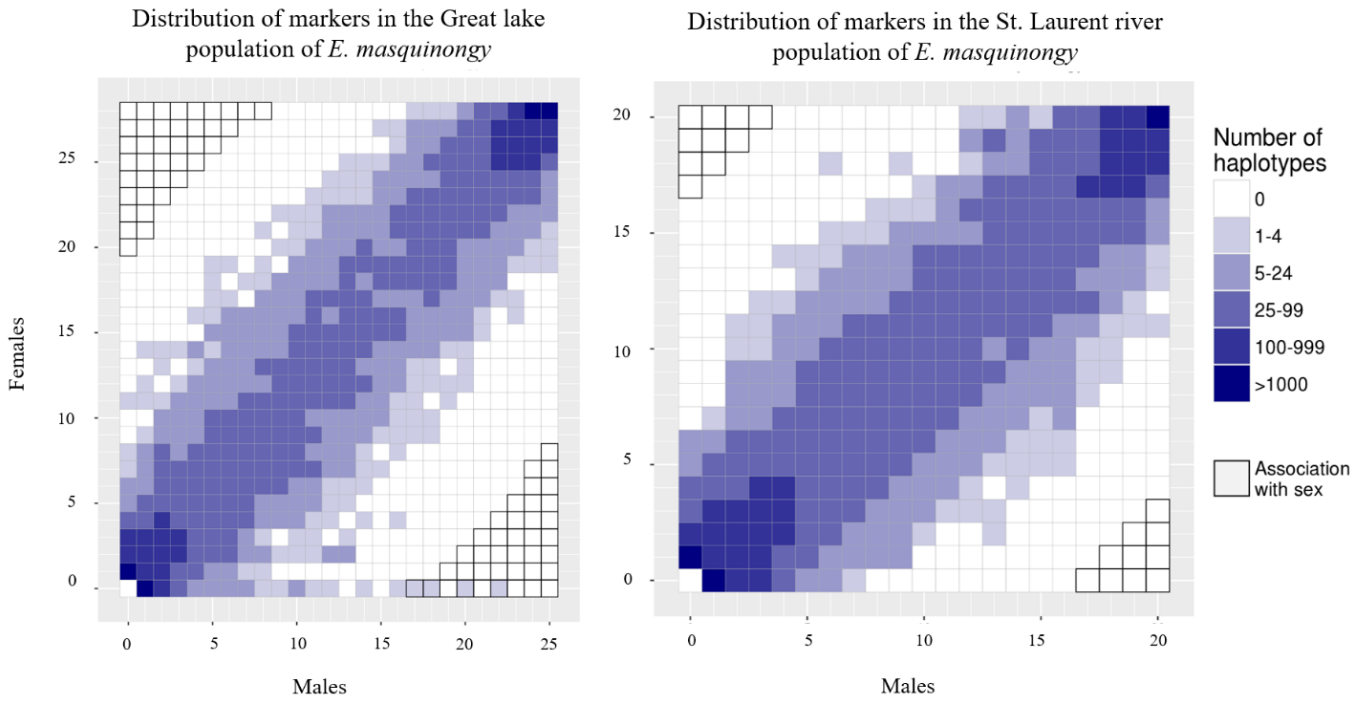


Figure 4: Heat map of markers shared by males and females of two populations of *E. masquinongy* from the RADSex pipeline. Positions that show significant association with sex are highlighted with the black outline on the lower right corner and upper left corner for male and female respectively. The intensity of the color corresponds to the number of markers.

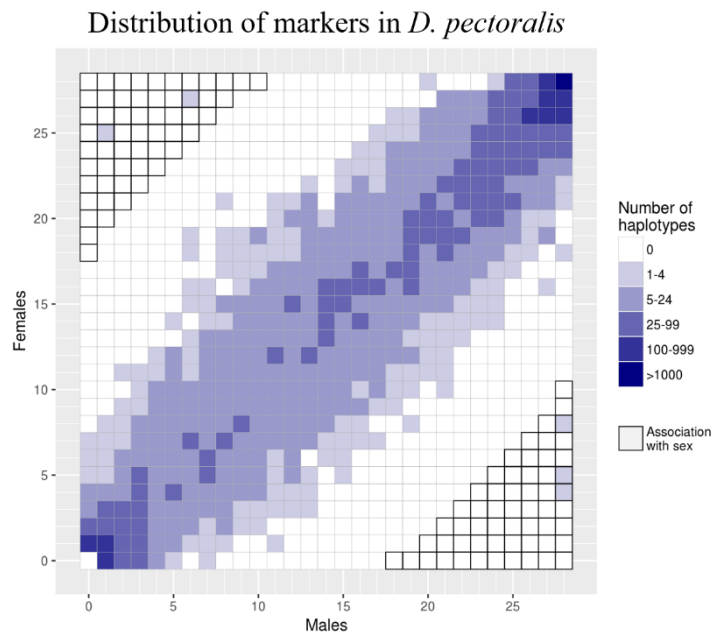


Figure 5: Heat map of markers shared by 28 males and 28 females of *D. pectoralis* from the RADSex pipeline. Positions that show significant association with sex are highlighted with the black outline on the lower right corner and upper left corner for male and female respectively. The intensity of the color corresponds to the number of markers.

Combined with a time-calibrated phylogeny of the *Esociformes* (Campbell *et al.*, 2013), these results suggest that the duplication of *amh* occurred between 65 and 90 Mya. However, *amhby* was not found in *D. pectoralis*; as gene loss was hypothesized to be the most common fate for duplicated genes (Conant and Wolfe, 2008), the simplest hypothesis would then be that *amhby* was quickly lost in the genus *Dallia* after its divergence from the rest of the *Esocidae*.

Knowing that *amhby* originated from an old duplication, the following task was then to determine the timing of its recruitment as a MSD gene. Genotyping on sexed samples showed significant association between the presence of *amhby* and the male phenotype in all *Esocidae* species in which *amhby* was found, supporting the hypothesis that *amhby* was recruited as a MSD gene after prior to the divergence of the *Novumbra* and *Esox* genera, around 56 Mya (Campbell *et al.*, 2013). The alternative hypothesis would be that *amhby* was recruited around 66 Mya in the common ancestor of the *Esocidae*, and was subsequently lost in the *Dallia* genus. Either scenario would make *amhby* the most ancient MSD gene identified in the teleosts to date.

As *amhby* acquired its MSD function through duplication and subsequent neofunctionalization, its evolution is expected to have first gone through a relaxation of purifying selection post-duplication, which enabled the acquisition of its new MSD function. It would then experience positive selection to optimize its newly acquired function, before its evolution became constrained by purifying selection (Ohno, 1970; Kimura and Ohta, 1974; Hughes, 1994; Conant and Wolfe, 2008). Comparison of the molecular evolution rates between *amhby* and *amha* orthologs did not show any signature of positive selection on *amhby*. Instead, we found that both *amhby* and *amha* experienced similar purifying selection on all sites and in all species. This absence of signature of post-duplication relaxation of purifying selection or of directional selection suggests that during the majority of its life span, *amhby* experienced a purifying selection similar to its paralogous gene *amha*, which preserved the ancestral function of *amh* after the duplication. In light of the fact that *amhby* is associated with male phenotype in the majority of the *Esocidae*, this result suggests that *amhby* evolved its MSD function early after the duplication event in the *Esocidae*. It is worth noting that the *amhby* truncated sequences in *E. niger* and in *N. hubbsi* were not included in this analysis, and thus the onset of purifying selection within the *Esociformes* lineage cannot be precisely inferred from our analysis. Furthermore, only coding sequences were included, while selection could equally occur in the regulatory regions (Force *et al.*, 1999). Comparing molecular evolution rates on regulatory regions of *amhby* orthologs could have complemented our analysis, but the combination of substantially fragmented draft genomes and the complexity of determining the exact range of these regulatory regions prevented us from using this approach.

Ortholog sequences of *amhby* with different levels of truncation compared to the sequence of *Esox lucius amhby* were found in *N. hubbsi*, *E. masquinongy*, and *E. niger*. Such truncations of the Amh protein sequence could lead to defects in its function and even to a complete loss of function. However, an *amh* duplicate lacking exon II and exon III has been identified as a strong candidate for MSD gene in cobaltcap silverside (*Hypoatherina tsurugae*), indicating that male specific *amh* duplicates could maintain their function as MSD gene even after deep structural modification (Bej *et al.*, 2017).

In *N. hubbsi*, the Amhby protein lacks the N-terminal part encoded by the first exon, with no additional deletions in the rest of the protein, including the TGF- β domain coded by the last exon. In this species, *amhby* was almost perfectly associated with males, with a few outliers. It is worth noting that misassignment of phenotypic sex is not unlikely in this species due to the sexing method during sample collection. Because *N. hubbsi* is endemic to western Washington state (USA) and is considered a state sensitive species (DeHaan *et al.*, 2014), the phenotypic sex of the samples used in this study was solely determined based on the external phenotype. As this external sex phenotyping method is more prone to mistakes than direct visual inspection or histology of the gonads, it is possible that these few outliers were misassigned phenotypes. In this case, the sex linkage of *amhby* would be complete in *N. hubbsi*, indicating that this truncated *amhby* is likely to be functional as the MSD in this species. In *E. niger*, the sequence of *amhby* also completely lacks the exon I and, more surprisingly, also part of the last exon, resulting in an impaired TGF- β domain. Yet, this *amhby* ortholog was found to be perfectly associated with male phenotype, indicating that this gene could still be a good candidate for MSD gene in this species.

These two cases raise the question of how such truncated Amhby proteins could still serve as functional MSDs. For a canonical AMH, removal of the N-terminus during posttranslational proteolytic processing is necessary to generate a mature and much smaller protein that can interact with its receptor (di Clemente *et al.*, 2010). The truncated Amhby in *N. hubbsi* still encodes the entire C-terminal sequence, which is sufficient to produce a normal mature protein able to interact with its receptor, and thus, remain functional. In comparison, defects in the TGF- β domain like that found in *E. niger* could severely affect the interaction between Amh and its receptor. However, the mechanism by which Amhby triggers sex determination in the *Esociformes* is still unknown. Indeed, the TGF- β superfamily, which seems to play a more central role in sex determination and differentiation in teleosts compared to mammals (Kikuchi and Hamaguchi, 2013), remains poorly studied in fish.

Because of this lack of knowledge on the precise mechanisms of action and on the signalization pathways of *Amh* in fish, the possibility of an unusual and yet unknown function of these truncated *Amhby* proteins remains. Moreover, in the salmonids a truncated duplicate of an immune-related gene acting as MSD, *sdY*, demonstrated that preservation of all domains is not always necessary for a neo-functionalized protein to assume the role of MSD gene (Yano *et al.*, 2012; Bertho *et al.*, *subm.*).

In *E. masquinongy*, despite being significantly male sex-linked, *amhby* was absent in some males and even present in some females. In addition, two alleles of *amhby* were found to co-exist in some individuals, which is not compatible with the expected hemizygous status of a Y sex locus. Furthermore, the *E. masquinongy* *Amhby* protein lacks the first cysteine of its TGF- β domain, which could impair its folding and thus potentially lead to defects in its function. Yet, the significantly predominant presence of *amhby* in males in two out of three populations of *E. masquinongy* indicates that this gene is likely linked with sex and located on the sex chromosome. One scenario that could explain this observed pattern in *E. masquinongy* is that a novel male MSD has arisen in the vicinity of *amhby*. This new male MSD gene would be strictly associated with the male phenotype, and as a result of the short distance between this new MSD gene and *amhby*, these two genes would not segregate completely independently, leading to a higher chance of *amhby* being found in males than females. Furthermore, in this scenario, due to the relaxation of purifying selection after the loss of the MSD role, *amhby* could start to degenerate, which would explain the observed changes in the protein structure. An alternative explanation would be that sex determination in *E. masquinongy* is polygenic, with *amhby* as the male sex determining gene on the Y chromosome and a dominant female determining gene on a W chromosome. Such multilocus polygenic systems have been previously described in the housefly (Hamm *et al.*, 2014; Meisel *et al.*, 2016) and in the cichlid fish *Astatotilapia burtoni* (Moore and Roberts, 2013; Roberts *et al.*, 2016). In this scenario, *E. masquinongy* individuals would then have two pairs of sex chromosomes and the resulting sex phenotype would be driven by a combination of XX/XY and ZZ/ZW systems. More precisely, males and females with *amhby* would have an XYZZ or an XYZW genotype respectively, and males and females without *amhby* would have an XXZZ or an XXZW genotype respectively. Although no female-biased marker was found in our study to support this hypothesis, such a female dominant sex determining locus could potentially be very small and thus undetectable with the resolution of RAD-Seq.

Interestingly, this multilocus polygenic scenario would better fit with the fact that *E. masquinongy* has been previously characterized as a potential ZZ/ZW species due to the presence of males in the offspring from a gynogenic female (Dabrowski *et al.*, 2000): under this hypothesized polygenic system, male offspring could still be produced by a gynogenic female with XYZW genotype. Furthermore, the XX/XY system found from the results of the RADSex pipeline could result from a stronger signal from the more differentiated X and Y chromosome pair. Further data is needed to elucidate the puzzle of the SD system in *E. masquinongy*; nevertheless, our data indicates that *amhby* is no longer the sole MSD gene in this species.

Surprisingly, despite having been clearly identified as the MSD in an XX/XY sex determination system in European populations of *E. lucius*, *amhby* was completely absent in North American populations. This result is corroborated by the complete absence of sex-biased RAD markers in the North American populations, even on LG24, the sex chromosome identified in the European populations (Pan *et al.*, *in prep*). This indicates that in these North American populations, the sex chromosome is currently at an autosomal state. The current distribution of the extant populations of *E. lucius* is the result of an expansion from three different glacial refugia dated around 0.18 to 0.26 Mya, with the North American populations belonging to the monophyletic circumpolar lineage with a Eurasian origin (Skog *et al.*, 2014). We found a complete sex linkage of *amhby* in some European populations belonging to this circumpolar lineage. It is possible that *amhby* was lost in this entire circumpolar lineage prior to its expansion and then regained in the European populations, for instance from a secondary contact with the other two sympatric European lineages. However, the most parsimonious hypothesis is the sex locus was lost in the ancestors of the North American populations after their divergence from the rest of the circumpolar lineage, less than 0.18 Mya. In this case, the complete loss of this sex locus in such a short evolutionary time was likely not driven by a gradual degeneration of the sex locus, but instead driven by a drastic loss event. Such event may have resulted from the combination of a bottleneck in the refugia and founder effects during postglacial recolonization, which was suggested as an explanation for the low level of genetic diversity observed in some freshwater fishes at northern latitudes (Bernatchez and Wilson, 1998). In fact, the lowest genetic diversity among populations from the circumpolar lineage of *E. lucius* was found in North American populations (Skog *et al.*, 2014), indicating that these populations experienced the strongest bottleneck and founder effects. In this context, the loss of the *amhby* sex locus can be explained by the hypothesis that some of the founder males were actually genetic females that experienced sex reversal. This hypothesis is supported by the fact that masculinization of XX females, likely due to stress, has been observed in captive *E. lucius* (Pan *et al.*, *in prep*).

These environmental effects on sex determination could have further enabled this abrupt sex determination transition by providing a mechanism to maintain a balanced sex ratio in the population even after the complete loss of its ancient sex locus.

It is possible that a new master sex determining gene has already taken the lead of a new genetic sex determination system in these North American populations. Although we did not find any marker associated with sex in the RADSex analysis, it may be that the resolution of RAD-Seq is too low to catch the signal of nascent sex chromosomes with little differentiation. Resequencing projects with increased resolution may be able to pinpoint the location of the potential young sex locus, and experiments are currently ongoing to provide high resolution genome-wide information on the North American populations of *E. lucius*.

Conclusion

In this study, we traced the evolution of the sex determination system in a monophyletic group of vertebrate with an old but dynamic MSD gene as a guide. We showed that the MSD gene *amhby*, the male specific duplicate of the anti-Müllerian hormone first identified in the Northern pike (*Esox lucius*), resulted from an ancient duplication event prior to the diversification of the *Esocidae* between 65 and 90 Mya. This *amh* duplicate then gained the function of MSD gene in a common ancestor of the genera *Esox* and the *Novumbra* around 56 Mya. Subsequently, this old MSD experienced varying fates among different species of the *Esocidae*. We identified two cases of truncated *Amhby* that remained significantly associated with male phenotype, sparking speculation on a new mechanism in the function of the protein. We also found evidence for a transition of sex determination system in the Musky pike (*Esox masquinongy*), where *amhby* showed dissociation from male phenotype, and more surprisingly, in the North American population of the Northern pike (*Esox lucius*), where a rapid and complete loss of the sex locus with *amhby* occurred within 0.2 million years. Finally, we found evidence suggesting a potential polygenic SD system operating in the Alaska blackfish (*Dallia pectoralis*), where *amhby* was lost after the duplication event. Collectively, these results provide the first empirical study of the branching evolutionary trajectory of a single, old MSD gene in a group of species experiencing transitions of sex determination systems.

References

- Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., & Postlethwait, J. H. (2011). Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. *Genetics*, 188(4), 799–808. <https://doi.org/10.1534/genetics.111.127324>
- Bachtrog, D., & Charlesworth, B. (2002). Reduced adaptation of a non-recombining neo-Y chromosome. *Nature*, 416(6878), 323–326. <https://doi.org/10.1038/416323a>
- Bachtrog, D., Hom, E., Wong, K. M., Maside, X., & de Jong, P. (2008). Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biology*, 9(2), R30. <https://doi.org/10.1186/gb-2008-9-2-r30>
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., ... The Tree of Sex Consortium. (2014). Sex Determination: Why So Many Ways of Doing It? *PLoS Biology*, 12(7), e1001899. <https://doi.org/10.1371/journal.pbio.1001899>
- Bej, D. K., Miyoshi, K., Hattori, R. S., Strüssmann, C. A., & Yamamoto, Y. (2017). A Duplicated, Truncated amh Gene Is Involved in Male Sex Determination in an Old World Silverside. *G3: Genes, Genomes, Genetics*, g3.117.042697. <https://doi.org/10.1534/g3.117.042697>
- Bernatchez, L., & Wilson, C. C. (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, 7(4), 431–452. <https://doi.org/10.1046/j.1365-294x.1998.00319.x>
- Beukeboom, L. W., & Perrin, N. (2014). *The Evolution of Sex Determination*. Oxford University Press.
- Campbell, M. A., López, J. A., Sado, T., & Miya, M. (2013). Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. *Gene*, 530(1), 57–65. <https://doi.org/10.1016/j.gene.2013.07.068>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes/Genomes/Genetics*, 1(3), 171–182. <https://doi.org/10.1534/g3.111.000240>
- Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., ... Kocher, T. D. (2008). Genetics of Sex Determination in Tilapiine Species. *Sexual Development*, 2(1), 43–54. <https://doi.org/10.1159/000117718>
- Conant, G. C., & Wolfe, K. H. (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*, 9(12), 938–950. <https://doi.org/10.1038/nrg2482>
- Dabrowski, K., Rinchar, J., Lin, F., Garcia-Abiardo, M. A., & Schmidt, D. (2000). Induction of gynogenesis in muskellunge with irradiated sperm of yellow perch proves diploid muskellunge male homogamety. *The Journal of Experimental Zoology*, 287(1), 96–105.
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772. <https://doi.org/10.1038/nmeth.2109>
- Davidson, W. S., Huang, T.-K., Fujiki, K., Schalburg, K. R. von, & Koop, B. F. (2009). The Sex Determining Loci and Sex Chromosomes in the Family Salmonidae. *Sexual Development*, 3(2–3), 78–87. <https://doi.org/10.1159/000223073>
- DeHaan, P. W., Adams, B. A., Tabor, R. A., Hawkins, D. K., & Thompson, B. (2014). Historical and contemporary forces shape genetic variation in the Olympic mudminnow (*Novumbra hubbsi*), an endemic fish from Washington State, USA. *Conservation Genetics*, 15(6), 1417–1431. <https://doi.org/10.1007/s10592-014-0627-7>

- Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture*, 208(3), 191–364. [https://doi.org/10.1016/S0044-8486\(02\)00057-1](https://doi.org/10.1016/S0044-8486(02)00057-1)
- di Clemente, N., Jamin, S. P., Lugovskoy, A., Carmillo, P., Ehrenfels, C., Picard, J.-Y., ... Cate, R. L. (2010). Processing of Anti-Müllerian Hormone Regulates Receptor Activation by a Mechanism Distinct from TGF- β . *Molecular Endocrinology*, 24(11), 2193–2206. <https://doi.org/10.1210/me.2010-0273>
- Earl, D., Bradnam, K., St. John, J., Darling, A., Lin, D., Fass, J., ... Paten, B. (2011). Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Research*, 21(12), 2224–2241. <https://doi.org/10.1101/gr.126599.111>
- Faber-Hammond, J. J., Phillips, R. B., & Brown, K. H. (2015). Comparative Analysis of the Shared Sex-Determination Region (SDR) among Salmonid Fishes. *Genome Biology and Evolution*, 7(7), 1972–1987. <https://doi.org/10.1093/gbe/evv123>
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151(4), 1531–1545.
- Fridolfsson, A.-K., Cheng, H., Copeland, N. G., Jenkins, N. A., Liu, H.-C., Raudsepp, T., ... Ellegren, H. (1998). Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proceedings of the National Academy of Sciences*, 95(14), 8147–8152.
- Futschik, A., & Schlötterer, C. (2010). The Next Generation of Molecular Markers From Massively Parallel Sequencing of Pooled DNA Samples. *Genetics*, 186(1), 207–218. <https://doi.org/10.1534/genetics.110.114397>
- Gharbi, K., Gautier, A., Danzmann, R. G., Gharbi, S., Sakamoto, T., Høyheim, B., ... Guyomard, R. (2006). A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, 172(4), 2405–2419. <https://doi.org/10.1534/genetics.105.048330>
- Graves, J. A. M. (2008). Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes. *Annual Review of Genetics*, 42(1), 565–586. <https://doi.org/10.1146/annurev.genet.42.110807.091714>
- Graves, J. A. M., & Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biology*, 11(4), 205.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hamm, R. L., Meisel, R. P., & Scott, J. G. (2014). The Evolving Puzzle of Autosomal Versus Y-linked Male Determination in *Musca domestica*. *G3: Genes/Genomes/Genetics*, 5(3), 371–384. <https://doi.org/10.1534/g3.114.014795>
- Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., ... Strussmann, C. A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences*, 109(8), 2955–2959. <https://doi.org/10.1073/pnas.1018392109>
- Herpin, A., & Scharf, M. (2015). Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Reports*, 16(10), 1260–1274. <https://doi.org/10.15252/embr.201540667>
- Heule, C., Salzburger, W., & Bohne, A. (2014). Genetics of Sexual Development: An Evolutionary Playground for Fish. *Genetics*, 196(3), 579–591. <https://doi.org/10.1534/genetics.114.161158>
- Hobza, R., Lengerova, M., Svoboda, J., Kubekova, H., Kejnovsky, E., & Vyskot, B. (2006). An accumulation of tandem DNA repeats on the Y chromosome in *Silene latifolia* during early stages of sex chromosome evolution. *Chromosoma*, 115(5), 376–382. <https://doi.org/10.1007/s00412-006-0065-5>

- Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, 9(9), 868–877.
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)*, 17(8), 754–755.
- Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proceedings. Biological Sciences*, 256(1346), 119–124. <https://doi.org/10.1098/rspb.1994.0058>
- Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution: Novel Sex-Determining Genes in Fish. *Developmental Dynamics*, 242(4), 339–353. <https://doi.org/10.1002/dvdy.23927>
- Kimura, M., & Ohta, T. (1974). On some principles governing molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 71(7), 2848–2852.
- Kondo, M., Hornung, U., Nanda, I., Imai, S., Sasaki, T., Shimizu, A., ... Scharl, M. (2006). Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Research*, 16(7), 815–826. <https://doi.org/10.1101/gr.5016106>
- Kondo, M., Nanda, I., Schmid, M., & Scharl, M. (2009). Sex Determination and Sex Chromosome Evolution: Insights from Medaka. *Sexual Development*, 3(2–3), 88–98. <https://doi.org/10.1159/000223074>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., ... Wang, D. (2015). 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(11), e1005678. <https://doi.org/10.1371/journal.pgen.1005678>
- Lubieniecki, K. P., Lin, S., Cabana, E. I., Li, J., Lai, Y. Y. Y., & Davidson, W. S. (2015). Genomic Instability of the Sex-Determining Locus in Atlantic Salmon (*Salmo salar*). *G3: Genes, Genomes, Genetics*, 5(11), 2513–2522. <https://doi.org/10.1534/g3.115.020115>
- Mank, J. E., & Avise, J. C. (2009). Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, 3(2–3), 60–67. <https://doi.org/10.1159/000223071>
- Mank, J. E., & Ellegren, H. (2007). Parallel divergence and degradation of the avian W sex chromosome. *Trends in Ecology & Evolution*, 22(8), 389–391. <https://doi.org/10.1016/j.tree.2007.05.003>
- Marshall, O. J. (2004). PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics (Oxford, England)*, 20(15), 2471–2472. <https://doi.org/10.1093/bioinformatics/bth254>
- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., ... Sakaizumi, M. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417(6888), 559–563. <https://doi.org/10.1038/nature751>
- Meisel, R. P., Davey, T., Son, J. H., Gerry, A. C., Shono, T., & Scott, J. G. (2016). Is Multifactorial Sex Determination in the House Fly, *Musca domestica* (L.), Stable Over Time? *The Journal of Heredity*, 107(7), 615–625. <https://doi.org/10.1093/jhered/esw051>
- Moore, E. C., & Roberts, R. B. (2013). Polygenic sex determination. *Current Biology*, 23(12), R510–R512. <https://doi.org/10.1016/j.cub.2013.04.004>

- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., ... Sakaizumi, M. (2012). Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, *Oryzias latipes*. *Genetics*, 191(1), 163–170. <https://doi.org/10.1534/genetics.111.137497>
- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., ... Scharl, M. (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proceedings of the National Academy of Sciences*, 99(18), 11778–11783. <https://doi.org/10.1073/pnas.182314699>
- Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-86659-3>
- Pan, Q., Guiguen, Y., & Herpin, A. (2017). Evolution of sex determining genes in Fish. In *Encyclopedia of Reproduction, Second Edition*.
- Peichel, C. L., Ross, J. A., Matson, C. K., Dickson, M., Grimwood, J., Schmutz, J., ... Kingsley, D. M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Current Biology: CB*, 14(16), 1416–1424. <https://doi.org/10.1016/j.cub.2004.08.030>
- Potrzebowski, L., Vinckenbosch, N., Marques, A. C., Chalmel, F., Jégou, B., & Kaessmann, H. (2008). Chromosomal Gene Movements Reflect the Recent Origin and Biology of Therian Sex Chromosomes. *PLOS Biology*, 6(4), e80. <https://doi.org/10.1371/journal.pbio.0060080>
- Rentería, M. E., Cortes, A., & Medland, S. E. (2013). Using PLINK for Genome-Wide Association Studies (GWAS) and data analysis. *Methods in Molecular Biology (Clifton, N.J.)*, 1019, 193–213. https://doi.org/10.1007/978-1-62703-447-0_8
- Roberts, N. B., Juntti, S. A., Coyle, K. P., Dumont, B. L., Stanley, M. K., Ryan, A. Q., ... Roberts, R. B. (2016). Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics*, 17(1). <https://doi.org/10.1186/s12864-016-3177-1>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., & Peichel, C. L. (2009). Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). *PLOS Genetics*, 5(2), e1000391. <https://doi.org/10.1371/journal.pgen.1000391>
- Schultheis, C., Böhne, A., Scharl, M., Volff, J. N., & Galiana-Arnoux, D. (2009). Sex Determination Diversity and Sex Chromosome Evolution in Poeciliid Fish. *Sexual Development*, 3(2–3), 68–77. <https://doi.org/10.1159/000223072>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539. <https://doi.org/10.1038/msb.2011.75>
- Skog, A., Vøllestad, L. A., Stenseth, N. C., Kasumyan, A., & Jakobsen, K. S. (2014). Circumpolar phylogeography of the northern pike (*Esox lucius*) and its relationship to the Amur pike (*E. reichertii*). *Frontiers in Zoology*, 11(1), 67.
- Stöck, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., ... Perrin, N. (2011). Ever-Young Sex Chromosomes in European Tree Frogs. *PLOS Biology*, 9(5), e1001062. <https://doi.org/10.1371/journal.pbio.1001062>
- Takehana, Y., Matsuda, M., Myosho, T., Suster, M. L., Kawakami, K., Shin-I, T., ... Naruse, K. (2014). Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nature Communications*, 5. <https://doi.org/10.1038/ncomms5157>
- Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S., & Sakaizumi, M. (2007). Evidence for Different Origins of Sex Chromosomes in Closely Related *Oryzias* Fishes: Substitution of the Master Sex-Determining Gene. *Genetics*, 177(4), 2075–2081. <https://doi.org/10.1534/genetics.107.075598>

- Weisenfeld, N. I., Yin, S., Sharpe, T., Lau, B., Hegarty, R., Holmes, L., ... Jaffe, D. B. (2014). Comprehensive variation discovery in single human genomes. *Nature Genetics*, 46(12), 1350–1355. <https://doi.org/10.1038/ng.3121>
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., ... Guiguen, Y. (2012). An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Current Biology*, 22(15), 1423–1428. <https://doi.org/10.1016/j.cub.2012.05.045>
- Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., & Guiguen, Y. (2013). The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications*, 6(3), 486–496. <https://doi.org/10.1111/eva.12032>

Supplementary: Rise and fall of an ancient master sex determining gene in the *Esociformes*

Additional analysis on *Novumbra hubbsi*

Material and Methods

RAD-seq was performed on 26 males and 24 females of *N. hubbsi* samples following the same protocol for library preparation and sequence as for the other *Esociformes*. The RAD-seq reads were processed by the RADSex pipeline with the same parameters as for other *Esociformes*, following the protocol by Feron et al (*in prep*).

Furthermore, a PCA was performed on the RAD-seq data of *N. hubbsi*. The analysis was performed using stacks version 1.34. Raw RADSeq reads were processed with the *process_radtags.pl* wrapper script and analyzed using the *denovo_map.pl* wrapper script with the following parameters: m=10, M=2, N=3, --gapped not set and H not set. Results of *denovo_map.pl* were run through the *populations* pipeline with minimum percentage of individuals in a population set to 0.5 (r=0.5), minimum stacks depth at a locus set to 10 (m=10), and minimum populations at a locus set to 2 (p=2), to generate PLINK input files (.map and .ped). PLINK input files were manually completed to add sex information as phenotype. PCA was performed in R (version 3.4.2) using the package SNPRelate (version 1.11.2). Reads were first converted to *gds* format with *snpgdsBED2GDS*, and the resulting *gds* file was analyzed using *snpgdsPCA* with the following settings: bayesian normalization and monomorphic SNPs removed. Samples were plotted along the first and second eigenvectors with the package ggplot2 (version 2.2.1) and samples were colored according to population.

Results

RADSeq data was also generated for 26 males and 24 females *N. hubbsi*. The resulting heatmap revealed 2 markers present in 19 males (73.1%) and 2 females (8.3%). However, their association with sex was not significant ($p=0.158$, chi-squared test with Bonferroni correction). In contrast, no marker was predominantly present in females supporting an XX/XY system as already suggested by *amhby* sex linkage in that species (**Figure SS1A**). This analysis also revealed multiple groups of markers that were shared by different subsets of individuals (**Figure SS1B**), suggesting an underlying strong population structure. In order to characterize this potential population structure, this RAD-Seq dataset

was re-analyzed to obtain polymorphic markers, which were used in a principal component analysis (PCA). When plotting the first two eigenvectors from this PCA (**Figure SS2**), individuals from the same location clustered together, confirming that a strong population structure was present in this dataset. Furthermore, results from the PCA showed that individuals from Stern Wetland and South Hanaford Creek formed a single group that was distant from the other populations. Genetic relatedness of *N. hubbsi* samples collected from several locations was also assessed with F_{st} from pairwise comparisons among populations (**Table SS1**). Samples collected from South Hanaford Creek and Stearns Wetland had the lowest divergence with a in between-population F_{st} of 0.0031, while showing the highest divergence from all other populations, with the exception of Powell Creek. Coincidentally, the three markers found predominantly in males were absent in these two populations. Consequently, the RAD-Seq data was re-analyzed excluding individuals from these two populations, and two markers were significantly associated with sex ($p=0.021$, chi-squared test with Bonferroni correction) (**Figure SS1B**).

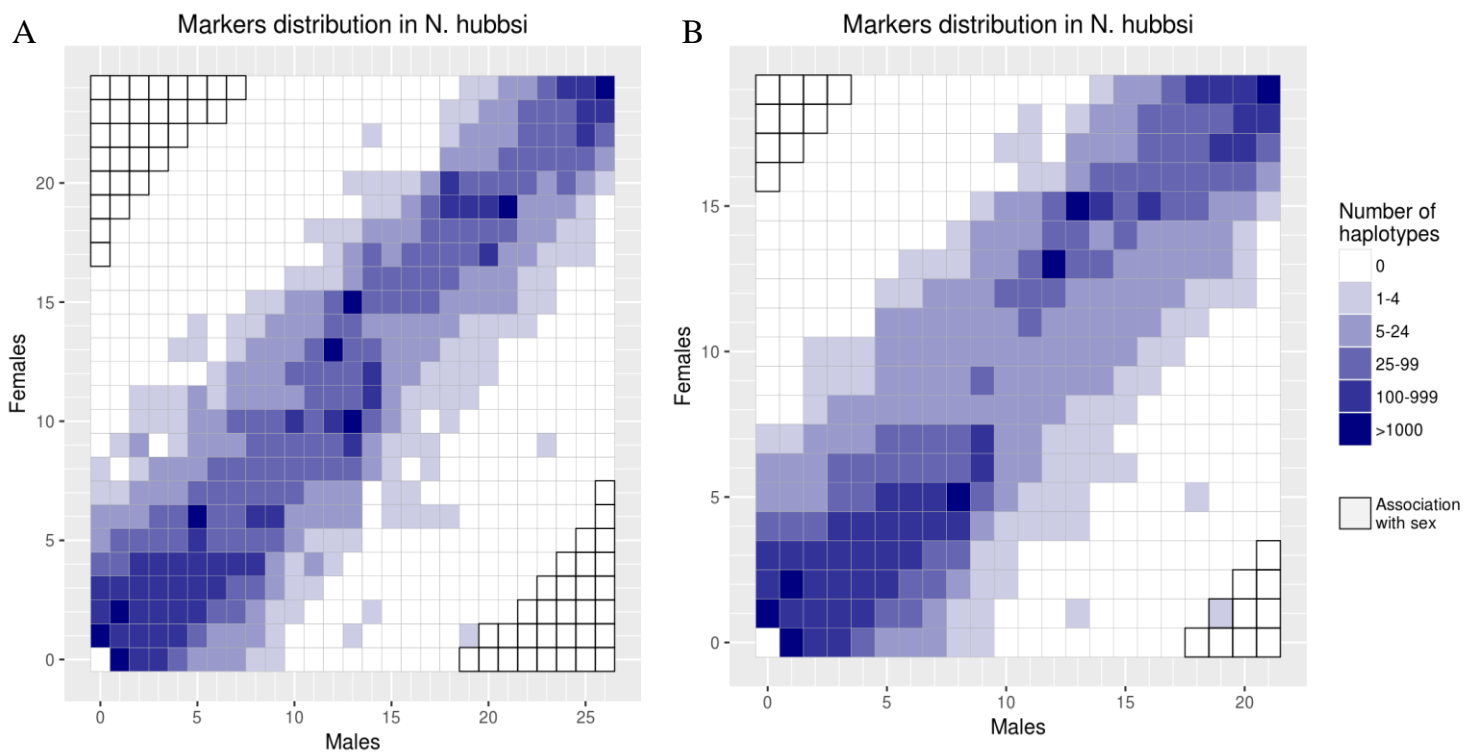


Figure SS1: Heat map of markers shared by males and females of *N. hubbsi* from the RADSex pipeline. Positions that show significant association with sex are highlighted with the black outline on the lower right corner and upper left corner for male and female respectively. The intensity of the color corresponds to the number of markers. **A:** all population included; **B:** excluding samples from South Hanaford Creek and Stearns Wetland.

	James Pond	Powell Creek	Stearns Wetland	S. Hanaford Creek	Conner Creek	Steamboat Bog
James Pond		0.246	0.491	0.424	0.301	0.169
Powell			0.243	0.205	0.075	0.248
Stearns Wetland				0.033	0.250	0.567
S. Hanaford Creek					0.213	0.475
Conner Creek						0.266

Table SS1: Pairwise comparison of F_{st} values between different populations of *Novumbra hubbsi* used in the RADSex analysis.

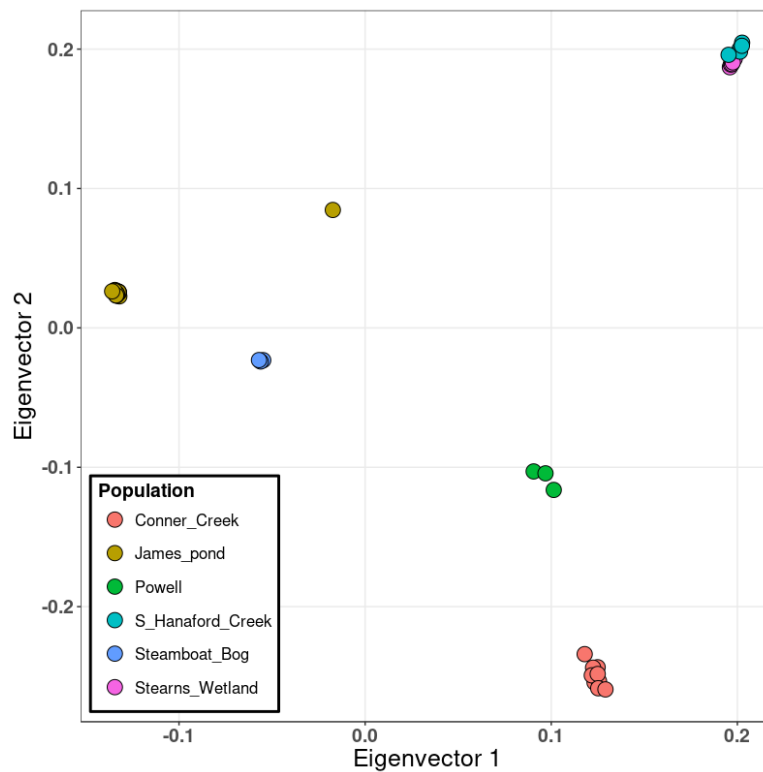


Figure SS2: PCA on RAD-Seq SNP data from *Novumbra hubbsi*. Samples were collected from six populations, and samples origins are indicated by color. Each point corresponds to one individual in the analysis.

