Molecular Analysis of the Sex-Determining Region of the Y Chromosome in the Platyfish *Xiphophorus maculatus*

DISSERTATION

zur Erlangung des naturwissenschaftlichen Doktorgrades
der Bayerischen Julius-Maximilians-Universität Würzburg

vorgelegt von

Qingchun Zhou

aus

Hu Bei, V. R. China

Würzburg, March 2005

Eingereicht am:..

Mitglieder der Promotionskommission:

Vorsitzender:......**Prof. Dr. Dr. h. c. U. Scheer** ....

Gutachter :..........**Prof. Dr. M. Schartl**..……

Gutachter :..........**Prof. Dr. G. Krohne**...

Tag des Promotionskolloquiums:...

Doktorurkunde ausgehändigt am:...

Acknowledgement

First of all, I would like to express all my gratitude to my mentor Prof. Dr. Manfred Schartl for providing me the opportunity to work in his Institute, and for his warm help and friendly concern. His many theories and researches are the basis and scientific guidance of my dissertation.

I am greatly indebted to Prof. Dr. Georg Krohne for his acceptance to be an examiner of my PhD. In addition, I am grateful to Prof. Dr. Manfred Gessler for his kind help.

I would specially thank Dr. Jean-Nicolas Volff for his patience, time and untiring work, including scientific guidance, motivation, encouragement and indispensable stimulation. I would like to extend my thank to Dr. Alexander Froschauer who always warmly and patiently helped me in my work and in my life in Würzburg; Cornelia Körting who helped me a lot at the beginning of my work and sequenced for me; Anne-Marie Veith from whom I learnt the *in situ* hybridization, enjoyed with a lot of scientific discussions; Christina Schultheis and Yvonne Selz for their friendly co-operation, help and cheerful talking; Dr. Agnes Dettai for her great help in the writing of this dissertation and many basic knowledge and skills of phylogenetic analysis, and other members in Dr. Volff ´s group such as Marina Vucic, Jürgen Brandt, Patrick Bienert and Matthias Mösl.

Dr. Christoph Winkler and his group are warmly acknowledged for their help for microscope analysis.

I wish to express my appreciation to Ute Hornung, Christine Froschauer, Dr. Mariko Kondo, Robin Wacker, Hugo Schwind and Dr. Dunja Lamatsch and all of the here unnamed colleagues in the department of Physiological Chemistry I for their kindly help and the friendship.

Dr. Indrajit Nanda is greatly thanked for scientific discussions and help in the writing of the dissertation.

I also acknowledge the Biofuture program from BMBF (to J.-N. Volff) for financial support.

Last but not least, I would like to give my personal thanks to my parents and all my Chinese friends for backing and cheering me up on many occasions. Finally, my greatest acknowledgement is reserved for my dear wife, Qin.

Content

CONTENT ... **1**

ABBREVIATIONS AND CHEMICAL SYMBOLS ... **4**

ABSTRACT ... **6**

1. INTRODUCTION ... **10**

1.1 *XIPHOPHORUS* AS A MODEL ORGANISM FOR MELANOMAGENESIS ... 10

1.2 *XMRK* AND ITS PROTOONCOGENIC ANCESTOR *EGFRB* ARE LINKED TO THE SEX-DETERMINING LOCUS OF *XIPHOPHORUS MACULATUS* ... 12

1.3 SEX DETERMINATION IN MAMMALS ... 13

1.3.1 SRY is the testis–determining gene in mammals ... *13*

1.3.2 Molecular dissection of the human Y chromosome ... *14*

1.4 SEX DETERMINATION IN FISHES ... 16

1.5 SEX DETERMINATION IN THE MEDAKA (*ORYZIAS LATIPES*) ... 18

1.6 SEX DETERMINATION IN THE PLATYFISH *XIPHOPHORUS MACULATUS* ... 21

1.7 GENOMIC PLASTICITY IN THE SEX DETERMINATION REGION OF *XIPHOPHORUS MACULATUS* ... 23

1.7.1 Gene deletions and duplications in the sex determination region of X. maculatus ... *24*

1.7.2 Accumulation of transposable elements and other repetitive sequences in the sex determination region of X. maculatus ... *25*

1.8 AIMS OF THIS WORK ... 28

2. MATERIALS ... **29**

2.1 FISHES ... 29

2.2 FISH CELL LINES ... 29

2.3 COMPETENT CELLS AND VECTORS ... 30

2.4 ENZYMES ... 30

2.5 MEDIA, ANTIBIOTICS AND SOLUTIONS ... 30

2.5.1 Media ... *30*

2.5.2. Antibiotics ... *31*

2.5.3. Solutions ... *31*

2.6 OLIGONUCLEOTIDES ... 32

2.7 INSTRUMENTS, HARD- AND SOFTWARES ... 33

2.7.1 Instruments ... *33*

2.7.2 Hard- and softwares ... *33*

3. METHODS ... **34**

3.1 DNA EXTRACTION METHODS ... 34

3.1.1 BAC DNA extraction methods ... *34*

3.1.2 Genomic DNA extraction methods ... *35*

3.1.3 Plasmid DNA extraction methods ... *36*

3.2 RNA EXTRACTION METHODS ... 36

3.3 DNA DIGESTION, LIGATION AND DEPHOSPHORYLATION ... 37

3.4 TRANSFORMATION OF *E. COLI* ... 37

3.5 PHENOL-CHLOROFORM EXTRACTION ... 39

3.6 DNA amplification by Polymerase Chain Reaction (PCR) 39
3.6.1 General PCR method *39*
3.6.2 RT-PCR *40*
3.6.3 Linker-PCR *40*
3.7 Electrophoresis techniques 41
3.7.1 DNA agarose gel electrophoresis *42*
3.7.2 Pulsed field gel electrophoresis *42*
3.8 DNA isolation from agarose gels 43
3.9 ´´Shot-gun`` and ´´transposon tagging`` sequencing 43
3.9.1 ´´Shot-gun`` sequencing *43*
3.9.2 ´´Transposon tagging`` sequencing *44*
3.10 Automated DNA sequencing 45
3.11 Southern blot hybridization 45
3.12 Chromosome walking 47
3.13 RNA whole-mount *in situ* hybridization 48
4. RESULTS 52
4.1 Chromosome walking on the sex determining region of the Y chromosome of *Xiphophorus maculatus* 52
4.1.1 Construction of the ps-criptY BAC contig of Xiphophorus maculatus *52*
4.1.2 Order of BAC clones in the ps-criptY contig *53*
4.1.3 Extension of the ps-criptY BAC contig *53*
4.1.3.1 Extension from the left side of the *ps-criptY* contig 54
4.1.3.2 Extension from the right side of the *ps-criptY* contig 58
4.1.3.3 The *ps-criptY* contig covers more than 1000 kb of the Y chromosome of *X. maculatus* 64
4.2 Sequence analysis of the *ps-criptY* contig of *X. maculatus* 67
4.3 Identification and analysis of gene candidates in the *ps-criptY* contig 67
4.3.1 Swimy, a new gene exclusively expressed in spermatogonia of X. maculatus adult testis *68*
4.3.1.1 *Swimy* encodes a putative DNA binding protein 68
4.3.1.2 *Swimy* is expressed in some spermatogonia of *X. maculatus* adult testis 70
4.3.1.3 *Swimy* is located in the *ps-criptY* contig of *X. maculatus* 72
4.3.1.4 *Swimy* distribution in the genus *Xiphophorus*, in the family Poeciliidae and in other ray-finned fishes 73
4.3.1.5 Sequences related to *swimy* in other fish 74
4.3.2 Fredi encodes a novel helix-turn-helix protein *76*
4.3.2.1 Expression pattern and location of *fredi* in *X. maculatus* 76
4.3.2.2 Is *fredi* a fish-specific gene? 77
4.3.3 A DNA polymerase type B related to organellar and viral genes is located in the sex determination region of X. maculatus *78*
4.3.3.1 A DNA polymerase type B gene in *X. maculatus* 78
4.3.3.2 The DNA_pol_B_2 domain of DNA polymerase type B 79
4.3.3.3 Location and expression analysis of *DNA polymerase type B* 81
4.3.3.4 Distribution of *DNA polymerase type B* in ray-finned fishes 83
4.3.4 Other gene candidates in the ps-criptY contig *84*
4.3.4.1 A multicopy *msh2* (pseudo) gene on the Y chromosome of *X. maculatus* 85
4.3.4.2 Sequence analysis of *ps-cript1* on the X chromosome and comparison with *ps-criptY* 86
4.3.4.3 Sequence analysis of *mc4rP, mc4rQ* and *mc4rR* 87
4.4 Structural analysis of the *ps-criptY* contig of *X. maculatus* 90

4.4.1 Genomic structure and instability of the ps-criptY contig *90*
4.4.2 Synteny analysis of the ps-criptY contig *91*
4.5 *HELITRON*, A ROLLING CIRCLE TRANSPOSON, ON THE SEX CHROMOSOMES OF *X. MACULATUS* 93
4.5.1 Xiphophorus helitrons *93*
4.5.2 At least four helitrons of X. maculatus are located in the sex determining region *95*
4.5.3 Transcripts of Xiphophorus helitrons *97*
4.5.4 Helitron distribution in the genus Xiphophorus, in the family Poeciliidae and in other ray-finned fishes *98*
4.5.5 The medaka (Oryzias latipes) helitron *99*
4.5.6 Phylogenetic analysis of helitrons *99*

5. DISCUSSION 101
5.1 PLATYFISH *X. MACULATUS* AS A MODEL TO STUDY EARLY SEX CHROMOSOME DIFFERENTIATION 101
5.1.1 Why choosing X. maculatus to study the mechanism of the sex determination? *101*
5.1.2 X. maculatus is suitable for study on early sex chromosomes differentiation *102*
5.1.3 BAC contigs close to the SD locus were set up to isolate the master SD gene in X. maculatus by positional cloning *102*
5.2 WHERE IS THE *PS-CRIPTY* BAC CONTIG? 103
5.3 GENOMIC PLASTICITY IN THE SEX DETERMINATION REGION AND INITIAL DIFFERENTIATION OF SEX CHROMOSOMES IN VERTEBRATES 104
5.3.1 Genomic plasticity in the sex determination region *104*
5.3.2 The initial differentiation of sex chromosomes in vertebrates *106*
5.4 *SWIMY* AND *FREDI* ARE CANDIDATES FOR SEX DETERMINATION OR SEX DIFFERENTIATION GENES IN *X. MACULATUS* 108
5.4.1 Swimy is exclusively expressed in some spermatogonia of X. maculatus *108*
5.4.2 Fredi in X. maculatus *110*
5.5 GENE CANDIDATES FOR PIGMENTATION IN THE SEX DETERMINATION REGION OF *X. MACULATUS* 110
5.6 CONSERVED SYNTENIES BETWEEN THE SEX DETERMINATION REGION OF *X. MACULATUS*, HUMAN AND OTHER FISHES 111
5.7 *DNA POLYMERASE TYPE B* IN FISH 112
5.7.1 DNA polymerase type B domain in fish *112*
5.7.2 DNA polymerase type B is also a novel W-specific marker in X. maculatus Usumacinta, along with swimy *113*
5.8 DNA TRANSPOSON *HELITRONS* IN *X. MACULATUS* 113
5.9 PERSPECTIVES 115

6. APPENDIX 117
6.1 TABLES 117
6.2 PROTEIN SEQUENCES OF SOME GENE CANDIDATES AND ALIGNMENT OF *HELITRONS* 122

7. REFERENCES 126

PUBLICATIONS 137

LEBENSLAUF 139

ERKLÄRUNG 140

Abbreviations and chemical symbols

AGRP	Agouti-related protein
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BC	Backcross
bp	Base pair(s)
BSA	Bovine serum albumin
CaCl2	Calcium chloride
cDNA	Complementary DNA
Ci	Curie (equal to 37 gigabecquerel)
CIAP	Calf intestinal alkaline phosphatase
cpm	Counts per minute
CRIPT	Cysteine-rich interactor of PDZ3
C-Terminus	Carboxylic terminus
CTP	Cytosine triphosphate
DEPC	Diethyl pyrocarbonate inhibitor of RNase
DEPC-H2O	Distilled water with diethylpyrocarbonat
dH2O	Distilled water
Dmrt	Doublesex and mab-3 related transcription factor
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleoside triphosphate
DTT	1,4-Dithiothreitol
e.g.	for example
EDTA	Ethylenediaminetetraacetic acid
egfrb	Epidermal growth factor receptor b
et al.	and others
EtBr	Ethidium bromide
EtOH	Ethanol
GTP	Guanosine triphosphate
h	Hour
HCl	Hydrochloric acid
Hepes	2-[4-(2-Hydroxyethyl)-1-piperazine]ethanesulfonic acid
IPTG	Isopropyl-beta-D-thiogalactopyranoside
K2HPO4	Potassium phosphate, dibasic; Dikalium hydrogen phosphate
Kb	Kilobase (s) ; kilobase pairs
KCl	Potassium chloride
KH2PO4	Potassium phosphate, monobasic; Kalium dihydrogen phosphate
L.	Liter
LB	Luria-Bertani medium
lof	loss-of-function
LTR	long terminal repeat
M	Mole
Mc4r	Melanocortin receptor type 4
Mdl	Macromelanophore-determining locus in *Xiphophorus*
MEM	Minimum essential medium
mg	Milligram
MgCl2	Magnesium chloride
min	Minutes
ml	Milliliter
mM	Millimole
mRNA	Messenger RNA

MW	Molecular weight
Na2CO3	Sodium carbonate
NaCl	Sodium chloride
NaOAc	Sodium acetate
NaOH	Sodium hydroxide
ng	Nanogram
nmol	Nanomole (=10^{-9} mole)
nt	Nucleotide
N-Terminus	Amino-terminus
ORF	Open reading frame
P	Puberty; puberty locus in *Xiphophorus*
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PGC	Primordial germ cell
pmol	Picomole (=10^{-12} mole)
PMSF	Phenylmethylsulphonyl fluoride
ps-x	Pseudogene Gene x
R	Tumor regulator locus in *Xiphophorus*
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rpm	Rotations per minute
RT	Reverse transcriptase
RY	Red-yellow; a pigmental locus in *Xiphophorus*
SD	Sex-determining gene
Sd	Spotted dorsal; a macromelanophore pattern in *Xiphophorus*
SDS	Natriumdodecylsulfate
spp.	Species
SRY	Sex-determining region on Y chromosome
SSC	Sodium chloride and sodium citrate buffer
Tab.	Table
TEMED	N,N,N',N' -Tetramethylethylene diamine
T_m	Melting temperature of DNA or RNA
TTP	Thymidine triphosphate
µg	Microgram
µl	Microliter
µmol	Micromole (=10^{-6} mole)
U	Units (measure for enzyme activity)
UTP	Uridine triphosphate
UTR	Untranslated region
vol	Volume
WT	Wildtype
Xgal	5-Brom-4-chlor-3-indolyl-beta-D-galactoside
Xmrk	Xiphophorus melanoma receptor kinase

Abstract

A large variety of sex determination systems have been described in fish. However, almost no information is available about sex determination in the classical fish models, the zebrafish *Danio rerio* and the pufferfish *Takifugu rubripes*. A DNA-binding protein gene called *dmrt1bY* (or *DMY*) has been recently described as an outstanding candidate for the primary sex-determining gene in the medaka fish *Oryzias latipes*. But this gene is not the universal master sex-determining gene in teleost fish, since *dmrt1bY* is not found in most other fishes. Hence, other fish models need to be examined including the platyfish *Xiphophorus maculatus*. *Xiphophorus maculatus* has three types of sex chromosomes (X, Y and W; females are XX, WX or WY; males are XY or YY). Its gonosomes are at an early stage of differentiation. The sex-determining locus on the sex chromosomes is flanked by two receptor tyrosine kinase genes, the *Xmrk* oncogene and its protooncogenic progenitor gene *egfrb*, which both delimit a region of about 0.6 centiMorgans. This situation should allow the positional cloning of the sex-determining gene (*SD*) of the platyfish.

For this purpose, Bacterial Artificial Chromosome (BAC) contigs were assembled from a BAC library of XY males constructed in our laboratory, using the oncogene *Xmrk*, *egfrb,* as well as a Y-specific pseudogene called *ps-criptY* as starting points. The *ps-criptY* sequence was found to be closely linked to the *SD* gene, since no recombination was observed between *SD* and *ps-criptY* in more than 400 individuals tested. Two major BAC contigs for the X chromosome (about 2.5 Mb) and three major BAC contigs for the Y chromosome (about 3.5 Mb) were built up and analyzed by strategic sequencing. These are some of the largest contigs ever assembled for the sex chromosomes of a non-mammalian vertebrate species. The molecular analysis of the *ps-criptY* contig was the major objective of this work.

The Y-specific *ps-criptY* contig has been extended over 1 Mb in this work with 58 identified molecular markers. Approximatively 700 kb of non-redundant sequences has been obtained from this contig by strategic sequencing.

Numerous Y-linked markers from the contig including *ps-criptY* were also detected on the X chromosome. Nevertheless, major structural differences were observed between the X and Y chromosomes. Particularly, a large region, which is present at one copy on the X chromosome and contains several candidate genes, was found to be duplicated on the Y chromosome. Evidence for an inversion in the sex-determining region and for the Y-specific accumulation of a repeated sequence called *XIR* was also obtained. Such events might correspond to an

initiation of differentiation between both types of gonosomes.

Accumulation of transposable elements was also observed in the *ps-criptY* contig. A DNA transposable element, *helitron*, was isolated from the sex-determining region of *X. maculatus*. Three copies of *helitron* are located on the *ps-criptY* contig and one copy on the X-linked contig (*helitron* has roughly 15 copies per haploid genome). No in-frame stop codon, truncation or intron was found in these four copies, which present high nucleotide identities to each other. This suggests that *helitron* elements might be active or have been recently active in *X. maculatus*. A consensus open reading frame of *helitron* was also assembled from medaka (*Oryzias latipes*) genomic sequences.

Two candidate genes from the *ps-criptY* contig are located on the W chromosome in the *X. maculatus* Usumacinta strain (female heterogamety). These markers show the relationship between the different types of gonosomes and allow comparing male and female heterogameties in the platyfish.

Several gene candidates have been identified in the *ps-criptY* contig. However, some of them such as *msh2*, *cript, igd1-2* and *acr* probably correspond to pseudogenes. Interestingly, a novel gene, called *swimy*, is exclusively expressed in spermatogonia of adult testis. *Swimy* is a gene encoding a DNA-binding protein with several putative DNA-binding domains. The data suggest that *swimy* is a very promising candidate for the master *SD* gene. Another novel gene, which is called *fredi* and encodes a novel helix-turn-helix protein, is predominately expressed in adult testis and currently under scrutiny.

There is no doubt that the master *SD* gene of *X. maculatus* will be identified by positional cloning. Further molecular analysis of the contigs built in this work will shed new light on the molecular mechanism of sex determination and the evolution of sex chromosomes in fish.

Zusammenfassung

In Fischen wurde eine große Anzahl Geschlechtsbestimmungssysteme beschrieben. Allerdings gibt es kaum Informationen über die Geschlechtsbestimmung der klassischen Modellorganismen, des Zebrafisches *Danio rerio* und des Pufferfisches *Takifugu rubripes*. Das für ein DNA-bindendes Protein kodierende Gen *dmrt1bY* (oder *DMY*) wurde kürzlich als ein herausragender Kandidat für das primäre Geschlechts bestimmungsgen im Medaka *Oryzias latipes* beschrieben. Dieses Gen ist jedoch nicht das universelle Geschlechtsbestimmungsgen der echten Knochenfische (Teleostei), da *dmrt1bY* in den meisten anderen Fischen nicht identifiziert werden konnte. Deshalb dienen andere Fische wie der Platy *Xiphophorus maculatus* als Modelle.

Xiphophorus maculatus besitzt drei Geschlechtschromosomen X, Y und W in einem frühen Stadium der Differenzierung (Weibchen sind XX, WX oder WY, Männchen XY oder YY). Der geschlechtsbestimmende Locus wird flankiert von zwei Rezeportyrosinkinase-Genen, dem Onkogen *Xmrk* und seinem Vorläufer, dem Proto-onkogen *egfrb*. Sie markieren eine Region von ca. 0.6 centiMorgan, was die positionelle Klonierung des geschlechtsbestimmenden Gens *SD* des Platys erlauben sollte.

Zu diesem Zweck wurden BAC- (Bacterial Artificial Chromosome-) Contigs der X- und Y-Chromosomen aus einer genomischen Bibliothek erstellt, wobei *Xmrk*, *egfrb* und das Y-spezifische Pseudogen *ps-criptY* als Startpunkte gewählt wurden. *Ps-criptY* ist eng an *SD* gekoppelt, wie die Analyse von über 400 Individuen zeigte. Zwei BAC-Contigs des X-Chromosoms (ca. 2.5 Mb) und drei BAC-Contigs des Y-Chromosoms (ca. 3.5 Mb) wurden erstellt und durch strategisches Sequenzieren analysiert. Dies sind einige der größten geschlechtschromosomalen Contigs, die je für eine Wirbeltierart außerhalb der Säuger erstellt wurden. Der Aufbau und die molekulare Analyse des BAC-Contigs um *ps-criptY* war Hauptziel dieser Arbeit.

Dieses Y-spezifische Contig wurde durch die Analyse von 58 molekularen Markern in dieser Arbeit um über eine Megabase erweitert. Fast 700 kb nicht-redundanter Sequenz konnten durch strategisches Sequenzieren analysiert werden.

Obwohl eine Vielzahl von Markern des Y-Chromosoms inklusive *ps-criptY* ebenfalls auf dem X-Chromosom detektiert wurden, konnten große strukturelle Unterschiede der Geschlechtschromosomen nachgewiesen werden. Im besonderen konnte die Duplikation einer großen Region des X-Chromosoms, die mehrere Genkandidaten enthält, auf dem Y-

Chromosom gezeigt werden. Außerdem konnte die Inversion dieser Region inklusive einer Akkumulation der repetitiven Sequenz *XIR* konnte belegt werden. Solche Ereignisse entsprechen einer beginnenden Differenzierung zwischen heteromorphen Geschlechtschromosomen.

Außerdem wurde die Akkumulation transposabler Elemente im *ps-criptY*-Contig beobachtet. Eines dieser Elemente, *helitron*, konnte aus der geschlechtsbestimmenden Region von *X. maculatus* isoliert werden. Von den vier Kopien der geschlechts-bestimmenden Region (3 Kopien im *ps-criptY*-Contig des Y-Chromosoms, 1 Kopie im *Xmrk*-Contig des X-Chromosoms, 15 im gesamten Genom) enthielt keine ein vorzeitiges Stop-Codon, Unterbrechung oder sonstige Störung des offenen Leserasters. Dies könnte darauf hinweisen, dass die *helitron*-Elemente in *X. maculatus* noch aktiv sind oder bis vor kurzem waren. Ein Konsensus-ORF des *helitron*-Elements konnte auch aus Datenbank-Sequenzen des Medaka (*Oryzias latipes*) erstellt werden.

Zwei Genkandidaten des *ps-criptY*-Contigs konnten auch auf dem W-Chromosom von *X. maculatus* (Rio Usumacinta, weibliche Heterogametie) nachgewiesen werden. Diese Marker zeigen die enge Beziehung zwischen den Geschlechtschromosomen des Platys und ermöglichen eine detaillierte Untersuchung von männlicher und weiblicher Heterogametie im Platy.

Verschiedene Genkandidaten konnten im *ps-criptY*-Contig identifiziert werden. Allerdings zeigte die Analyse, dass einige davon, wie *msh2*, *cript, igd1-2* und *acr* Pseudogene darstellen. Interessanterweise ist eines dieser Gene, *swimy*, ausschließlich in Spermatogonien exprimiert. Dieses neue Gen kodiert für ein Protein mit mehreren DNA-bindenden Domänen. Diese Daten machen *swimy* zu einem vielversprechenden Kandidaten für *SD*. Ein weiteres neues Gen, *fredi*, kodiert für ein Helix-Loop-Helix Protein, ist ebenfalls im adulten Hoden exprimiert und wird gerade eingehender analysiert.

Zweifellos wird das geschlechtsbestimmende Gen in *X. maculatus* durch positionelle Klonierung identifiziert werden. Weitergehende molekulare Analysen der geschlechtschromosomalen Contigs werden Licht in die molekularen Grundlagen der Geschlechtsbestimmung und die Evolution der Geschlechtschromosomen in Fischen bringen.

1. Introduction

1.1 *Xiphophorus* as a model organism for melanomagenesis

Xiphophorus are Poeciliid fish that inhabit freshwater drainages in eastern Mexico, Guatemala, Belize and Honduras. Twenty-three species are described presently and are roughly distributed within three broad taxonomic groups: the northern and southern swordtails and the platyfishes (Meyer and Schartl, 2002). Fish from the genus *Xiphophorus* has been used for the study of sex determination, formation of spontaneous and hereditary melanoma cancer, behavioral ecology, DNA repair, genomics as well as transposable elements (Froschauer *et al*., 2001; Kazianis *et al*., 2004a, and references therein).

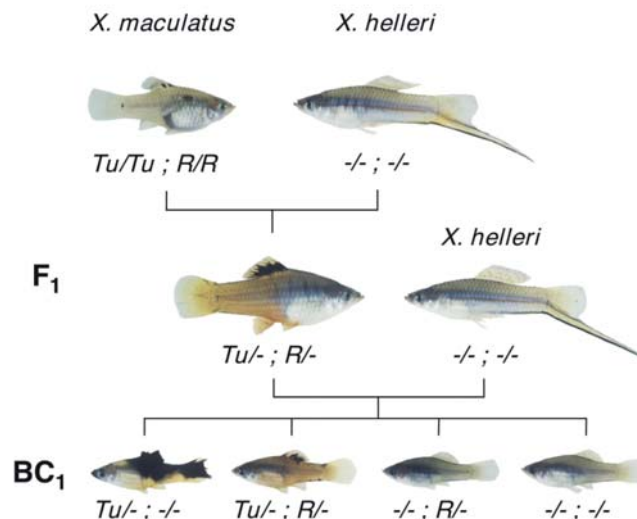


Fig. 1. Classical crossing experiment leading to hybrids that develop malignant melanoma. For detail see text. *Tu* represents the tumour locus; *R* represents the tumour suppressor locus. Note that the parental *Xiphophorus maculatus* carries macromelanophore pigment spots in the dorsal fin that develop into melanoma in the hybrids (Schartl and Schartl, 1990).