Additional figures and tables

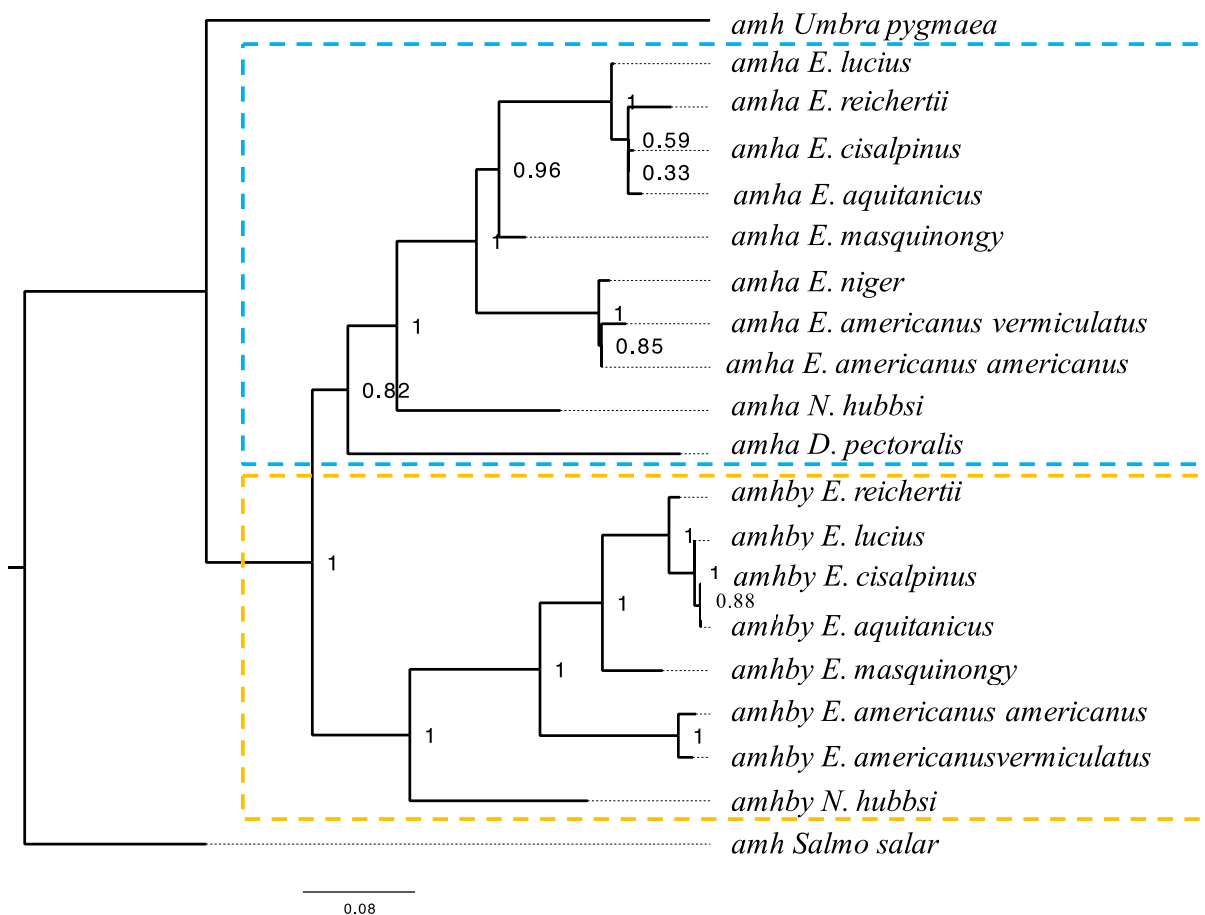


Figure S1: Phylogenetic reconstruction of Amha and Amhby protein orthologs from the *Esociformes* with Amh sequence from *Salmo salar* as an outgroup. This Phylogeny tree was constructed with the Maximum likelihood method. Bootstrap values are given on each nod of the tree. Amha ortholog cluster is highlighted with yellow dotted lines and Amhby ortholog cluster is highlighted with blue dotted

Primer name	Target gene	Sequence	Sequence used for primer design
SeqAMH1Fw1	<i>amhby</i>	GAAAGACACTGGCTCACAG	<i>E. lucius amhby</i>
SeqAMH1Fw2	<i>amhby</i>	TGGCACCATCTCTTTAC	<i>E. lucius amhby</i>
SeqAMH1Fw3	<i>amhby</i>	TCCAGTCCCGTTGTTCTC	<i>E. lucius amhby</i>
SeqAMH1Fw4	<i>amhby</i>	CAACATGGTGGAACCTAAGTG	<i>E. lucius amhby</i>
SeqAMH1Rev1	<i>amhby</i>	TTTCCCTCTGATGGCATTCC	<i>E. lucius amhby</i>
SeqAMH1Rev2	<i>amhby</i>	CTGTTGGTGGTCTTTGC	<i>E. lucius amhby</i>
SeqAMH1Rev3	<i>amhby</i>	TGGCAGTGAAGATATTAGC	<i>E. lucius amhby</i>
SeqAMH1Rev4	<i>amhby</i>	GGTAATATTTGTGCCCTGTG	<i>E. lucius amhby</i>
seqAMH2_Forward1	<i>amha</i>	TGTTAGATGCAGACGTGAG	<i>E. lucius amhby</i>
seqAMH2_Reverse1	<i>amha</i>	CACTTTCTTGTTCCCAACC	<i>E. lucius amhby</i>
seqAMH2_Forward2	<i>amha</i>	CCATCGTTCCCAACACTTCTC	<i>E. lucius amhby</i>
seqAMH2_Reverse2	<i>amha</i>	AGGACATCACTGAGGAACC	<i>E. lucius amhby</i>
seqAMH2_Forward3	<i>amha</i>	GTAATTGTCTCCATCAGAGCGT	<i>E. lucius amhby</i>
seqAMH2_Reverse3	<i>amha</i>	CACTCACACTTCGTTGCCAC	<i>E. lucius amhby</i>
ConserveAMH1_F1	<i>amhby</i>	GTTACTTTTCTCGCTAGCGTGA	Conserved region on <i>amhby</i> orthologs
ConserveAMH1_F2	<i>amhby</i>	GTGATAGGCACTGTAAACACCA	Conserved region on <i>amhby</i> orthologs
ConserveAMH1_R1	<i>amhby</i>	CTATTACTAGTGTGGATAAGGCCG	Conserved region on <i>amhby</i> orthologs
ConserveAMH1_R2	<i>amhby</i>	CCTCTTCACCCTCAGTGAGC	Conserved region on <i>amhby</i> orthologs
ConserveAMH1_F3	<i>amhby</i>	TATATCCTGCATGTTACTTTTTCTG	Conserved region on <i>amhby</i> orthologs
ConserveAMH1_F4	<i>amhby</i>	ACCAGACCACCAAGGAACAG	Conserved region on <i>amhby</i> orthologs
ConserveAMH1_R3	<i>amhby</i>	CAACCTCCCATGAACCTT	Conserved region on <i>amhby</i> orthologs
ConserveAMH1_R4	<i>amhby</i>	GGTGGATGGTGATTTTCTTCTC	Conserved region on <i>amhby</i> orthologs
AMH1ConENDRev1	<i>amhby</i>	ATTCCTGCACTACCAACCTCC	Conserved region on <i>amhby</i> orthologs
AMH1ConENDRev2	<i>amhby</i>	CAGTCTGCAGAGCCTTCAGTAG	Conserved region on <i>amhby</i> orthologs
MuskyAMH1_Gap_F1	<i>amhby</i>	GTAATAGAAGCCAACAGACAGAG	<i>E. masquinongy amhby</i>
MuskyAMH1_Gap_R1	<i>amhby</i>	CTTTCATCCATTCTCGATCCA	<i>E. masquinongy amhby</i>
Con_Amha_F1	<i>amha</i>	GTCACTATAGGTTGGTTCATC	Conserved region on <i>amha</i> orthologs
Con_Amha_R1	<i>amha</i>	CGATTTGCCATTTTCAGGTG	Conserved region on <i>amha</i> orthologs
Con_Amha_F2	<i>amha</i>	CACGATCTGCAGCCTTACAA	Conserved region on <i>amha</i> orthologs
Con_Amha_R2	<i>amha</i>	CTTCAGTAGTAGCAGGGCACG	Conserved region on <i>amha</i> orthologs
ESNI_Amhby_gapF1	<i>amhby</i>	GGATATGCTTTCAGCACTTCG	<i>E. niger amhby</i>
ESNI_Amhby_gapR1	<i>amhby</i>	GATCTGCCATTCCAGGTGAC	<i>E. niger amhby</i>
ESNI_Amhby_spcR1	<i>amhby</i>	GAATAACATCACTGACCAGCC	<i>E. niger amhby</i>
ESAA_Amhby_gapF1	<i>amhby</i>	TAGTAACGGCCTTATCCACAC	<i>E. niger amhby</i>
ESAA_Amhby_gapR1	<i>amhby</i>	GGCAGGAGTCATGTATCAACAG	<i>E. niger amhby</i>
ESAA_Amhby_spcR1	<i>amhby</i>	GTCACAGCTAAATCAGGTCAC	<i>E. niger amhby</i>
Novum_Amha_F1	<i>amha</i>	GTTTATTCACTGATGTGGGAC	<i>N. hubbsi amha</i>
Novum_Amha_R1	<i>amha</i>	GAAAGCATATCGTCAACAACC	<i>N. hubbsi amha</i>
Novum_Amhby_F1	<i>amha</i>	CAAGATGCCAACAAACAGAG	<i>N. hubbsi amhby</i>
Novum_Amhby_R1	<i>amha</i>	TATGGATAATAGCCTGAACCTG	<i>N. hubbsi amhby</i>
NigerBY_end_F1	<i>amhby</i>	GCACAGCTTAACCGTTTCCC	<i>E. niger amhby</i>
NigerBY_end_R1	<i>amhby</i>	CCTTGACCATAAAGCATTTCAGAG	<i>E. niger amhby</i>
NigerBY_front_F1	<i>amhby</i>	TGTCCATACCATTACAGATCC	<i>E. niger amhby</i>
NigerBY_front_R1	<i>amhby</i>	TCAGAGAACCATCTAAACGG	<i>E. niger amhby</i>
Con_Amhby_end_F2	<i>amhby</i>	GTAGTGCAGGAATGCACAGGC	Conserved region on <i>amhby</i> orthologs
Con_Amhby_end_R2	<i>amhby</i>	CAGCGGCATCCACACTTAGTTGC	Conserved region on <i>amhby</i> orthologs
Niger_Amha_F	<i>amha</i>	CTACTACTGAAGGCTCTACAGAC	<i>E. niger amha</i>
Niger_Amha_R	<i>amha</i>	ATCATAGTCCACAGGTACACAG	<i>E. niger amha</i>
Niger_Amhby_F	<i>amhby</i>	AAGCCGTTTAGATGGTTCTCTG	<i>E. niger amha</i>
Niger_Amhby_R	<i>amhby</i>	CAATGGTATCAACAATGCTCAGG	<i>E. niger amha</i>

Table S1: primers used to obtain *amhby* and *amha* ortholog sequences from the sister species of *E. lucius* in the *Esociformes*.

Species	Sex/genotype	Total contig length	N50 contig length	L50 contig count
<i>Esox lucius</i> (European population)	male	1.01G	17641	15128
<i>Esox masquinongy</i>	male	1.14G	19504	15247
<i>Esox masquinongy</i>	female	1.10G	22589	12838
<i>Esox niger</i>	amhby +	0.86G	30384	7354
<i>Novumbra hubbsi</i>	male	0.76G	17942	10955
<i>Dallia pectoralis</i>	male	0.86G	5062	35093

Table S2: Assemblathon's metrics of the genome assemblies obtained in six *Esociformes*.

Model name in PALM	Log-likelihood of the model
M0	-5121.89
M1a	-5123.61
M2	-5123.61
M7	-5123.41
M8	-5123.28
Free-ratio model	-5139.74
Branch-site model	-5122.80

Table S3: Log-likelihood of different selection models tested on *amha* and *amhby* orthologs of the *Esociformes*.

Sex determination and karyotype evolution in the Eastern mudminnow, *Umbra pygmaea*

Abstract

To complete our survey of sex determination (SD) systems in the *Esociformes*, we investigated the SD system and karyotype evolution in the Eastern mudminnow, *Umbra pygmaea*. This species belongs to the *Umbridae* family that diverged from the *Esocidae* prior to the birth of *amhby*, a strong candidate for the master sex determining (MSD) gene in many species of the *Esocidae*. We first identified and confirmed an XX/XY SD system in *U. pygmaea*, consistent with most SD systems found in the sister clades *Esocidae* and *Salmoniformes*. We also constructed a genetic linkage map of *U. pygmaea* and used it to identify a well differentiated sex locus in the sex chromosome, LG1. Using data from this genetic map, we also compared the chromosomes of *U. pygmaea* with those of *E. lucius*, and showed that the drastic reduction of karyotype number in *U. pygmaea* is a result of one-to-one fusion of ancestral karyotypes. In addition, a draft genome of *U. pygmaea* was sequenced and assembled, but the large size of this genome and its lack of continuity prevented the identification of a candidate MSD gene. Finally, we characterized the early molecular and morphological differentiation of the gonads in *U. pygmaea*, paving the way for further studies of the sex determination in this non-model teleost fish.

Introduction

In the salmonids, an unusual master determining (MSD) gene named *sdY* was characterized (Yano *et al.*, 2013), and it was hypothesized that this MSD gene was a salmonid innovation. The sister group of salmonids, the *Esociformes*, contains two families, the *Umbridae* and the *Esocidae*, which diverged around 90 million years ago (Campbell *et al.*, 2013). In the *Esocidae*, a male specific duplicate of *amh* was identified as the MSD in most species (Pan *et al.*, *in prep*). However, a phylogenetic analysis of *amh* orthologs in *Esociformes* indicated that the *Umbra* genus diverged before the duplication of *amh*, and thus, the *Umbridae* is likely to have a different MSD gene. Yet, to date, little is known about sex determination and reproduction in this clade. Characterizing the sex-determination mechanism of the *Umbridae* would improve our knowledge on the evolution of sex determination systems in the *Esociformes*, and represents a crucial step in the determining the potential ancestral MSD of this clade. In addition, while the European mudminnow (*Umbra krameri*) have 22 pairs of chromosomes (2N= 44) (Rab and Mayr, 1987) similar to that of the *Esox* (Beamish *et al.*, 1971), the other two American mudminnow species, *Umbra pygmaea* and *Umbra limi*, have only 11 pairs of chromosomes (2N= 22) (Beamish *et al.*, 1971). This substantial karyotype evolution within *Umbridae* further increase the interest in studying their genome and sex determination evolution.

To start our investigations on sex determination in *Umbridae*, we focused mainly on the Eastern mudminnow *Umbra pygmaea*, a small bodied freshwater species, which is native to the East coast of the United states but which has been introduced in Europe in the past century (Kuehne and Olden, 2014). We used RAD-Sequencing (Amores *et al.*, 2011) data to find markers associated with sex in a wild-caught population and we also built a genetic linkage map from a captive breeding (Pan *et al.*, *in revision*) family panel. These analyses revealed an XX/XY sex determination system with a large, well-differentiated sex locus located on the first Linkage Group (LG1). Results from comparative genomics approaches suggest that this linkage group was formed by a chromosomal fusion of the ancestral karyotypes of LG8, LG12 and LG22 in *E. lucius*. To complement these genomic analysis, we performed molecular analyses to delineate the period of gonadal sex differentiation and we provide guidelines for future functional studies of the MSD in this species.

No candidate MSD gene has been identified yet, but a high quality sequencing of *U. pygmaea* genome is under way, which will improve the assembly of the sex locus and could lead to the identification of the MSD gene. In addition, our results provided many sex markers which can be used to sex *Umbra pygmaea* samples with high confidence.

Material and Methods

Fish collection

The population of eastern mudminnows (*U. pygmaea*) used in this study was sampled in Flanders, Belgium. Eastern mudminnows were captured using electric fishing gear in a pond and connecting ditches in a nature reserve in the vicinity of Genk (Belgium). Exact coordinates (latitude and longitude) of the pond are 50°57'12.84"N and 5°27'16.54"E. *U. pygmaea* is a non-native species in Belgium and is not protected by any legislation there. These *U. pygmaea* were transported alive to an indoor experimental facility located in Rennes (France).

Laboratory rearing conditions

Mudminnows were maintained indoor in 1 m² tanks filled with 200 liters of dechlorinated local water (pH=8) filtered with a recirculating system. Tanks were enriched with artificial aquarium plants and polyvinyl chloride (PVC) plastic tubes. Rearing temperature followed the trend of external natural temperature. Controlled breeding were performed by hormonal induction followed by *in vitro* fertilization (Pan *et al.*, in revision).

Rad library preparation and sequence

For both the wild population of *U. pygmaea* and the family panel including two parents and 92 of their offspring, the Restriction Associated DNA sequencing (RADSeq) libraries were constructed from genomic DNA extracted from fin clips, digested with *SbfI* following standard protocols (Amores *et al.*, 2011), and sequenced in one lane of Illumina HiSeq 2500.

Genome sequencing and assembly

A first genome of *U. pygmaea* was sequenced using a whole genome shotgun strategy with small Illumina reads. Libraries were built using the Truseq nano kit (Illumina, ref. FC-121-4001) following the manufacturer's instruction. 200 ng of gDNA was briefly sonicated using a Bioruptor (Diagenode). The DNA 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 DNA 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. The libraries were sequenced on a HiSeq2500 using a paired end 2x250 bp v2 rapid mode according to the manufacturer's instruction. Image analysis was performed by the HiSeq Control Software and base calling by the RTA software provided by Illumina. Raw sequencing reads were assembled using the DISCOVAR *de novo* software (Weisenfeld *et al.*, 2014) with the following assembly parameters: MAX_MEM_GB=256, MEMORY_CHECK=False, and NUM_THREADS=16. The resulting assembly metrics were calculated with the assemblathon_stats.pl script (Earl *et al.*, 2011) and are given in **Table S4**.

A second genome was sequenced with the 10X Genomics Chromium technology. Libraries were prepared by the MGX platform (Montpellier, France) following the manufacturer's instructions. Sequencing was done on one lane of Illumina HiSeq 3000 generating 2x150 bp reads. These reads were assembled with the supernova assembler with default settings (10X Genomics).

RADSex pipeline and allele depth analysis

Raw RAD-Seq reads were analyzed with *stacks* version 1.44 (Catchen *et al.*, 2011; Catchen *et al.*, 2013). Quality control and demultiplexing of the raw reads were performed with the *process_radtags.pl* wrapper script with default settings. This step was done in a consistent way for all species to associate reads with the individuals by their barcodes, creating one read file per individual. For all species, RAD-Seq reads were analyzed with a custom RADSex analysis pipeline (Feron *et al.*, *in prep*). This analysis was performed in order to analyze both polymorphic and non-polymorphic markers, including potentially hemizygous markers. In this analysis, demultiplexed reads were grouped into stacks using the *denovo_map.pl* wrapper script with minimum number of reads to form a stack set to 10 (m=10), maximum number of mismatches allowed when forming stacks in one individual set to 0 (M=0), maximum number of mismatches allowed when merging stacks in the catalog set to 0 (N=0), gapped assembly deactivated (--gapped not set), calling haplotypes from secondary reads disabled (H not set), *ustacks* max locus per stack set to 1 (--max_locus_stacks=1), and *ustacks* lumberjack step deactivated (--keep_high_cov set).

These parameter values ensured that each stack represented a single allele (or haplotype) from a single locus. The distribution of haplotypes between males and females was calculated using a custom pipeline, and results were presented in a heatmap with number of males in which a haplotype is present as abscissa (x-axis), number of females in which a haplotype is present as ordinate (y-axis), and number of haplotypes present in x males and y females as color of the tile at coordinates (x , y). For each combination of number of males and number of females in which a haplotype is present, probability of association with sex was estimated with a chi-square test, applying Bonferroni correction.

Remapping of *U. pygmaea* markers to the genome of *E. lucius*

All sequences from the *stacks* catalog obtained with the RADSex pipeline were mapped to the genome assembly of *E. lucius* (Elu_V3, GenBank assembly accession: GCA_000721915.3) using BWA (version 0.7.15-r1140, Li and Durbin, 2009) with all settings set to default. Sequences with mapping quality > 15 and a unique mapping position were retained for this analysis.

DNA extraction and genotyping with sex markers from RADSex analysis

All fin clips were collected and stored at 4°C in 75-100% ethanol until DNA extraction. Two DNA extraction protocols were used depending on the subsequent procedures. 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. After brief centrifugation, the supernatant containing genomic DNA was transferred to clean tubes without the Chelex beads (Gharbi *et al.*, 2006). For sequencing preparation, DNA extraction was performed using NucleoSpin Kits for Tissue (Macherey-Nagel, Duren, Germany) following the producer's protocol. When using this protocol, DNA concentration was quantified with NanoDrop ND2000 spectrophotometer (Thermo scientific, Wilmington, Delaware).

Primers for genotyping were designed using Perl-Primer (version 1.1.21, Marshall, 2004) based on the surrounding genomic sequences of four sex markers identified by the RADSex analysis. Primer sequences and corresponding experiments can be found in **Table S2**.

RNA extraction, cDNA synthesis and qPCR

Six males and six females were sampled at 3, 6, 10, 14, and 20 weeks post fertilization (wpf). The whole trunk was collected for RNA extraction as it was not feasible to isolate the gonads at these stages. The genotypic sex of each animal was determined based on the confirmed sex markers from the RADSex analysis. Samples 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, Delaware). Reverse transcription (RT) was performed by denaturing a mix of 1µg of RNA and 5 µL of 10mM 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, 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.

qPCR primers were designed using Perl-Primer (version 1.1.21, Marshall, 2004) on intron-exon junctions to avoid genomic DNA amplification (**Table S3**). Primer pairs were diluted to 6 µg/µL for qPCR. qPCR 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 1µL 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. Samples 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 (Vandesompele *et al.*, 2002). The differentiation of expression between males and females was tested with a Wilcoxon signed rank test implemented in R (version 3.2.2).

Visualizing primordial germ cells during embryogenesis

GFP-nanos-1 including the mmGFP5 open reading frame (ORF) (Siemering *et al.*, 1996) fused to the 3'-UTR of the nanos-1 gene were constructed as described by Köprunner *et al.*, (2001) and Herpin *et al.*, (2007). Injection of RNA generated from this construct can result in specific GFP expression in the PGCs in teleost species (Saito *et al.*, 2006). The RNA was administrated to *U. pygmaea* embryos at the 1-cell stage, within 2 hours post fertilization. Confocal images were acquired on each day post fertilization until hatching at 9 days post-fertilization (dpf). Imaging of the embryos was performed using a laser microscope (Eclipse C1 laser-scanning, Nikon, Tokyo, Japan) with a 60x Nikon objective (PL APO, 1.4 NA) and the Nikon image software.

Histology

Whole trunks of *U. pygmaea* 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). Cross sections of each embedded sample were cut serially into slices of 7 μ m 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).

RADSeq based genetic map

RADSeq reads were generated from a full-sib family of x males and y females to create a linkage map for *U. pygmaea* including 56 female offspring, 35 male offspring and both parents. Raw reads were analyzed with the Stacks program version 1.34 (Catchen *et al.*, 2011; Catchen *et al.*, 2013). Quality control and demultiplexing of the 2,198,433,345 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, 1,501,898,810 reads were retained after this filtering step (68%), including 3,307,826 sequences from the mother, 2,978,238 sequences from the father, and between 440,000 and 2,000,000 sequences from each offspring.

Stacks of reads were built using the *denovo_map.pl* wrapper script with the minimal number of reads to form a stack set to 5 ($m=5$), high bound set to 0.05 (`--bound_high 0.05`) and all other settings set to default. In total, 119,412 loci were obtained, among which 9288 markers were polymorphic. Afterwards, the *genotypes* program was used to export loci found in at least 80 of the 92 offspring (87%), and 1713 loci were used to build the linkage map using Lep-Map 2 (Rastas *et al.*, 2015). First, markers showing significant segregation distortion ($P < 0.01$) were excluded using the *Filtering* module with the option “dataTolerance=0.01”. As previous cytological study had determined the chromosome number to be 22 ($2n$), markers were separated into 11 linkage groups using the modules *SeparateChromosomes* and *JoinSingles*, with the minimum logarithm of odds (LOD) value set to 5 (“lodLimit=5”) and the minimal number of markers in each linkage group set to 20 (“sizeLimit=20”). Then, the female informative markers were masked with the setting value “informativeMask=13” in order to generate a male map, and the male informative markers were masked with the setting value “informativeMask=23” in order to generate a female map. Finally, a consensus map was generated with the setting value “informativeMask=123”.

To determine the order and distance between markers in each linkage group, the Kosambi mapping function from the *OrderMarkers* module (“useKosambi=1”) was ran 10 times, and the run with the highest likelihood score was retained for each linkage group. Finally, the genetic maps were drawn and aligned using Genetic-Mapper v0.6 (Bekaert, 2016). Markers used in the linkage map were analyzed with the *populations* program of Stacks (version 1.34) using the `--fstats` setting to obtain population genetic statistics between males and females.

Remapping of the markers placed on the genetic map of *U. pygmaea* on the genome of *E. lucius*

In this analysis, mapping was done using BWA (version 0.7.15-r1140, Li and Durbin, 2009) with all settings set to default, and markers were considered uniquely mapped when they had a mapping quality > 15 . As the direct remapping of RAD-Seq tags did not provide a significant anchorage on the *E. lucius* genome we use the draft genomes of *U. pygmaea* as a stepping-stone to recover specific genomic sequences around the marker sequences and remapping these *U. pygmaea* genomic sequences on the genome of *E. lucius*. All markers from the genetic map of *U. pygmaea* were mapped to the Discovar *de novo* genome assembly and to the 10x genome assembly. Contigs from these assemblies that were hit by uniquely mapped markers were extracted.

These contig sequences were then used as proxy references to remap all markers from the *stacks* catalogs obtained in the RADSex analysis of the wild population and the RADSex analysis of the family panel. All markers uniquely mapped on these contigs were then mapped to the genome of *E. lucius*, and only markers uniquely mapped to the genome were retained. A correspondence table was generated for all resulting markers storing their position in the genetic map of *U. pygmaea* as well as their position on the genome of *E. lucius*. For markers recovered from the RADSex data, the position in the genetic map was determined as the position of the marker from the genetic map which mapped to the same contig. If more than one marker from the genetic map was mapped to the same contig, the one giving the position was selected randomly.

A circos plot showing all LGs from the genetic map of *U. pygmaea* and all chromosomes from the genome of *E. lucius* was generated using the *circlize* package (version 0.4.2, Gu *et al.*, 2014) in R (version 3.4.2), with links showing the correspondence between position on the genetic map of *U. pygmaea* and position on the genome of *E. lucius* for each marker.

Results

Genome sequencing and assembly of *U. pygmaea* confirms the absence of MSD from sister clade

Two draft genomes were sequenced and assembled for *U. pygmaea* with either a whole genome shotgun strategy with Illumina reads, which is referred to in the later text as the ‘discover genome’, or a 10X genome sequencing strategy, which is referred to in the later text as the ‘10X genome’. The size of resulted assembly of the discover genome was around 1.9 Gb (**Table S4**), about 80% of the estimated 2.4 Gb genome size of *U. pygmaea* (Beamish *et al.*, 1971). The size of the resulted assembly of 10X genome was around 1.5 Gb, about 63% of the estimated genome size. However, the continuity of the assembly was better in the 10X genome than the discover genome (**Table S4**).

Homologous sequences to the salmonid MSD gene *sdY* was not identified in the draft genomes of *U. pygmaea*. Furthermore, only one homozygous *amh* sequence was identified in *U. pygmaea* genome, confirming previous findings that the *amh* duplication, which originated the *Esocidae* MSD gene *amhby*, took place after the divergence of the *Umbridae* (Pan *et al.*, *in prep*).

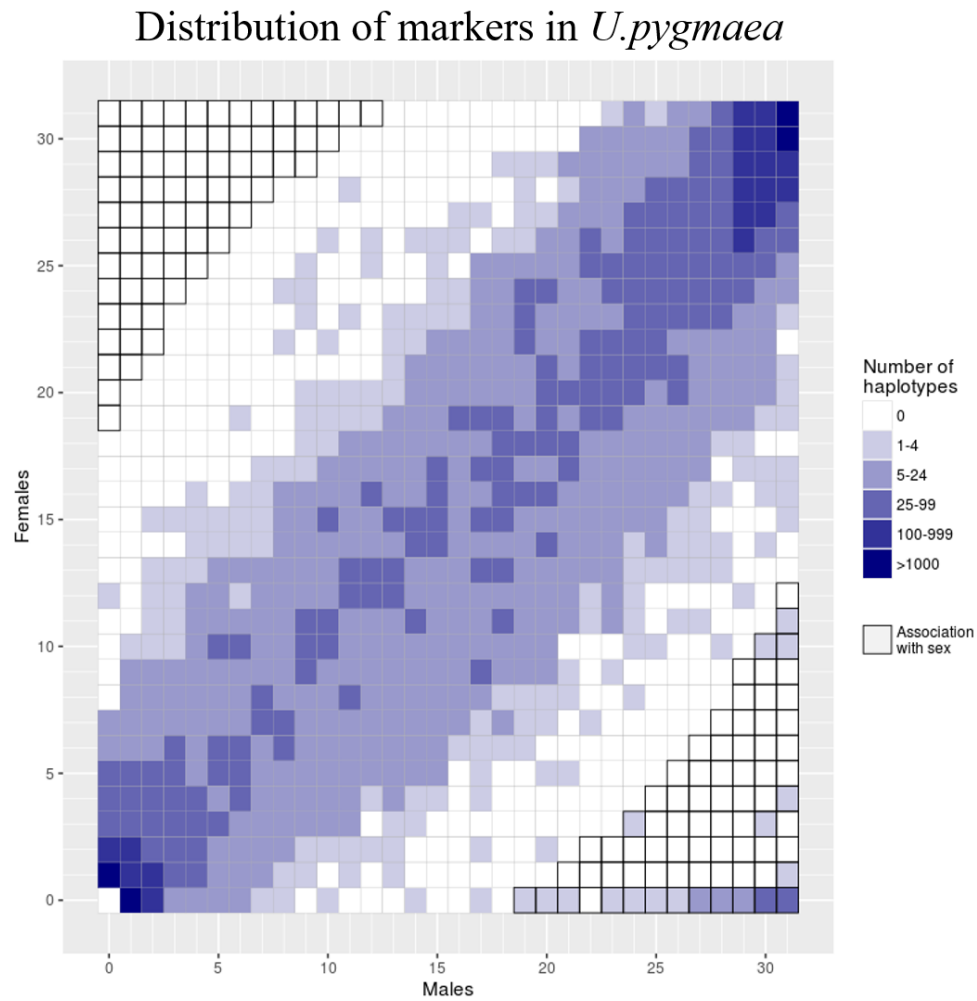


Figure 1: Heat map of RADSex markers in 31 males and 31 females of *U. pygmaea*. Positions with significant association with sex are highlighted in black on the lower right corner and upper left corner for male and female respectively. The intensity of the blue color corresponds to the number of markers.

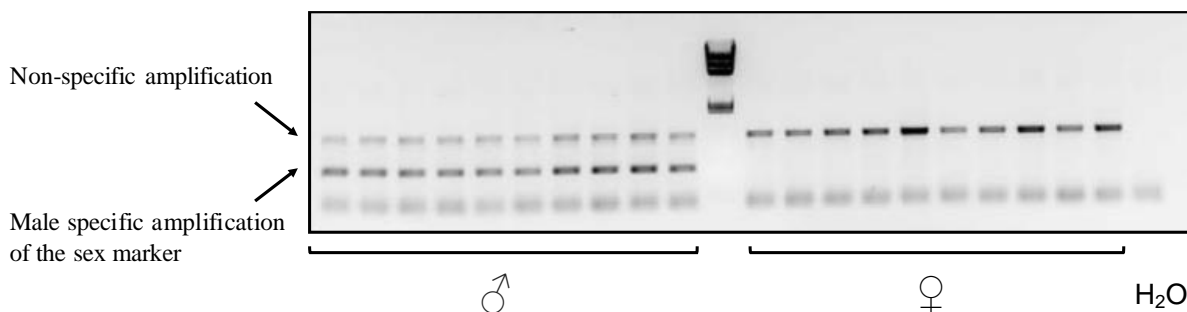


Figure 2: PCR amplification of a male specific sex marker in male (n=10) and female (n=10) genomic DNA samples from *U. pygmaea*.

***Umbra pygmaea* has an XX/XY sex determination system**

Analysis of the RADSeq data from 31 males and 31 females *U. pygmaea* using the RADSex pipeline revealed a clear XX/XY sex determination system (**Figure 1**) with 137 markers significantly associated with the male phenotype ($p < 0.05$, Chi-squared test with Bonferroni correction). Sequences homologous to these 137 markers were recovered and filtered based on coverage to determine whether markers were in a male specific locus or in a locus with fixed polymorphism between males and females. In total, 67 (49%) markers were determined as polymorphic between sexes and 70 (51%) markers were determined as specific to the males.

Mapping of the markers significantly associated with the male phenotype on to the two genome assembly resulted in 40 contigs with a total length of about 0.4 Mb in the discovar genome and 52 contigs with a total length of 2.5 Mb in the 10X genome. No candidate for MSD gene was identified in these sequences.

Male markers from RADSex were validated in *U. pygmaea*, but were not conserved in the sister species *U. limi*

Primers were designed on the contigs containing four of the male-specific markers. These primers were tested in 29 females and 20 males independent from the ones used in the RADSex analysis. In all these samples, specific amplification was obtained, with either a single amplification in males or multiple amplifications with a male specific band (**Table S2**), confirming a perfect correlation with male phenotype (**Figure 2**). These primers were also tested on gDNA extracted from sexed samples from a closely related species, *U. limi*, but not amplification were obtained.

The genetic map of *U. pygmaea* revealed a large and strongly differentiated sex chromosome (LG1)

A genetic map was generated from RAD-Seq data from a single full-sib family of *U. pygmaea*, for which genotypic sex was inferred using the markers from the RADSex analysis confirmed to be perfectly associated with sex. In total, 714 markers were retained as informative with 351 markers specifically for the males and 398 markers for the females. The total length of the sex-averaged map was 965 cM with an average distance of 1.36 cM between markers.

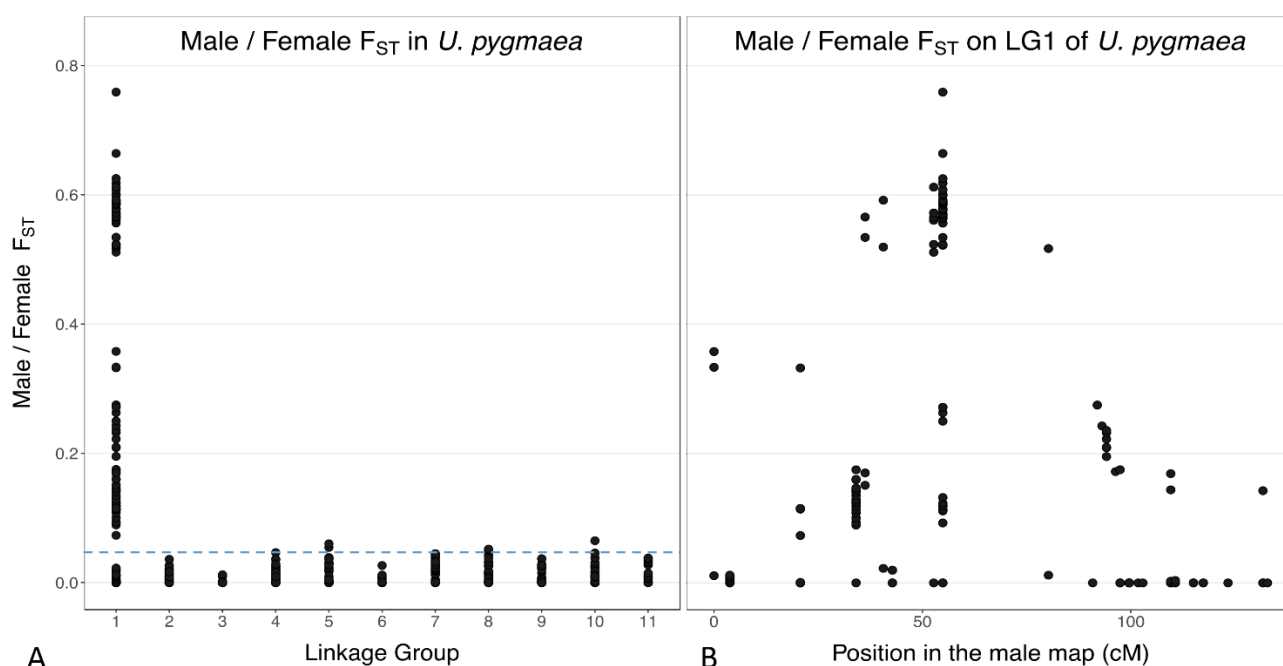


Figure 3: F_{ST} between males and females for all markers across the 11 linkages groups (LG) of *U. pygmaea* and along LG1. **A:** The F_{ST} between sex is plotted for all markers used in the genetic linkage map against the linkage group they belong to. The blue dotted line indicates the genome average F_{ST} (0.046). **B:** F_{ST} between males and females for all markers on LG1 is plotted against their position on the LG1 in the male map.

LG	Number of Markers	Male map length (cM)	Female map length (cM)	Sex-averaged map length (cM)
1	130	132.79	35.99	122.45
2	84	177.34	145.35	154.47
3	79	103.92	171.26	117.53
4	81	123.83	49.56	93.06
5	67	172.35	38.87	72.69
6	67	140.13	105.99	115.95
7	58	82.76	59.31	91.57
8	47	98.67	39.35	73.63
9	44	33.79	32.7	42.22
10	35	50.81	34.19	59.99
11	20	0	21.89	21.83
Total	714	1116.39	735.06	965.39

Table 1: Numbers of makers and the length of the 11 linkage groups of *U. pygmaea* in the male map, the female map, and the sex-averaged map.

The total length of the male map was 1,116 cM with an average distance of 1.56 cM between markers, and the total length of the female map was 735 cM with an average distance of 1.03 cM between markers. The average recombination rate was 1.51 times higher in females than in males. The total length and number of markers for each linkage group (LG) are given in **Table 1** and the sex-averaged map can be found in **Figure S2**. F_{ST} between males and females was calculated for each marker; the genome average F_{ST} was 0.046. The highest average F_{ST} was found on LG1 with a value of 0.23 (**Figure 3A**), much higher than the genome average and much higher than the average F_{ST} of any other LG, indicating that LG1 is the sex chromosome. On this LG, the F_{ST} was the highest at 54.9 cM on the male map with a value of 0.76 and gradually decreased towards both telomeres, locating the sex locus in the central region of this LG1 chromosome (**Figure 3B**).

The RADSeq data used to construct the linkage map was also analyzed with the RADSex pipeline and yielded 158 markers significantly associated with male phenotype ($p < 0.05$, Chi-squared test with Bonferroni correction), confirming the results from the analysis performed on the wild population.

The karyotype of *U. pygmaea* experienced Robertsonian translocation and the sex chromosome is a fusion of the ancestral karyotypes of LG08, LG12 and LG22 of *E. lucius*

All 14,142 markers that could be placed on the genetic map were mapped to the genome of *E. lucius*, and only markers with unique hits were retained. Due to the evolutionary distance between *U. pygmaea* and *E. lucius*, only 1821 markers (13%) were mapped. For 22 chromosomes in the genome of *E. lucius*, the majority of mapping markers (>70%) originated from a single linkage group in the genetic map of *U. pygmaea* (**Figure 4A; Table S1**). These results indicate that the reduction in karyotype number in *U. pygmaea* was likely the result of a series of mostly one-to-one chromosomal fusion.

Among 132 markers that mapped to LG08 of *E. lucius*, 125 (%) originated from the sex chromosome of *U. pygmaea*. Among 156 markers that mapped to LG12, 132 (85%) originated from the sex chromosome. Among 45 markers that mapped to LG22, 44 (98%) originated from the sex chromosome. Together, these markers added up to 88 % of the markers that can be placed on the sex chromosome of *U. pygmaea* (**Figure 4B**).

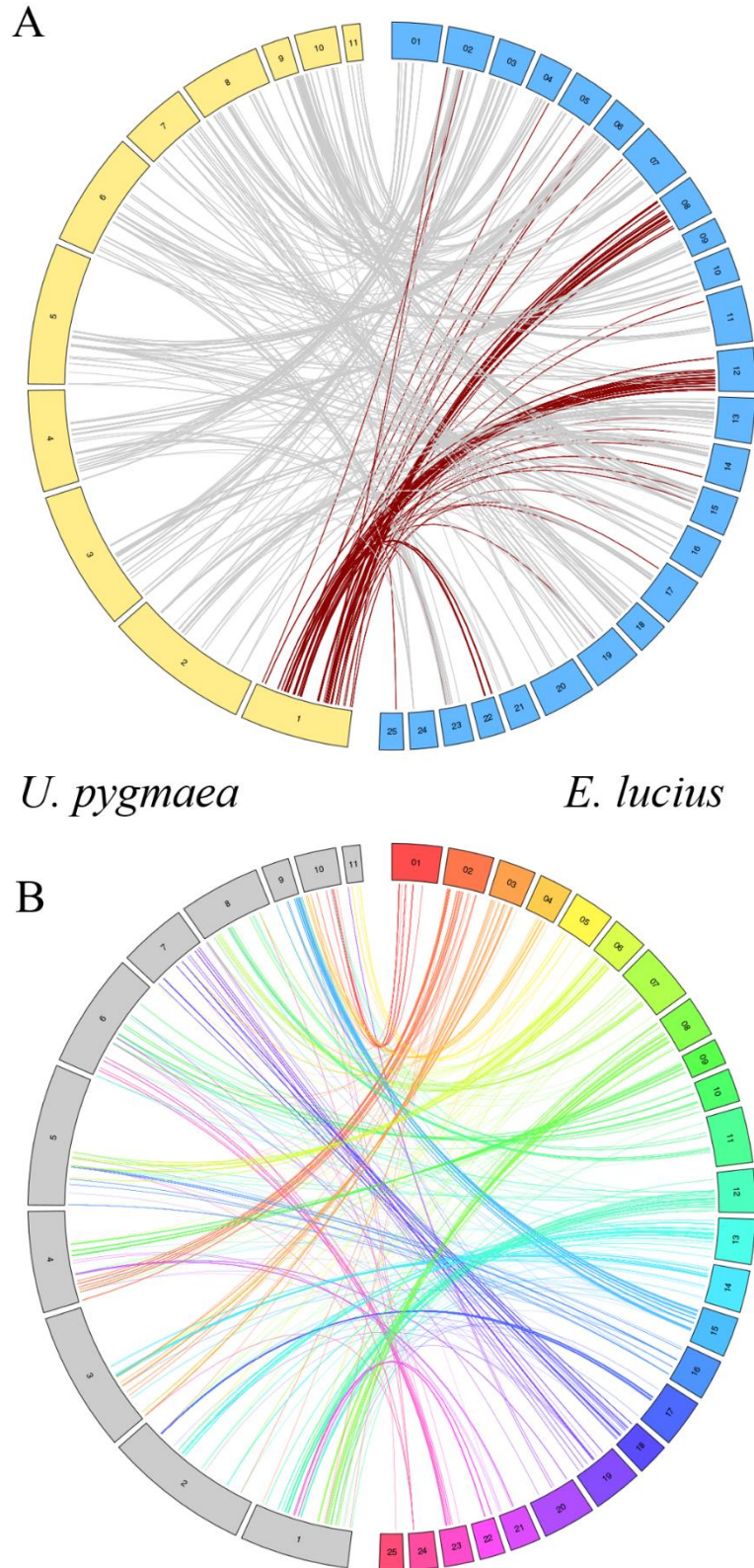


Figure 4: Circos plots showing markers linking LGs of *E. lucius* to LGs of *U. pygmaea*. The LGs of *U. pygmaea* are places on the left side and the LGs of *E. lucius* are place on the right side. **A:** The markers from the sex chromosome (LG1) of *U. pygmaea* is highlighted with red to show its correspondence with LG8, LG12 and LG22 of *E. lucius*. **B:** The markers from each LG of *U. pygmaea* is colored differently to show each of their correspondence with the LGs of *E. lucius*.