The study of melanoma formation in *Xiphophorus maculatus* can be dated back to 1920s. It was reported roughly 80 years ago that certain hybrids, which derived from crossing between the platyfish *Xiphophorus maculatus* and the unspotted swordtail *Xiphophorus hellerii*, develop malignant melanoma (Gordon, 1927; Kosswig, 1928). Melanoma development in the hybrids was defined as a regulatory imbalance between a dominant tumor-inducing locus (*Tu*) and a *Tu*-repressing regulatory locus (*R*, aka *Diff* or R_{Diff}), which are sex chromosomal and autosomal, respectively (Schartl and Schartl, 1990; Schartl, 1988, and references therein).

In the typical Gordon-Kosswig crossing experiment (Fig, 1), a female *Xiphophorus maculatus* which carries macromelanophore pigmentation spots in the dorsal fin is mated with *X. hellerii* which does not carry the corresponding macromelanophore locus and exhibits the uniform wild-type pigmentation. The F1 hybrid is heterozygous for both *Tu* and *R* loci, and shows enhancement of the spotted dorsal fin phenotype. Backcrossing of the F1 hybrid to *X. hellerii* produces offspring (BC1) that segregate into 50% of fish not having the *Tu* phenotype similar to the *X. hellerii* parental strain and 50% carrying *Tu* and developing melanoma. Half of the *Tu*-bearing BC1 offspring is heterozygous for *Tu*, but devoid of *R* [Tu/-; -/-]. In this case, *Tu* is out of control by *R* in the pigment cell lineage. This results in the overexpression of *Tu* and the formation of highly malignant, invasive and exophytic melanomas that are lethal to the individuals. The other half of *Tu*-bearing BC1 offspring is heterozygous for both *Tu* and *R* loci [Tu/-; *R*/-]. Like F1-hybrids, these genotype offspring develop noninvasive, superficially spreading neoplastic pigment cell lesions, which are apparently nonmalignant (so called ´benign melanoma`) (Fig. 1). This is assumed to be due to the loss of one copy of *R*.
The *Xiphophorus* melanoma receptor kinase gene (*Xmrk*) behaves as a dominant oncogene inducing the formation of melanoma and is encoded by *Tu* (Schartl, 1995). *Xmrk* is located at the subtelomeric region of the *X. maculatus* sex chromosomes (Nanda *et al*., 2000a). The molecular nature of the autosomal regulator locus *R* has not been identified so far, but it might correspond to a member of the *CDKN2* gene family (Kazianis *et al*., 1998; Nairn *et al*., 1996).
Xmrk is expressed at low levels in eyes, brain and gills, barely detectable in skin, muscle and kidney of the wild-type *X. maculatus* and not expressed in liver (Schartl *et al*., 1999). In certain hybrids, *Xmrk* is overexpressed a particular type of pigment cells, the macromelanophores.
Identification and functional analysis of *Xmrk* loss-of-function mutations further demonstrated the oncogenic potential of *Xmrk* (Schartl, 1988; Schartl, 1990; Schartl *et al*., 1999; Wittbrodt *et al*., 1989). *Xmrk* loss-of-function mutants were isolated from backcrossing-experiments in the laboratory. They contained the *X. maculatus* X chromosome normally bearing *Xmrk* in the genetic background of *X. hellerii*, but could not develop malignant melanoma (Wittbrodt *et al*., 1989). In the lof-1 (loss of function-1) mutant, the wild-type *Xmrk* was disrupted by a retrotransposable element called *Tx-1*. *Tx-1* was assigned to the class of non-autonomous long terminal repeat (LTR) retrotransposons and was recently active in *X. maculatus.* Insertion of *Tx-1* caused the disruption of the *Xmrk* oncogene, subsequently encoding a truncated putative protein lacking the carboxyterminal domain of the receptor tyrosine kinase. *Tx-1* is presents at

> 50-100 copies per genome in *Xiphophorus* and was not detected outside of the family of Poeciliidae. In the lof-2 mutant, the entire *Xmrk* oncogene has been deleted (Schartl *et al*., 1999).

1.2 *Xmrk* and its protooncogenic ancestor *egfrb* are linked to the sex-determining locus of *Xiphophorus maculatus*

Xmrk is a gene encoding a receptor tyrosine kinase from the epidermal growth factor receptor (EGFR) family. In contrast to worms and flies, where only one *egfr*-like gene has been described, four genes are present in mammals: *EGFR/ErbB1*, *HER2/ErbB2*, *HER3/ErbB3* and *HER4/ErbB4.* Probably due to an event of genome duplication having occurred early in the ancestral lineage leading to the present-day ray-finned fishes, fish have generally at least seven *egfr*-like genes, which include two *egfr* (*egfra* and *egfrb*), one *erbB2*, two *erbB3* (*erbB3a* and *erbB3b*) and two *erbB4* (*erbB4a* and *erbB4b*) (Volff and Schartl, 2003; Gomez *et al*., 2004).

The *Xmrk* oncogene is an eighth much recent *egfr*-related gene, identified in certain species of the genus *Xiphophorus* (Schartl, 1995; Wittbrodt *et al*., 1989). *Xmrk* was generated from *egfrb* (formerly called INV-*Xmrk*) by a located event of gene duplication at least 5 million years ago before the divergence of most *Xiphophorus* species. *Xmrk* was not detected outside of the genus (Adam *et al*., 1993; Weis and Schartl, 1998; Volff and Schartl, 2003).

The overexpression of *Xmrk* in certain pigment cells has been explained by the presence of a sequence directly adjacent to the 5` untranslated region of *Xmrk,* but not of *egfrb.* This sequence is derived from a repeated element called the *D* locus (Adam *et al*., 1993; Fornzler *et al*., 1996).

Xmrk and *egfrb* were both located in the subtelomeric region of the long arm of the X and Y chromosomes of *X. maculatus* where they are separated by about 1 Mb, suggesting that *Xmrk* was formed by intrachromosomal segmental duplication (Gutbrod and Schartl, 1999; Nanda *et al*., 2000a).

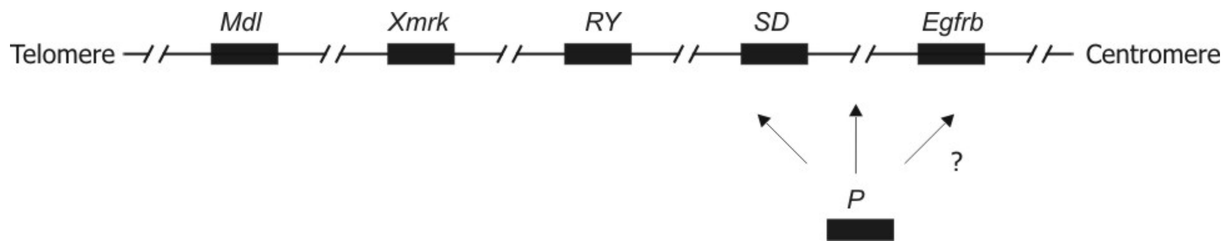


Fig. 2. Putative gene order in the sex-determining region of the *X. maculatus* sex chromosomes according to Gutbrod and Schartl (1999). *Mdl* represents the macromelanophore-determining locus, *Xmrk* the *Xiphophorus* melanoma receptor kinase oncogene, *RY* the red-yellow pigment locus, *SD* the sex-determining locus, *egfrb* the protooncogenic ancestor of *Xmrk, P* the puberty locus. The position of the *P* locus has not been precisely determined.

Interestingly, both *Xmrk* and *egfrb* genes are closely linked to the master sex determining (*SD*) locus of the platyfish on the X and Y chromosomes (Schartl, 1990). The putative gene order in the sex determination region has been established by analysis of sex chromosomal crossovers (Gutbrod and Schartl, 1999; Fig. 2). The melanoma phenotype is determined by the *Xmrk* oncogene in combination with the macromelanophore-determining locus (*Mdl*). Macromelanophores are large melanin-containing pigment cells that are the precursor cells of melanoma in *Xiphophorus*. The *Mdl* locus is also located in this region, but not identified so far at the molecular level. It contains the genetic information determining the onset and location of both macromelanophore patterns and melanomas as well as the severity of tumors (Weis and Schartl, 1998). Other gene loci located in this subtelomeric region are the *RY* (red/yellow) locus, which is responsible for red, brown, orange and yellow pigmentation patterns in the iris, on the body and on the fins, as well as the *P* (puberty) locus, which influences the onset of sexual maturation of the fish (Kallman and Schreibman, 1973).

1.3 Sex determination in mammals

1.3.1 *SRY* is the testis–determining gene in mammals

Most mammals have an XX/XY sex determination system, in which a gene on the Y chromosome induces testis development, the first step in the male developmental pathway. The sex-determining region Y gene (*SRY*) has been cloned from the sex-determining region of the human Y chromosome (Sinclair *et al*., 1990) and identified to act as the testis-determining gene by mutation analysis (Berta *et al*., 1990; Jager *et al*., 1990) and transgenesis (Koopman *et al*., 1991). The *SRY* gene encodes a transcription factor that is a member of the HMG (high mobility group)-1/-2 family of DNA binding proteins, is located on the short arm of the Y chromosome and maps to the locus-Yp11.3. The group of the HMG-1/-2 box proteins shows

both sequence-specific and DNA structure-specific binding (Ferrari *et al*., 1992; Harley *et al*.). The main function of the SRY protein is to initiate male sex determination as a transcriptional activator, which regulates a genetic switch in male development. The SRY HMG-box recognizes DNA by partial intercalation in the minor groove of the DNA helix. The SRY protein then binds to the DNA and distorts its shape (an active structure of the SRY protein). In doing so, it regulates genes that control the development of the testis.

Although *SRY* has been identified in both placental and marsupial mammals (Foster *et al*., 1992; Sinclair *et al*., 1990), this gene is absent from monotremes (egg-laying mammals) (Graves, 2002) and has been lost in at least two mole voles (Just *et al*., 1995). The wood lemming has three sex chromosomes, an X, a Y, and a modified X*, which functions to suppress the action of the Y (Fredga, 1988). The duck-billed platypus, which does not contain *SRY*, has been recently reported to have five Y-like chromosomes ($Y_{1\text{-}5}$) and five X-like chromosomes ($X_{1\text{-}5}$) in males as well as 10 X-like chromosomes in females (Gruetzner *et al*., 2004; Rens *et al*., 2004). Interestingly, the largest X chromosome (X_1) has homology to the human X chromosome and another X chromosome (X_5) has homology to the bird Z sex chromosome. The *DMRT1* gene (encoding Doublesex and Mab-3-related transcription factor 1), which is located on the bird Z chromosome and corresponds to the sex-determining gene in a fish, is found on the platypus fifth X chromosome (X_5). A mixture of mammalian and bird-style features (a mammalian-style X_1Y_1 pair and a bird-style X_5Y_5 pair) coexists in the platypus genome, suggesting an evolutionary link between mammal and bird sex chromosome systems. Whereas, gene mapping of the chicken Z chromosome has shown a conserved synteny with the human chromosome 9 including the male regulatory gene *DMRT1* (Nanda *et al*., 2000b).

1.3.2 Molecular dissection of the human Y chromosome

Humans have an XX/XY sex determination system. The roughly 60 Mb-large Y chromosome and the 160 Mb-large X chromosome are morphologically distinguishable. There are relatively short pseudoautosomal regions at both ends of the Y chromosome that are homologous to the corresponding regions of the X chromosome, reflecting the frequent exchange of DNA between these regions in male meiosis (Burgoyne, 1982). However, more than 95% region of the Y chromosome is constituted by a male specific region (MSY), consisting of 23 megabases (Mb) of euchromatin, which contains most of the Y-chromosomal genes, and about 30 Mb of heterochromatin, which is rich in repetitive DNA.

Recently, the sequence of the 23 Mb euchromatic segment in the MSY has been completed (Skaletsky *et al*., 2003). There are 156 transcription units identified in this euchromatic region, comparing to the roughly 1000 transcription units on the 160 Mb of the X chromosome. Moreover, only 78 of the 156 transcription units seem to encode proteins that belong to 27 gene families. The density of protein-encoding genes is low in the MSY (~10% of the genome-average density). These 78 protein-encoding genes fall into two categories. The first category comprises 27 genes with homology to genes on the X chromosome, betraying the common origin of these two types of sex chromosomes. 13 genes within this category have degenerated into pseudogenes. The remaining 14 active singe-copy genes are expressed in many different tissues. *SRY* belongs to this group. The second category consists of nine Y-linked gene families. These genes are organized into repeats of two or more units and show testis-specific expression. Two of these families have been formed by transposition from the X chromosome, and the others have been transposed from autosomes. Furthermore, the recent human-specific transposition of a 3.4 Mb segment from the X to the Y chromosome has added only two genes. An inversion of part of the short arm of the Y chromosome subsequently split this transposed segment into two non-contiguous blocks. How human sex chromosomes acquired their peculiar gene content has been well reviewed (Vallender and Lahn, 2004).

Surprisingly, the human Y chromosome repeats are organized as palindromes that read the same sequences on both strands of the DNA double strands, with two ´´arm`` stretching out from a central point of mirrored symmetry (Skaletsky *et al*., 2003). The most prominent features are eight massive palindromes, comprising 25% of the Y euchromatin and their arms ranging from 9 kb to 1.45 Mb, at least six of which contain testis genes. Such palindromes have also been found on the Y chromosome in chimpanzees and gorillas (Rozen *et al*., 2003). Some palindromes of chimpanzee are older than the minimal divergence between human and chimpanzee, suggesting that palindromes are not very recent in origin. These palindromes are proposed to result from concerted evolution driven by abundant gene conversion, which is the non-reciprocal transfer of sequence from one homologous sequence to another (Jackson and Fink, 1981; Szostak *et al*., 1983). The very little sequence divergence (< 0.5%) between these palindrome arms suggests ongoing gene conversion on the Y chromosome, indicating that this region is very clearly recombining (gene conversion). This is the reason why the term ´´MSY`` is used now to denote the region once referred to as the NRY, for non-recombining region of the Y chromosome. These palindromes might protect essential genes against the

degeneration that is the inevitable consequence of haploidy, and to enhance male fertility through the accumulation of most testis-specific genes within the palindromes (Skaletsky *et al*., 2003).

Taken together, the human males are likely to keep on their fantastic genetic make-up for the foreseeable evolutionary future (Charlesworth, 2003).

1.4 Sex determination in fishes

In several flies and in the nematode *Caenorhabditis elegans*, the sex-determining genes and mechanisms are known and reasonably well understood (Capel, 1998, and references therein). In mammals, a few of the key components of the sex determination cascade have been identified. Their position in the hierarchy of genes and their modes of interaction are still under investigation (Capel, 1998; Knower *et al*., 2003, and references therein). However, both the *Sxl* gene of *Drosophila* and *xol* gene of *C. elegans*, which are at the top of the sex determination cascade in these organisms, have no orthologues in more divergent species (Marin and Baker, 1998; Wilkins, 1995; Zarkower, 2001). Similarly, *SRY* has not been found in monotremes, birds, reptiles, amphibia, and fish. Thus, there appears to be a large gap between the mechanisms of sex determination in invertebrates (nematodes and flies) and mammals (Hornung *et al*., 2004).

Orthologues of most genes that function downstream of *SRY* have also been found in non-mammalian vertebrates. However, lower vertebrates, especially reptiles, amphibians and fish, have an enormous variety of sex determination mechanisms, resulting putatively from an extremely dynamic process of divergent evolution of sex determination mechanisms and sex chromosomes (Graves and Shetty, 2001; Schartl, 2004b, and references therein). This view has turned over the old model, in which the sex determination genes (such as *SRY* for most mammals) represent a common and stable developmental mechanism.

Fishes are by far the largest group of vertebrates, with roughly 25,000 species. Fish are an attractive group of organisms to study the evolution of sex determination because they present a broad range of various types of sexuality from hermaphroditism to gonochorism and from environmental to genetic sex determination (Baroiller *et al*., 1999; Devlin and Nagahama, 2002). Almost all different forms of genetic sex determination are found in fish, including male heterogamety (males are XY and females are XX), female heterogamety (males are ZZ and females are ZW), multiple sex chromosomes (e.g. X, Y and W chromosome in the platyfish), polygenic sex determination (e.g. the *SD* factors are distributed over several

chromosomes) and autosomal influence (e.g. XX males in the medaka, Nanda *et al*., 2003). Unlike the situation in higher vertebrates, where the male and female sexes are always presented by two different individuals (gonochorism), several hundred species of fish are known to be hermaphrodites, most of them from the order Perciforms (the largest order in fish). These fishes become male and female at the same time (synchronous), are male first and become female subsequently (protandrous), or vice versa (protogynous). Age, social factors or temperature can control such a sex changing. The most bizarre case known so far is a self-fertilizing hermaphrodite in a single species, the clonal killifish *Rivulus marmoratus* (Harrington, 1963). Sex determination in gonochorism can also be influenced by environmental factors including temperature, hormones and pH of the water (Devlin and Nagahama, 2002). It was recently reported in zebrafish that the primordial germ cells (PGCs) might have a key role in female development (Slanchev *et al*., 2005).

No sex-linked markers and no sex chromosomes have been identified so far in the zebrafish (*Danio rerio*) and smooth pufferfishes (*Takifugu rubripes* and *Tetraodon nigroviridis*). Hence, alternative models like medaka, platyfish, salmonids, tilapia and sticklebacks are utilized to analyze sex determination and sex chromosome evolution. It is no doubt that the master *SD* gene of some of these fishes will be identified by positional cloning. This will reveal the molecular mechanisms driving the evolution of sex determination and sex chromosomes in fish. In the medaka fish *Oryzias latipes*, the *dmrt1bY* (aka *DMY*) has been found to be an excellent candidate for the master male *SD* gene (Nanda *et al*., 2002; Matsuda *et al*., 2002). In salmonids (XX/XY system), comparative mapping of sex-linked microsatellite markers has suggested that Arctic charr, brown trout, Atlantic salmon and rainbow trout have developed different sex chromosomes with respect to the location of their *SD* loci or that the *SD* locus can transpose (Woram *et al*., 2003). Sex-linked markers have been identified in the Nile tilapia *Oreochromis niloticus* (XX/XY) (Lee *et al*., 2003) and the blue tilapia *O. aureus* (ZW/ZZ) (Lee *et al*., 2004). In the threespine stickleback *Gasterosteus aculeatus*, the complete sequence of a 250 kb-large region linked to the *SD* locus from the proto-X and proto-Y chromosomes revealed the extensive divergence between X and Y sequences. However, the extent of divergence is variable across this region, with high X-Y similarity in coding regions, but little similarity in many intergenic regions (Peichel *et al*., 2004). In the platyfish *Xiphophorus maculatus*, megabase-sized bacterial artificial chromosome contigs covering the sex determination region of the X and Y chromosomes have been constructed and partially sequenced (Froschauer *et al*., 2002; Volff and Schartl,

2001).

1.5 Sex determination in the medaka (*Oryzias latipes*)

The medaka (*Oryzias latipes*) is a small, egg-laying, freshwater bony fish that is native to Asian countries (primarily Japan, Korea and China). In this recently emerging small aquarium fish model species, the situation for the molecular analysis of sex determination is extremely favorable (Wittbrodt and Schartl, 2002). In fact, the medaka was the first vertebrate that was observed to cross-over between the X and Y chromosomes (Aida, 1921). Medaka has an XX/XY sex determination system like mammals. The Y chromosome carries a male-determining locus, which triggers the processes leading to male development of the undifferentiated primordial gonad. Male and female medaka can be easily distinguished by a lot of secondary sex-linked characteristics, such as the shape and size of the dorsal and anal fin (Yamamoto, 1975). A linkage map of the sex chromosomes with several molecular and phenotypic markers is available. Especially useful is the *quart* strain, where the sex chromosomes express the different alleles of the *lf* pigment marker. The presence of leucophores in the males and their absence in the females allow differentiation of both sexes as early as at 2-3 days of embryonic development (Wada *et al*., 1998).

The X and Y chromosomes of the medaka are considered to be at a very early stage of evolution. They are morphologically indistinguishable (Matsuda *et al*., 1998) and crossing-over are possible over the entire length of each chromosome (with the obvious exception of the region containing the *SD* locus). The linkage map of the sex chromosomes (LG1, represented by the second largest pair of chromosomes) shows, for most regions, an even distribution of markers, similar to the autosomes. A region around the *SD* locus, however, presents a largely reduced recombination between the sex chromosomes (Kondo *et al*., 2001; Matsuda *et al*., 1999; Naruse *et al*., 2002). In addition, full sex reversals can be obtained by treatment with certain steroid hormones (such as estrone or testosterone) during embryogenesis or the early larval period. By this and other means YY males, XY females, XX males and even YY females have been produced (Yamamoto, 1975). Such experiments revealed three important phenomena in the medaka. First, sex can be artificially reverted at an early stage of ontogenesis. Second, the viability of YY fish indicates that no essential genes have been inactivated or deleted on the Y chromosome during sex chromosome evolution. Third, without the Y chromosome, testis development and spermatogenesis can occur as revealed by the existence of fully fertile XX males (Hornung *et al*., 2004; Volff and Schartl,

2002). This is different from the situation in mammals, where YY individuals are not viable because the Y chromosome has lost most genes present in the X and XX males are infertile because the Y chromosome carries several spermatogenesis genes that have no alleles on the X (Lahn *et al*., 2001).

A gene with high similarity to *dmrt1*, called *dmrt1bY* or *DMY*, was identified as an excellent candidate for the master male *SD* gene in the medaka by independent studies using candidate gene (Nanda *et al*., 2002) and positional cloning (Matsuda *et al*., 2002) approaches. This gene is designated as *dmrt1bY*, because it is concordant with the generally accepted fish gene nomenclature (Volff *et al*., 2003c). Dmrt1 (Doublesex and Mab-3-related transcription factor 1) is a putative transcription factor apparently ubiquitously involved in sex determination or differentiation in vertebrates, and is a member of a family of proteins containing a conserved DNA-binding motif called DM (Raymond *et al*., 1999; Volff *et al*., 2003c). *Dmrt1bY* is also a member of this gene family. Isolation of the different *dmrt* gene family members from the medaka, chromosome mapping, and sequence analysis of the corresponding regions (Brunner *et al*., 2001; Kondo *et al*., 2002; Nanda *et al*., 2002) revealed that *dmrt1bY* arose by a duplication event of an autosomal segment from linkage group (LG) 9, which contains *Dmrt1*. The whole duplicated fragment from LG9 was firstly inserted into another chromosome (LG1). This insertion made LG1 became the proto-Y, whereas its wild-type homolog (without insertion) became the proto-X. This is perfectly consistent with Ohno`s hypothesis that sex chromosomes are derived from a pair of ancestral autosomes (Ohno, 1967). Hence, the medaka fish has two different copies of *dmrt1*, which will be called *dmrt1a* (the autosomal copy) and *dmrt1bY* (the ′Y` suffix indicating the presence of this gene specifically on the Y chromosome of the medaka).

This insertion created a 260 kb-large chromosomal region on the Y chromosome that is absent from the X chromosome. In this Y-specific region, *dmrt1bY* is the only functional gene. Other genes became inactive, and repetitive DNA as well as transposable elements accumulated around *dmrt1bY*. In addition, numerous deletions and insertions were observed in this region, compared to the parental locus on LG9 (Nanda *et al*., 2002). These accumulations, deletions and insertions lead to recombinational isolation of the *SD* gene. This process is a prerequisite for keeping the identity of a sex chromosome.

Dmrt1bY expression during embryonic development precedes morphological differentiation of the gonads. Its transcripts can be detected only in male embryos. Furthermore, *Dmrt1bY* mRNA and protein have been detected in the somatic cells surrounding the primordial germ

cells (PGCs) in the early gonadal primordium (Kobayashi *et al*., 2004). In the adult fish, *dmrt1bY* is expressed exclusively in Sertoli cells. Even in sex-inversal XY females, *dmrt1bY* is transcribed in the ovary (Nanda *et al*., 2002). The autosomal *dmrt1a* gene, however, is not expressed during embryonic and larval development (Winkler *et al*., 2004). Both a point mutation in *dmrt1bY* and a natural mutant with strongly reduced expression of *dmrt1bY* led to XY male to female sex reversal, indicating that *dmrt1bY* is necessary for male development in the medaka (Matsuda *et al*., 2002). These findings support that *dmrt1bY* is indeed the male-determining gene of the medaka. However, evidence by transgenesis or knock-down methods that this gene is also sufficient for the determination of the gonad primordium towards the male pathway is still lacking.

The identification of *dmrt1bY* as the master regulator of male development in the medaka immediately led to the speculation that this gene might perform a similar function in other fish. However, *dmrt1bY* was not found in two closely related species of the genus *Oryzias*, namely, *O. celebensis* and *O. mekongensis* (Kondo *et al*., 2003). The absence of *dmrt1bY* from these fishes is most probably not due to a secondary loss of the gene, because it was also not detected by extensive Southern blot hybridization analysis in other more divergent species, such as tilapia, guppy, platyfish, and zebrafish (Kondo *et al*., 2003; Veith *et al*., 2003). Importantly, phylogenetic analysis of all available fish *dmrt1* sequences clearly shows that *dmrt1bY* is more closely related to the autosomal *dmrt1* gene of the medaka than to the *dmrt1* gene from more distantly related fish species (Volff *et al*., 2003). Hence, the duplication event that generated the medaka male-determining gene occurred within the genus *Oryzias*. *Dmrt1bY* was predicted to originate by this duplication event approximately 10 million years ago (Kondo *et al*., 2004). The situation of the *dmrt1bY* in medaka is compatible with the hypothesis that genes at the top of the sex determination cascade has evolved relatively rapidly, while the downstream genes are quite conserved (Marin and Baker, 1998; Wilkins, 1995; Zarkower, 2001). Thus, a new question arises, namely how a gene, like the *dmrt1bY* in the medaka, which acts in other organisms as an effector at a downstream position of the sex determination cascade, can become the master regulator at the top. This has been explained by the hypothesis that a duplicated copy of a gene at a downstream position of the sex determination cascade could make all upstream components unnecessary and become the ´´new master *SD* gene`` (Schartl, 2004b).