In addition, all 52,258 markers used in the RADSex analysis on the wild population and all 50,133 markers used in the RADSex analysis on the family panel were mapped to the genome of *E. lucius*, and only markers with unique hits were retained. In total, 11,402 (22%) markers from the wild population and 10,949 (22%) markers from the family panel were mapped. Among the 6 mapped markers from the wild population showing significant association with male phenotype, 3 (50%) were mapped to LG22, 1 (15%) was mapped to LG17, and 2 (30%) were mapped to unplaced scaffolds 0312 and 1154 (**Figure 5**). Among the 12 mapped markers from the family panel showing significant association with male phenotype, 4 (30%) were mapped to LG22, 4 (30%) were mapped to LG08, 2 (15%) were mapped to LG12, and 2 (15%) were mapped to unplaced scaffolds 0312 and 1154 (**Figure 5**).

Taken together, these results indicate that the sex chromosome of *U. pygmaea*, LG1, results from the fusion of the ancestral karyotypes of LG8, LG12 and LG22 of *E. lucius*.

Early molecular differentiation of the gonads in *U. pygmaea*

The migration of primordial germ cells (PGCs) was visualized from 4 dpf to 9 dpf (**Figure 6**), just before hatching. The PGCs had already reached the animal pole of the egg by 4 dpf, but they were still clustered together (**Figure 6A**). From 7 dpf to 9 dpf, the PGCs were no longer clustered, but instead lined along the ventral side of the embryos on either side of the body (**Figure 6D-6F**), a location that is likely to be close to the genital ridges of *U. pygmaea*.

The expression kinetics of *amh*, *cyp19a1a*, *dmrt1*, *foxl2*, *nr5a1 (sf1)*, and *fst* were measured by qPCR at 3 weeks post-fertilization (wpf), 6 wpf, 10 wpf, 14 wpf, and 20 wpf (**Figure 7**). The expression of all six genes had already started at 3 wpf. None of these genes, with the exception of *cyp19a1a*, showed an increase in their expression over the sampled developmental periods from 3 weeks to 20 wpf. In addition, histological analysis at 14 wpf showed clear morphological differentiation of the gonads (**Figure S1**). However, none of the surveyed genes showed sexual dimorphic expression at this time point; in fact, none of these genes showed a sustained sexual dimorphic expression pattern during the entire sampled developmental periods.

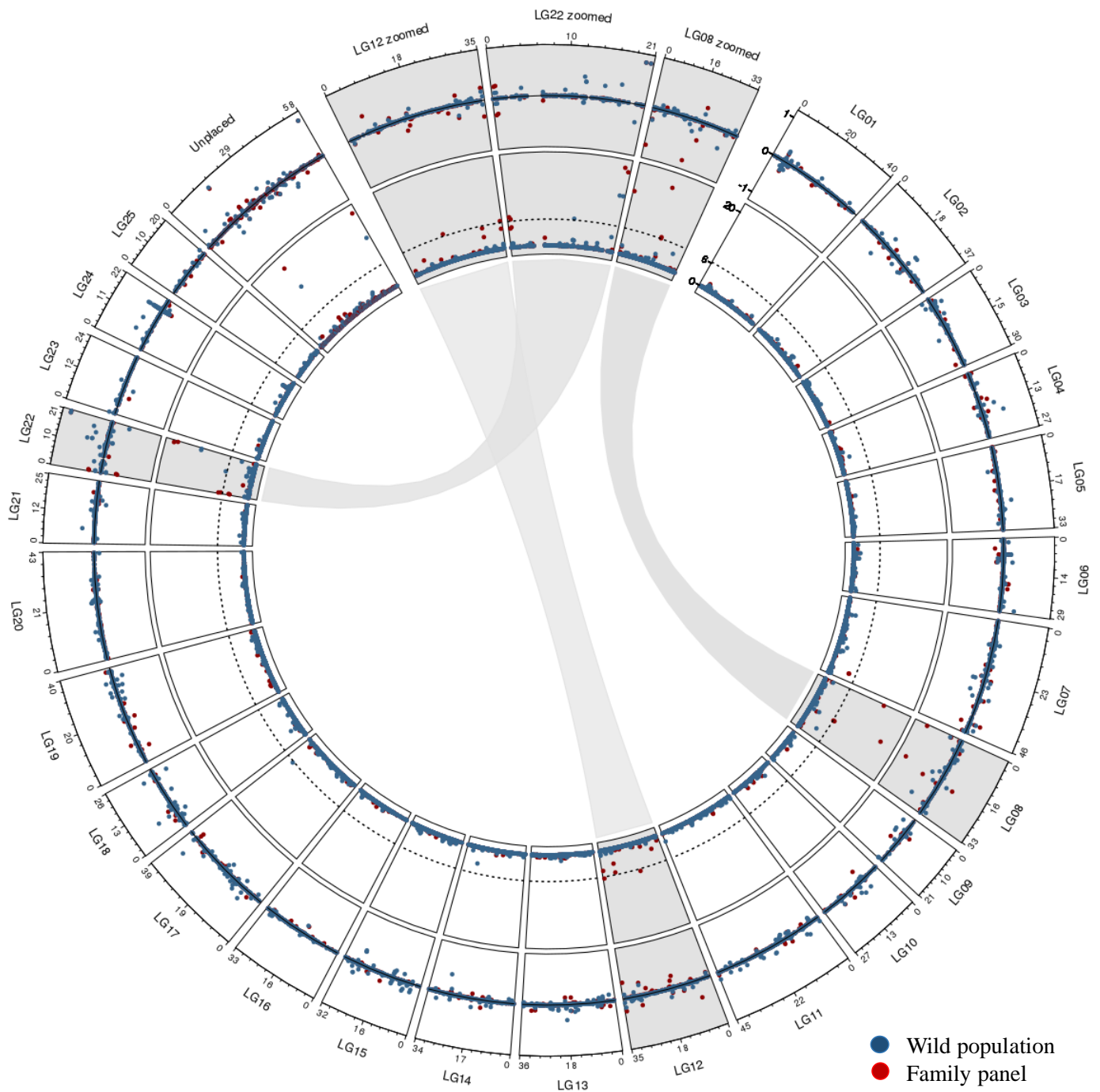


Figure 5: circular plot showing remapping of *U. pygmaea* RAD-Seq makers on *E. lucius* genome. The inner track plots the p-value of marker association with sex phenotypes, with the red-dotted line denotes significant level of $p = 0.05$, with Bonferroni correction. The outer track plots marker segregation pattern in each linkage group. For each marker, the ratio of difference in number of males and females (number of males – number of females) over total number of samples (number of males + number of females) is plotted against the mapped position of this marker. Having a ratio approaching 1 indicates male bias of a given marker, and having a ratio approaching -1 indicated female bias of a given marker. A zoomed view of remapping results on LG08, LG12, and LG22 is presented. Blue dots indicate sequence origin in the wild population, and red dots indicate sequence origin from the family panel.

Discussion

Our RAD-Seq approach yielded 714 informative SNP markers from a family panel of *U. pygmaea*, from which we were able to construct the first genetic linkage map for this non-model teleost species. Mapping construction was performed with stringent parameters and resulted in 11 robust linkage groups, in accordance with the karyotype of the species. The sex specific maps revealed differences in recombination rate between males and females, which have been reported in a few teleost species such as the Atlantic salmon (Lien *et al.*, 2011) and the Japanese flounder (Shao *et al.*, 2015). This genetic map was used to scaffold the draft genomes of *U. pygmaea*. However, due to the large size of the genome of *U. pygmaea*, the fragmentation of the draft genome, and the limited number of markers in this genetic map, only about 10% of the genome could be anchored on the linkage groups. Nevertheless, we were able to use information from this map to identify the sex chromosome of *U. pygmaea* and to study the evolution of the genome of this species in comparison with the genome of *E. lucius*.

Our analysis of a wild population of *U. pygmaea* using the RADSex pipeline revealed an XX/XY genetic sex determination system in this species. By anchoring male-specific RAD markers on the draft genomes, we were able to identify several contigs, whose size added to 2.5 Mb, potentially located in the sex locus. However, we were not able to identify candidate MSD genes in these contigs. In addition, the sex markers from *U. pygmaea* could not be amplified in its sister species *U. limi*, which has a short divergence time of 3.7 Mya from *U. pygmaea* (Marić *et al.*, 2017). Taken together, these results indicate that the sex-specific sequences identified in *U. pygmaea* originate from non-coding sequences, which evolve faster than protein-coding regions, and are thus more likely to be specific to *U. pygmaea*. This is not too surprising, as most of the sex locus apart from the MSD gene is expected to be degenerated, non-coding sequences (Charlesworth and Charlesworth, 2000).

In the genomes of *U. pygmaea*, we did not find homologous gene to neither *sdY*, the MSD gene of the salmonids (Yano *et al.*, 2012, 2013), nor *amhby*, the MSD gene of the *Esocidae* (Pan *et al.*, *in prep*), indicating that another gene is responsible for triggering sex determination in this species. The absence of *sdY* in *U. pygmaea* provides support for the hypothesis that this unusual MSD gene emerged after the divergence between the *Esociformes* and the salmonids from their last common ancestor around 113 Mya (Campbell *et al.*, 2013), and before the diversification of the salmonids lineages around 55 Mya (Campbell *et al.*, 2013). Interestingly, although the MSD gene is not conserved between these two orders, the majority of species within the *Esociformes* and the salmonids share an XX/XY sex determination system.

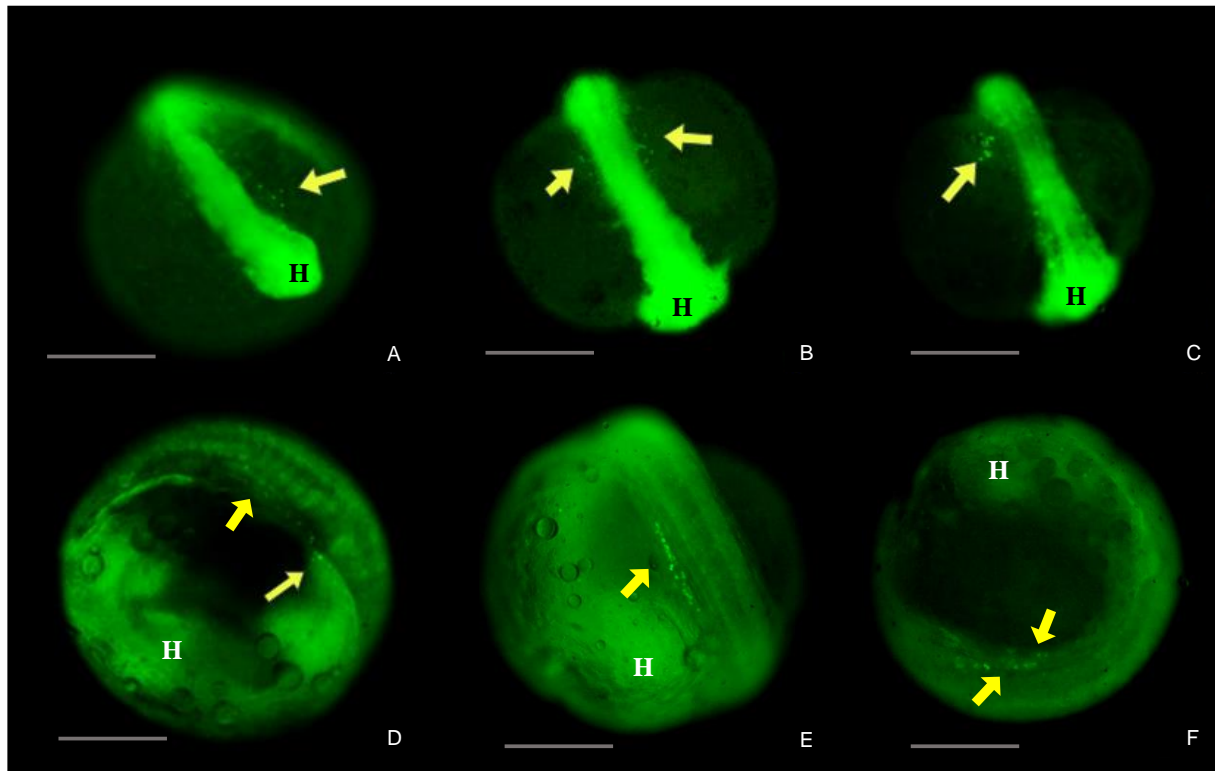


Figure 6: Visualization of PGCs (primordial germ cells) labelled by GFP-nanos1 3'UTR mRNA during embryogenesis of *U. pygmaea* from 4 dpf (days post fertilization) to 9 dpf. The yellow arrows point to the position of the PGCs and the letter 'H' indicate the position of the head of the embryo. **A:** 4 dpf, **B:** 5 dpf, **C:** 6 dpf, **D:** 7 dpf, **E:** 8 dpf, and **F:** 9 dpf. Scale bars denote 500 nm.

The RADSex analysis yielded 137 male-specific markers in the wild population and 158 markers in the family panel. In *E. lucius*, where the region with suppression of recombination was found to span around two thirds of the sex chromosome (Pan *et al.*, *in prep.*), the same RADSex analysis yielded 79 male-specific markers in a family panel. As the genomes of *E. lucius* and *U. pygmaea* have a roughly similar frequency of RAD *SbfI* cutting sites (1 site every 0.052 Mb in *E. lucius* and every 0.045 Mb in *U. pygmaea*), these results suggest that the sex locus in *U. pygmaea* is bigger than that of *E. lucius*. In agreement with this estimation, the genetic map shows that suppression of recombination occurs over most of the large LG1. Taken together, these results indicate that the sex locus in *Umbra pygmaea* is large and well-differentiated. In fact, this sex locus is by far the largest we detected in all *Esociformes*, which could suggest that the differentiation between the X and Y chromosomes started earlier in *U. pygmaea* than in the *Esocidae*. Therefore, it is possible that the ancestral MSD gene of the *Esociformes*, prior to the duplication of *amh* in the *Esocidae*, could still be operating in *U. pygmaea*. On the other hand, the genome of *U. pygmaea* is estimated to be 2.4 Gb in size (Beamish *et al.*, 1971), more than twice the average size of the genomes in the *Esocidae* (Hardie and Hebert, 2004). As we did not detect any signature of whole genome duplication based on our own analysis and on the absence of ohnologs inferred from RNA-Seq cDNA assemblies (Pasquier *et al.*, 2016), this genus-specific genome size expansion is likely caused by an accumulation of repetitive elements as already found in a few plant species, such as *Capsicum annuum* and the genus *Panax* (Park *et al.*, 2012; Lee *et al.*, 2017). In this case, the accumulation of extra repetitive elements in the region with suppression of recombination could have inflated the size of the sex locus. A better genome assembly, which is on the way, will help to better determine the size of the sex locus relative to LG1 and to the entire genome of *U. pygmaea*.

U. pygmaea has 11 pairs of chromosomes (N=22), fewer than half the number of chromosomes in its sister species from the *Esox* (N=50) and the *Novumbra* (N=48) genera (Beamish *et al.*, 1971), and lower than the average number of chromosomes in teleost fishes (N=50, Nirchio *et al.*, 2014; Braasch *et al.*, 2016; Amores *et al.*, 2017). Our comparison of the genetic map of *U. pygmaea* with the genome of *E. lucius* showed that most chromosomes from *E. lucius* are mapped to a single linkage group in *U. pygmaea*, indicating that the reduction of the number of chromosomes in *U. pygmaea* most likely resulted from a series of simple one-to-one chromosomal fusions. The occurrence of such fusion events without prior whole genome duplication has only been reported in *Notothenia coriiceps*, another teleost species (Amores *et al.*, 2017).

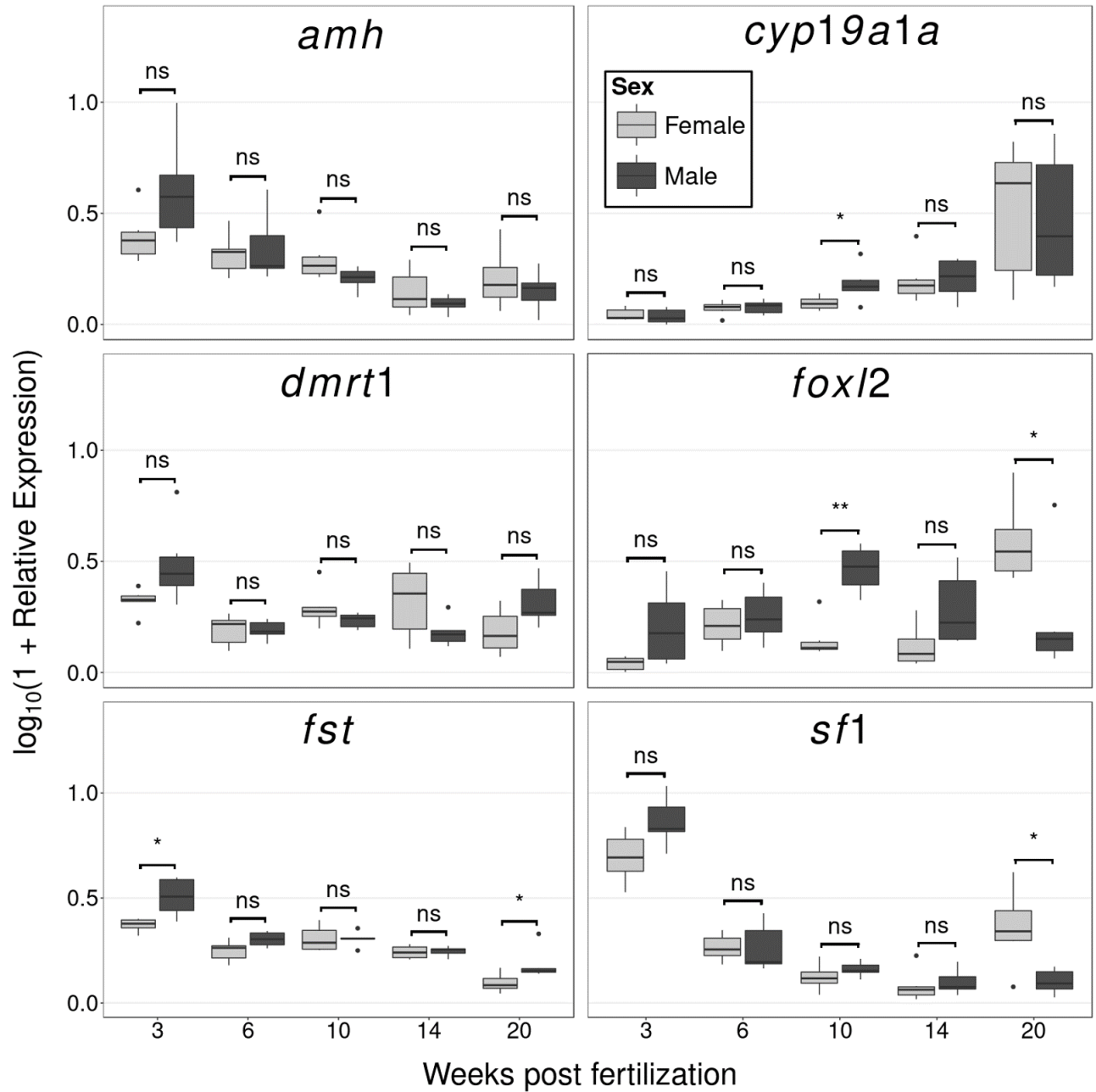


Figure 7: Boxplots showing the first quantile, median, and the third quantile of the temporal expression of *amh*, *dmrt1*, *cyp19a1a*, *foxl2*, *fst* and *sf1*, measured by qPCR. Outliers are displayed as a dots. The \log_{10} of relative expression of mRNA of these four genes were measured at 3, 6, 10, 14 and 20 weeks post fertilization in male and female samples of *U. pygmaea*. P-value significance level is 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 $P > 0.05$.

In particular, our results show that the sex chromosome in *U. pygmaea*, LG1, results from the fusion of the ancestral karyotypes of LG8, LG12, and LG22 of *E. lucius*. As the LG8 of *E. lucius* harbors *amh*, the duplicate of which determines the sex in most *Esocidae*, these results open the possibility of *amh* as the MSD gene in *Umbra pygmaea*. However, we did not detect evidence of gene duplication nor allelic diversification of *amh* in this species; furthermore, *amh* was not differentially expressed during gonadal development.

The gonadal development of *U. pygmaea* started much earlier than that of *E. lucius*. In *E. lucius*, the MSD gene, *amhby*, started its expression about 8 wpf, while other gonadal differentiation markers such as *dmrt1* and *cyp19a1a* started their expression even later, around 14 wpf (Pan *et al.*, *in prep*). In comparison, the expression of six gonadal differentiation markers had already started at 3 wpf in *U. pygmaea*. The expression of these markers did not increase from 3 wpf to 14 wpf, at which stage clear morphological differentiation of the gonads could be observed, suggesting that their expression peaked before 3 wpf. Furthermore, the PGCs of *U. pygmaea* had already migrated to the position of the genital ridge at 7 dpf, enabling an early formation of functional gonads (Wylie, 1999; Starz-Gaiano and Lehmann, 2001). Taken together, these results suggest that gonadal differentiation occurs earlier in *U. pygmaea* than in *E. lucius*, and that earlier sampling before three wpf would be needed to capture the molecular differentiation of the gonads in *U. pygmaea*.

Conclusions and perspectives

In this study, we investigated the sex determination of *U. pygmaea*, a member of the *Umbridae*, to complement our work on the evolution of sex determination in its sister group, the *Esocidae*. We analyzed RAD-Seq data from a wild population and also provided two genome assemblies, a genetic linkage map, and early molecular differentiation pattern of the gonads of *U. pygmaea* to facilitate the study of sex determination in this species. We found that *U. pygmaea* has an XX/XY sex determination system and identified the sex chromosome with a well-differentiated sex locus. Although we did not find candidate MSD genes, we identified sex markers that could be readily used as a sexing tool. In addition, we found a much earlier engagement of gonadal differentiation in *U. pygmaea* comparing to *E. lucius*, and provided guidelines for further molecular studies on the sex determination of *U. pygmaea*. Lastly, comparison between the genetic linkage groups of *U. pygmaea* and the chromosomes of *E. lucius* revealed considerable chromosomal fusions in the karyotype evolution of *U. pygmaea*.

The sequencing of a *U. pygmaea* genome with better continuity is ongoing, which will help unveil the identity of the MSD gene in the near future. We will also incorporate more markers, as well as protein synteny in the genome comparison analysis to better understand the mechanism underlying the drastic karyotype reduction in the *U. pygmaea* genome.

References

- Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., & Postlethwait, J. H. (2011). Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. *Genetics*, 188(4), 799–808. <https://doi.org/10.1534/genetics.111.127324>
- Amores, A., Wilson, C. A., Allard, C. A. H., Detrich, H. W., & Postlethwait, J. H. (2017). Cold Fusion: Massive Karyotype Evolution in the Antarctic Bullhead Notothen Notothenia coriiceps. *G3 (Bethesda, Md.)*, 7(7), 2195–2207. <https://doi.org/10.1534/g3.117.040063>
- Beamish, R. J., Merrilees, M. J., & Crossman, E. J. (1971). Karyotypes and DNA values for members of the suborder Esocoidei (Osteichthyes: Salmoniformes). *Chromosoma*, 34(4), 436–447.
- Bekaert, M. (2016). Genetic-Mapper: vectorial genetic map drawer. *F1000Research*, 5. <https://doi.org/10.7490/f1000research.1112266.1>
- Braasch, I., Gehrke, A. R., Smith, J. J., Kawasaki, K., Manousaki, T., Pasquier, J., ... Postlethwait, J. H. (2016). The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature Genetics*, 48(4), 427–437. <https://doi.org/10.1038/ng.3526>
- Campbell, M. A., López, J. A., Sado, T., & Miya, M. (2013). Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. *Gene*, 530(1), 57–65. <https://doi.org/10.1016/j.gene.2013.07.068>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes/Genomes/Genetics*, 1(3), 171–182. <https://doi.org/10.1534/g3.111.000240>
- Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1403), 1563–1572. <https://doi.org/10.1098/rstb.2000.0717>
- Gharbi, K., Gautier, A., Danzmann, R. G., Gharbi, S., Sakamoto, T., Høyheim, B., ... Guyomard, R. (2006). A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, 172(4), 2405–2419. <https://doi.org/10.1534/genetics.105.048330>
- Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics (Oxford, England)*, 30(19), 2811–2812. <https://doi.org/10.1093/bioinformatics/btu393>
- Hardie, D. C., & Hebert, P. D. (2004). Genome-size evolution in fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(9), 1636–1646. <https://doi.org/10.1139/f04-106>

- Herpin, A., Rohr, S., Riedel, D., Kluever, N., Raz, E., & Schartl, M. (2007). Specification of primordial germ cells in medaka (*Oryzias latipes*). *BMC Developmental Biology*, 7, 3. <https://doi.org/10.1186/1471-213X-7-3>
- Köprunner, M., Thisse, C., Thisse, B., & Raz, E. (2001). A zebrafish nanos-related gene is essential for the development of primordial germ cells. *Genes & Development*, 15(21), 2877–2885. <https://doi.org/10.1101/gad.212401>
- Kuehne, L. M., & Olden, J. D. (2014). Ecology and conservation of mudminnow species worldwide. *Fisheries*, 39(8), 341–351.
- Lee, J., Waminal, N. E., Choi, H.-I., Perumal, S., Lee, S.-C., Nguyen, V. B., ... Yang, T.-J. (2017). Rapid amplification of four retrotransposon families promoted speciation and genome size expansion in the genus *Panax*. *Scientific Reports*, 7(1), 9045. <https://doi.org/10.1038/s41598-017-08194-5>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lien, S., Gidskehaug, L., Moen, T., Hayes, B. J., Berg, P. R., Davidson, W. S., ... Kent, M. P. (2011). A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics*, 12, 615. <https://doi.org/10.1186/1471-2164-12-615>
- Marić, S., Stanković, D., Wanzenböck, J., Šanda, R., Erős, T., Takács, P., ... Snoj, A. (2017). Phylogeography and population genetics of the European mudminnow (*Umbra krameri*) with a time-calibrated phylogeny for the family Umbridae. *Hydrobiologia*, 792(1), 151–168. <https://doi.org/10.1007/s10750-016-3051-9>
- Marshall, O. J. (2004). PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics (Oxford, England)*, 20(15), 2471–2472. <https://doi.org/10.1093/bioinformatics/bth254>
- Nirchio, M., Rossi, A. R., Foresti, F., Oliveira, C., Nirchio, M., Rossi, A. R., ... Oliveira, C. (2014). Chromosome evolution in fishes: a new challenging proposal from Neotropical species. *Neotropical Ichthyology*, 12(4), 761–770. <https://doi.org/10.1590/1982-0224-20130008>
- Park, M., Park, J., Kim, S., Kwon, J.-K., Park, H. M., Bae, I. H., ... Choi, D. (2012). Evolution of the large genome in *Capsicum annuum* occurred through accumulation of single-type long terminal repeat retrotransposons and their derivatives. *The Plant Journal: For Cell and Molecular Biology*, 69(6), 1018–1029. <https://doi.org/10.1111/j.1365-3113X.2011.04851.x>
- Pasquier, J., Cabau, C., Nguyen, T., Jouanno, E., Severac, D., Braasch, I., ... Bobe, J. (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics*, 17, 368. <https://doi.org/10.1186/s12864-016-2709-z>
- Rab, P., & Mayr, B. (1987). Chromosome Banding Studies in European Esocoid Fishes: Localization of Nucleolar Organizer Regions in *Umbra krameri* and *Esox lucius*. *Copeia*, 1987(4), 1062–1067. <https://doi.org/10.2307/1445579>
- Rastas, P., Calboli, F. C. F., Guo, B., Shikano, T., & Merilä, J. (2015). Construction of Ultradense Linkage Maps with Lep-MAP2: Stickleback F2 Recombinant Crosses as an Example. *Genome Biology and Evolution*, 8(1), 78–93. <https://doi.org/10.1093/gbe/evv250>
- Saito, T., Fujimoto, T., Maegawa, S., Inoue, K., Tanaka, M., Arai, K., & Yamaha, E. (2006). Visualization of primordial germ cells in vivo using GFP-nos1 3'UTR mRNA. *The International Journal of Developmental Biology*, 50(8), 691–699. <https://doi.org/10.1387/ijdb.062143ts>
- Shao, C., Niu, Y., Rastas, P., Liu, Y., Xie, Z., Li, H., ... Chen, S. (2015). Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis. *DNA Research: An International Journal for*

Rapid Publication of Reports on Genes and Genomes, 22(2), 161–170.

<https://doi.org/10.1093/dnares/dsv001>

- Siemering, K. R., Golbik, R., Sever, R., & Haseloff, J. (1996). Mutations that suppress the thermosensitivity of green fluorescent protein. *Current Biology: CB*, 6(12), 1653–1663.
- Starz-Gaiano, M., & Lehmann, R. (2001). Moving towards the next generation. *Mechanisms of Development*, 105(1–2), 5–18.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3, research0034.
<https://doi.org/10.1186/gb-2002-3-7-research0034>
- Wylie, C. (1999). Germ Cells. *Cell*, 96(2), 165–174. [https://doi.org/10.1016/S0092-8674\(00\)80557-7](https://doi.org/10.1016/S0092-8674(00)80557-7)
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., ... Guiguen, Y. (2012). An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Current Biology*, 22(15), 1423–1428.
<https://doi.org/10.1016/j.cub.2012.05.045>
- Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., & Guiguen, Y. (2013). The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications*, 6(3), 486–496.
<https://doi.org/10.1111/eva.12032>

Supplementary: Sex determination and karyotype evolution in the Eastern mudminnow, *Umbra pygmaea*

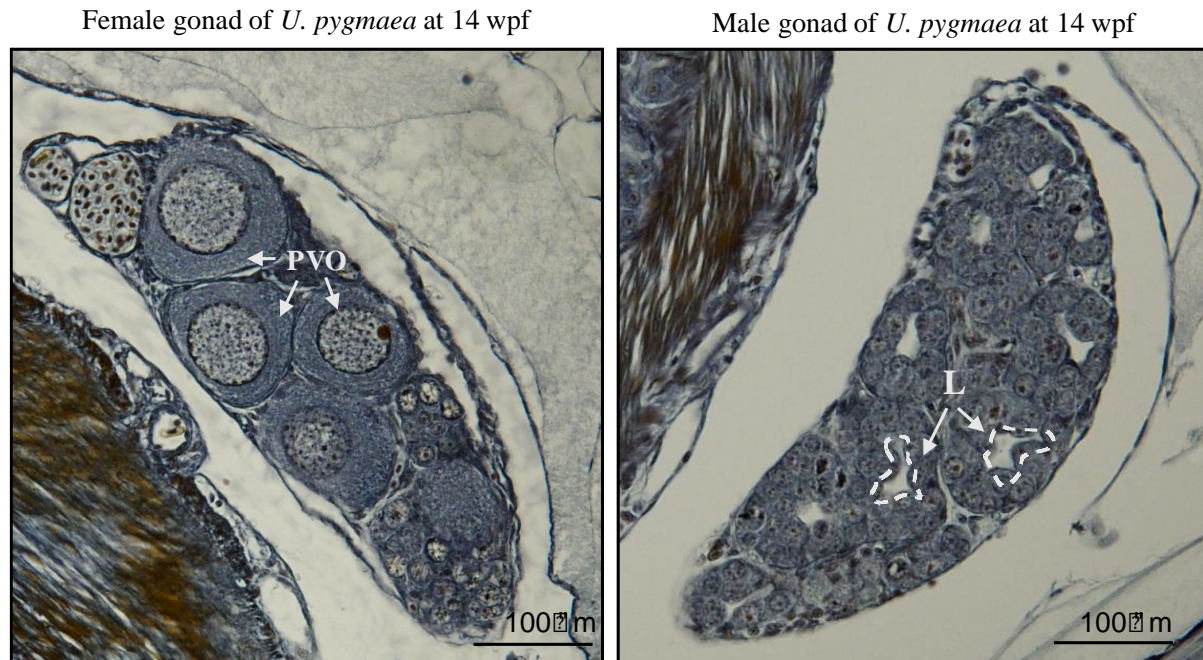


Figure S1: Gonadal phenotypes histology of male and female *U. pygmaea* gonads at 14 wpf (weeks post fertilization). **PVO:** previtellogenic oocytes; **L:** testicular lumen. Scale bars denote 100 μm.

Primer name	Primer sequence	Amplification pattern
Up_Mtag3_Fw	AAAGAAGGACAGAAAGCGAC	amplification only obtained in male samples
Up_Mtag3_Rv	CTCATGTTTCCATGTGTTCCC	
Up_Mtag8_Fw	GTTATCCACTACTGCCGTTCC	two bands amplified in males, one band in female
Up_Mtag8_Rv	GTTGCTCACATCCAATTTGAC	
Up_Mtag10_Fw	AACTGAAGTGTGGATATGGG	two bands amplified in males, one band in female
Up_Mtag10_Rv	CTTACAACATCCTCATTGACCA	
Up_Mtag12_Fw	CCAAGCATTTCATACCATGTCCT	amplification only obtained in male samples
Up_Mtag12_Rv	GTGTTCCCTGTCGTTCAATGTC	

Table S1: Primers to genotype *U. pygmaea* sex based on RADSex markers. Primers were designed on contigs containing the RADSex markers, which are included in the resulted amplicons from PCR on genomic DNA.



Figure S2: A sex-averaged map of the 11 linkage groups of *U. pygmaea*. For each sex, numbers at the left side of the linkage map indicate cumulative distances in cM and ID of the markers are listed on the right side.

Primer name	Primer sequence	Target gene
UP_dmrt1_F	TGCGTCGTCAACAAGCACAG	<i>dmrt1</i>
UP_dmrt1_R	GCTCATTCTTCACCACCACATC	
UP_foxl2_F	GGCACAACCTTGAGTCTGAACG	<i>foxl2</i>
UP_foxl2_R	CAGGTAGCCGTATCCATCCC	
UP_amh_F	ACAAGACCCTTAGGCACCAT	<i>amh</i>
UP_amh_R	TCAACAATAATTCGCCATCCC	
UP_SF1_1_F	GTGACCGAGCCCTGAAACAG	<i>sf1</i>
UP_SF1_1_R	TCCGTGGCAGTGATGGAGAT	
UP_cyp19a1a_F	GTTCTTCCTCAAGCCCAATG	<i>cyp19a1a</i>
UP_cyp19a1a_R	TACCGTGACAACAAAGTTACCAG	
UP_Fst_2_F	AAACGGAAGGTGTCAGGTGC	<i>fst</i>
UP_Fst_2_R	CATCCATCGGAACAAAGTGC	
UP_βactin_F	CCTCGGTATGGAGTCTTGCG	<i>β-actin</i>
UP_βactin_R	CTGTCAGCGATGCCTGGGTA	
UP_EF1α_F	ATCCGTCTGGTAATGTGGC	<i>ef1-α</i>
UP_EF1α_R	TGAACTTGCAGGCGATGTGA	
UP_gapdh_F	CTGTCACCGCCACCCAGAAG	<i>gapdh</i>
UP_gapdh_R	TGCCAGTGATCTTGCCGTTT	
UP_eftud2_F	TACTAAGAAAGCGCCAACC	<i>eftud2</i>
UP_eftud2_R	TGTCACAGTCACCCACCACC	
UP_ubr2_F	AGGAAGCGTAAGGCGGAGAT	<i>ubr2</i>
UP_ubr2_R	AGTGAGGAAGAGGTGGTGGC	
UP_ccdc6b_F	AAACCAGGCGTGATTTCT	<i>ccdc6b</i>
UP_ccdc6b_R	CTGCTCTGGCTTGGATTGTGA	

Table S2: Primers used for qPCR, with primer name, sequence and target genes.

Genome	Sex	Total length	N50 contig length	L50 contig count
Discover genome	male	1.95	7736	61111
10x genome	male	1.51	19366	1642

Table S3: Assemblathon's metrics of the two genome assemblies for *U. pygmaea*.

Section IV:

Perspectives

The importance of AMH in teleost sex determination

In my thesis project, we identified and characterized a male-specific duplicate of Anti-Müllerian hormone, that we named *amhby*, as the MSD gene in the Northern pike, *Esox lucius*. We demonstrated that *amhby* is the result of an ancient duplication event that occurred prior to the diversification of the *Esocidae* between 65 and 90 Mya, and we found a tight association of *amhby* with the male phenotype in most *Esocidae*. Prior to our studies, other members of the *amh* pathway were already identified as MSD genes in teleosts (Bej *et al.*, 2017; Hattori *et al.*, 2012; Li *et al.*, 2015; Rondeau *et al.*, 2016). Our discovery of *amhby* in Northern pike and its conservation as a MSD gene in several *Esociformes* strengthen the prevalence of the members of the *amh* pathway as important SD genes in teleosts.

This prevalence could be explained in part by the difficulties to identify MSD genes that do not belong to the ‘usual suspects’ families, namely, DMRT, SOX, and TGF- β , which contains AMH (Herpin and Schartl, 2015; Heule *et al.*, 2014; Kikuchi & Hamaguchi, 2013; Pan *et al.*, 2017). Yet, among these usual suspects, members of the *amh* pathway are still the most recurrent identified MSD genes in teleosts, which could suggest a central role of this pathway in sex determination in this group of vertebrates. However, most of the current knowledge on *amh* pathway members and their interactions comes from research on therian mammals in which AMH does not contribute to the determination of gonadal fate (di Clemente *et al.*, 2003; di Clemente *et al.*, 2010; Josso *et al.*, 1998; Josso *et al.*, 2001; Josso *et al.*, 2005). It is thus difficult to directly apply this mammalian knowledge to the understanding of the MSD function of the members of the *amh* pathway in other vertebrates. In particular, the teleost clade experienced a whole genome duplication, and several teleost lineages even experienced additional rounds of whole genome duplication (Bernot *et al.*, 2004; Christoffels *et al.*, 2004; Glasauer & Neuhauss, 2014). The genetic redundancy introduced by these whole genome duplications provided opportunities for gene neofunctionalization. In the TGF- β super family, to which AMH belongs, neofunctionalization could then generate novel interactions and regulations between TGF- β members, resulting in a more complex regulatory network than what is known in mammals. Therefore, there is a real need to comprehensively study the interactions and regulations of the TGF- β super family in a model teleost species to better understand why this protein family plays such an important role in sex determination in teleosts.

Despite the absence of a teleosts-specific framework describing the *Amh* pathways, several questions raised by our results could already be addressed using specific approaches to understand how *amh* duplicates could acquire a MSD function.

First, one unique characteristic of the duplication of *amh* in *Esox lucius*, as opposed to *amh* duplications in other teleosts, is the level of sequence divergence between the paralogs which lead us to hypothesize that the two proteins may have diverged in their functions. To investigate this hypothesis, we already performed *in vitro* transfection experiments to determine if *Amha* and *Amhby* use different downstream pathways for signal transduction. However, results from these preliminary experiments were very difficult to interpret and were thus not included in this thesis. One reason for this difficulty could be that the experiments were performed using mammalian cell lines, that do not account for potential fish specific regulations. To address this problem, such transfection experiment could be performed in fish cell lines instead. Beside these transfection experiments, several approaches could provide evidence for a divergence in the functions of the two *amh* paralogs. Using yeast two-hybrid screening (Y2H) in complement with immunoprecipitation, we could identify whether each *amh* paralog protein binds to different receptors and characterize these receptors, which would set the foundation to study a potential divergence in the downstream signaling pathways of these proteins. In addition, 3-dimensional modeling could potentially reveal how structural differences between the two proteins lead to the observed divergence in their functions.