An atypical case in the medaka should be mentioned with respect to the evolution of sex determination. The genetic analysis of several laboratory strains of medaka (Nanda *et al*.,

2003) demonstrated that a reasonable number of males have an XX genotype and are lacking *dmrt1bY*. The frequency of such male was high up to 12% in some strains (e.g. Carbio strain) and 0% in others (e.g. i-3 strain). The offspring of XX males and XX females show a strong bias towards females and still have XX males at varying number. This relatively high frequency of spontaneous sex-reversed XX males suggests that the medaka can develop as a male without *dmrt1bY* (Nanda *et al*., 2003). This is consistent with the obtention of fertile XX males by artificial sex-reversion. These data indicate that there might be so far unidentified autosomal modifiers taking over the function of *dmrt1bY*.

1.6 Sex determination in the platyfish *Xiphophorus maculatus*

Even though *Xiphophorus* fish is viviparous and is almost impossible to be analyzed functionally *in vivo* by classical embryological methods, several important factors led it to be a good model for the study of vertebrate sex determination. The primary sex determination locus of *Xiphophorus* is tightly linked to gene loci involved in the development of normal and malignant pigment cells, including a tumor locus (*Tu*) equating the *Xmrk* oncogene, a macromelanophore-determining locus (*Mdl*), a locus (*RY*) regulating the formation of red (erythrophore) and yellow (xanthophore) pigment patterns as well as a sex maturity locus (*P* for ´´puberty``) (Gordon, 1927; Kallman, 1975; Volff and Schartl, 2001). A genetic linkage map is also available for *Xiphophorus* (Kazianis *et al*., 2004b; Morizot *et al*., 1991; Walter *et al*., 2004; http://www.xiphophorus.org). In *Xiphophorus*, adult males are typically distinguished from females by the presence of a highly modified anal fin called a gonopodium. Certain species also display pronounced sexual dimorphism including secondary characteristics such as caudal fin extensions (swords).

One of the most thoroughly studied group of fish for the genetics of sex determination is the genus *Xiphophorus* including platyfishes and swordtails (for reviews, see Kallman, 1984; Kazianis *et al*., 2004a; Volff and Schartl, 2001). The diversity of genetic sex determination observed in fish is nicely demonstrated in these fishes.

The platyfish, *Xiphophorus maculatus*, has three types of sex chromosomes: X, Y and W. Females have WY, WX or XX genotypes. Males are either YY or XY. The platyfish is therefore an excellent model to investigate the relationships between both types of heterogamety (XY male and ZW female). The W, X and Y chromosomes are homologous

(Kallman and Schreibman, 1973). Interestingly, YY males in *X. maculatus* are fully viable and fertile, suggesting that no essential genes have been deleted during the evolution of the Y chromosome like in the medaka. WW females have not been found in natural habitats, but can be produced in laboratory and are viable as well. The W chromosome is absent in certain natural population and from the Rio Jamapa strain used in this work. Sex determination in the three chromosome situation is explained at present by two different hypotheses: one of them proposes a network of male-determining genes and regulatory genes distributed over the different types of sex chromosomes and autosomes (Kallman, 1984). Another hypothesis involves a dosage-dependent mechanism based on one sex-determining gene being present at different copy numbers on the three types of sex chromosomes (Volff and Schartl, 2001).
Spontaneous sex-reversal was also observed in some *Xiphophorus* natural populations. XX, XW, YW and WW males as well as XY and YY females have been all found (Anders and Anand, 1963; Kallman, 1984), albeit the overall frequency of this atypical sex determination (ASD) in platyfish is very low (roughly 1%). Autosomal modifiers were speculated to be responsible for this phenomenon. Additionally, the sex determination of the platyfish can be overridden by exogenous factors, e.g. X-ray (Anders *et al*., 1969) and steroid hormone (Dzwillo and Zander, 1967).
The X and Y chromosomes of the platyfish are genetically well studied. The absence of differences by karyotype and synaptonemal complex analysis and comparative genomic hybridization indicates that they are at an early stage of differentiation (Traut and Winking, 2001). Nevertheless, the X and Y chromosomes can be easily distinguished by fluorescent *in situ* hybridization (FISH), using as a probe a retrotransposon-like repeat called *XIR* preferentially accumulating on the Y chromosome (Nanda *et al*., 2000a). Recombination has been observed over the almost entire length of the X and Y chromosomes in *X. maculatus* (Bellamy and Queal, 1951; Gutbrod and Schartl, 1999; Kallman, 1975; Morizot *et al*., 1991). However, X/Y crossovers close to the *SD* locus, where pigmentation genes are located, are rare. This may be due to suppression of recombination in the sex determination region. The W chromosome is peculiar. Except for *egfrb*, none of the X- and Y-chromosomal loci studied so far have alleles on the W chromosome (Gutbrod and Schartl, 1999; Kallman, 1975). Crossing-overs between the W and Y chromosomes have been observed in laboratory breeding. Only in one reported case, a crossover resulted in a chromosome that contained the W-*SD* allele and an *RY* locus that originally was on a Y chromosome (Kallman, 1984). The absence of pigment patterns on the W chromosome remains an open question.

Dmrt1bY has been found to be an excellent candidate for the master male *SD* gene in the medaka. However, *dmrt1bY* is not present in the platyfish (Veith *et al*., 2003), although platyfish and medaka are reasonably related. Therefore the master *SD* gene of the platyfish has still to be identified. Of particular interest is the situation of this gene on the three different types of gonosomes.

The platyfish population used in this work (Rio Jamapa, strain WLC1274) has a stable and reproducible XX/XY sex determination system in the absence of W chromosome. However, as soon as a candidate of the master *SD* gene is identified in this population, this candidate can be easily and rapidly investigated on the W chromosome in other populations of *X. maculatus*. Importantly, the *SD* gene locus in the *X. maculatus* population (Rio Jamapa) used in this work is flanked by the *Xmrk* and *egfrb* genes, which are both well characterized at the molecular level and delimit a relatively small region (about 1 Mb) (Gomez *et al*., 2004; Gutbrod and Schartl, 1999; Volff *et al.*, 2003) (Fig. 2). Thus it becomes feasible to identify the master *SD* gene of *X. maculatus* through a positional cloning approach.

For the swordtail, *X. hellerii*, at least two strains have been shown to utilize a WY/YY mechanism (Kazianis *et al*., 2004a), while others exhibit a so-called polyfactorial system of male- and female-determining genes (Peters, 1964). The latter system has been explained by the presence of several autosomal modifiers in a population with each modifier locus having several alleles of different strength with regard to their influence on the gonosomal sex determination gene (Kallman and Bao, 1987).

Attempts towards the isolation of the sex determining gene(s) in *Xiphophorus* are underway. Bacterial artificial chromosome (BAC) contigs starting from *Xmrk* and *egfrb* that flank the sex determination region on the X and Y chromosomes of *X. maculatus* (Rio Jamapa) have been established covering approximately 1Mb on each sex chromosome (Froschauer *et al*., 2002).

1.7 Genomic plasticity in the sex determination region of *Xiphophorus maculatus*

The molecular and phenotypic analysis of X/Y crossover mutants of *X. maculatus* showed that roughly half of the crossovers are located within a defined segment of the 5` region of the *Xmrk* oncogene, suggesting that a hotspot of recombination is present in the sex determination region of *X. maculatus* (Gutbrod and Schartl, 1999). Most traits encoded or controlled by gene loci located within the sex determination region of *X. maculatus* are subjected to a high level

of genetic variability (Volff and Schartl, 2001). For instance, platyfish populations are highly polymorphic for the pituitary *P* locus and display a wide variety of phenotypes that range from very early to very late maturating animals (Kallman, 1989). Nine *P* alleles have been described so far in *X. maculatus* (Schreibman *et al*., 1994). Moreover, both macromelanophore pigmentation patterns and melanoma phenotypes are extremely variable in *X. maculatus* (Gordon, 1927). They are controlled by different allelic forms of *Mdl* and *Xmrk*. Both genes probably act in combination to determine the onset, compartment, intensity and extension of macromelanophore patterns as well as the onset, location, and severity of melanoma (Schartl and Wellbrock, 1998). Thus, a high level of genomic plasticity may be responsible for the genetic variability of the traits encoded or controlled by gene loci linked to the sex determination region of *X. maculatus*.

1.7.1 Gene deletions and duplications in the sex determination region of *X. maculatus*

The sex determination region of *X. maculatus* apparently undergoes frequently rearrangements including duplications, deletions, amplifications, and transpositions (Froschauer *et al*., 2001; Volff *et al*., 2003a). One deletion event in the sex determination region has been reported in the *X. maculatus* lof-2 mutant, losing the ability to generate melanoma through the deletion of the entire *Xmrk* oncogene (Schartl *et al*., 1999).

The first reported example of gene duplication in the sex determination region of *X. maculatus* was the duplication of *egfrb* that led to the formation of the *Xmrk* oncogene (Adam *et al*., 1993). Molecular analysis of BAC contigs spanning the *Xmrk* oncogene region revealed that other genes have frequently undergone duplication in this region as well (Froschauer *et al*., 2002). One of these genes is *Xmrk*-linked and encodes a conceptual protein sequence highly similar to the N-terminal part of the mammalian Cript protein. Cript is a post-synaptic protein found in rat and highly conserved in mammals and plants. Its described function in mammals is the linking of the PSD-95 protein to the cytoskeleton in excitatory synapses (Niethammer *et al*., 1998; Passafaro *et al*., 1999). Three copies of *cript* have been identified in the *X. maculatus* genome. The intact *cript* is on autosomes; a second truncated copy is on the X and Y chromosomes; a third truncated copy is only detected on the Y chromosome (Froschauer, 2003). In addition, a gene called *mc4r*, which encodes a transmembrane type 4 melanocortin receptor that is involved in the control of body weight (Jordan and Jackson,

1998), has been duplicated several times in the sex determination region of *X. maculatus*. As many as eleven copies have been identified so far on the BAC contigs from the X and Y chromosomes. In contrast, gene encoding members of the melanocortin receptor family are normally single-copy in other fish species as well as in higher vertebrates (Selz, 2003).

1.7.2 Accumulation of transposable elements and other repetitive sequences in the sex determination region of *X. maculatus*

Transposons are mobile DNA elements found in the genome of a variety of organisms, including bacteria, fungi, plants and animals (Berg and Howe, 1989). These “jumping genes”, first described by McClintock (McClintock, 1956a; McClintock, 1956b), are discrete segments of DNA capable of moving from one site to another within the genomes of their hosts by one of several possible mechanisms. Autonomous transposable elements usually encode the proteins necessary to carry out their movement (including transposases and integrases).

As repetitive sequences, transposable elements can be substrate for recombination and generate inversions, duplications, deletions, amplifications and transpositions. They can also induce the transposition of non-autonomous segments and influence gene expression. Furthermore, transposable elements have potential important applications in fish biotechnology, including the construction of genomic libraries (Kirchner *et al*., 2001), DNA sequencing (Fischer *et al*., 1996), fingerprinting, mapping, gene isolation, gene tagging, gene transfer (Izsvak and Ivics, 2003; Izsvak *et al*., 1997).

Based on their structure and mechanisms of transposition, transposable elements are usually divided into two major classes (Craig, 1995; Curcio and Derbyshire, 2003). Class I elements, or retrotransposable elements, move via an RNA intermediate which is reverse-transcribed and integrated by element-encoded proteins including a reverse transcriptase. They include reverse transcriptase retrotransposons (LTR or non-LTR retrotransposons depending on the presence of flanking long terminal repeats), retroviruses (reverse transcriptase LTR elements with an envelope gene) and various categories of nonautonomous retroelements like the short interspersed nuclear elements (SINEs). For class II elements, or DNA transposable elements, transposition is catalyzed by either a transposase, the so-called ´´cut-and-paste`` mechanism that results in target site duplications (TSD) flanking the element, or catalyzed by an element-encoded rolling circle replication initiator like protein, a putative mechanism that does not

result in TSD (Kapitonov and Jurka, 2001).

Molecular analysis of cosmid and BAC contigs containing the *Xmrk* oncogene from the X and Y chromosomes of *X. maculatus* revealed numerous retrotransposon elements and other repetitive sequences (Nanda *et al*., 2000a; Volff *et al*., 2001a; Volff *et al*., 2001b; Volff *et al*., 2003b; Volff *et al*., 2001c; Volff *et al*., 2000; Volff *et al*., 1999). About 20 solo long-terminal-repeats (LTRs) of a putative non-autonomous retrotransposon called *XIR* (Roushdy *et al*., 1999) were identified between the *Xmrk* oncogene and the next telomere (Nanda *et al*., 2000a; Volff *et al*., 2003b). A probably recent high-copy-number amplification of the *XIR* LTR was detected by fluorescent *in situ* hybridization in the sex determination region on the Y but not on the X chromosome in the Rio Jamapa population. However, in other populations of *X. maculatus*, this amplification was not observed (Nanda *et al*., 2000a). This event might correspond to an early stage of molecular differentiation between both types of gonosomes. In addition, there are three other *XIR*-associated repeat clusters in the *Xmrk* region (Volff *et al*., 2003b).

Several non-long terminal repeat (non-LTR) retrotransposons such as *Rex1, Rex3, Rex2*, *Rex5* and *Rex6* were identified around the *Xmrk* oncogene. *Rex1* retroposons and the *Babar* elements, which might be related to *CR1* elements from the chicken, present several ancient lineages in fish genomes (Volff *et al*., 2000). They frequently and independently invaded fish genomes with varying extent (from < 5 to 500 copies per haploid genome). There are approximately 200 copies of *Rex1* in *Xiphophorus*. These copies belong to three different lineages. Interestingly, *Rex1* might have been transferred horizontally between divergent fish species (Volff *et al*., 2000). The *Rex3* element is a member of the so-called *RTE* clade of retroposons and highly conserved in teleosts. *Rex3* was very active and underwent numerous independent bursts of retrotransposition during teleost evolution (Volff *et al*., 2001c; Volff *et al*., 1999). There are about a thousand copies of *Rex3* in the *Xiphophorus* haploid genome. Most of them are truncated at their 5` end, due to incomplete reverse-transcription of the RNA intermediate involved in retrotransposition. Two other elements called *Rex2* and *Rex5* have been identified in the sex determination region as well (Volff *et al*., 2001a, Volff person communication; Volff *et al*., 2003b). They are both related to the *Maui* element of *Takifugu rubripes* (Poulter *et al*., 1999). The *Rex6* element, which was also active in many fish species, belongs to the *R4* family of retrotransposons containing *Dong* and *R4* from nematodes and insects. More than 150 copies per haploid genome were detected in *X. maculatus*. Strikingly, *Rex6* encodes besides a reverse transcriptase an endonuclease related to type IIS restriction

enzymes. In contrast, most of the other non-LTR retrotransposons encode an apurinic-apyrimidinic (AP) endonuclease (EN), which is necessary for cleaving the target sequence before integration (Volff *et al*., 2001b).

Several LTR retrotransposons were also detected in the *Xmrk* oncogene region such as *Rex7* and *Jule* (Volff *et al*., 2001a). These two LTR retrotransposons belong to the *Ty3/Gyspy* class. *Rex7* is related to *Sushi* from *Takifugu rubripes* and to retrotransposon fossils found in mammalian genome (Poulter and Butler, 1998; Volff *et al*., 2001a). *Jule* is the first vertebrate LTR retrotransposon belonging to the *Mag* family of *Ty3/Gyspy* retrotransposons. This element is 4.8 kb in length and is flanked by two 202 bp long terminal repeats. *Jule* also encodes the classical proteins Gag (structural core protein) and Pol (protease, reverse transcriptase, RNase H and integrase, in this order), but no envelope. Although only three to four copies of *Jule* were detected in the haploid genome of *X. maculatus*, two of them are located the sex determination region. One almost intact copy of *Jule* is found in the first intron of the X-chromosomal allele of *egfrb*, and another more corrupted copy is present only 56 nucleotides downstream from the polyadenylation signal of the *Xmrk* oncogene on both X and Y chromosomes. The high nucleotide identity between both copies strongly suggests a recent retrotransposition activity of *Jule* in the genome of *Xiphophorus* (Volff *et al*., 2001a). In a spontaneous mutant of *X. maculatus* (lof-1), the nonautonomous LTR retrotransposon *Tx-1* has jumped into the *Xmrk* oncogene on the X chromosome. Thereby the oncogene lost its ability to induce melanoma (Schartl *et al*., 1999). This is a strong experimental evidence for an active retrotransposon in the sex determination region of *X. maculatus*.

A single copy of a repetitive sequence called the *D* locus containing sequences reminiscent of DNA transposons is located directly upstream from the transcriptional start of ONC-*Xmrk* (Adam *et al*., 1993). This *D* locus is present at more than 20 copies in the genome of *X. maculatus* (Fornzler *et al*., 1996), and most of them are clustered at one unique autosomal site (Nanda *et al*., 2000a). A *piggyBac*-like sequence, which is related to the human *piggyBac* transposable element-derived-4 sequence (Accession No. NP_689808), is inserted in the *D* locus (Volff *et al*., 2003b).

Furthermore, partially sequencing and Southern blot hybridization analysis of an over 1Mb large *SD*-linked Y-specific contig (*ps-criptY* contig) from *X. maculatus* uncovered additional events of duplication, inversion, transposition and amplification as well as insertions of several retroelements and DNA transposable elements (such as *helitron*, *Tc-1*) (this work, see results). These described elements might be involved in the formation of rearrangements in

the sex determination region of *Xiphophorus maculatus* through their repeated nature and mobility (Volff *et al*., 2003a).

1.8 Aims of this work

In order to identify the master *SD* gene of *X. maculatus*, a BAC library of *X. maculatus* (Rio Jamapa) has been set up. BAC contigs starting from *Xmrk* and *egfrb* that flank the sex determination region on the X and Y chromosomes have been established, covering at present approximately 1Mb on each sex chromosome (Froschauer *et al*., 2002).

Three copies of a *Xmrk*-linked gene encoding a conceptual protein sequence highly similar to the mammalian Cript protein have been identified in the *X. maculatus* genome. The intact *cript* is autosomal; a second truncated copy is closed to the *Xmrk* oncogene on the X and Y chromosomes; a third truncated copy called *ps-criptY* is only detected on the Y chromosome (Volff *et al*., 2003b). Interestingly, no recombination was observed between the *SD* gene and *ps-criptY* in more than 400 fishes tested (Vucic, M. *et al.*, pers. commun.), indicating a tight linkage between *ps-criptY* and the master *SD* gene of *X. maculatus.* Hence, a roughly 500kb-BAC contig containing *ps-criptY* has been assembled for positional cloning of the master *SD* gene of *X. maculatus* (Froschauer, A., pers. commun.).

The major objective of this work is to extend and molecularly analyze the *ps-criptY* contig in order to contribute to identify the master *SD* gene of *X. maculatus* via positional cloning. The second aim is to identify other gene candidates (e.g. those involved in melanomagenesis) in the sex determination region. The third aim is to elucidate molecular mechanisms during the initial differentiation of sex chromosomes in fish through comprehensive structural comparisons and analyses between the *ps-criptY* contig and the other contigs established on the X and Y chromosomes.

2. Materials

2.1 Fishes

The following fishes were from stocks maintained at the University of Würzburg: *Xiphophorus maculatus* (Rio Jamapa strain WLC1274, Rio Usumacinta strain WLC1372), *Xiphophorus hellerii* (Rio Lantecilla strain hIII and Rio Santecomapan), *Xiphophorus milleri* (Catemaco 96), *Poecilia mexicana* (Media Luna), *Poecilia formosa* (Tampico), *Poecilia latipunctata, Gambusia affinis* (Pena Blanca), *Girardinus falcatus* (aquarium stock), *Poeciliopsis gracilis* (Rio Jamapa), *Heterandria bimaculata* (Tierra Blanca), *Phallichthys amates* (aquarium stock), *Belonesox belizanus, Priapella olmecae, Fundulus heteroclitus*, *Oryzias latipes* (medaka fish strain HB32c) and *Danio rerio* (zebrafish strain m14).

Rainbow trout (*Oncorhynchus mykiss*), pike (*Esox lucius*), common carp (*Cyprinus carpio*), European eel (*Anguilla anguilla*) and sturgeon (*Acipenser sturio*) were obtained from a local fish farm near Würzburg. Genomic DNA was generated from the Nile tilapia (*Oreochromis niloticus,* Göttingen, Germany).

Backcross hybrids (Sd/XhIII Bcn) were obtained between *Xiphophorus maculatus* (Rio Jamapa strain WLC1274) and *Xiphophorus hellerii* (Rio Lantecilla strain hIII) using *X. hellerii* as the recurrent parent (third backcross generation). Fishes carrying the wild-type X chromosome of *X. maculatus* with Mdl^{Sd}-ONC-X*mrk* develop, depending on the presence or absence of *R*, benign or malignant melanoma. They are known as Sd-melanoma fish from the classical Gordon-Kosswig cross (Anders, 1991; Kazianis *et al*., 1998).

Lof-2 is a spontaneous loss-of-function mutant for the Sd locus that occurred in our laboratory in a backcross brood of Sd melanoma fish with *X. hellerii* (Rio Lantecilla strain hIII) as the recurrent parent. It carries a Sd/*Xmrk* loss-of-function X chromosome, derived from the wild-type Mdl^{Sd}-ONC-X*mrk* chromosome, in the genetic background of *X. hellerii* (Schartl *et al*., 1999).

MAD is derived from male *Xiphophorus maculatus* Rio Jamapa strain WLC1235. In this fish, a recombination between *Xmrk* oncogene and *egfrb* has been identified. Its genotype is $X^{DrSd/INV2}/Y^{ArSr/INV2}$ (Froschauer, 2003).

2.2 Fish cell lines

The cell line PSM (´´Platy Swordtail Melanoma``) is derived from melanoma of a hybrid of *Xiphophorus maculatus* and *Xiphophorus hellerii* (Wakamatsu, 1981). The A2 cell line is an

immortalized cell line from *Xiphophorus xiphidium* embryos (Kuhn *et al*., 1979). Both of them have been kindly provided by Petra Fischer (Physiologische Chemie I, University Würzburg).

2.3 Competent cells and vectors

The competent bacteria used in this work for transformation were *Escherichia coli* XL1-blue (Bullock *et al*., 1989), *E. coli* DH5α (Sambrook *et al*., 1989) and Electromaxx® *E. coli* DH10β, (INVITROGEN™), while the cloning vector was pBluescript II KS+/- (STRATAGENE, Amsterdam).

For sequencing by ´´transposon tagging``, competent bacteria including *E. coli* JM109 F' Lac[::TnSeq2] and *E. coli* HB101 as well as the vector pJOE2114 (Fischer *et al*., 1996) were obtained from Dr. Josef Altenbuchner (Stuttgart University).

2.4 Enzymes

Superscript™ II reverse transcriptase, *Taq* DNA polymerase, large fragment of polymerase I (Klenow-Polymerase) and restriction endonucleases were obtained from INVITROGEN (Karlsruhe) except for *Sca*I and *Eco*47III (MBI FERMENTAS, St. Leon-Rot).

T4 DNA ligase, DNaseI (RNase-free), and alkaline phosphatase were obtained from ROCHE DIAGNOSTICS GMBH, Mannheim. Proteinase K and RNase A were from MERCK KGAA, Darmstadt.

2.5 Media, antibiotics and solutions

2.5.1 Media

LB Medium	bacto tryptone	10 g
	yeast extract	5 g
	NaCl	10 g

The pH of the medium was adjusted to 7.0 with 10 N NaOH. The volume was completed to 1000 ml with dH_2O and autoclaved at 121°C, 15 bar for 20 minutes.

dYT Medium	bacto tryptone	16 g
	yeast extract	10 g
	NaCl	5 g

These were dissolved in 1000 ml dH_2O and autoclaved at 121°C, 15 bar for 20 minutes.

SOB (Gibco)	bacto tryptone	2%
	yeast extract	0.55 %
	NaCl	10 mM
	KCl	10 mM
	Mg^{++}	10 mM

The pH was adjusted to 7.6 with KOH if necessary (for plates add 15 g/l agar), autoclaved, stored at 4 °C. Before use, 2 % (vol/vol) sterile 1M $MgSO_4$ (1 ml / 50 ml) was added.

SOC	glucose are added up to 20 mM into SOB medium

2.5.2. Antibiotics

Antibiotic	**Stock solution**	**Working concentration in medium**
Ampicillin	50 mg/ml in 50% ethanol	100 µg/ml
Chloramphenicol	12.5 mg/ml in 100% ethanol	12.5 µg/ml
Kanamycin	10 mg/ml in dH_2O	30 µg/ml
Streptomycin	200 mg/ml in dH_2O	200 µg/ml
Tetracyclin*	15 mg/ml in 50% ethanol	15 µg/ml

Antibiotics were prepared as stock solutions and stored at -20°C.

*Store solution in light-tight containers

2.5.3. Solutions

0.5 M EDTA (pH 8)	93 g/500 ml EDTA · 2 H_2O in dH_2O, autoclaved
1 M glucose (20 %)	19.8 g/100 ml D(+)Glucose · H_2O in dH_2O, autoclaved
40 % glycerol	40 ml glycerol in 60 ml dH_2O, autoclaved
2 M HEPES (pH 6.6)	47.6 g/100 ml in dH_2O, pH adjusted with 5 N NaOH, autoclaved
1 M KOAc (pH 7.5)	9.8 g/100 ml in dH_2O, pH adjusted with HCl, autoclaved
1 M KCl	7.45 g/100 ml in dH_2O, autoclaved

1 M $MgCl_2$	20.3 g/100 ml $MgCl_2 \cdot 6\ H_2O$ in dH_2O, autoclaved
1 M $MgSO_4$	24 g/200 ml in dH_2O, autoclaved
5 N NaOH	20 g/100 ml in dH_2O
3 M NaOAc (pH 5.2)	40.82 g NaAcetate • 3 H_2O in 50 ml dH_2O, pH adjusted with 11 ml acetic acid, completed with dH_2O up to 100 ml, autoclaved
0.5 M Na_2HPO_4 (pH 6.9)	44.5 g $Na_2HPO_4 \cdot 2\ H_2O$ in 450 ml dH_2O, pH adjusted with ca. 3 ml H_3PO_4, completed with dH_2O up to 500 ml, autoclaved
7.5 M NH_4OAc	231.2 g ammoniumacetate in 180 ml dH_2O, completed with dH_2O up to 400 ml and autoclaved
2 M NH_4OAc	61.6 g in 300 ml dH_2O, completed with dH_2O up to 400 ml, autoclaved
20 % SDS [wt/vol]	1000 g in 4 l dH_2O, pH 7.2 adjusted with HCl_{conc}, completed with dH_2O up to 5 l
TE 10·1 [mM]	1.0 ml 1 M Tris + 200 µl 0.5 M EDTA, completed with 100 ml dH_2O, autoclaved
TE 10·0.1	1.0 ml 1 M Tris+ 20 µl 0.5 M EDTA, completed with 100 ml dH_2O, autoclaved
1 M Tris-HCl (pH 8)	60.6 g in 400 ml dH_2O, pH adjusted with HCl_{conc} (ca 20 ml), completed with dH_2O up to 500 ml, autoclaved
2.5 M Tris-Cl, pH 8	47.6 g in 90 ml dH_2O
0.25 M HCl	43 ml in 2 l dH_2O
1.5 M NaCl / 0.5 M NaOH	219 g NaCl + 50 g NaOH, completed with 2500 ml dH_2O
X-gal	X-gal 40 mg/ml was dissolved into dimethylformamide, kept in light-tight containers and stored at -20°C. completed with 50 µl into each plate
IPTG	IPTG 200 mg/ml was dissolved into 50% ethanol and stored at -20°C, supplemented with 50 µl into each plate

All chemicals of molecular biology research grade were obtained from respective manufacturers (Merck, Roth, Serva and Sigma). All solutions were prepared using distilled water. When necessary, solutions were sterile filtered or autoclaved.

2.6 Oligonucleotides

All oligonucleotides used in this work were obtained from INVITROGEN™ and Biomers.net GmbH. They are listed in Tab. A2 of the appendix (Chapter 6.1). Primers are 19-22 nt in length and their melting temperature ($T_m \approx [4^*\ (G\ or\ C) + 2^*\ (A\ or\ T)]°C$) is between 55 and 62°C.

2.7 Instruments, hard- and softwares

2.7.1 Instruments

The instruments used are listed as below:

Autoclave	Webeco
Cold centrifuge	Eppendorf
Superspeed centrifuge	Du Pont
CEQ™ 2000XL	Beckman Coulter
Gel dryer	BioRad 1125 B
Heating blocks	Hartenstein
Ice machine	Scotsman
Microliter pipettes	Eppendorf, Gilson
Microcentrifuge	Eppendorf
PCR machine	Bio-med, Thermocycler 60
PH meter	WTW pH523, Metrohm-Herisau
DNA/RNA Calculator	Gene Quant II, Amersham
Phosphorimager	Taifun, Molecular Dynamics
Refrigerators (-20°C; -80°C)	Privileg
Shaking incubator	Hartenstein
Electrophoresis apparatus	Bio-Rad, Hartenstein
Spectrophotometer	Amersham
Vortexer	Hartenstein
Incubator	Heraeus
Magnetic stirrer	BIOSAN
Photoelectric colorimeter	Klett MEG
Autoradiography machine	Kodak M 35 X-OMAT Processor
Laminar flow	NUAIR
Gs Gene Linker®	BIORAD™
Speedvac-concentrator	Univapo 150 H, Savant

2.7.2 Hard- and softwares

Text and pictures were edited using the MICROSOFT® Office 2000 and COREL ESSENTIALS® 9.0. Sequences were analyzed using the Wisconsin Package Version 10.3, Accelrys Inc., San Diego, CA and the attached PAUP4* software (Swofford, 2003), which were accessible from an UNIX-based central computer in the computer center of the University of Würzburg.

Sequences were compared in databanks (http://www.ncbi.nlm.nih.gov/BLAST/; http:// fugu. hgmp.mrc.ac.uk/blast/; http://www.ensembl.org) and analyzed using the NIX-software (http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/).

3. Methods

3.1 DNA extraction methods

3.1.1 BAC DNA extraction methods

1) QIAGEN large-construct kit

A single colony was isolated from a freshly streaked selective plate and inoculated in a starter culture of 2-5 ml LB medium containing the appropriate selective antibiotic. After incubation about 8 h at 37°C with vigorous shaking (~ 300 rpm), 0.5-1.0 ml of the culture was diluted into 500 ml LB medium (1:500 to 1:1000 dilution) with antibiotic. The culture was incubated at 37°C for 12-16 h. Bacterial cells were harvested by centrifugation at 6000 g for 15 min at 4°C. The pellet was resuspended in 20 ml buffer P1 (50 mM Tris•Cl, pH 8.0; 10 mM EDTA; 100 µg /ml RNase A). The suspension was supplemented with 20 ml buffer P2 (200 mM NaOH, 1% SDS), mixed gently but thoroughly by inverting 4-6 times, and incubated at room temperature for 5 min. Then the viscous lysate was supplemented 20 ml chilled buffer P3 (3.0 M potassium acetate, pH5.5), mixed gently but thoroughly by inverting 4-6 times, and incubated at room temperature for 10 min. Before loading into the centrifuge, the sample was mixed again. Centrifugation was performed in non-glass tubes at ≥ 20,000 g for 30 min at 4°C. The supernatant containing the BAC DNA was removed and filtered promptly. DNA was precipitated by adding 0.6 volume room-temperature isopropanol and centrifuged immediately at ≥ 15,000 g for 30 min at 4°C. The supernatant was carefully decanted. DNA pellet was washed with 5 ml room-temperature 70% ethanol, and centrifuged at ≥ 15,000 g for 15 min. The supernatant was carefully decanted. The tube containing the DNA pellet was placed upside down on a paper towel to allow the DNA to air-dry for 2-3 min. The DNA was carefully redissolved in 9.5 ml buffer EX (50 mM Tris•Cl; 10 mM $MgCl_2$, pH 8.5), until the DNA was completely dissolved. Afterwards, 200 µl ATP-dependent exonuclease and 300 µl ATP solution were added to the dissolved DNA, mixed and incubated in a water bath at 37°C for 60 min. Before addition of 10 ml buffer QS (1.5 M NaCl; 100 mM MOPS, pH 7.0; 15% isopropanol) to the DNA sample and transfer into the QIAGEN tip, the QIAGEN tip was equilibrated by 10 ml buffer QBT (750 mM NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol; 0.15% triton X-100). After the sample entered the resin by gravity follow, the QIAGEN tip was washed twice with 30 ml buffer QC (1.0 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol). DNA was eluted with 15 ml prewarmed (at 65°C) buffer QF (1.25 M NaCl; 50

mM Tris•Cl, pH 8.5; 15% isopropanol). The DNA sample was precipitated by adding 10.5 ml room-temperature isopropanol to the eluted DNA, mixed and centrifuged immediately at ≥ 15,000 g for 30 min at 4°C. The pellet DNA was washed by adding 5 ml room-temperature 70% ethanol (without disturbing the pellet) and centrifuged at ≥ 15,000 g for 30 min at 4°C. Finally, the pellet was air-dried for 5-10 min, redissovled in a suitable volume of buffer (e.g. TE, pH 8.0 or 10 Mm Tris•Cl, pH 8.5) and stored at 4 °C.

2) Midi-preparation

5 ml overnight culture was centrifuged at 3000 g for 10 min and the pellet was resuspended in 300 µl buffer 1 (50 mM Tris, 10 mM EDTA, pH8). 300 µl buffer 2 (0.2 M NaOH, 1% SDS) was added. The sample was mixed and incubated at room temperature for 5 min. After addition of 300 µl buffer 3 (3 M NaOAc, pH 5.2) and mixing, the sample was incubated for 10 min on ice and centrifuged for 15 min at 14,000 g. The supernatant was transferred into a new 2 ml tube. The DNA was precipitated with isopropanol and washed with 70% ethanol. The DNA was redissolved in 50 µl TE 10.0.1 and stored at -20°C.

3.1.2 Genomic DNA extraction methods

Extraction buffer	EDTA	0.1 M
	SDS	0.2%
	NaCl	0.2 M
	Proteinase K	200 µg/ml
	(proteinase K was added immediately before use)	
TE buffer	Tris (pH 8.0)	10 M
	EDTA	1 mM

500 µl of extraction buffer were added to 80-100 mg of the respective tissue, the sample was homogenized by up- and down strokes and transferred to a 2 ml Eppendorf tube with a cut tip. The homogenizer was washed twice with 500 µl of extraction buffer and incubated for 3 h at 80°C with gently shaking. Then, 750 µl of phenol was added and incubated 3-5 min under shaking. 750 µl of chloroform/isoamylalcohol was added and incubated 3-5 min under shaking. The sample was centrifuged for 10 min at 2000 rpm. The supernatant was transferred into a 15 ml Falcon tube containing 3.75 ml of 100% ethanol. The tube was incubated on ice for 15 min. The DNA was wrapped around a glass capillary and washed twice by submerging the capillary into a Falcon tube filled with 10 ml 70% ethanol. The DNA was dried on air, dissolved in 500 ml TE overnight at room temperature and stored at 4°C.

3.1.3 Plasmid DNA extraction methods

Mini preparation of plasmid DNA (Bio-Feedback-Method):

5 ml overnight culture was centrifuged at 3000 g for 10 min and the pellet was lysed with 200 µl of lysis buffer (50 mM glucose, 25 mM Tris, 10 mM EDTA; supplemented with 10 mg/ml lysozyme and 100 µg/ml RNase A directly before use). The reaction was incubated at 37°C for 10-20 min and, after addition of 400 µl of 0.2 N NaOH/1% SDS and shaking by hand, was kept on ice for 5 min. 300 µl 7.5 M chilled NH_4OAc was added and the reaction was incubated on ice for 5 min. After centrifugation at 13,000 g for 10 min, the supernatant was removed to a new tube. 500 µl of isopropanol was added and the reaction was incubated at room temperature for 10 min. After centrifugation at 13,000 g for 10 min, the supernatant was removed. 300 µl 2 M NH_4OAc was added to the pellet. The reaction was kept at room temperature for 5 min and then on ice for 5 min. After centrifugation at 13,000 g for 10 min, the supernatant was transferred into a new tube and 200 µl isopropanol was added. After centrifugation at 13,000 g for 10 min, the pellet was washed with 100 µl ethanol, dried on air, and dissolved in 50 µl TE.

3.2 RNA extraction methods

The trizol reagent (Invitrogen) was used for RNA isolation from tissues.

50-100 mg of tissue was stored shortly in a 1.5 ml tube on dry ice and then transferred into a glass homogenizer. 1 ml of Trizol reagent was added. After homogenization, the tissue was transferred to a 1.5 ml tube and incubated at 15 to 30°C for 5 min to permit the complete dissociation of nucleoprotein complexes. After addition of 200 µl of chloroform, the sample was shaken vigorously for 15 sec and incubated at 15 to 30°C for 5 min. The sample was centrifuged at 12,000 g for 15 min at 2 to 8°C. Following centrifugation, the mixture separated into a low red phenol-chloroform phase, a white interphase and a colorless upper aqueous phase. RNA remained exclusively in the aqueous phase. The upper phase was carefully transferred into a new 1.5 ml tube. RNA was precipitated by adding 500 µl isopropanol. After vortexing and incubation at room temperature for 10 min, the RNA was precipitated by centrifugation at 12,000 g for 10 min at 2 to 8°C. The supernatant was then carefully removed. RNA pellet was washed twice by adding 1 ml 75% ethanol and dried briefly (the RNA should not be dried by centrifugation under vacuum; it was important not to let the RNA pellet dry completely as this will greatly decrease its solubility). RNA was

dissolved in 100 µl RNase-free water and then incubated at 60°C for 5 min. RNA was stored at -70°C.

3.3 DNA digestion, ligation and dephosphorylation

Restriction endonucleases and corresponding buffers were purchased from Pharmacia and Stratagene, respectively, and used as recommended by the manufacturers.

DNA ligation (enzyme and buffer from New England Biolab) was performed as follows:

	Blunt ends	**Sticky ends**
Volume	20 µl	20 µl
Insert:vector (molar ratio)	5:1	5:1
Insert (nmol)	45-180	9-90
Vector (nmol)	15-60	3-30
DNA total (I+V in µg)	0.1-1	0.01-0.1
T4 ligase	1 U	0.1 U
Incubation time and temperature	16°C overnight	16°C overnight or 1 h at 24°C-28°C

A general formula for calculating the concentration of vector and DNA fragment to be inserted is: ng of insert = (ng of vector x kb size of insert / kb size of vector) x molar ratio of (insert/vector).

Dephosphorylation of DNA fragments with alkaline phosphatase:

In some cases, it was advantageous to treat the vector with Calf Intestinal Alkaline Phosphatase (CIAP, Promega) to remove the phosphate groups from the 5´-ends to prevent self-ligation of the vector.

DNA solution (1-20 pmol DNA)	10-40 µl
reaction buffer (10x)	5 µl
dH_2O up to	49 µl
alkaline phosphatase (20 U)	1 µl

The reaction was incubated at 37°C for 45 min, then stopped by heating at 85°C for 15 min and DNA was extracted with phenol-chloroform.

3.4 Transformation of *E. coli*

<u>**DnD** (DMSO and DTT solution)</u>	1 M DTT (Dithiothreit)	1.53 g
	90 % DMSO	9 ml
	10 mM KOAc, pH 7.5	100 µl (1 M,

sterilely filtrated through 0.22 µm filter)
dH_2O up to 10 ml

The solution was filtered through 0.45 µm filter and stored as 200 µl aliquots at –20 °C.

TFB	100 mM KCl	7.4 g
	45 mM $MnCl_2$ x 4 H_2O	8.9 g
	10 mM $CaCl_2$ x 2 H_2O	1.5 g
	3 mM Hexamine Co (III) chloride	0.8 g
	10 mM K-MES	20 ml (0.5 M)
	dH_2O up to 1000 ml	

The 0.5 M solution of MES (2-[N-Morpholino]ethanesulfonic acid) was equilibrated to pH 6.3 using high concentration KOH solution, then sterilized by filtration through a 0.22 µm membrane and stored in aliquots at - 20 °C. The prepared TFB solution was filtered through a 0.22 µm prerinsed membrane, then transferred into sterile tubes (45 ml each aliquot) and stored at 4 °C, where it is stable for approximately 1 year.

1) Preparation of competent *E. coli* cells (Hanahan, 1995):
0,5-2 ml of overnight culture was inoculated into 50 ml SOB (with 1 ml 1M $MgSO_4$) in a flask and grown at 37°C for about 3h with vigorous agitation until an OD_{600} of 0.4 – 0.7 was reached. The bacterial cells were transferred to a sterile 50 ml polypropylene tube and centrifuged at 3500 rpm for 5 min at 4°C. The pellet was carefully washed with 4-5 ml TFB and then resuspended in 2-3 ml of TFB. After incubation of cells on ice for about 30 minutes, the cells were suspended with 35 µl DnD per ml TFB and incubated on ice for 10-20 min. The cells were then supplemented with the same amount of DnD as before, swirled and incubated on ice for 10-20 min. Bacteria were ready to use (handle carefully), they can be left on ice up to 24 hrs.

2) Transformation of competent *E. coli* cells
DNA was put into sterile tubes, including positive and negative controls (e.g. the pBS plasmid vector as positive, no DNA as negative control). 200 µl of competent bacteria suspension were added into each tube, swirled and incubated on ice for 10-40 min. Each tube was heated at 42 °C for 90 sec, chilled on ice for 2 min. 1 ml SOB (or LB) was added and incubated at 37 °C for 30-60 min with gentle agitation (50 rpm). 100 µl of the culture were plated on plates with selective antibiotics (10 % of transformed bacteria), the rest was spun down (2 min x

4000 rpm). The pellet was resuspended in 100 µl SOB (or LB) and plated as before (90 % of transformed bacteria). Plates were incubated overnight at 37 °C.

3.5 Phenol-chloroform extraction

Using phenol-chloroform extraction, proteins can be removed from DNA solutions. This is a standard method in many molecular biology protocols, including DNA extraction from bacteria, eukaryotic cells or tissue, to remove unwanted enzymes such as restriction enzymes or alkaline phosphatase before DNA is further processed.

½ reaction volume phenol and ½ reaction volume chloroform- isoamylalcohol (24:1, v/v) were added. The suspension was then mixed by vortexing and centrifuged at 20,000g 10 min. After the supernatant was transferred into a new tube, it was supplemented with 1 volume isopropanol and 0.1 volume 3 M NaOAc (pH 5.4), mixed and centrifuged again to precipitate the DNA. The DNA pellets were washed with 70% EtOH and resuspended in dH_2O or buffer.

3.6 DNA amplification by Polymerase Chain Reaction (PCR)

3.6.1 General PCR method

PCR is an enzymatic method for *in vitro* synthesis of multiple copies of specific sequences of DNA. The reaction mixture for general PCR contained the following components:

PCR reaction mixture		
	10x PCR buffer	5 µl
	20 mM dNTPs	0.5 µl
	primer 1 (0,5-1 µg/µl)	0.5 µl
	primer 2 (0,5-1 µg/µl)	0.5 µl
	template-DNA (0.1-0.25 µg)	2 µl
	Taq polymerase (5 U/µl)	0.2 µl
	dH_2O	up to 50 µl

The 10x PCR buffer was made of 200 mM Tris-Cl, 500 mM KCl, 15 Mm $MgCl_2$ and 1% triton, filtered through 0.45 µm filter and stored at - 20°C. The PCR mixture was immediately incubated in a PCR machine for amplification using the cycling program as follows:

Program	"first delay"	5 min	94°C
30 cycles (steps 1-3)	1. denaturing	30 sec	95°C
	2. annealing	30 sec	55°C
	3. elongation	90 sec	72°C
	"last delay"	5 min	72°C

The annealing temperature is dependent on the composition of primers, normally 5-10°C below the melting temperature (T_m) of the primer. The time of elongation is determined by the polymerase activity and the length of the desired DNA fragment. The *Taq* polymerase can synthesize 1 kb of DNA in 30 seconds, while the *Pfu* polymerase can do it in 2 min.

3.6.2 RT-PCR

RT-PCR (reverse transcription-polymerase chain reaction) is a sensitive technique for mRNA detection and quantitation.

Reverse transcription was done using the Superscript II kit (Invitrogen). From the isolated total RNA, single stranded cDNA were generated by using oligo dT primers and the Superscript II reverse transcriptase. In a nuclease-free microcentrifuge tube, 1 µl of oligo dT primers (500 µg/ml) were mixed with 1-5 µg of total RNA and 1 µl of 10 mM dNTP Mix (10 mM each dATP, dCTP, dGTP and dTTP at neutral pH). RNAase-free water was used to adjust the volume to 12 µl. After 5 min heat shock at 65°C to denature any RNA secondary structures, the mixture was quickly chilled on ice and briefly centrifuged. The samples were supplemented with 4 µl 5x first-strand buffer, 2 µl 0.1 M DTT and 1 µl RNaseOUT™ recombinant ribonuclease inhibitor (40 units/µl). Then, the reaction was mixed gently and incubated at 42°C for 2 min. 1 µl of Superscript II reverse transcriptase (200 units) was added. The sample was incubated 50 min at 42°C and inactivated by heating at 70 °C for 15 min. The cDNA can was ready to be used as a template for amplification in a PCR reaction. In order to remove the RNA complementary to the cDNA and amplify large PCR targets (> 1 kb), 1 µl (2 units) of *E. coli* RNase H was added and incubated at 37°C for 20 min.

3.6.3 Linker-PCR

This method is utilized for the amplification of unknown end-fragments from cosmid-, BAC- or YAC clones and was modified on the base of the method of Riley (Riley *et al.*, 1990). The amplified end fragments can be directly used as probes for Southern blot hybridization-based chromosome walking without sequence analysis. BAC clones were digested in parallel with

several blunt-cutting restriction enzymes. The digested DNA was precipitated with isopropanol, washed with 70% ethanol and redissolved in dH_2O. DNA was then ligated with the partial double-strand linker-mix (1 µM of oligonucleotide LIS1 + 1 µM of oligonucleotide LIS2, annealed at 68°C and slowly cooled down at room temperature, stored at –20°C). The subsequent PCR was performed with 20 pmol of a vector-specific primer and 2 pmol of primer LIS1 (Table 1). Five enzymes, *Rsa*I, *Hae*III, *Eco*RV, *Hin*cII and *Pvu*II were routinely used. However, if no end fragment was obtained or the size of amplification was not within a 400-1,000 bp range, *Alu*I (not for BAC clones, because *Alu*I is within the cloning site and would separate the vector from the end fragment), *Sca*I, *Sma*I, *Stu*I, *Dra*I and *Eco*47III were used. The PCR program of the linker-PCR is: 1* [3'00'' at 95°C], 35* [0'30'' at 95°C/ 0'30'' at 60 °C/ 1'00'' at 72°C], 1* [5'00'' at 72°C], 4°C.

Table 1. Oligonucleotides (INVITROGEN) used for the linker-PCR.

Name	Sequence	Characteristics
SP6	5′-CGATTTAGGTGACACTATAGA-3′	Primer for vector Lawrist 7
T8	5′-CGGAAGCTGCGGCCGCGGTA-3′	Primer for vector Lawrist 7
FOR(-47)	5′-CGCCAGGGTTTTCCCAGTCAC G-3′	Primer for vector pBAC-lac
FOR-*Hin*d	5′-AAAACGACGGCCAGTGCCAAG-3′	Primer for vector pBAC-lac
REV(-48)	5′-AGCGGATAACAATTTCACACAGGA-3′	Primer for vector pBAC-lac
REV-*Hin*d	5′-GTCGACCTGCAGGCATGCAAG-3′	Primer for vector pBAC-lac
LIS 1	5′-GCGGTGACCCGGGAGATCTGAATTC-3′	Primer for linker-PCR
LIS 2	5′-GAATTCAGATC-3′	oligonucleotide for linker-mix
LIS1::LIS2	5′-GCGGTGACCCGGGAGATCTGAATTC-3′ 3′-CTAGACTTAAG-5′	linker-mix (LIS 1 annealed with LIS 2)

3.7 Electrophoresis techniques

Generally used electrophoresis buffers:

50x TAE		
	Tris	242 g
	acetic acid	57.1 ml
	EDTA (0.5 M, pH 8)	100 ml
	dH_2O	up to 1 Liter

The pH was adjusted to 8.0. 1x TAE is usually used as the working solution.

5x TBE	Tris	54 g
	boric acid	27.5 g
	EDTA (0.5 M, pH 8)	20 ml
	dH_2O	up to 1 Liter

The pH was adjusted to 8.3. 1x TBE is usually used as the working solution.

EtBr	stock concentration	10 mg/ml
	working concentration	10 mg/l

3.7.1 DNA agarose gel electrophoresis

6x DNA loading buffer	bromophenol blue	0.01 g
	xylene cyanol	0.02 g
	sucrose	4.00 g
	dH_2O	up to 10 ml

The loading buffer is stored at 4°C.

Agarose gel electrophoresis was employed to check the progression of a restriction enzyme digestion, to quickly determine the yield and purity of a DNA isolation or PCR reaction, and to size-fractionate DNA molecules, which then could be eluted from the gel. Agarose gels were prepared according to the protocol described by Sambrook and Russell (2001). DNA samples were electrophoresed in 1x TAE (or 1xTBE) buffer and ethidium bromide was added to a final concentration of 0.5 µg/ml.

3.7.2 Pulsed field gel electrophoresis

Pulsed field gel electrophoresis (PFGE) is a technique allowing the separation of large DNA molecules (up to several Mb) (Anand, 1986; Schwartz and Cantor, 1984).
In this special type of gel electrophoresis, the direction of the electrical field is changed mostly in 120° angle through the moving of the longitudinal axis of gel within certain time interval and tension. The DNA molecule is considered as a flexible rod which ´´ratchets`` its way through an agarose gel. One end of the rod serves as the front of the molecule as it migrates through the gel in one field orientation, and then as the tail, as the field orientation is switched, and the molecule moves off in the new direction. Separation is achieved because smaller DNA molecules reorient themselves more rapidly than larger ones following a switch

in field direction, and are able to move faster than larger DNA molecules in a net forward direction (Southern *et al*., 1987).

In this work, the STRATAGENE RAGE™ (Rotating Agarose Gel Electrophoresis System) was used to determine the size of BAC clones. All parameters were following the attached protocol (Program C4: 3-160 sec pulse time, exponential raising, 220-150 V, logarithmic falling, 20 h at 10°C).

3.8 DNA isolation from agarose gels

DNA fragments were cut from agarose gel and isolated by different methods for further cloning, sequencing or as hybridization probes. In this work, commercial kits like QIAEX® II Gel Extraction Kit (QIAGEN) and Concert™ Rapid Gel Extraction System (INVITROGEN) were used for DNA isolation from agarose gel.

There are other methods for DNA isolation from agarose gel. One is to freeze the gel slices and manually squeeze out the liquid containing the DNA (Thuring *et al*., 1975). Another is to centrifuge through glass wool (Heery and Powell, 1990).

3.9 ´´Shot-gun`` and ´´transposon tagging`` sequencing

3.9.1 ´´Shot-gun`` sequencing

Shotgun cloning / sequencing is a strategy used for the sequencing of large DNA molecules. The target DNA is first fragmented by enzymatic digestion or physical shearing to fragment sizes in the range of 0.5 – 5 kb. Subfractions with a narrower size range (e.g. 0.8 – 1.5 kb) are subcloned into either a plasmid or a M13 vector. The subclones can then be sequenced from standard primer binding sites in the flanking vector DNA. By doing this for numerous subclones from a same large DNA molecule (400-600 subclones are required to sequence a 40 kb large cosmid insertion and 1500-1800 may be required to sequence a 100-120 kb large BAC clone), overlapping fragments are generated. Finally, computer programs align these overlapping sequences and determine the sequence of the original molecule.

In this study, the target DNA (500-1000 ng of BAC DNA) and the plasmid vector pBS or pJOE2114 (1 µg) were first digested with a same enzyme (10 U). The DNA fragments were then extracted by phenol-chloroform and precipitated by isopropanol. Before the fragments of target DNA were inserted into the vector, the ends of the vector were treated with alkaline phosphatase to remove the phosphate groups from the 5´ ends to prevent self-ligation. After

ligation, the reaction mixture was transferred into *E. coli* competent cells. Positive clones were picked up from selective plates and inserts were sequenced using primers matching the vector.

3.9.2 ´´Transposon tagging`` sequencing

The ´´Transposon tagging`` sequencing strategy allows to precisely sequence total DNA molecular without subcloning small fragments (Fischer *et al*., 1996). The plasmid vector pJOE2114 used in this strategy contains two 180 bp DNA sequences in inverted orientation (*res*-sites), which are necessary to resolve the cointegrate of the plasmid and transposon TnSeq2. Between these palindromic sequences, there is a spacer fragment (100 bp), which can be replaced by the fragment to be sequenced. The plasmid pJOE2114 did not replicate in *E. coli* without the fragment between the inverted repeats. The transposon TnSeq2 is derived from Tn*1721*, which has a low insertion specificity and belongs to the family of Tn*3* transposons. New restriction enzyme sites (*Eco*RI and *Sca*I) and a M13 binding site have been added to the transposon.