Another result showing that mechanisms differing from the canonical mammalian ones could be involved in the MSD function of *amh* duplicates in teleosts is the identification of male-associated truncated *amhby* in *Esox niger* and *Novumbra hubbsi*. Prior to our study, a truncated male-specific *amh* duplicate was identified as strong candidate for male MSD gene in another teleost species (Bej *et al.*, 2017). However, the truncations identified in our studies are particularly intriguing, because they lead to strong modifications of the TGB- β domain, suggesting that a canonical *Amhby* protein is not required to trigger the sex-determination pathway. It is worth noting that these results are derived from genomic assemblies, and the first step to validate them would be to characterize the mRNA sequences and expression pattern of *amhby* in *Esox niger* and in *Novumbra hubbsi*. One foreseeable difficulty in this experiment is that both species are restricted to specific areas of North America. In addition, *Novumbra hubbsi* is a protected species found in only one national park, and it is very difficult to obtain live samples for this species.

In summary, our findings provide opportunities to study whether *amh* duplicates with MSD function use the canonical *amh* pathway in the teleosts, which would offer insights on the factors underlying the prevalence of *amh* as MSD gene in this group of vertebrates. Nevertheless, a comprehensive study of the *amh* pathway and of the TGF- β family in the teleosts is necessary to fully understand this evolutionary trend.

Evolution of sex determination (SD) systems

Our study constitutes the first nearly complete survey of SD systems in an order of teleosts. Our results revealed a dynamic evolution of SD systems with several suspected transitions, intertwined with the branching fate of an old MSD gene, *amhby*. These results from a single order of fishes mirror our knowledge of the fast turnover of SD systems in teleost fishes in general, but our ability to connect these transitions with changes in the MSD gene provided novel insights on the mechanisms driving this turnover.

The most surprising and interesting finding in this thesis is probably the rapid loss of the entire sex locus containing *amhby* in the North American populations of *E. lucius*. This finding is especially intriguing because even though the North American populations diverged from the European populations less than 200,000 years ago, the chromosomes pair which harbors *amhby* and shows strong differentiation between sexes in the European populations resembles autosomes with no differentiation between sexes in the North American populations.

Based on our current data, we think the most likely scenario to explain this complete loss of sex locus is that the small founder population of North American *E. lucius* was composed of mostly genetic females, some of which experienced sex reversal and became phenotypic males. In this scenario, the sex locus was either absent from the North American population entirely or was initially rare and then lost by drift. In the absence of a sex determiner, environmentally induced sex reversal could have served as a mechanism to generate phenotypic males in the population. Under this hypothesis, the turnover of SD system was not driven by the emergence of a new MSD gene, but rather by a sudden and complete loss of the old MSD gene. This scenario was never documented empirically, and is not predicted by the current models of sex chromosome evolution and transition. In addition, this scenario would highlight the importance of the interactions between environmental SD and genetic SD in the evolution and transition of SD systems in teleost fishes and other Poikilotherms.

Following the characterization this loss of the ancestral sex locus in the North American populations of Northern pike, the naturally following question is what SD system and potential new MSD gene are currently operating in these North American populations? To begin answering this question, we recently resequenced whole genomes of pooled males and pooled females from one North American population, an approach that provides a much better resolution than our initial RAD-Seq experiments.

Our preliminary analysis of these pooled sequencing experiments suggests an even more surprising evolution as we found two small genomic regions strongly enriched in female-specific heterozygous sites, indicating that a ZZ/ZW SD system could be operating in this North American population of *E. lucius*. If these results are confirmed, they would constitute the most rapid natural transition from XX/XY system to ZZ/ZW system documented to date. In addition, this scenario would provide a unique opportunity to study the competition between two SD systems by crossing individuals from the two populations. Results from such crossing experiments would provide unique insight on the mechanisms of transitions between different SD systems.

Beside these surprising finding on North American populations of *Esox lucius*, our results suggest that some other *Esociformes* species rely on non-monofactorial SD systems either with *amhby*, as in *Esox masquinongy*, or without *amhby*, as in *Dallia pectoralis*. In *E. masquinongy*, *amhby* showed varying levels of dissociation from male phenotype in different populations, as well as multiple *amhby* alleles, sometimes even in the same individual; both results could be explained by a polygenic system with two pairs of sex chromosomes. In *D. pectoralis*, we found both male-biased and female-biased markers, also suggesting a polygenic system, and *amhby* was not found in the genome of this species. Interestingly, theoretical work predicts that polygenic SD systems are not stable, and would most likely represent transitional stages of sex determination (van Doorn, 2014). Because many teleost fishes show a fast turnover of SD systems, it is probably not surprising to capture one of these ongoing transitions of SD systems. However, although our results support this hypothesis of potential polygenic SD systems in *E. masquinongy* and *D. pectoralis*, we need to confirm these findings on new biological samples and with additional resequencing experiments. After validating the sex association of these markers, crosses of breeders with different combinations of genotypes could be performed to isolate the sex determining regions and understand their epistasis.

In *Umbra pygmaea*, we found a monofactorial XY SD system, but this species diverged from the *Esocidae* before the birth of *amhby*, and thus another gene is regulating sex determination in this species. Because of the large size of the sex locus identified in *U. pygmaea*, we hypothesized that the sex chromosomes have been evolving for a long time, with an old MSD gene which could potentially be the ancestral MSD gene in the *Esociformes* prior to the recruitment of *amhby* in the *Esocidae*. However, the genome of *U. pygmaea* is twice the size of the genome of other *Esocidae* which probably due to an accumulation of transposable elements. The assembly of a genome with better continuity, which has already started, will allow us to compare the accumulation of transposable elements in the sex locus compared to the rest of the genome, and will thus help us estimate more precisely the cause of the large size of the sex locus, and potentially even the age of the sex locus.

References

- Bej, D. K., Miyoshi, K., Hattori, R. S., Strüssmann, C. A., & Yamamoto, Y. (2017). A Duplicated, Truncated amh Gene Is Involved in Male Sex Determination in an Old World Silverside. *G3: Genes, Genomes, Genetics*, g3.117.042697. <https://doi.org/10.1534/g3.117.042697>
- Bernot, A., Segurens, B., Vacherie, B., Birren, B., Fischer, C., Dossat, C., ... Skalli, Z. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431(7011), 946. <https://doi.org/10.1038/nature03025>
- Christoffels, A., Koh, E. G. L., Chia, J., Brenner, S., Aparicio, S., & Venkatesh, B. (2004). Fugu Genome Analysis Provides Evidence for a Whole-Genome Duplication Early During the Evolution of Ray-Finned Fishes. *Molecular Biology and Evolution*, 21(6), 1146–1151. <https://doi.org/10.1093/molbev/msh114>
- di Clemente, N., Jamin, S. P., Lugovskoy, A., Carmillo, P., Ehrenfels, C., Picard, J.-Y., ... Cate, R. L. (2010). Processing of Anti-Müllerian Hormone Regulates Receptor Activation by a Mechanism Distinct from TGF- β . *Molecular Endocrinology*, 24(11), 2193–2206. <https://doi.org/10.1210/me.2010-0273>
- di Clemente, N., Josso, N., Gouédard, L., & Belville, C. (2003). Components of the anti-Müllerian hormone signaling pathway in gonads. *Molecular and Cellular Endocrinology*, 211(1–2), 9–14. <https://doi.org/10.1016/j.mce.2003.09.005>
- Glasauer, S. M. K., & Neuhauss, S. C. F. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics: MGG*, 289(6), 1045–1060. <https://doi.org/10.1007/s00438-014-0889-2>
- Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., ... Strussmann, C. A. (2012). A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences*, 109(8), 2955–2959. <https://doi.org/10.1073/pnas.1018392109>
- Herpin, A., & Scharl, M. (2015). Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Reports*, 16(10), 1260–1274. <https://doi.org/10.15252/embr.201540667>
- Heule, C., Salzburger, W., & Bohne, A. (2014). Genetics of Sexual Development: An Evolutionary Playground for Fish. *Genetics*, 196(3), 579–591. <https://doi.org/10.1534/genetics.114.161158>
- Josso, N., Belville, C., di Clemente, N., & Picard, J.-Y. (2005). AMH and AMH receptor defects in persistent Müllerian duct syndrome. *Human Reproduction Update*, 11(4), 351–356. <https://doi.org/10.1093/humupd/dmi014>
- Josso, N., di Clemente, N., & Gouédard, L. (2001). Anti-Müllerian hormone and its receptors. *Molecular and Cellular Endocrinology*, 179(1–2), 25–32.
- Josso, N., Racine, C., di Clemente, N., Rey, R., & Xavier, F. (1998). The role of anti-Müllerian hormone in gonadal development. *Molecular and Cellular Endocrinology*, 145(1–2), 3–7.
- Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution: Novel Sex-Determining Genes in Fish. *Developmental Dynamics*, 242(4), 339–353. <https://doi.org/10.1002/dvdy.23927>
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., ... Wang, D. (2015). 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(11), e1005678. <https://doi.org/10.1371/journal.pgen.1005678>
- Pan, Q., Guiguen, Y., & Herpin, A. (2017). Evolution of sex determining genes in Fish. In *Encyclopedia of Reproduction, Second Edition*.

- Rondeau, E. B., Laurie, C. V., Johnson, S. C., & Koop, B. F. (2016). A PCR assay detects a male-specific duplicated copy of Anti-Müllerian hormone (amh) in the lingcod (*Ophiodon elongatus*). *BMC Research Notes*, 9(1). <https://doi.org/10.1186/s13104-016-2030-6>
- van Doorn, G. S. (2014). Evolutionary transitions between sex-determining mechanisms: a review of theory. *Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, 8(1–3), 7–19. <https://doi.org/10.1159/000357023>

Appendix

Induced spawning of Eastern mudminnow, *Umbra pygmaea*

Q. Pan¹, A. Patinote¹, F. Borel¹, H. Verreycken², A. Herpin¹, Y. Guiguen^{1§}, J. Bobe¹, H. Rime¹.

¹ INRA, Fish Physiology and Genomics laboratory, 35042 Rennes, France

² Research Institute for Nature and Forest (INBO), Kliniekstraat 25, 1070, Brussels, Belgium

Abstract

Objective

Umbra pygmaea (Eastern mudminnow) is invasive in Europe and holds potential ecological importance in aquatic ecosystems characterized by increased acidity. The reproduction of this species has only been documented in natural habitats, while information on how to breed them in captivity is scarce.

Results

In our study, we used partially purified gonadotropins (PPGs) from Chinook salmon to induce final gamete maturation in male and female *U. pygmaea*. Subsequently, we performed *in vitro* fertilization that allowed us to follow the development of the embryos, and we present here a time series of embryonic development from the first cell division to hatching. Additionally, we noticed that under our treatment, captive *U. pygmaea* females and males did not have a synchronized reproductive season. Overall, about 30% of the treated females showed a reproductive response and produced either ovules or immature oocytes. We also investigated factors influencing the success of female spawning in our protocol; both female weight and timing of the PPGs treatment during the reproductive season were positively correlated with spawning success. These results could potentially facilitate the development of *U. pygmaea* as a model species for studying other *Esociformes* species.

Introduction

The eastern mudminnow is one of the five species of *Umbridae* (*Esociformes*). It is a small freshwater fish from the east coast of the U.S.A [1, 2]. It was introduced in Europe at the beginning of the 20th century [1, 3], and is considered low-medium risk as a result of its moderate invasiveness [3]. It is also an exceptionally acid-tolerant species [2, 4, 5] that might play an important role in aquatic ecosystems characterized by increased acidity [5]. Reproduction of *U. pygmaea* has only been documented in a few ecological reports [5, 6], which describe spawning from April to May [6] after migration to spawning sites, male courtship displays, and nest guarding by both males and females [1, 7]. However, the ability to breed this species in captivity has not been successful. Spawning induction has been performed in many teleost species, and hormonal induction provides direct control over the final stages of the reproductive cycle [8], facilitating downstream practices, such as *in vitro* fertilization. The main goal of our study was then to develop an effective spawning induction protocol in *U. pygmaea* in order to facilitate its usage as an experimental model for other *Esociformes* species, which could be hard to maintain in captivity, i.e. the northern pike (*Esox lucius*), or are ecologically venerable, i.e. the European mudminnow (*Umbra krameri*).

Material and Method

Fish collection

The population of eastern mudminnows (*U. pygmaea*) used in this study was sampled in Flanders, Belgium. Eastern mudminnows were captured using electric fishing gear in a pond and connecting ditches in a nature reserve in the vicinity of Genk (Belgium). Exact coordinates (latitude and longitude) of the pond are 50°57'12.84"N and 5°27'16.54"E. *U. pygmaea* is a non-native species in Belgium and is not protected by any legislation there. These *U. pygmaea* were transported alive to an indoor experimental facility located in Rennes (France).

Week number	Dosage of PPGs	Number of female	Number of male
1	100µg/Kg	5	5
1	500µg/Kg	5	0
1	1000µg/Kg	5	0
2-5	100µg/Kg	0	5
2-5	500µg/Kg	5	0
2-5	1000µg/Kg	5	0
6	100µg/Kg	0	5
6	500µg/Kg	4	0
7-9	100µg/Kg	0	5
7-9	1000µg/Kg	10	0

Table 1: The concentration of PPG (µg/kg of fish weight) used and the number of animals that received PPG treatment each week since the 1st week at the beginning of March 2016.

Laboratory rearing conditions

Mudminnows were maintained indoor in 1 m² tanks filled with 200 liters of dechlorinated local water (pH=8) filtered with a recirculating system. Tanks were enriched with artificial aquarium plants and polyvinyl chloride (PVC) plastic tubes. Rearing temperature followed the trend of external natural temperature. The photoperiod scheme used also mirrored the natural conditions. A detailed table can be found in the supplementary material (**Additional file 1, Table S1**). Mudminnows were fed with bloodworm larvae (Berkley Powerbait Micro Blood Worms) three times per day. To facilitate individual identification, each mudminnow was implanted in early January 2016 with a magnetic microchip from the National Nanotechnology Center (Nanotec). Around the end of February 2016, fish of different sex were separated in different tanks to prevent spontaneous natural reproduction.

Spawning induction

Animals scheduled to be treated in the upcoming week were maintained in separation in smaller 5 L tanks 2 days prior to the experiment. Feeding was stopped for these animals to facilitate the collection of clean eggs and sperm. From preliminary experiences, we knew that spermiation was relatively easy to obtain in males and that induced spawning of females was the limiting step in achieving effective controlled reproduction of captive *U. pygmaea*. As a result, we decided to include more females in each induction experiment, and to treat them with higher concentrations of hormone. We then used 5 males and 10 females for weekly induction treatments, and individuals of different sexes were always kept separately. In total, 59 females and 30 males were used in the first round of experiments starting from the beginning of March.

Each treatment cycle lasted for about one week. On experiment day 1, after anesthetizing the animals with phenoxyethanol (0.3 ml/l), we administered PPGs directly into the abdomen of both male and female fish using a 5 mL syringe with a 0.45×12 mm needle. The treatment used in these experiments was PPGs from Chinook salmon containing around 1.8% follicle-stimulating hormone (FSH) and 98.2% luteinizing hormone (LH) [8]. All fish were weighed before the experiment, and the amount of PPGs used was proportional to the body weight of the animals; males were treated with PPGs at a ratio of 100 µg/kg, and half of the females were randomly treated with PPGs at a ratio of 500 µg/kg while the other half received 1000 µg/kg. Afterwards, animals were released back into the same small tanks and feeding was not resumed. The concentration of PPGs used and the number of animals that received PPGs treatment each week are detailed in **Table 1**.

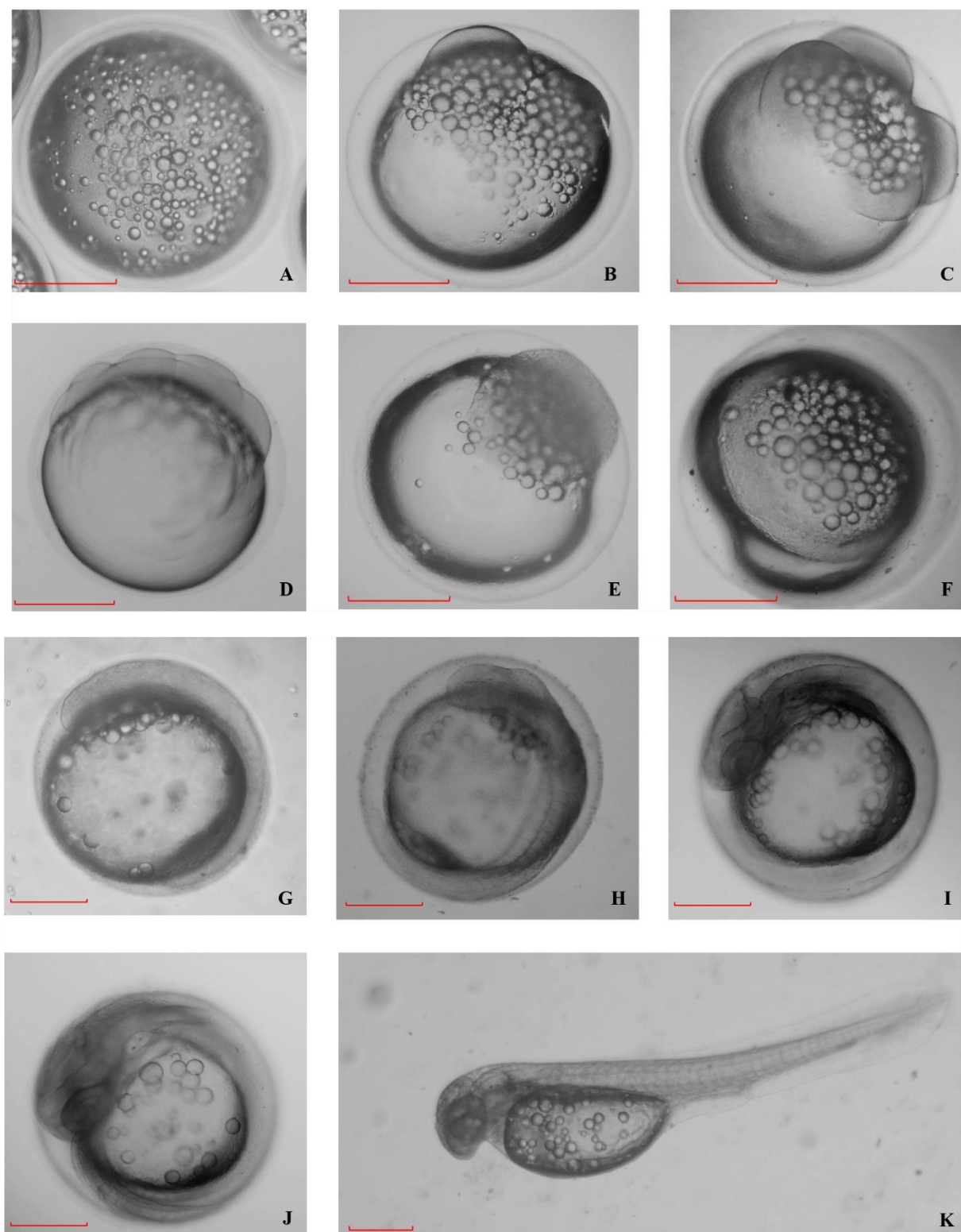


Figure 1: Developing *U. pygmaea* embryos at 12°C. **A:** 30 minutes after fertilization, 1-cell stage; **B:** 2 hours after fertilization, 2-cell stage; **C:** 3 hours after fertilization, 4-cell stage; **D:** 5 hours after fertilization, 8-cell stage; **E:** 24 hours after fertilization, blastula period; **F:** 28 hours after fertilization, gastrula period, germ ring stage; **G:** 3 days post-fertilization, segmentation period and eyes started to form; **H:** 4 days post-fertilization; **I:** 7 days post-fertilization; **J:** 8 days post-fertilization; **K:** 9 days post-fertilization, hatched embryo. Scale bars denote 500 nm.

On day 3, a second PPGs treatment at the same concentration and volume was delivered. On day 5, females were checked for ovulation by gentle abdominal pressure while applying a sliding motion towards the vent. When ovulation was observed, the stream of eggs was directed towards a clean and dry petri dish. Sperm was collected from a single male to perform *in vitro* fertilization using Actifish medium (IMV Technologies, France) according to the producer's protocol. Afterwards, individuals were returned to the main tanks with the rest of the population and feeding was resumed.