After target DNA fragment was cloned in pJOE2114 (ampicillin-resistant) and proliferated in *E. coli* XL1-blue, the plasmid DNA was transformed into *E. coli* JM108 F' Lac (containing one F' Plasmid with tetracyclin-resistance and the transposon TnSeq2). Then the transformation suspension was incubated at 30°C overnight on ampicillin agarose plate to select the transformants. At this temperature, fusionproducts (F'::TnSeq2 and pJOE2114) were formed. The selected clones were suspended in 5 ml LB medium. This suspension and an overnight-culture of *E. coli* HB101 as conjugation recipient were mixed at 1:1 (according to the OD_{600}) in a total volume of 100 µL. The mixture (JM108 F' Lac and *E. coli* HB101) was incubated overnight in 5 mL LB medium without antibiotic at 30°C on a slow shaker. At this step, only cointegrates of F'::TnSeq2 and pJOE2114 could be transferred into HB101. In HB101, cointegrates were then resolved, removing most of the transposon and part of the insertion by a site-specific recombination event between the *res* site of the vector pJOE2114 and the *res* site of TnSeq2. After this deletion, only the restriction enzyme cutting sites (*Eco*RI and *Sca*I) and the sequencing primer binding site of the transposon TnSeq2 were left, together with partial insertion in pJOE2114. In order to pick up the clones with the resolved cointegrate, the culture was diluted from 10^{-1} to 10^{-4} (100 µL for each). The dilution was incubated overnight on plates with streptomycin (*E. coli* HB101 is streptomycin-resistant) and ampicillin (pJOE2114 is ampicillin-resistant).

The size of insertions were analyzed by *Eco*RI digestion. The orientation of transposon insertions was identified by *Sca*I restriction. After the insertion map was set up, some positive clones were selected to be sequenced. Sequence contigs were established *in silico* and the sequence of the original fragment was determined.

3.10 Automated DNA sequencing

DNA sequencing was done using an automatic sequencer (CEQ^{TM} 2000XL, Beckman Coulter) based on a method derived from the chain-termination method developed by Sanger and coworkers in the late 1970s (Sanger *et al*., 1977). Sequencing of DNA was achieved by generating fragments through the controlled interruption of enzymatic replication. DNA polymerase I was used to copy a particular sequence of a single–stranded DNA. The synthesis was primed by a complementary oligonucleotide. In addition to the four deoxyribonucleoside triphosphates (ddNTP), the incubation mixture contains a 2`, 3`-dideoxy analogue of one of them. The incorporation of this analogue blocks further extension of the new chain because it lacks the 3`- hydroxyl terminus needed to form the next phosphodiester bond. A fluorescent tag is attached to the dideoxynucleotides (dye terminator). The reaction mixtures are combined and electrophoresed on a denaturing high-resolution polyacrylamide gel. The separated bands of DNA are detected by their fluorescence as they pass out the bottom of the tube, and the sequence of their colors directly yields the base sequence. A high sensitivity fluorescence detector measures the amount of each fluorophore as a function of time and hence sequence is determined from the order of peaks of four different dyes. The thermal-cycle reaction for sequencing includes 15 µl of DNA template (200-1000 ng), 1 µl of 5 pmol primer and 4 µl of sequencing mix (provided by Beckman Coulter in the CEQ Dye terminator cycle sequencing kit). The thermal cycling is: 96°C/20s; 50°C/20s; 60°C/4 min, for 30 cycles followed by holding at 4°C. The reactions were cleaned using sodium acetate (pH 5.2) and ethanol precipitation as recommended by the manufacturer.

3.11 Southern blot hybridization

Southern blot hybridization, invented by the British biologist E.M. Southern (Southern, 1975), is a method for the permanent immobilization of single-stranded nucleic acids to a solid support. The procedure in this work is as follows:

1) Radioactive labeling of the DNA probe

The DNA probes can be obtained through different ways, including PCR (generated from linker-PCR or target sequence amplification) or cutting of cloned fragments from vectors by restriction enzymes. Probes can be purified from electrophoresis gel.

50-100 ng DNA probe is denatured in 33µl volume (dH_2O) at 95°C for 10 min. After 10 µl OLB-Mix [1 M Hepes, 250 mM Tris-Cl, 25 mM $MgCl_2$, 0,35% Mercaptoethanol, 100 µM dATP, 100 µM dGTP, 100 µM dTTP, OD 27/ml Hexanucleotide ($pd(N)_6$, AMERSHAM™)] was added, the tube with the probe was put on ice for 5 min to allow the annealing of the hexanucleotides to the probe. Then 3 µl of BSA (5 mg/ml), 3 µl of [α^{32}P]dCTP (nominative 30 µCi, HARTMANN ANALYTIC™) and 4 U Klenow-Polymerase were added and the reaction was subsequently incubated at 37°C for 2h.

The synthesized DNA fragments were purified by centrifugation using Sephadex-G50 beads. The efficiency of labeling can be measured using a scintillation counter (MR 300, KONTRON).

2) Alkaline transfer from DNA to nylon membrane

DNA was digested with restriction enzymes and separated on a 0.8-1% agarose gel after electrophoresis. A nylon membrane (Hybond™-N+, AMERSHAM) was cut and firstly immerged in distilled water for DNA transfer.

The agarose gel was incubated twice in 0.25 M HCl for 15 min on a slow shaker to make the long DNA fragments easier to transfer through depurination. Then the agarose gel was incubated twice in 0.5 M NaOH/ 1.5 M NaCl for 15 min to denature DNA. The gel was put on a glass plate and was covered with the membrane and gel-blotting papers (SCHLEICHER & SCHÜLL™), which have been equilibrated in the same buffer. 10-15 dry gel-blotting papers and pulp-papers are added to create capillarity. Another glass plate with weight was placed on the top.

For the fast transfer of plasmid-, cosmid-, or BAC-DNA, two hours are sufficient and buffer in the gel (250-300 ml/ 400 cm^2) allows the transfer.

For genomic DNA transfer, the gel was put on 2-3 gel-blotting paper sheets, which ends were immerged in the transfer buffer (0.2 N NaOH) to keep a continual fluid. The transfer time was at least 6 hrs or overnight.

The wells of the gel were marked on the membrane with a pencil after transfer. The wet membrane was exposed to UV-light (150 mJ, 254 nm, Gs Gene Linker®, BIORAD™) to enhance DNA binding. The membrane was then dried between two gel-blotting papers.

3) DNA-DNA hybridization

In order to eliminate non-specific hybridization, the membrane was prehybridized with hybridization buffer (35-50% formamide, 5x Denhardt's solution, 5x SSC, 1% SDS, 0.1% Na-Pyrophosphat, 50 mM Tris-Cl, pH 8, 0.1 µg/ml denatured calf sthymus DNA) in a rolling cylinder at 42°C overnight. The membrane was then hybridized overnight with 5-10 ml fresh hybridization buffer containing the labeled and denatured probe.

Hybridization specificity was controlled through several wash steps with different stringencies: after hybridization the buffer was removed, 10 ml 2x SSC/1% SDS was added and incubated in the same rolling cylinder at 42°C for 15 min. The following wash steps were performed with increasing stringency with 50-55°C in 2x SSC/1% SDS for heterologous probes or 55-65°C in 1x SSC/1% SDS for homologous probes. After each wash step, the hybridized membrane was exposed to a X-ray film to check the intensity of the signal. The probes can be washed away from the membrane in 0.1x SSC/1% SDS at 68-72°C for 30 min.

3.12 Chromosome walking

This method is used to move systematically along a chromosome from a known location and to identify overlapping genomic clones covering parts of a chromosome. Chromosome walking is used to identify gene adjacent to a molecular marker (positional cloning) and to analyze long stretches of genomic DNA.

A short segment of DNA from one end of a given genomic subclone is used as a probe to isolate overlapping clones. The end sequences of this second clone are used to isolate new overlapping clones and so forth until a series of overlapping clones are isolated (Fig. 3). It is preferable to use single-copy DNA probes and avoid repeated sequences. A chromosome walking can be bidirectional. The end sequences of the BAC clones used in this work were generated by linker-PCR (see Chapter 3.6.3).

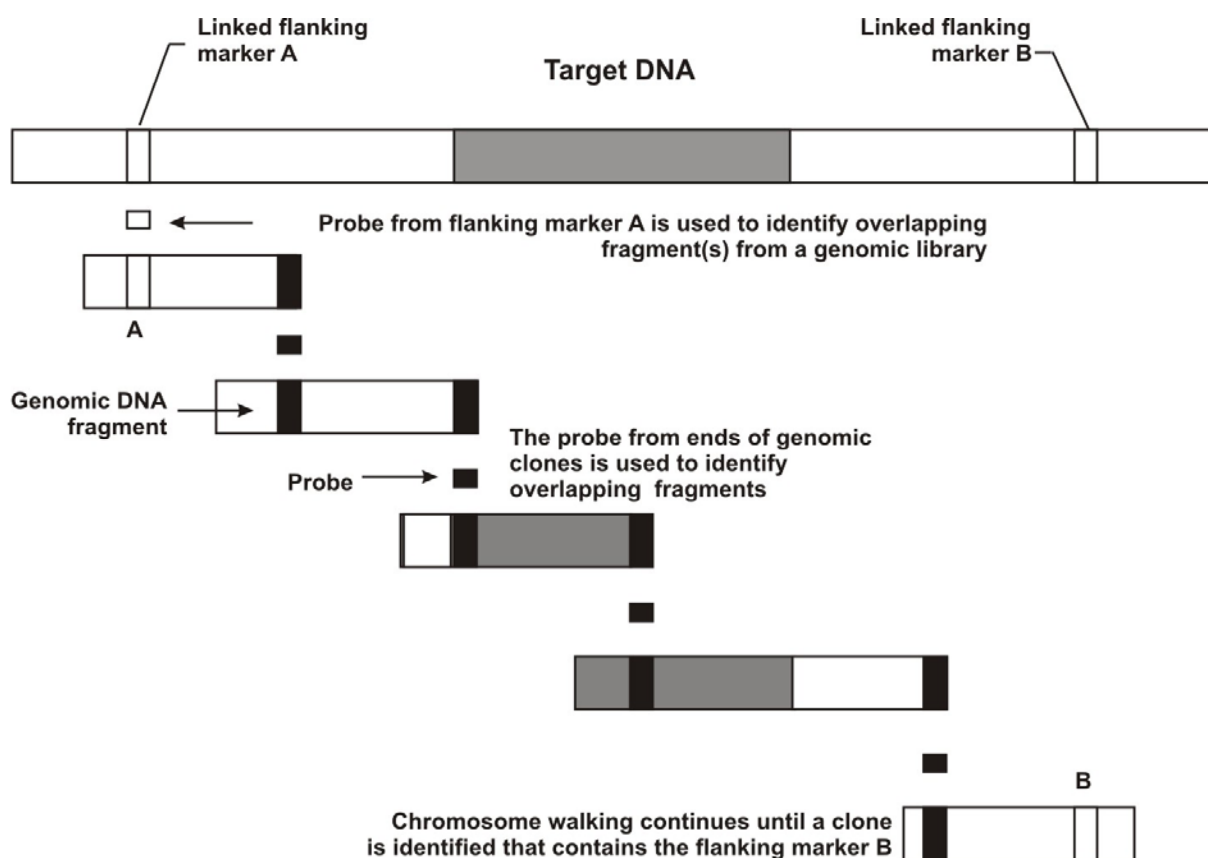


Fig. 3. Schematic chromosome walking by clone to clone hybridization. In this chart, the starting point is a flanking marker A and a monodirectional chromosome walk (towards the right) is shown by using end fragments as hybridization probes for screening a genomic DNA library.

3.13 RNA whole-mount *in situ* hybridization

In situ hybridization is used to detect specific mRNA sequences in tissue sections or cell preparations by hybridizing the complementary strand of a nucleotide probe to the sequence of interest.

1) Preparation of riboprobes (antisense RNA) for *in situ* hybridization

This procedure is used to obtain antisense RNA strands that contain digoxigenin (Dig) or fluorescein (Flu) labeled nucleotides. The antisense probes bind to complementary endogenous sense mRNA strands, and can be detected using *in situ* hybridization. Using an antibody directed against the labeled nucleotides, immunostaining is used to detect the expression of the desired gene. To obtain riboprobe RNAs, the following components were used in a final volume of 20 µl: 2 µl of the linearized template (approximately 1 µg) were mixed with 2 µl 100 mM DTT, 2 µl Dig/Flu labeling NTP mix, 0.5 µl RNase inhibitor, 4 µl transcription buffer, 8.5 µl H_2O and 1 µl T7, T3 or SP6 RNA polymerase (20 U/µl) depending on the probe. The sample was gently mixed and incubated for 2 h at 37°C. Then 1 µl DNase

RQ1 was added and the mix was incubated for 15-30 min at 37°C. After the RNA was cleaned up using the RNA probe purification kit (Peqlab Biotechnologie GmbH), the purified RNA was precipitated by adding 1/10 volume of 2 M NaOAc (5 µl), and three volumes of absolute ethanol (150 µl), and incubated for at least 30 min at -80°C. After centrifugation for 30 min (13000 rpm) at 4°C, the pellet was washed using 70% ethanol and dried. The pellet was dissolved in 26 µl RNase-free water. 1 µl was analyzed on a 1% TAE agarose gel. The remaining 25 µl were mixed with 75 µl hybridization mix and stored at -20°C.

2) Preabsorption of antibody

Preabsorption is a procedure blocking unspecific binding activities of antibodies used in *in situ* hybridization. For this procedure, different stages of zebrafish embryos were collected. Usually 24 h old embryos were used. The embryos were fixed in 4% paraformaldehyd (PFA) at 4°C overnight, then washed 4 x 5 min with PBST, and stored in MetOH at -20°C. Without dechorionating, embryos were rehydrated by rinsing 3 x 5 min in PBST. Then embryos were transferred to 2 ml eppendorf tubes, homogenized using a pestle and the homogenate was adjusted to 1 ml PBST. 10 µl of antibody (anti-Dig or anti-Flu) were added and incubated for at 4°C. After spinning down, the supernatant was sterile filtered using a cellulose acetate filter. The pellet of embryonic debris was resuspended in PBST, centrifuged and the supernatant sterile filtered again. The filtered solution was adjusted with PBST to a final 1:2000 dilution of antibody (20 ml final volume) and stored at 4°C for a maximum of four weeks. The antibody solution can be reused twice.

3) RNA whole mount *in situ* hybridization

Adult testes were removed to 8 ml glass vials and fixed in 4% PFA in PBST at 4°C overnight or up to 6 h at room temperature. After fixation, testes were washed 4 times for 5 min in PBST at room temperature. For dehydration, testes were washed three times in 100% methanol for 5 min and stored at -20°C for at least overnight.

Adult testes were then rehydrated through three consecutive washing steps with 75%, 50% and 25% methanol in PBST. Afterwards, the testes were washed 4 times for 5 min in PBST and were treated with proteinase K in PBST (5 µg/ml) for 30 min. The testes were then briefly rinsed twice in 2 mg/ml glycine in PBST to stop digestion and fixed in 4% PFA in PBST for 20 min at room temperature. The PFA was removed by washing 5 times for 5 min in PBST.

Subsequently, testes were prehybridized in 1 ml hybridization buffer at 65°C for at least 1 h. In the meantime, the hybridization mix containing the labeled probes at 1:100 to 1:300 dilutions was heated at 80°C for 10 min and then chilled on ice to remove secondary structures. The hybridization solution was replaced by the probe solution and the testes were hybridized overnight at 65°C.

The hybridization probe was removed and stored at -20°C for later use. All following steps were performed at 65°C for 30 min. The testes were subsequently washed twice in 50% formamide in 2x SSCT, once in 2x SSCT and twice in 0.2% SSCT. The testes were briefly rinsed in PBST. Then, blocking solution (5% sheep serum in PBST) was added and incubated for at least 1 h on a rocking shaker. The blocking solution was replaced by 500 µl of preabsorbed sheep anti-digoxigenin (DIG) or anti-fluorescein (FLU) Fb antibody fragments conjugated with alkaline phosphatase in a 1:2000 dilution in PBST. The testes were then incubated for 2 h at room temperature or overnight at 4°C with gentle shaking. Afterwards, the antibody solution was removed and stored at 4°C for later use. The testes were washed 6 times for 20 min each in PBST. For a red color reaction, the testes were then washed in 0.1 M Tris-Cl pH 8.2/0.2% Tween and incubated in staining solution prepared by dissolving one fast-red tablet in 2 ml Tris-Cl buffer, pH 8.2. For blue staining, the testes were washed 5 min in SB staining buffer and then the SB solution was replaced by the staining solution, which was prepared by dissolving one NBT/BCIP tablet in 10 ml water. Both staining reactions were performed under protection from light at room temperature or 4°C. The reaction was stopped by washing the testes 3 times in PBST (hybridization buffer: Formamide, 50%; SSC, 5x; Heparin, 50 µg/ml; Tween-20, 0.1%; Torula RNA, 5 mg/ml and staining buffer: NaCl, 0.1M; MgCl , 0.05M; Tris-Cl, pH 9.5, 0.1 M; Tween-20, 0.1%).

4) Embedding and section of the stained samples

The stained tissue samples were fixed in 4% PFA overnight and then dehydrated with ethanol. About half of the fix solution was removed and an equal amount of 98% ethanol was added. The sample was incubated at room temperature for 5-10 min. This removing and adding procedure was repeated three times. Then all solution was removed and replaced by 100% ethanol. After incubation for 10 min, about half ethanol was removed and replaced by an equal amount of clearing agent (such as xylol). After incubation for 15 min, all mixed solution was removed and replaced by the clearing agent xylol. After incubation for 10-20 min, all xylol was removed. The sample was paraffin-embedded by a pre-warmed wax-xylol mixture

(1:1, vol/vol) and incubated at 45°C for 15 min. Then all wax mixture was replaced by pre-warmed fresh wax and the sample was incubated at 60°C for 30-60 min. The wax with sample was allowed to solidify. The wax block with the sample was put on the microtome stage and cut into sections at 2-4 µM. Sections were flattened on 45°C water and floated onto a clean microscope slide. Slides were put into an oven to dry at 45-50°C overnight. The dried sections were placed in xylol for 5 min. After three times of this washing procedure, wax was totally washed away. Finally, sections were removed from the xylol and mounted under a cover glass using suitable mounting media (such as pertex and roti histokitt). After sections were dry, they could be viewed under the microscope.

4. Results

4.1 Chromosome walking on the sex determining region of the Y chromosome of *Xiphophorus maculatus*

In order to identify the master sex-determining gene of *Xiphophorus maculatus* through positional cloning, a Bacterial Artificial Chromosome (BAC) library of XY males was constructed. BAC contigs were assembled using the oncogene *Xmrk* and the proto-oncogene *egfrb* as starting points, which both flank the *SD* locus of *X. maculatus* (Froschauer *et al.*, 2002).

4.1.1 Construction of the *ps-criptY* BAC contig of *Xiphophorus maculatus*

In the region 3` from *Xmrk*, a truncated *cript* gene has been identified. Subsequent analysis revealed two additional copies of *cript*, one autosomal and one Y-specific (*ps-criptY*). *Ps-criptY* is a male-specific pseudogene and is closely linked to *SD*, since no recombination between the *SD* locus and the *ps-criptY* pseudogene was observed in more than 400 individuals tested (Vucic, M. *et al*., pers. commun.). Therefore, a new Y-specific BAC contig was assembled using *ps-criptY* as a starting point (Fig. 4).

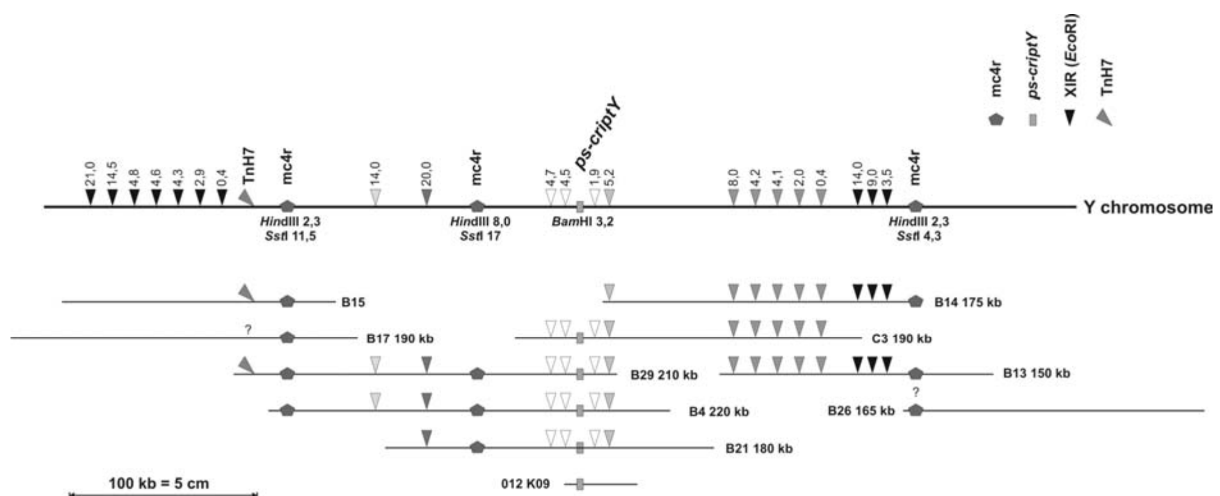


Fig. 4. The initial putative structure of the *ps-criptY* BAC contig. The size of Y-specific fragments is indicated in kb for *mc4r* and *ps-criptY*. The size of characteristic *Eco*RI fragments containing the XIR repeat is also given.

4.1.2 Order of BAC clones in the *ps-criptY* contig

The order of BAC clones in the *ps-criptY* contig was determined by Southern blot hybridization based on the restriction fragment length polymorphism (RFLP) analysis using probes generated from the terminal regions of BAC B17, B29, B14 and B26 (Fig. 5). The *ps-criptY* contig appears contiguous, since the BAC clones B17 and B29, B29 and B14, B14 and B26 overlap each other.

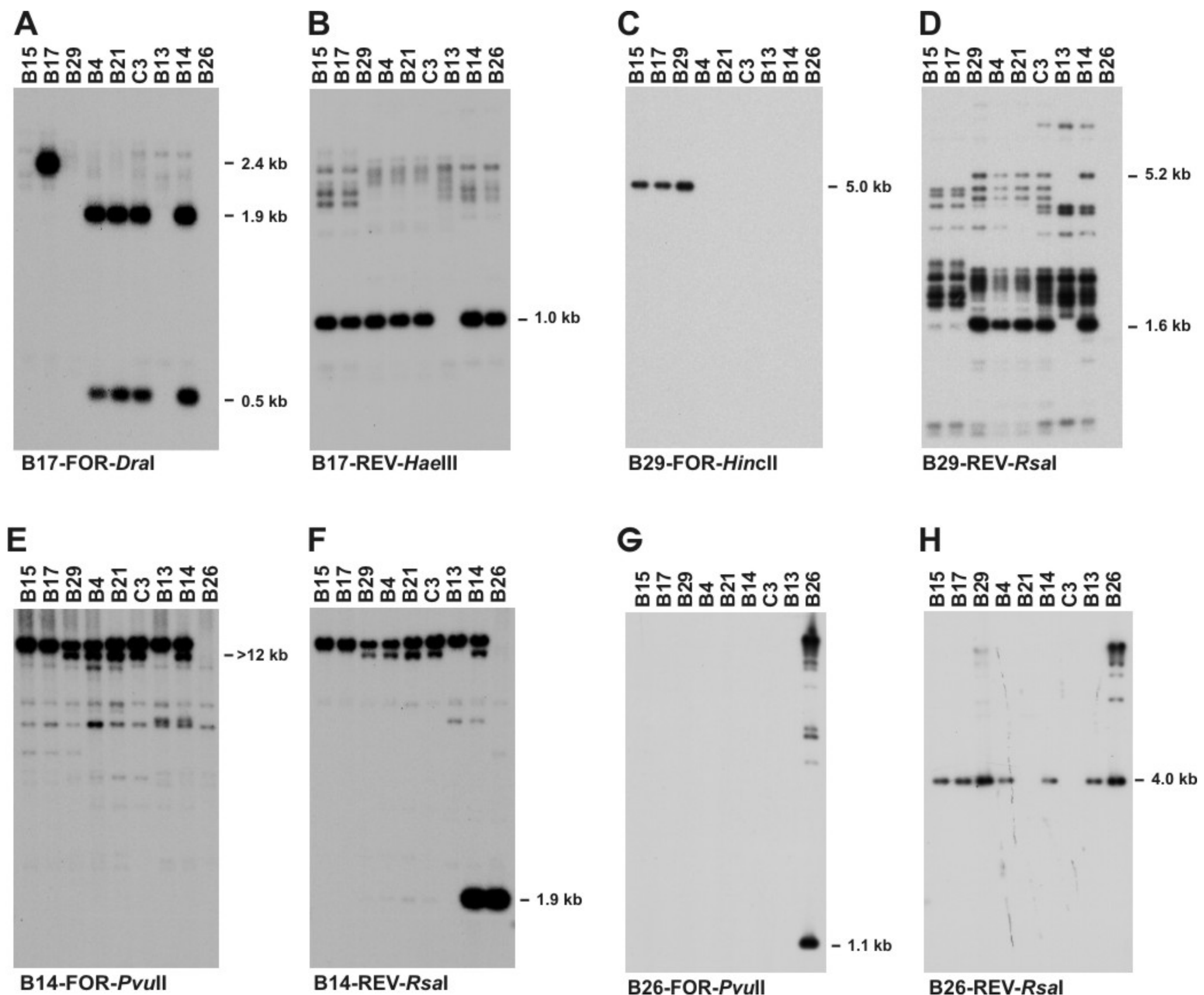


Fig. 5. Analysis of the order of BAC clones in the *ps-criptY* contig of *Xiphophorus maculatus* by end-probe DNA hybridization. BAC clones from the *ps-criptY* contig were hybridized with probes generated from the terminal segments of B17 (A, B), B29 (C, D), B14 (E, F) and B26 (G, H). BAC DNAs were digested with *Hin*dIII and *Eco*RI (A-D) and *Hin*dIII (E-H).

4.1.3 Extension of the *ps-criptY* BAC contig

In order to extend the *ps-criptY* contig, probes were generated from end fragments of BAC clones from the contig. These probes were tested by Southern blot hybridization on male and female genomic DNA and used to screen the BAC library of *Xiphophorus maculatus*. After

identification of positive clones, their terminal sequences were amplified by linker-mediated PCR and subsequently used for the next screening and identification of overlapping clones. Contigs were built and extended after analysis of different data sets including Southern blot hybridization, PCR- and sequence analysis of overlapping clones.

4.1.3.1 Extension from the left side of the *ps-criptY* contig

The terminal fragment B17-FOR-*Dra*I was chosen as a probe for the next chromosome walking from the results of RFLP analysis (Fig. 4 and 5A). 45 new BAC clones (named N series) were selected from the BAC library screening, 25 of them showed the same size band as the BAC B17 using B17-FOR-*Dra*I as a probe. However, none of them contained the 2.1 kb-long Y-specific band detected in male genomic DNA but also absent from B17 (Fig. 6).

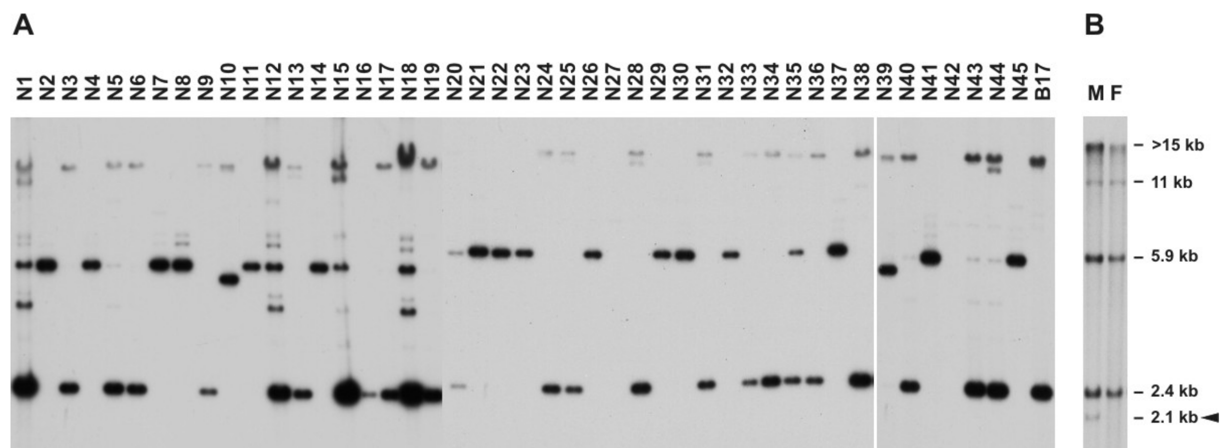


Fig. 6. Southern blot hybridization of BAC clones (A) and genomic DNA from male (M) and female (F) of *X. maculatus* (B) with the probe B17-FOR-*Dra*I. The BAC clones in (A) are positive for B17-FOR-*Dra*I in the BAC library screening. The Y-specific band is indicated by a black triangle (2.1 kb). All BAC clones and genomic DNA were digested with *Hin*dIII.

Since the terminal fragment B17-FOR-*Dra*I is not Y-specific, another probe (named Hel-protease from BAC B17, Körting, C., unpublished data) was used to identify the overlapping of BAC B17 and N-series BAC clones (Fig. 7A-B). This probe revealed a Y-specific fragments of 9.0 kb also present in B17. The RFLP result indicated the overlapping between BAC N20 and B17, which both contain the 9.0 kb fragment. This overlapping was confirmed using the terminal fragments of BAC N20 as a probe (Fig. 7C-D). Additionally, BAC clone B14 shows a band of the same size as N20 using the end fragment N20-REV-*Rsa*I as a probe (Fig. 7D).

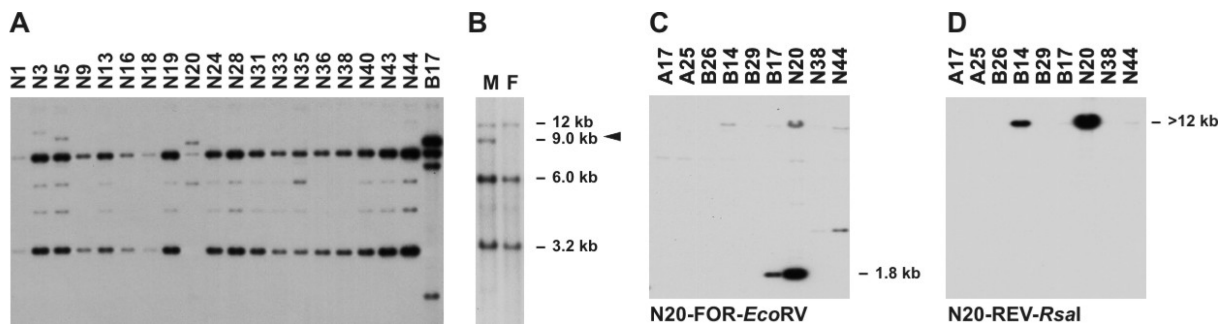


Fig. 7. Demonstration of the overlapping between BAC clones B17 and N20. All BAC clones (A, C-D) and genomic DNA from male (M) and female (F) *X. maculatus* (B) were digested with *Hin*dIII and hybridized with a 1.2kb-insertion subclone containing the Hel-protease sequence (A), the Hel-protease probe (B) and terminal fragments of BAC N20 (C-D). The BAC clones in (A) are positive for the B17-FOR-*Dra*I probe in the BAC library screening and show the same size band as BAC B17 (Fig. 6). The Y-specific band is indicated by a black triangle (9.0 kb). Except for the 9.0 kb-long band, the other bands on the BAC clone B17 were resulted from the probe containing the TOPO vector.

The chromosome walking could be continued using the terminal fragment N20-REV-*Rsa*I as a probe. 24 new BAC clones (R series of BAC clones) and some BAC clones from N series were positive in the BAC library screening with this probe. Among them, BAC R1 and R19 were identified to contain the same Y-specific bands as those identified in BAC N20 (Fig. 8A-B).

Both terminal fragments of BAC R1 hybridized with fragments of the same size in BAC N20 (Fig. 8D-E). BAC clone R1 is shorter than N20 and shows the same size bands as N20 after restriction enzyme analysis (Fig. 8C). Therefore, BAC R1 is contained in BAC N20. Similarly, BAC R19 DNA is contained in BAC B14 (Fig. 8C, F-G). Thus, no BAC clone, which overlaps with BAC N20 and extends from the *ps-criptY* contig, could be obtained from the *X. maculatus* BAC library.

In order to obtain a genomic clone extending the *ps-criptY* contig, the *X. maculatus* cosmid library (Burgtorf *et al*., 1998) was screened with the N20-REV-*Rsa*I probe. 14 new cosmids were isolated and 8 of them were shown to contain one Y-specific band also present in N20 (Fig. 9).

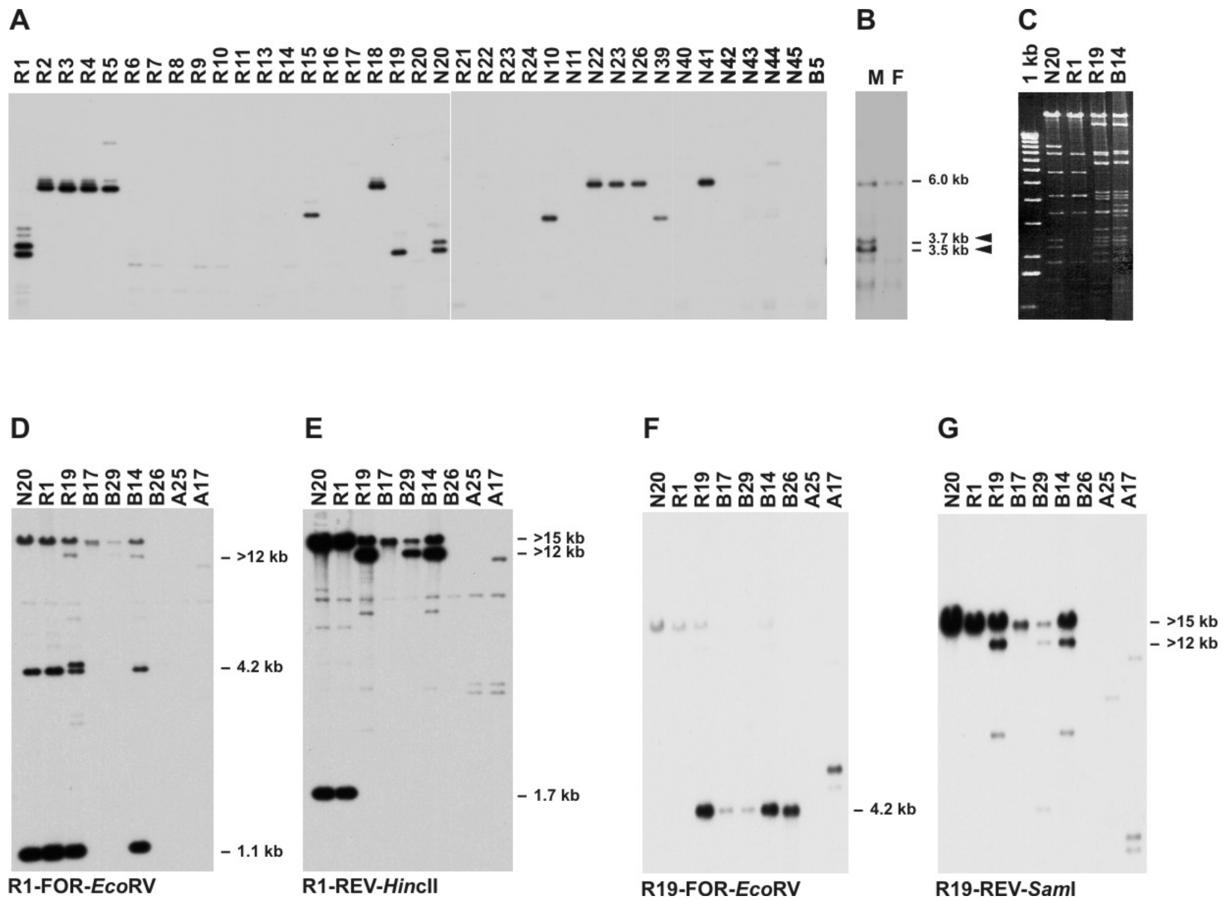


Fig. 8. Southern blot hybridization analysis of BAC clones (A, D-G) and genomic DNA from male and female *X. maculatus* (B) with the N20-REV-*Rsa*I probe (A-B), end fragments of R1 (D-E) and R19 (F-G). The BAC clones in (A) were digested with *Hin*dIII and *Eco*RI, and positive for the N20-REV-*Rsa*I in the BAC library screening. The genomic DNA was digested with *Hin*dIII and *Eco*RI. The Y-specific bands are indicated by black triangles (3.5 and 3.7 kb). The other BAC DNAs (D-G) were digested with *Hin*dIII. (C) is the restriction profile of BAC N20, R1, R19 and B14 digested with *Hin*dIII. 1 kb represents 1 kb ladder.

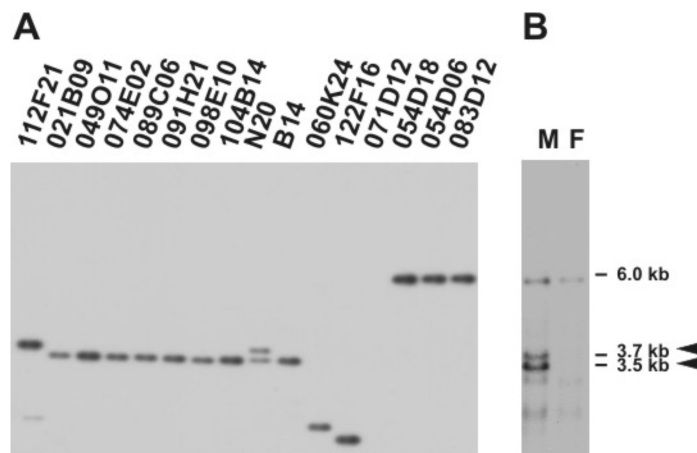


Fig. 9. Southern blot hybridization analysis of cosmid clones (A) and genomic DNA from male (M) and female (F) *X. maculatus* (B) with the probe N20-REV-*Rsa*I. The cosmid clones in (A) were digested with *Hin*dIII and *Eco*RI, and positive for N20-REV-*Rsa*I in the cosmid library screening. The genomic DNA was digested with *Hin*dIII and *Eco*RI. The Y-specific bands are indicated by black triangles (3.5 and 3.7 kb).

BAC N20 has two Y-specific *Hin*dIII-*Eco*RI bands of the end fragment N20-REV-*Rsa*I (Fig. 9). Which band is at the terminus of N20? The RFLP analysis indicated BAC B14 shows the same size *Hin*dIII fragment (> 12 kb, Fig. 7D) and *Hin*dIII-*Eco*RI fragment (3.5 kb, Fig. 9A) of the end probe N20-REV-*Rsa*I as N20 and has not the 3.7 kb *Hin*dIII-*Eco*RI fragment (Fig. 9A). That is to say, the end fragment N20-REV-*Rsa*I was duplicated on BAC N20, but not on B14. Hence, the 3.5 kb *Hin*dIII-*Eco*RI fragment, which was shared by B14 and N20, must be at the terminus of N20 reversal region and another *Hin*dIII-*Eco*RI fragment (3.7 kb) is away from the terminus but within the *Hin*dIII fragment (> 12 kb, Fig. 7D) of N20. Thus, 7 cosmids containing the 3.5 kb band might overlap either with BAC N20 or with B14 (Fig. 9A). The cosmid 112F21 with the 3.7 kb *Hin*dIII-*Eco*RI fragment is contained in BAC N20 for the absence of the 3.5 kb *Hin*dIII-*Eco*RI fragment. But apart from cosmids 021B09 and 049O11, the terminal fragments of the other cosmids either could not be obtained by linker-PCR (such as 074E20, 089C06, 091H21 and 098E10) or contained repeated sequences (such as 104B14). Unfortunately, the RFLP analysis of end fragments of the cosmids 049O11 and 021B09 do not show that cosmids 049O11 and 021B09 overlap with BAC clone N20 (Fig. 10).

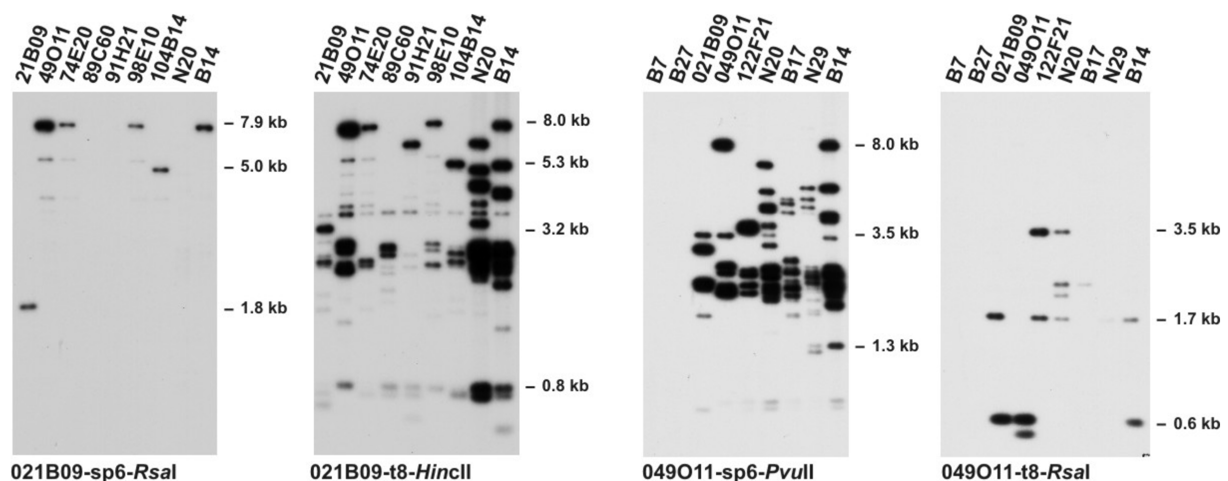


Fig. 10. Southern blot hybridization analysis of cosmid clones with the probes generated from the terminal fragments of cosmid 021B09 and 049O11. The cosmid clones were digested with *Hin*dIII and *Eco*RI and positive for the N20-REV-*Rsa*I in the cosmid library screening.

To conclude, no overlapping clone was found that extends the left side of the *ps-criptY* contig using the N20-REV-*Rsa*I probe (Fig. 19, left part).

4.1.3.2 Extension from the right side of the *ps-criptY* contig

Starting with the terminal fragment B26-FOR-*Pvu*II (Fig. 5G), the *ps-criptY* contig was extended through four steps of chromosome walking (Fig. 19, right part).

At the first step of chromosome walking, no Y-specific band was recognized with the terminal fragment B26-FOR-*Pvu*II as a probe for RFLP analysis (Fig. 11B). Nonetheless, B26-FOR-*Pvu*II probe, which is moderately repetitive, was used to screen the BAC library. Hundreds of new clones and 19 known clones were obtained. Four new BAC clones (named O series), which shared the same signal with one end fragment of the X-linked BAC clone B33 as a probe (data not shown), were isolated from the BAC library. All known positive BAC clones were tested on one membrane. The RFLP results suggested that B26 might overlap with F11 and H93 (Fig. 11A, black triangle).

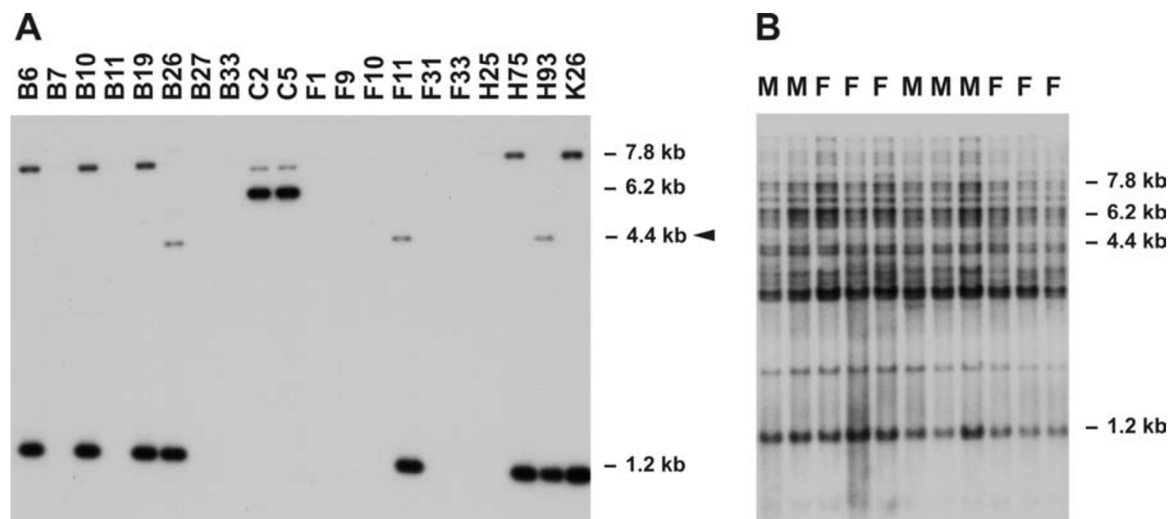


Fig. 11. Southern blot hybridization analysis of BAC clones (A) and genomic DNA from male (M) and female (F) *X. maculatus* (B) with the probe B26-FOR-*Pvu*II. All BAC clones were digested with *Hin*dIII and positive for the B26-FOR-*Pvu*II probe in the BAC library screening. Genomic DNAs were also digested with *Hin*dIII. The 4.4 kb band (black triangle) suggests that B26, F11 and H93 overlap.

The RFLP analysis results using the terminal fragments of BAC F11 and O2 as probes indicated that BAC B26 overlaps with F11, O2 and H93 (Fig. 12). In addition, BAC F11, F1 and F6 are known from a library screening with the *ps-rnase*III probe, which is predicted to encode a ribonuclease III and has several copies in the sex determination region of the platyfish (Froschauer, 2003). BAC F11 and O2 also contain Y-specific bands using other markers as probes, such as *T-Rex* (a non-LTR retrotransposon), *igd1* (encoding a putative immunoglobulin V-type domain protein) (Schultheis, 2003) and *swimy* (encoding a putative

DNA-binding protein) (Fig. 25). These data indicated that BAC B26 overlaps with BAC F11, O2 and H93.

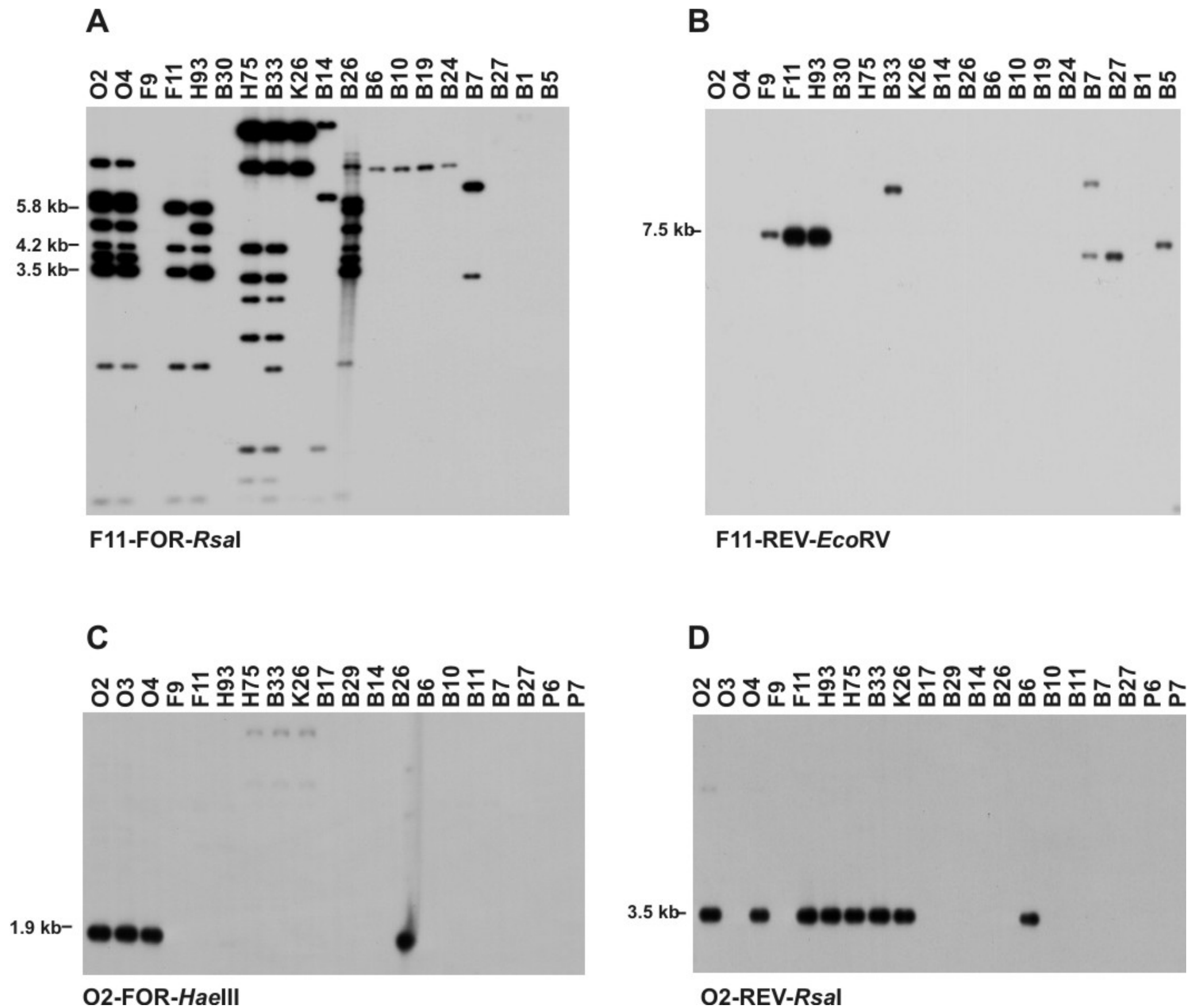


Fig. 12. Identification of BAC clones overlapping with BAC B26. The Southern blot hybridization results are shown for the end fragments of BAC F11 (A-B) and O2 (C-D). All BAC clones were digested with *Hin*dIII.

For the second step of chromosome walking, the probe F11-REV-*Eco*RV was used to screen for overlapping clones. The RFLP result indicated that this probe revealed a 3.0 kb Y-specific band in the *X. maculatus* genome (Fig. 13B). After BAC library screening, two BAC clones, F1 and H93, were found to contain this Y-specific band (Fig. 13). Using the end fragment of F1 as a probe, BAC F11 was shown to overlap with F1 and H93 (Fig. 13C-D).

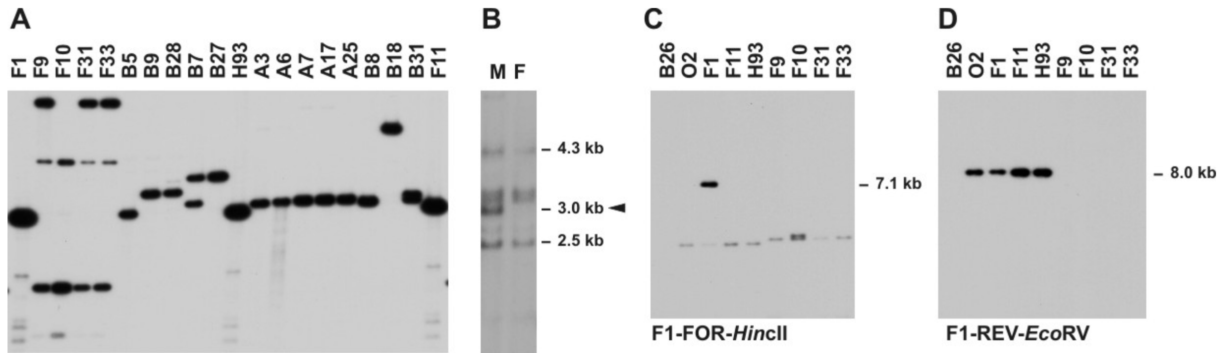


Fig. 13. Demonstration of the overlapping between BAC clones F1 and F11. BAC clones (A, C, D) and genomic DNA from male (M) and female (F) *X. maculatus* (B) were hybridized the terminal fragment F11-REV-*Eco*RV (A-B) and the terminal fragments of F1 (C-D). BAC clones in (A) were digested with *Hin*dIII and *Pst*I and positive for F11-REV-*Eco*RV in the BAC library screening. The *X. maculatus* genomic DNA (B) was also digested with *Hin*dIII and *Pst*I. The Y-specific band is indicated by a black triangle (3.0 kb). Other BAC clones (C-D) were digested with *Hin*dIII.