For females that did not show response during the first experimental round from March to April, we carried out a second round of induction treatment from the end of April to the end of May. We followed the procedure used during the first experiment, with the exception that all females were treated with gonadotropins at 1000 µg/kg and all males at 100 µg/kg. *In vitro* fertilization was performed when eggs were obtained.

Embryo development characterization

We followed the development of a single clutch of eggs. Embryos were placed in 1x Yamamoto's embryo-rearing medium [9] and kept in an incubator at a temperature of 12°C. The medium was changed and dead eggs were discarded every day. To record the development, we took pictures at different developmental stages from 30 minutes post-fertilization until hatching (**Figure 1**).

Statistical analyses

To understand the factors influencing the success of female spawning with our protocol, we analyzed our results using a logistic regression model, with female ovulation as a binary response variable with three explanatory variables: female weight, concentration of PPGs used, and timing of treatment. We only considered data from the first experimental round of spawning induction, which included the first 6 weeks of induction experiment. We represented the timing of treatment as the week in the experiment during which a female received the induction treatment (see **Additional file 1, Table S2**). We fitted this model to our data using the *glm* function from the *stats* package in R (version 3.2.2).

Female ID	Weight (g)	Number of treatment received	Spawning date	Concentration of PPG	Week number
5	6.0	1	4/4/2016	500µg/kg	3
12	14.0	1	11/4/2016	500µg/kg	4
26	14.5	1	13/04/2016	500µg/kg	4
48	12.2	1	18/04/2016	500µg/kg	5
84	4.3	1	18/04/2016	1000µg/kg	5
13	5.0	1	25/04/2016	500µg/kg	6
170	5.3	1	27/04/2016	1000µg/kg	6
55	5.6	2	6/5/2016	1000µg/kg	7
43	7.9	2	17/05/2016	1000µg/kg	8
66	5.0	2	17/05/2016	1000µg/kg	8
100	9.6	2	27/05/2016	1000µg/kg	9

Table 2: Weight, the number of treatments, spawning date, treatment concentration, and the week of treatment of the 11 females that successfully spawned and produced viable eggs after fertilization during our experiment.

Results

Induction of spawning in both males and females of *U. pygmaea*

During the entire experimental period from early March to late May, we did not observe any case of spontaneous spawning in females. Additionally, during each week's induction experiment, none of the females showed any sign of ovulation upon egg stripping until after a second administration of PPGs. During the first and second weeks of the first induction experimental round (third and fourth weeks of March 2016), some females already produced either immature or mature oocytes after the treatment. However, oocytes produced by these responsive females could not flow freely from the vent upon stripping, and did not develop normally after fertilization. Starting from the third experimental week (first week of April), we were able to obtain viable ovules from a few females every week.

In order to maximize ovule production, we performed a second round of PPGs induction. Starting from the 7th week of the experiment (late April), we obtained ovules from five PPG-treated females that did not respond to the first induction. None of the females that responded to the first round of treatment were able to produce ovules again after the second treatment. Surprisingly, although females continued to respond to the treatment in the second round of PPGs induction until the end of May, males stopped producing sperm before the end of the experimental period.

Overall, we applied PPGs treatment on 59 females and 31 males. Among them, 19 females (32.2 %) responded to the treatment by producing immature oocytes or ovules, and 11 of these females (18.6 %) produced eggs that developed normally after fertilization (**Table 2**).

Statistical analysis on factors influencing the success of induced spawning in female *U. pygmaea*

We examined potential factors that could influence the outcome of induced spawning in female *U. pygmaea* which included: 1) weight, 2) concentration of PPGs, and 3) timing of the treatment. The logit model shows that two factors were positively correlated with the success of induced spawning: female weight ($p = 0.018$) and the timing of injection ($p = 0.001$). The concentration of PPGs did not seem to have a significant effect on the response to the treatment ($p=0.397$). A regression plot for each explanatory variable can be found in **Figure S1**. Data used for this analysis is presented in **Table S1**.

Development of *U. pygmaea* embryos

We followed the development of a single clutch of *U. pygmaea* embryos from fertilization until hatching, and a chronological timeframe for different developmental stages is presented in **Figure 1**. The newly fertilized eggs stayed at the one-cell stage for about two hours before the first cleavage occurred. The embryos reached the blastula stage 1 day post-fertilization (dpf), and reached the gastrula stage the next day. After three days, the embryos reached the segmentation period, and eyes started to form. On the 8th day post-fertilization, some embryos started to hatch. Asynchrony of hatching was observed, but all embryos had hatched by the 11th day.

Discussion

U. pygmaea is a small-sized, easy to maintain member of the *Esociformes* that is suitable to be developed as an experimental model to study other *Esociformes* species. In this study, we designed a protocol to successfully induce spawning in a wild-caught population of *U. pygmaea* under laboratory captivity conditions. We also provide a time-series of embryonic development of this species to facilitate future studies on the development of *U. pygmaea*. We noted that hatching started around 8 dpf when incubated at 12°C, and this was comparable to the development timings reported for its European sister species *U. krameri* [10, 11].

Additionally, we investigated the factors that could influence female spawning success in our protocol, which included: 1) weight as a proxy for age, 2) concentration of PPGs treatment, and 3) the timing of PPGs treatment. We found that female weight and timing of treatment were the two parameters that had the most impact on the success of induced spawning in our protocol, while the concentration of PPGs did not show a strong influence. In our analysis, weight was used as a proxy for age as the precise age of the wild-caught individuals was unavailable. We found a positive correlation between the weight of the female and the success of spawning. This result agrees with the previous observation by Dederen and colleagues [5], who reported a positive correlation between fecundity and weight in *U. pygmaea*. It is worth noting that while we identified factors impacting the success of induced spawning, the genetic background of the females could have also contributed to the success of reproduction in captivity.

Another interesting point to note was that the periods of spawning of males and females under PPGs induction were not exactly synchronized. As complicated reproductive behavior is observed in this species in nature, inhibition of behavioral stimuli due to separation of the two sexes in captivity could be the reason for this observation, which has also been documented in other teleost species [12].

Limitations

- In our study, we did not have replicates in testing parameters in our protocol, as our main goal was to induce spawning instead of investigating these factors.
- We also did not identify the peak of reproductive season for *U. pygmaea* as the small number of animals used in each week's treatment could not confer statistical power.
- Lastly, our results were generated within one reproductive season, leaving yearly variation unclear.

References

- [1] Jenkins RE, Burkhead NM. Freshwater fishes of Virginia. Am Fish Soc; 1994.
- [2] Kuehne LM, Olden JD. Ecology and conservation of mudminnow species worldwide. Fisheries. 2014 Aug 3;39(8):341-51.
- [3] Verreycken H, Geeraerts C, Duvivier C, Belpaire C. Present status of the North American *Umbra pygmaea* (DeKay, 1842) (eastern mudminnow) in Flanders(Belgium) and in Europe. Aquat Invasions. 2010 Mar 1;5(1):83-96.
- [4] Bonga, Sjoerd E. Wendelaar, *et al.*, "Prolactin cell activity and sodium balance in the acid-tolerant mudminnow *Umbra pygmaea* in acid and neutral water." *Gen Comp Endoc* 78.3 (1990): 421-432.
- [5] Dederen LH, Leuven RS, Wendelaar SE, Oyen FG. Biology of the acid-tolerant fish species *Umbra pygmaea* (De Kay, 1842). *J Fish Biol*. 1986 Mar 1;28(3):307-26.
- [6] Panek FM, Weis JS. Age, Growth, and Reproduction of the Eastern Mudminnow (*Umbra pygmaea*) at the Great Swamp National Wildlife Refuge, New Jersey. *Northeast Nat*. 2012 Jul 17;19(2):217-28.

- [7] Breder, C.M., Jr., and D.E. Rosen. Modes of Reproduction in Fishes. American Museum of Natural History, Natural History Press; 1966.
- [8] Govoroun M, Chyb J, Breton B. Immunological cross-reactivity between rainbow trout GTH I and GTH II and their α and β subunits: application to the development of specific radioimmunoassays. *Gel Comp Endoc.* 1998 Jul 31;111(1):28-37.
- [9] Rembold M, Lahiri K, Foulkes NS, Wittbrodt J. Transgenesis in fish: efficient selection of transgenic fish by co-injection with a fluorescent reporter construct. *Nat Protoc.* 2006 Aug 1;1(3):1133.
- [10] Wanzenböck J. Current knowledge on the European mudminnow, *Umbra krameri* WALBAUM, 1792 (Pisces: Umbridae). *Annalen des Naturhistorischen Museums in Wien. Serie B für Botanik und Zoologie.* 1995 Jan 1:439-49.
- [11] Kucska B, Kabai P, Hajdú J, Várkonyi L, Varga D, Müllerné-Trenovszki M, Tatár S, Urbányi B, Zarski D, Müller T. Ex situ protection of the European mudminnow (*Umbra krameri* Walbaum, 1792): Spawning substrate preference for larvae rearing under controlled conditions. *Arch Biol Sci.* 2016 Jun 24;68(1):61-6.
- [12] Liley NR, Breton B, Fostier A, Tan ES. Endocrine changes associated with spawning behavior and social stimuli in a wild population of rainbow trout (*Salmo gairdneri*): I. Males *Gel Comp Endoc.* 1986 Apr 30;62(1):145-56.

Supplementary: Induced spawning of Eastern mudminnow, *Umbra pygmaea*

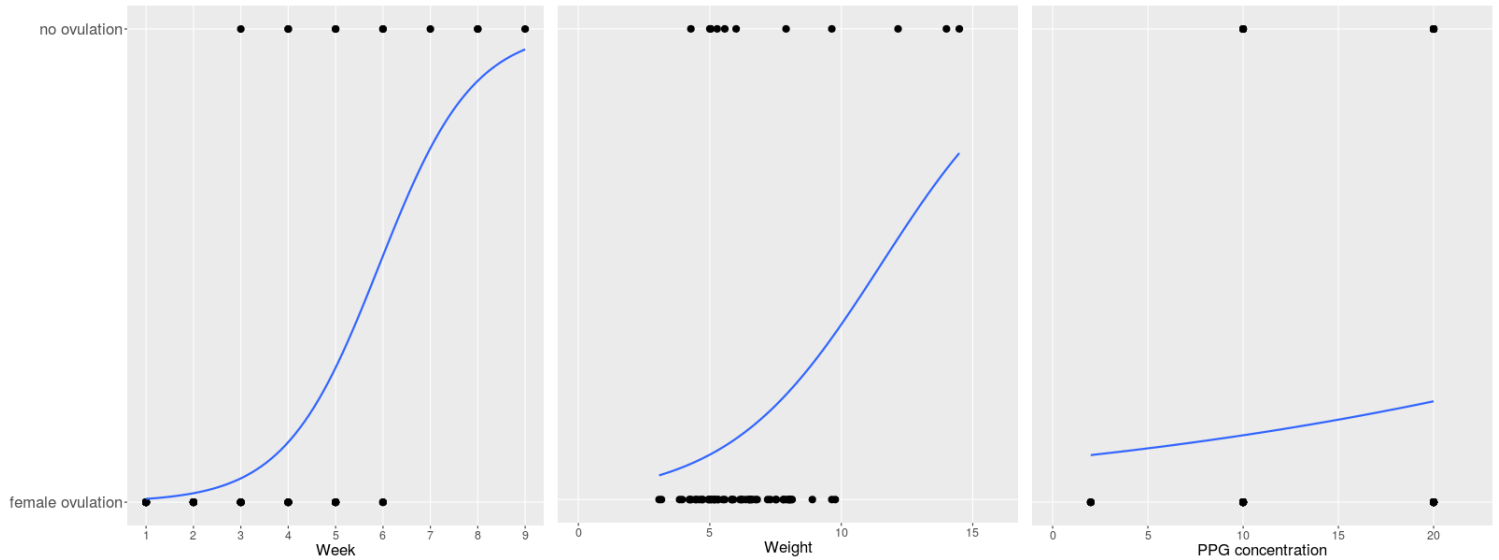


Figure S1: Female ovulation as a response variable with three explanatory variables: week (timing of treatment), weight (proxy for female age), and concentration of PPG administrated.

Experiment week ID	Time of light on	Time of light off
1	06:45	19:15
2	06:30	19:30
3	06:15	19:30
4	06:00	19:45
5	05:45	20:00
6	05:30	20:00
7	05:15	20:15
8	05:00	20:30
9	05:00	20:46

Table S1: Photoperiod scheme applied on *Umbra pygmaea* in captivity during the experimental period of induced spawning.

ID	Weight	PPGs Concentration	Eggs production	Week number
5	6	10	Y	3
12	14	10	Y	4
26	14.49	10	Y	4
48	12.16	10	Y	5
84	4.28	20	Y	5
13	5	10	Y	6
170	5.28	20	Y	6
55	5.57	20	Y	7
43	7.9	20	Y	8
66	5.04	20	Y	8
100	9.64	20	Y	9
158	6.5	2	N	1
123	3.85	2	N	1
56	5.84	2	N	1
2	6.5	2	N	1
71	6.73	2	N	1
197	5.83	10	N	1
174	7.8	10	N	1
140	8.13	10	N	1
133	4.29	10	N	1
136	5.2	10	N	1
118	9.77	20	N	1
17	8	20	N	1
69	4.7	20	N	1
74	7.8	20	N	1
194	3.07	20	N	1
170	5.3	10	N	2
129	7.5	10	N	2
111	7.2	10	N	2
91	8.1	10	N	2
92	6.8	10	N	2
21	6.5	20	N	2
159	6.6	20	N	2
165	5.9	20	N	2
148	8	20	N	2
121	5.15	20	N	2
184	4.23	10	N	3
178	6.37	10	N	3
144	5.32	10	N	3
38	5.05	10	N	3
55	5.57	20	N	3
156	8.9	20	N	3

155	7.3	20	N	3
150	8.05	20	N	3
35	3.16	20	N	3
15	4.49	10	N	4
157	4.95	20	N	4
43	7.9	10	N	4
153	5.5	10	N	4
66	5.04	20	N	4
130	5.17	20	N	4
72	6.2	20	N	4
33	6.15	20	N	4
97	3.95	10	N	5
100	9.64	10	N	5
88	6.24	10	N	5
68	7.53	10	N	5
64	8.02	20	N	5
77	4.23	20	N	5
177	4.6	20	N	5
99	6.58	20	N	5
28	4.96	10	N	6
200	4.45	10	N	6
57	4.73	10	N	6

Table S2: Data of female ID, female weight, timing of PPGs treatment, concentration of PPGs used and success of egg production as a binary factor.

Summary:

Teleost fishes, the most species-rich clade among vertebrates, employs an astonishing diversity of sex-determining (SD) mechanisms, including environmental and genetic systems. Recent studies identified many new genetic regulators in fish sexual development that lead to the notion of an 'evolutionary playground'.

This thesis project aims to provide a complete picture of the evolutionary dynamic of SD systems within a small order of teleost, *Esociformes*. In our focal species *Esox lucius*, we identified a duplicate of a member of the TGF- β family - a family which has emerged to provide crucial pathways regulating sex in teleost to be the master SD gene. We obtained functional proof of its role in SD, its interaction with environmental factors and insights on the evolutionary processes in its acquisition of a new function after duplication. While the gene is lost in another population of the same species rapidly possibly during post-glacial recolonization process, it is well conserved among different species. Meanwhile, additional transitions of SD system have also been identified in a sister *Esox* species. Moreover, in the most distant genera *Dallia* and *Umbra*, the *Esox* master SD gene is not present and we found different SD mechanisms with novel SD genes that adds additional layers of complexity to this group, which mirrors the observed high genetic plasticity in teleost SD.

Résumé:

Les téléostes, le clade possédant le plus d'espèces parmi les vertébrés, emploient une saisissante diversité de mécanismes de détermination du sexe, incluant des mécanismes génétiques et environnementaux. Des études récentes ont identifié de nombreux nouveaux régulateurs génétiques du développement sexual des poissons et ont introduit la notion de "terrain de jeux évolutif" pour qualifier ce domaine.

Dans ce projet de thèse, nous avons étudié l'ensemble des dynamiques évolutives des déterminants du sexe dans un ordre de téléostes, les *Esociformes*. Dans notre espèce focale, *Esox lucius*, nous avons identifié qu'une duplication d'un membre de la famille des TGF- β , une famille qui a émergé comme fondamentale dans la régulation du sexe chez les téléostes, agit comme gène contrôlant la détermination du sexe (MSD). Nous avons obtenu des preuves fonctionnelles du rôle de ce gène en tant que MSD ainsi que de nouvelles informations sur les processus évolutifs régissant l'acquisition de fonction d'un MSD après duplication. Ce gène a disparu dans une population de la même espèce en Amérique du Nord, malgré sa générale conservation parmi les autres *Esociformes*. Dans certains *Esociformes*, d'autres systèmes de détermination du sexe ont été identifiés. De plus, dans des clades plus distants comme *Umbra* et *Dallia*, le MSD d'*Esox* n'est pas présent, et d'autres mécanismes impliquants de nouveaux MSD ont été identifiés. Ces résultats complexifient l'histoire évolutive de la détermination du sexe dans ce groupe, qui reflète la plasticité génétique observée dans les téléostes en général.

RÉSUMÉ

Utiliser ou ne pas utiliser un gène de détermination du sexe : évolution des systèmes de détermination du sexe chez les Esociformes.

Les téléostes, le clade possédant le plus d'espèces parmi les vertébrés, emploient une saisissante diversité de mécanismes de détermination du sexe, incluant des mécanismes génétiques et environnementaux. Des études récentes ont identifié de nombreux nouveaux régulateurs génétiques du développement sexuel des poissons et ont introduit la notion de "terrain de jeux évolutif". Le but de ce projet de thèse est de donner une vue complète des dynamiques évolutives des déterminants du sexe dans un ordre de téléostes, les Esociformes. Dans notre espèce focale, *Esox lucius*, nous avons identifié une duplication d'un membre de la famille des TgfB – une famille qui a émergé en tant que fondamentale dans la régulation du sexe chez les téléostes – comme MSD (gène contrôlant la détermination du sexe). Nous avons obtenu des preuves fonctionnelles du rôle de ce gène en tant que MSD, de son interaction avec des facteurs environnementaux, ainsi que des nouvelles informations sur les processus évolutifs régissant l'acquisition de fonction après duplication. Ce gène est perdu dans une population de la même espèce en Amérique du Nord, mais il est conservé parmi les autres Esociformes. En parallèle, d'autres systèmes de détermination du sexe ont été identifiés dans des espèces proches. De plus, dans des clades plus distants comme *Umbra* et *Dallia*, le MSD d'*Esox* n'est pas présent, et d'autres mécanismes impliquants de nouveaux MSD ont été identifiés, complexifiant l'histoire évolutive des MSD dans ce groupe, qui reflète la plasticité génétique observée dans les téléostes en général.

Mots-clefs : détermination du sexe, chromosomes sexuels, AMH, Esociformes, évolution

ABSTRACT

To use or not use a Master sex determining gene: Evolution of sex determination system in the Esociformes

Teleost fishes, the most species-rich clade among vertebrates, employs an astonishing diversity of sex-determining (SD) mechanisms, including environmental and genetic systems. Recent studies identified many new genetic regulators in fish sexual development that lead to the notion of an 'evolutionary playground'.

*This thesis project aims to provide a complete picture of the evolutionary dynamic of SD systems within a small order of teleost, Esociformes. In our focal species *Esox lucius*, we identified a duplicate of a member of the TGF- β family - a family which has emerged to provide crucial pathways regulating sex in teleost - to be the master SD gene. We obtained functional proof of its role in SD, its interaction with environmental factors and insights on the evolutionary processes in its acquisition of a new function after duplication. While the gene is lost in another population of the same species rapidly possibly during post-glacial recolonization process, it is well conserved among different species. Meanwhile, additional transition of SD system have also been identified in a sister *Esox* species. Moreover, in the most distant genera *Dallia* and *Umbra*, the *Esox* master SD gene is not present and we found different SD mechanisms with novel SD genes that adds additional layers of complexity to this group, which mirrors the observed high genetic plasticity in teleost SD.*

Keywords: Sex determination, Anti-Müllerian hormone, Northern Pike, Esociformes, transition of sex determination systems, master sex-determining genes



AGROCAMPUS OUEST • Institut supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage
65 rue de Saint-Brieuc – CS84215 – F-35042 Rennes Cedex
Tél. : 02 23 48 50 00
www.agrocampus-ouest.fr