For the third step of chromosome walking, F1-FOR-*Hin*cII was used as a probe, revealing a 7.0 kb Y-specific band in the *X. maculatus* genome (Fig. 14B). RFLP results indicated that BAC F6 overlaps with F1 and contains the Y-specific band (Fig. 14). F1-FOR-*Hin*cII apparently has only two copies in the male diploid genome.

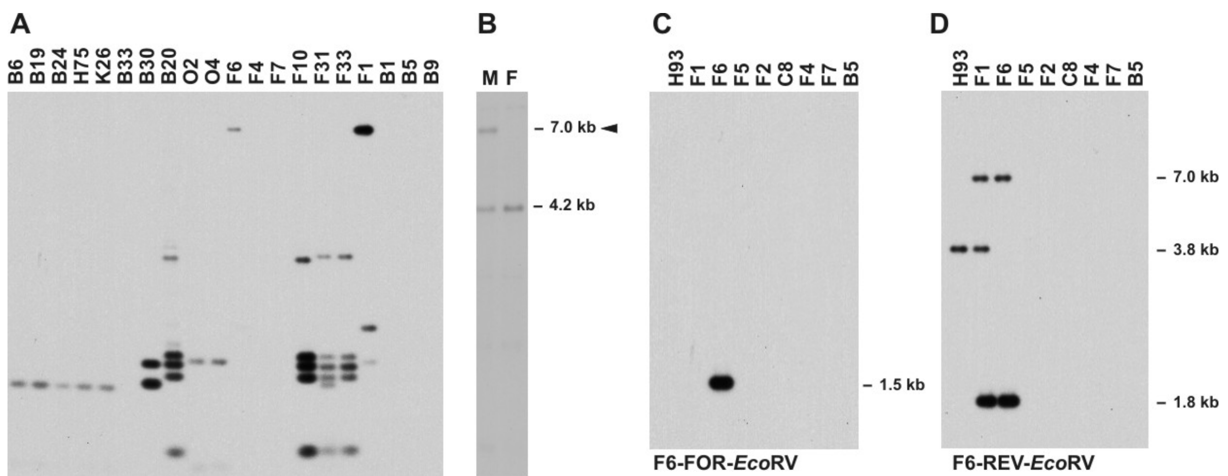


Fig. 14. Demonstration of the overlapping between BAC clones F6 and F1. The BAC clones (A, C, D) and genomic DNA from male (M) and female (F) *X. maculatus* (B) were hybridized with the terminal fragment F1-FOR-*Hin*cII (A-B) and the terminal fragments of F6 (C-D). BAC clones in (A) were digested in *Hin*dIII and *Sst*I and positive for F1-FOR-*Hin*cII in the BAC library screening. The *X. maculatus* genomic DNA (B) was also digested with *Hin*dIII and *Sst*I. The Y-specific band is indicated by a black triangle (7.0 kb). Other BAC clones (C-D) are digested with *Hin*dIII. The genomic DNA (B) did not show small bands (< 1.5 kb, A) observed in some BAC clones (A), due to the different amount of genomic DNA and BAC DNA, different exposure time or different stringent washing.

For the fourth step of chromosome walking, F6-FOR-*Eco*RV was used as a probe. This end fragment showed a 4.5 kb Y-specific band in the *X. maculatus* genome digested with *Eco*RI (Fig. 15B). In addition, this probe has apparently only two copies in the diploid genome.

Besides BAC F6, 14 new clones of the W series were positive for this probe in the BAC library screening (Fig. 15A). The RFLP analysis suggested that BAC W3, 7, 8, 9, 11, 12, 13 and 14 belong to the Y chromosome and overlap each other, whereas W1, 2, 4 and 5 overlap each other and might belong to the X chromosome or to an autosome. However, BAC F6 displays an exceptional 9.5 kb band (Fig. 15A). After the restriction map of the vector was checked, the 9.5 kb-*Eco*RI-digested fragment of F6 might be expected from the following analysis. As we know, there is no *Eco*RI cutting site in the cloning region of the BAC vector (approximately 8 kb in length). However, one *Eco*RI cutting site is 1 kb away from the reverse primer binding site of the vector. Thus, the 9.5 kb *Eco*RI fragment might contain the 7 kb-BAC vector fragment. The remaining 2.5 kb-fragment might be derived from the terminal fragment of BAC F6, which is part of the 4.5 kb-Y-specific fragment in the W series of BACs when digested with *Eco*RI (Fig. 15D). Furthermore, F6 has a 1.5 kb-*Hin*dIII-digested end fragment (Fig. 14C) and F6-FOR-*Eco*RV only reveals one 1.5 kb band in the genome digested either with *Hin*dIII or with both *Hin*dIII and *Eco*RI (Fig. 15C). These indicated that the Y-specific 4.5 kb-*Eco*RI-fragment and the terminal fragment of BAC F6 share the same 1.5 kb-*Hin*dIII-fragment. Hence, the 9.5 kb-*Eco*RI-digested fragment of BAC F6 is probably the sum of the 7 kb-vector fragment and the 2.5 kb-terminal fragment of F6.

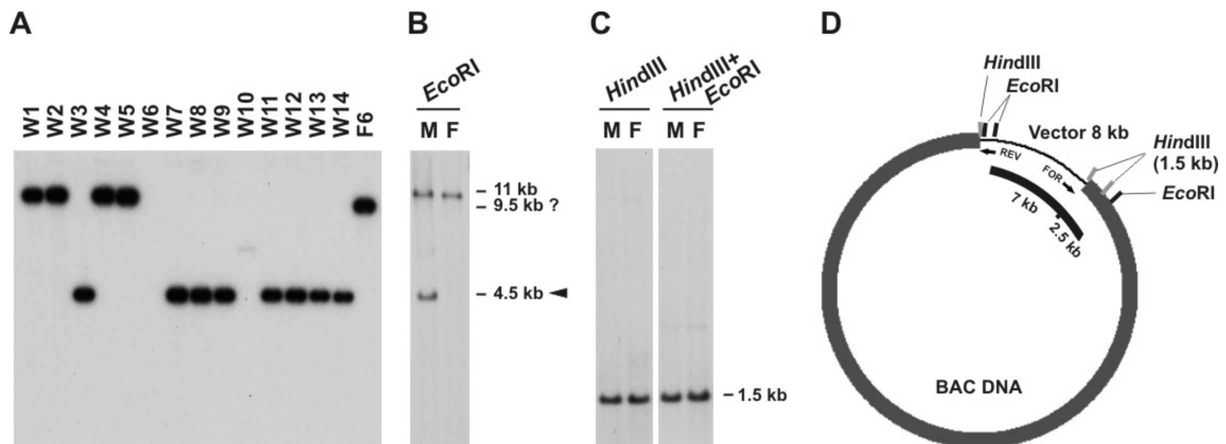


Fig. 15. Southern blot hybridization analysis of the W series of BAC clones positive for the probe F6-FOR-*Eco*RV and restriction map of BAC clone F6. The BAC clones and genomic DNA from male (M) and female (F) *X. maculatus* were hybridized with the terminal fragment F6-FOR-*Eco*RV (A-C). All BAC clones were digested with *Eco*RI (A) and positive for F6-FOR-*Eco*RV in the BAC library screening. The Y-specific band is indicated by a black triangle (4.5 kb). The black arch bar in D represents the 9.5 kb-*Eco*RI-digested fragment of BAC clone F6.

In order to identify the overlap between BAC F6 and the W-BAC clones containing the Y-specific band (4.5 kb), the terminal fragments of BAC W9 and W7 were used as probes for Southern blot hybridization analysis. The RFLP analysis indicated that BAC W9 and W7

overlap with F6 (Fig. 16). The overlap between BACs F6 and W9 was confirmed again by the present of a 4.0 kb Y-specific band (Fig. 16E-F).

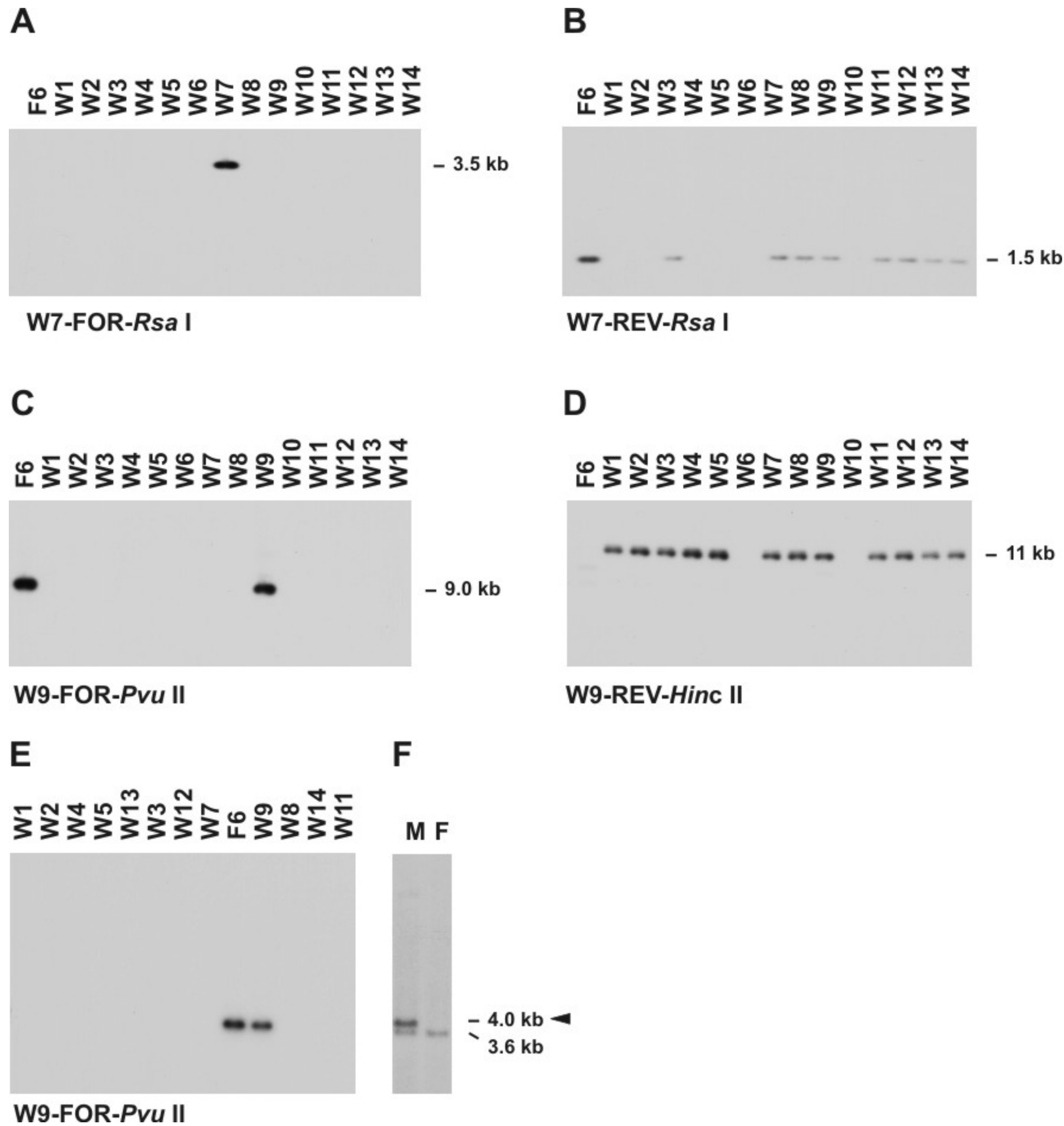


Fig. 16. Demonstration of the overlap between BAC clones W7, W9 and F6. The BAC clones (A-E) and *X. maculatus* male and female genomic DNA (F) were hybridized with the terminal fragments of W7 (A-B) and W9 (C-D) as well as the terminal fragment W9-FOR-*Pvu*II (E-F). All BAC clones were digested either by *Hin*dIII (A-D) or by *Hin*dIII and *Sst*I (E) and positive for F6-FOR-*Eco*RV using as a probe in the BAC library screening. The genomic DNA was digested with *Hin*dIII and *Sst*I. The Y-specific band is indicated by a black triangle (4.0 kb).

The terminal fragment W7-FOR-*Rsa*I was chosen for the next chromosome walking. However, no Y-specific band was observed by the RFLP analysis of male and female genomic DNA cut with different restriction enzymes (Fig. 17).

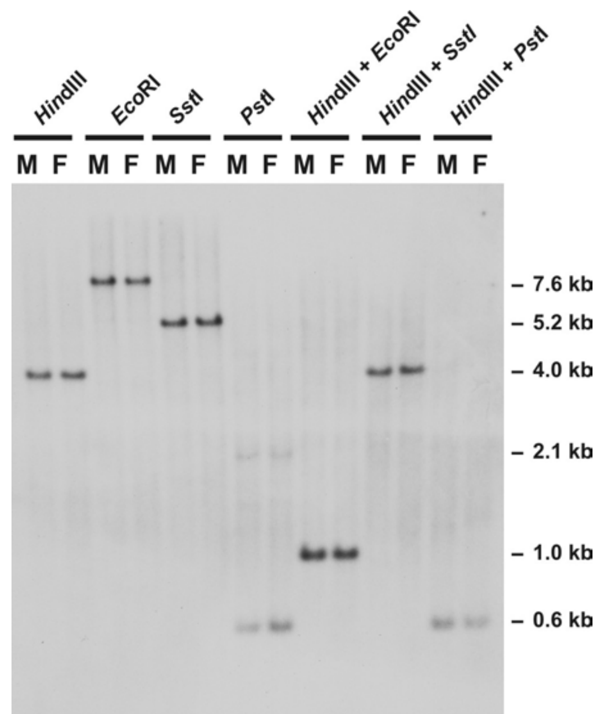


Fig. 17. Southern blot hybridization analysis of genomic DNA from male (M) and female (F) *X. maculatus* with the probe W7-FOR-*Rsa*I.

In order to analyze the relationships between BAC clones W1, 2, 4 and 5 and the other BAC clones from the W series in the *ps-criptY* contig, the terminal fragments of BAC W1 and W2 were used as probes for Southern blot hybridization analysis. RFLP results indicated that the terminal fragments of W1 and W2 reveal similar fragments on BAC F6 and BAC clones from W series belonging to the *ps-criptY* contig (Fig. 18). In addition, the terminal fragments W7-REV-*Rsa*I and W9-REV-*Hin*cII also revealed similar fragments on BACs W1, 2, 4 and 5 (Fig. 16). Taken together, these results strongly suggested that BAC W1, 2, 4 and 5 belong to a region of the X chromosome, homologous to the region of the Y chromosome containing the *ps-criptY* contig. However, there is no direct evidence so far that BACs W1, 2, 4 and 5 come from the X chromosome. They might also belong to a region on an autosome. Hence, the right part of the *ps-criptY* contig shows homology with another region located either on the X chromosome or on an autosome (Fig. 19, right part).

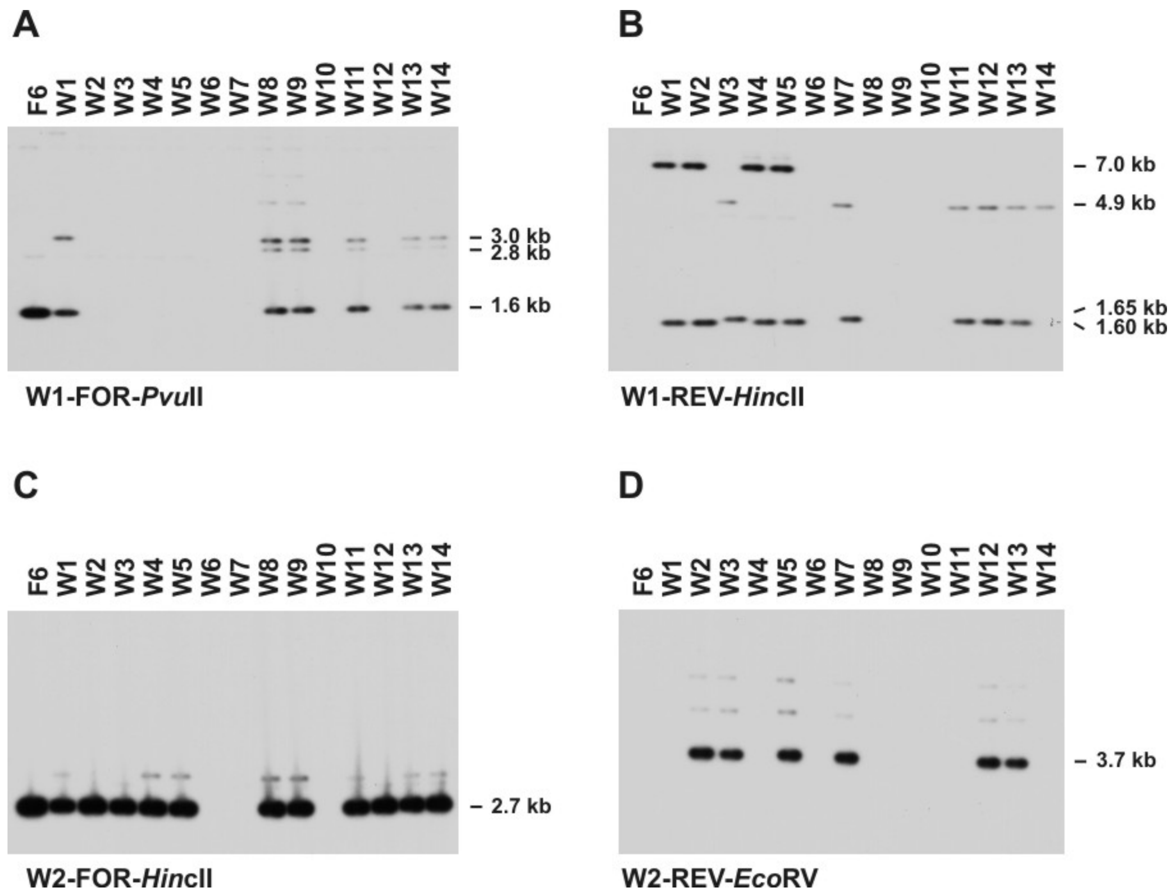


Fig. 18. Southern blot hybridization analysis of the BAC clones from the W series. All BAC clones were digested with *Hin*dIII, hybridized with the terminal fragments of W1 (A-B) and W2 (C-D) and positive for F6-FOR-*Eco*RV in the BAC library screening.

4.1.3.3 The *ps-criptY* contig covers more than 1000 kb of the Y chromosome of *X. maculatus*

To determine the order of BAC clones in the *ps-criptY* contig and to extend it, 41 different terminal fragments or sequences were used as probes in Southern blot hybridization (Tab. 2). Thirteen of them revealed Y-specific band(s). The *ps-criptY* contig was extended up to over 1000 kb and one side might reach the putative homologous region of the X chromosome (Fig. 18 and 19).

Table 2. Terminal fragments of BAC clones and sequences or genes used as probes in Southern blot hybridization analysis and RNA whole-mount *in situ* hybridization. All of them are from the *ps-criptY* contig.

Name of probes	Position of markers	Characteristics	References
B14-*Pvu*II-FOR	3´ region of *ps-criptY*		Fig. 5
B14-*Rsa*I-REV	3´ region of *ps-criptY*		Fig. 5
B26-*Pvu*II-FOR	3´ *Xmrk* (X) and 3´ *ps-criptY*		Fig. 5, 11
B26-*Rsa*I-REV	3´ region of *ps-criptY*		Fig. 5
B29-*Hinc*II-FOR	5´ region of *ps-criptY*	*Helitron*	Fig. 5
B29-*Rsa*I-REV	3´ region of *ps-criptY*		Fig. 5
B17-*Dra*I-FOR	3´ region of *ps-criptY*	Y specific fragment and *helitron*	Fig. 5, 6, 7
B17-*Hae*III-REV	5´ region of *ps-criptY*		Fig. 5
N20-*Eco*RV-FOR	5´ region of *ps-criptY*		Fig. 7
N20-*Rsa*I-REV	5´ region of *ps-criptY*	Y specific fragment	Fig. 7, 8, 9
R1-*Eco*RV-FOR	5´ region of *ps-criptY*		Fig. 8
R1-*Hinc*II-REV	5´ region of *ps-criptY*		Fig. 8
R19-*Pvu*II-FOR	3´ region of *ps-criptY*		Fig. 8
R19-*Hinc*II-REV	3´ region of *ps-criptY*		Fig. 8
021B09-*Rsa*I-sp6	Not identified		Fig. 10
021B09-*Hinc*II-t8	Not identified		Fig. 10
049O11-*Pvu*II-sp6	Not identified		Fig. 10
049O11-*Rsa*I-t8	Not identified		Fig. 10
O2-*Hae*III-FOR	3´ region of *ps-criptY*		Fig. 12
O2-*Rsa*I-REV	3´ region of *ps-criptY*		Fig. 12
F11-*Rsa*I -FOR	3´ region of *ps-criptY*	Repeats	Fig. 12
F11-*Eco*RV-REV	3´ region of *ps-criptY*	Y specific fragment	Fig. 12, 13
F1-*Hinc*II-FOR	3´ region of *ps-criptY*	Y specific fragment	Fig. 13, 14
F1-*Eco*RV-REV	3´ region of *ps-criptY*		Fig. 13
F6-*Eco*RV-FOR	3´ region of *ps-criptY*	Y specific fragment	Fig. 14, 15
F6-*Eco*RV-REV	3´ region of *ps-criptY*		Fig. 14
W7-*Rsa*I-FOR	3´ region of *ps-criptY*		Fig. 16, 17
W7-*Rsa*I-REV	3´ region of *ps-criptY*	Y specific fragment	Fig. 16
W9-*Pvu*II-FOR	3´ region of *ps-criptY*	Y specific fragment	Fig. 16
W9-*Hinc*II-REV	3´ region of *ps-criptY*		Fig. 16
W1-*Pvu*II-FOR	Not identified		Fig. 18
W1-*Hinc*II-REV	Not identified		Fig. 18
W2-*Hinc*II-FOR	Not identified		Fig. 18
W2-*Eco*RV-REV	Not identified		Fig. 18
Tnh7	5´ region of *ps-criptY*	Y-specific fragment	Fig. 19
Swimy	3´ region of *ps-criptY*	Y specific fragment	Fig. 23, 24, 25, 26
Fredi	Diverse RFLPs		Fig. 29
DNA polymerase type B	3´ region of *ps-criptY*		Fig. 33, 35
Msh2	Diverse RFLPs	Y specific fragment	Fig. 37
Hel_protease	3´ *Xmrk* (X), 5´ and 3´ *ps-criptY*	Y specific fragment	Fig. 43, 44, 45
Hel_ helicase	3´ *Xmrk* (X), 5´ and 3´ *ps-criptY*	Y specific fragment	Fig. 43, 44

Fig. 19. Schematic structure of the *ps-criptY* contig from the sex-determining region on the Y chromosome of *X. Maculatus*. Thicker line represents the Y chromosome. Dotted line shows the putative corresponding region of the X chromosome. All probes used for Southern blot hybridization analysis are described in Table 2.

4.2 Sequence analysis of the *ps-criptY* contig of *X. maculatus*

After ordering the BACs in the *ps-criptY* contig, representative BAC clones from this contig were ´´shotgun`` sequenced. These sequences were analyzed with the appropriate bioinformatic tools. A total of 733 kb of non-redundant sequences was generated for the *ps-criptY* contig, covering more than 50% of the *ps-criptY* contig (Tab. 3). Several gene candidates (Tab. 4) and various markers were mapped on the different X- and Y-chromosomal contigs by Southern blot hybridization and PCR analysis (Fig. 19). A total of 58 molecular markers have been mapped on the *ps-criptY* contig.

Table 3. Statistics of the *ps-criptY* contig sequencing project.

BAC DNAs	Length of total ´´shotgun`` sequences (kb)	Length of non-redundant sequences (kb)	Length of BACs (kb)	Part of clone sequenced (%)	References
N20	250	75	160	47.1	This work
B26	229	105	165	63.8	This work
F11	191	130	150	87.2	This work
B29	291	137	210	65.6	This work and sequences from A. Froschauer, C. Körting and M. Vucic
B14	243	87	175	50.1	This work and P. Bienert, C. Körting
F6	201	88	145	60.9	This work and C. Körting
B17	357	108	190	56.9	This work and C. Körting
Total	1764	733			

4.3 Identification and analysis of gene candidates in the *ps-criptY* contig

Several gene candidates were identified in the *ps-criptY* contig as well as other *SD*-linked contigs (Tab. 4 and Fig. 19). Some of them are probably pseudogenic duplicates, but others are likely to be functional. About 16 putative gene candidates were located in the *ps-criptY* contig. Syntenic regions were also found in the genome of fugu (*Takifugu rubripes*) and human (Tab. 4). Six novel gene candidates from the *ps-criptY* contig have been identified in this work.

Table 4. Gene candidates in the sex determination region of the platyfish *X. maculatus.*

Genes	Genes identified in this work§	Localization # (copy number)	Fugu	Human	Conceptual product
*swimy**	+	X(?)/Y(1)/?	M001299	5q33	DNA binding protein
*fredi**	+	X(2)/Y(4)/?(N)	M000027	nd	DNA binding protein (helix-turn-helix)
*pol**	+	X(?)/Y(1)/?	10 copies	nd	DNA polymerase type B
*tmpl **	+	X(?)/Y(1+?)/?	M000178/ M000033/ M000921	17q25	tripartite motif protein like
*trail **	+	X(?)/Y(1+?)/?	M000683	3q26	tumor necrosis factor related apoptosis inducing ligand
*gch1**	+	X(?)/Y(1)/?	M000263/ M003100	14q22	GTP cyclohydrolase I
*igd1**		X(1)/Y(1)/?	M007839/ M003343/ M002928	3q13/1p12	immunoglobulin V-type domain protein
*igd2**		X(1)/Y(1)/?	M007839/ M003343/ M002928	3q13/1p12	immunoglobulin V-type domain protein
*rnaseIII**		X(Nps)/Y(Nps)/?	M000622	5p13	ribonuclease III (microRNA processing)
*mc4r**		X(9)/Y(9)/?(N)	M000622	18q22	melanocortin receptor
*dc2**		X(9ps)/Y(9ps)/ ?(N)	M006132	1p34/6q11/ 4q25/9q13	predicted membrane protein
*acr**		X(1)/Y(1)/?(1)	multiple	2q33-34	acetylcholine receptor, delta chain
*bdg**		X(1)/Y(1)/?	M003906	2q37/7q21	DNA binding protein
*cript**		X(1ps)/Y(2ps)/ A(1)	M003100	2p21	postsynaptic protein
*msh2**	+	X(1ps)/Y(2ps+?)/ ?(N)	M004124	2p22-21	DNA mismatch repair protein
*tnh7**		X(1)/Y(1)/?(2)	M001903	X/Y	DNA-binding protein with BED-zinc-finger-motif
*tnh2**		X(3)/Y(3)/?(N)	M000967	13q12-13 /1p34	DNA-binding protein with zinc-finger-motif
fbx11		X(1ps)/Y(1ps)/ ?(?)	M004124/ M000049	2p21	F-box protein
tmeff		X(1)/Y(1)	M003910	9q31/2q32	transmembrane protein, EGF-like domain
egfrb		X(1)/Y(1)	M000399	7p12	epidermal growth factor receptor
Xmrk		X(1)/Y(1)	nd	nd	epidermal growth factor receptor-like
cide		X(1)/Y(1)	M001355/ M000374	18p11/3p26	cell death activator

§ gene candidates (+) identified in this work, * gene candidates located in the *ps-criptY*. ´´nd`` for not detected. ´´?`` under study. # gene candidates location and copy number on the X chromosome/Y chromosome/autosome of *X. maculatus* (Volff, JN. *et al*., pers. commun.).

4.3.1 *Swimy*, a new gene exclusively expressed in spermatogonia of *X. maculatus* adult testis

4.3.1.1 *Swimy* encodes a putative DNA binding protein

When analyzing BAC F11 ´´shotgun`` sequences, I found a predicted exon (645 nt in length) encoding a putative translation product with a SWIM Zn_finger domain. This domain has a

$CXCX_nCXH$ motif (n= 6-16 residues) with predicted zinc-chelating residues. The SWIM Zn_finger domain is found in a variety of prokaryotic and eukaryotic proteins, including: 1) bacterial ATPases of the SWI2/SNF2 family; 2) plant MuDR transposases and transposase-derived Far1 nuclear proteins; 3) vertebrate mitogen-activated protein kinase kinase kinase-1 (MAPKKK-1) and 4) several hypothetical proteins. The SWIM domain is named after SWI2/SNF2 and MuDR, and is predicted to have DNA-binding and protein-protein interaction functions (Makarova *et al*., 2002).

An approximatively 11 kb large genomic consensus sequence containing the above predicted exon was assembled by ´´shotgun`` sequencing. The gene containing the SWIM Zn_finger domain was designated under the name *swimy.*

Using a primer pair flanking the SWIM Zn_finger domain, the expression pattern of *swimy* was analyzed by RT-PCR. Interestingly, *swimy* is exclusively expressed in the testis of male *X. maculatus* and has two transcripts (Fig. 20). A transcript was also detected in stage 21 embryos of hybrids between *Xiphophorus maculatus* and *Xiphophorus hellerii*, in an embryonic cell line of *X. xiphidium* and in the PSM melanoma cell line, although the signal was weak in the two cell lines. No signal was observed in stage 20 embryos. The two transcripts of *swimy* in adult testis might result from two different *swimy* genes with different repeat number.

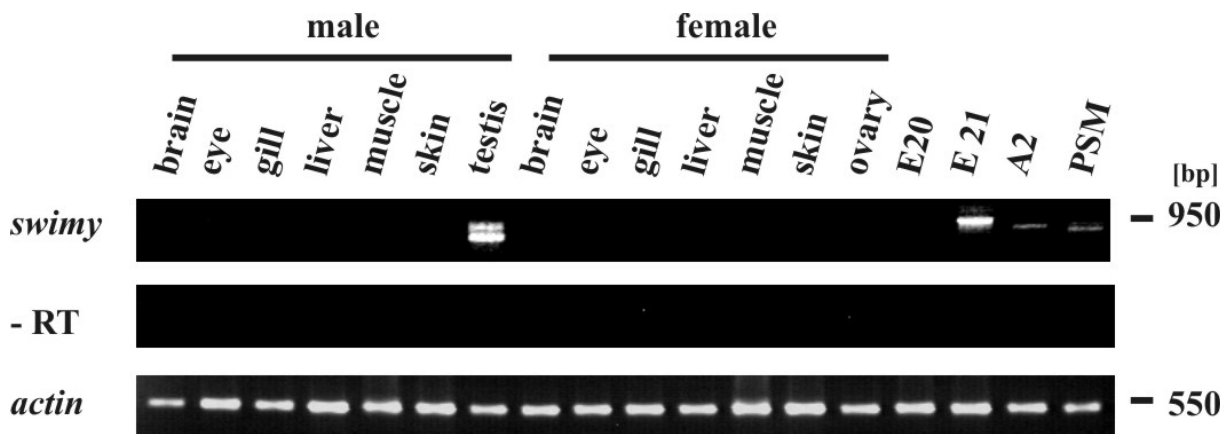


Fig. 20. RT-PCR-based expression analysis of *swimy* in *X. maculatus.* The primer pair allows the amplification of the sequence encoding the SWIM domain of *swimy* (Probe I in Fig. 21). E20 are stage 20 embryos of *X. maculatus* Jp163A. E21 are stage 21 embryos of hybrids (named BCn407) between *Xiphophorus maculatus* and *Xiphophorus hellerii*. A2 is an embryonic cell line of *Xiphophorus xiphidium*. PSM is a melanoma cell line derived from a hybrid between *Xiphophorus maculatus* and *Xiphophorus hellerii.* PCR program: 1* [3'00'' at 95°C], 35* [0'40'' at 95°C/ 0'40'' at 55 °C/ 1'00'' at 72°C], 1* [5'00'' at 72°C], 4°C.

A 2551 bp cDNA sequence of *swimy* was assembled by sequencing RT-PCR products from adult testis. From there, a protein sequence of 939 aa was predicted by conceptual translation.

After comparing the cDNA sequence with the 11 kb genomic sequence containing *swimy*, the result indicated that *swimy* has 7 exons (Fig. 21). This 11 kb genomic sequence has no stop codon or frameshift in its coding region and has 99.1% nucleotide identity with the above cDNA sequence.

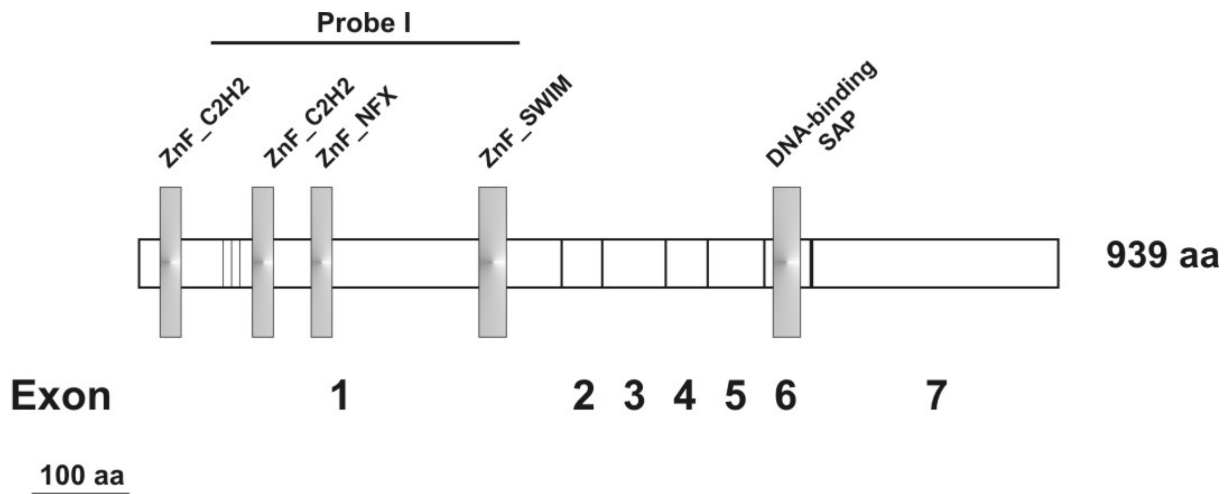


Fig. 21. Structure and putative domains of the Swimy protein. ZnF_C2H2 are C2H2 type Zn-finger domains; ZnF_NFX is a NFX type Zn-finger domain; ZnF_SWIM is a SWIM Zn-finger domain; DNA-binding SAP is a SAP DNA-binding domain. Probe I is a region of about 1 kb derived from RT-PCR product, and the probes for Southern blot hybridization and *in situ* hybridization are generated from the same region. White boxes represent putative exons of Swimy. Straight fine lines represent repeats of a sequence of 14 residues.

The aa sequence encoded by exon 1 of Swimy was predicted to contain four Zn-finger domains such as two C2H2 type (accession number IRP007086), one NFX type (accession number IRP000967) and one SWIM type (accession number IRP007527) Zn-finger domains. The aa sequence encoded by exon 6 of Swimy was predicted to contain one putative SAP DNA-binding domain (accession number IRP003034) (Fig. 21). These structural features suggest that the predicted Swimy protein is a DNA binding protein. The N terminal region between the two C2H2 Zn-finger domains of Swimy contains about 3 repeat sequences in tandem, which are 14 residues in length.

4.3.1.2 *Swimy* is expressed in some spermatogonia of *X. maculatus* adult testis

After RT-PCR using primers amplifying probe I, the PCR products were cloned into the TOPO vector. 14 positive clones were picked and sequenced. On the basis of their repeat numbers, these clones could be divided into four groups (Fig. 22). The RT-PCR clones swimy 02, 06, 11, 35, 41 and 63 belong to one group with the conceptual protein corresponding to that of BAC F11 genomic DNA, containing three repeats of a sequence of 14 residues. The

second group contains PCR clones swimy 21, 07, 54 and 38 with two repeats. The third group contains PCR clone swimy 09 with four repeats. The fourth group contains PCR clones swimy 33, 28 and 58 with six repeats. Furthermore, the sequences of RT-PCR clones swimy 35 and 11 are completely identical to the conceptual protein sequence from the BAC F11 genomic DNA. Taken together, these results suggest the presence of different swimy gene in the genome of Xiphophorus. This was confirmed by Southern blot hybridization analysis (see below). Alternatively, difference in repeat number might reflect cloning or sequencing artefacts. The RT-PCR clone swimy 35 with three repeats was selected to prepare sense and anti-sense RNA probes for whole-mount RNA *in situ* hybridization.

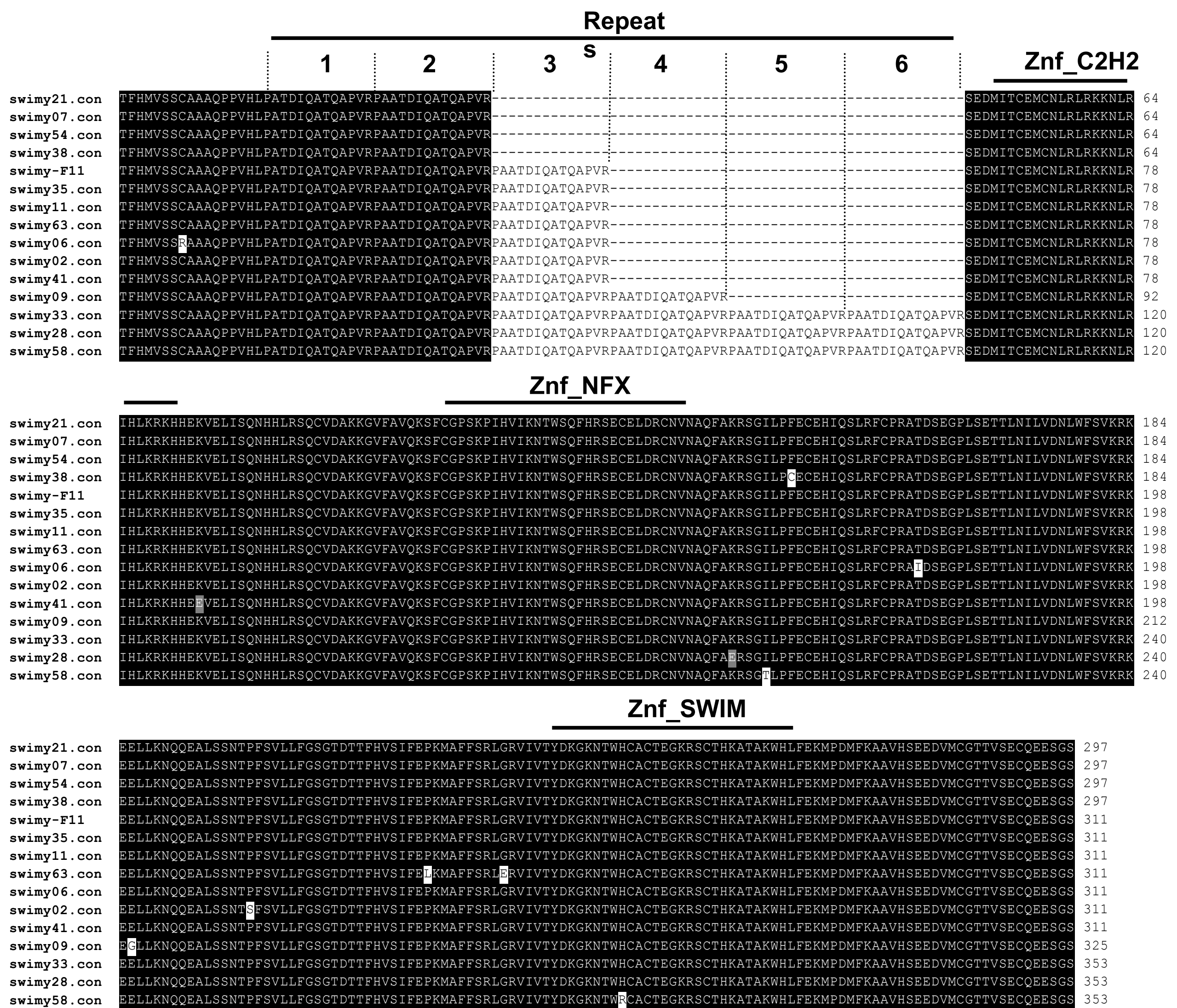


Fig. 22. Sequence comparison of the deduced aa sequence from RT-PCR products of *swimy* and conceptual protein sequence from the BAC F11 genomic DNA. ZnF_C2H2 is a C2H2 type Zn-finger domain; ZnF_NFX is a NFX type Zn-finger domain; ZnF_SWIM is a SWIM Zn-finger domain. The repeat consists in a sequence of 14 residues. These RT-PCR products are divided into 4 groups depending on their repeat number.

Using the *swimy* anti-sense RNA probe, *X. maculatus* adult testis was stained in RNA whole-mount *in situ* hybridization (Fig. 23). Furthermore, Hybridization on sections of the stained adult testis indicated that *swimy* is expressed in some spermatogonia (Fig. 23D), suggesting that *swimy* might be involved in spermatogenesis. This is the first gene of *X. maculatus* found to be exclusively expressed in spermatogonia.

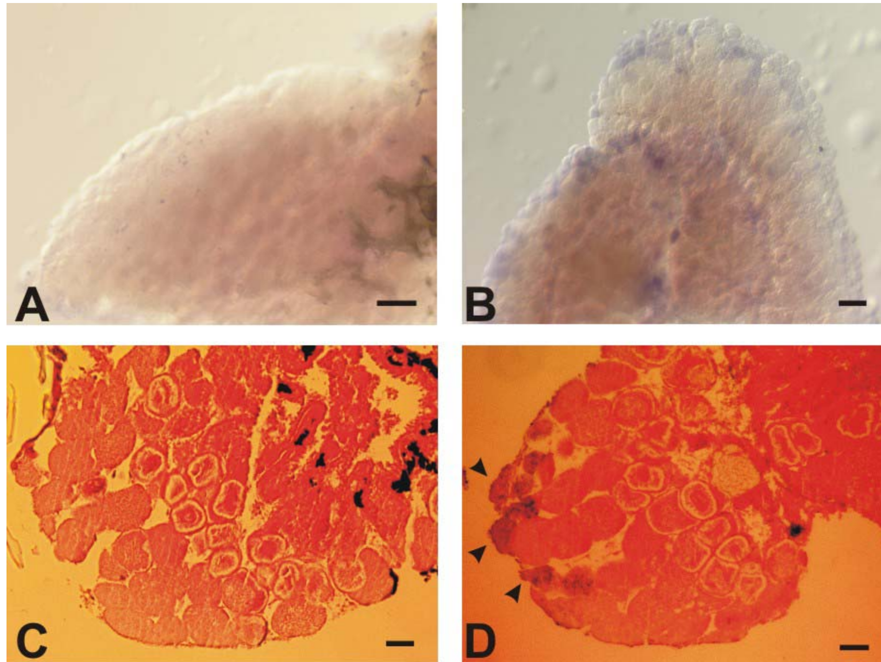


Fig. 23. *In situ* hybridization-based expression pattern of *swimy* in some spermatogonia of *X. maculatus* adult testis. *Swimy* sense control: A and C. Some cysts comprising spermatogonia (D, arrow heads) were stained by RNA whole-mount *in situ* hybridization using *swimy* anti-sense RNA probe generated from probe I. Bar scale=125 µm: A and B; 31 µm: C; 40 µm: D.

4.3.1.3 *Swimy* is located in the *ps-criptY* contig of *X. maculatus*

According to Southern blot hybridization using probe I, *swimy* has about 7 copies per haploid *X. maculatus* genome. There is only one 9.5 kb-Y-specific band in the genome (Fig. 24).

This 9.5 kb-Y-specific band was located in BACs F11 and O2 of the *X. maculatus ps-criptY* contig (Fig. 25A). With the same probe, no signal was observed in BAC clones from *Xmrk* contigs on the X and Y chromosomes (Fig. 25). Hence, only one copy of *swimy* was found in the sex determination region of *X. maculatus* so far, although about 7 copies of *swimy* are present per haploid genome. From sequence comparison of the cDNA and genomic DNA of BAC F11, this Y-specific copy does not contain any deletion, insertion or stop codon in the coding region, implicating that it might be functional.

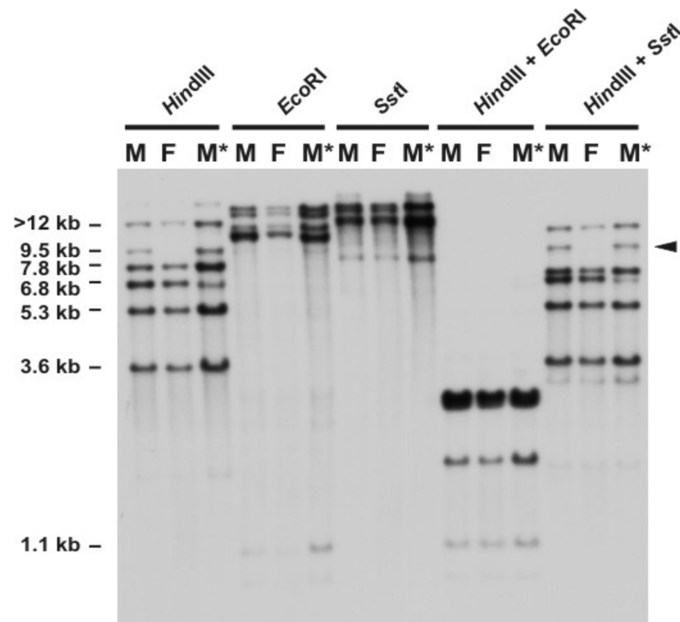


Fig. 24. Southern blot hybridization-based RFLP analysis of genomic DNA from male (M) and female (F) *X. maculatus* with the probe I of *swimy*. M* is the MAD strain of *X. maculatus* described in MATERIALS. The Y-specific band is indicated by the black triangle (9.5 kb).

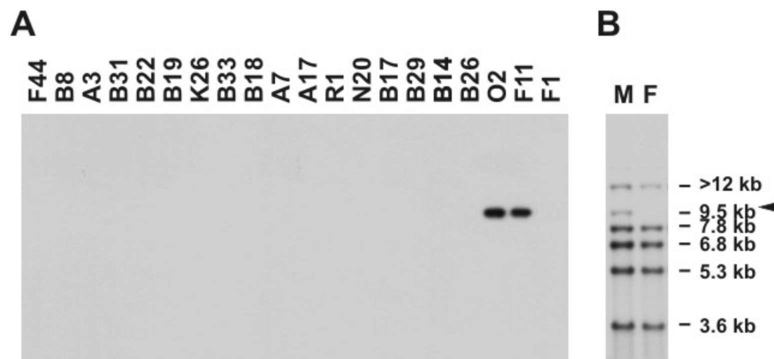


Fig. 25. Southern blot hybridization-based analysis of BAC clones from the sex-determining region for the presence of *swimy*. BAC clones (A) and genomic DNA from male (M) and female (F) *X. maculatus* (B) were digested with *Hind*III and *Sst*I and hybridized with the probe I of *swimy*. The BAC clones (A) cover the *ps-criptY* contig and the *Xmrk* contigs on the X and Y chromosomes. The Y-specific band is indicated by a black triangle (9.5 kb).

4.3.1.4 *Swimy* distribution in the genus *Xiphophorus*, in the family Poeciliidae and in other ray-finned fishes

Using the probe I of *swimy* for Southern blot hybridization analysis on genomic DNA of different fish species, *swimy* was found to be present in all *Xiphophorus* and some poeciliids tested (except *G. falcatus and B. belizanus*) with different copy numbers (Fig. 26A-B). The copy number of *swimy* per haploid genome of *Xiphophorus* is higher than in some other poeciliids. *Swimy* presents RFLPs between different populations of the same species (for example in *X. maculatus* and *X. hellerii*) as well as between males and females within a same species (such as all tested *Xiphophorus*). Sex-specific RFLPs between males and females

within a same species were not observed in other fishes from the Poeciliidae family. Moreover, interindividual RFLPs within a same species were observed in female *X. hellerii* (Rio Lantecilla strain hIII). Interestingly, *swimy* is apparently also located on the W chromosome of the Rio Usumacinta (RU) strain of *X. maculatus* (Fig. 26A, white triangles). No signal could be observed in the closely related family Fundulidae (*Fundulus heteroclitus*, Fig. 26B) and in more divergent bony fish species using the same probe (Fig. 26C).

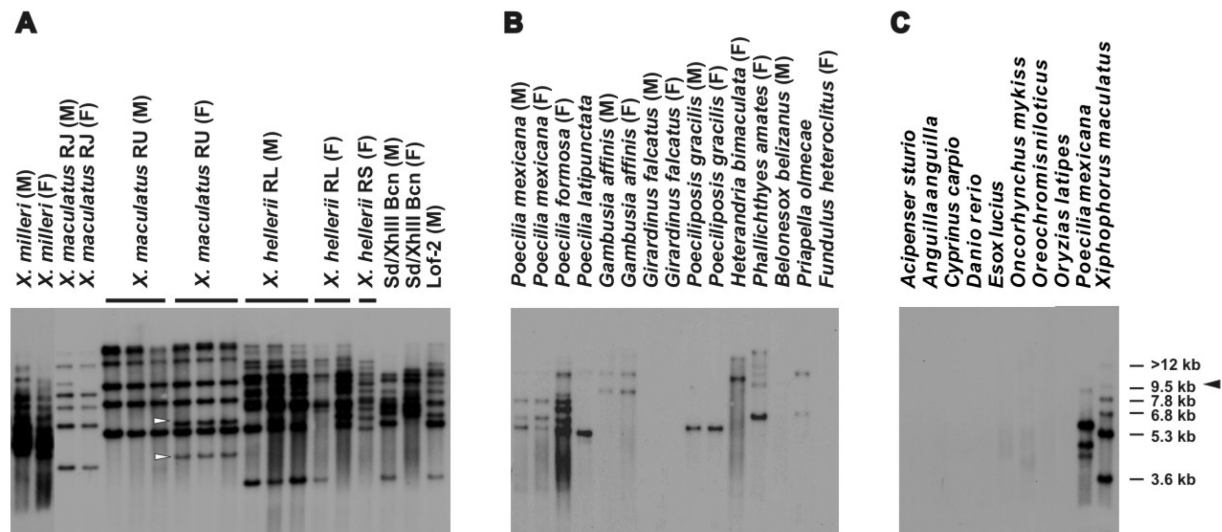


Fig. 26. Distribution of *swimy* in *Xiphophorus* (A), poeciliids (B) and other bony fishes (C). Genomic DNAs were digested with *Hin*dIII and hybridized with the probe I of *swimy*. The tested fish are described in MATERIALS. The Y-specific band is indicated by a black triangle (9.5 kb) and the putative W-specific bands are indicated by white triangles (5.3 and 4.0 kb). *Fundulus heteroclitus* belongs to the Fundulidae family that is closely related to Poeciliidae family.

4.3.1.5 Sequences related to *swimy* in other fish

In silico search for *swimy* in public sequence databases revealed similar sequences in the fugu, zebrafish, medaka fish, chicken, mouse and human genomes. An alignment of Swimy-like conceptual protein sequences was constructed (Fig. 27), except for sequences from chicken, mouse and human due to alignment ambiguities. It shows that sequences from medaka and fugu share all conserved regions with Swimy of the platyfish, and sequences from zebrafish share several conserved regions downstream of the SWIM domain with the platyfish sequence. All Swimy-like sequences in fish contain the SAP DNA binding domain. Except for zebrafish, Swimy-like sequences in the fugu and medaka fish contain the ZnF_SWIM domain. The similarities of Swimy-like protein sequences between platyfish and medaka, platyfish and fugu as well as platyfish and zebrafish are 30-51%, 43% and 29-40%, respectively. The fugu, medaka and zebrafish have one copy, three copies and three copies of

sequences similar to *swimy*, respectively. So far no EST (expressed sequence tag) sequence similar to *swimy* was identified by tblastx in the NCBI database.

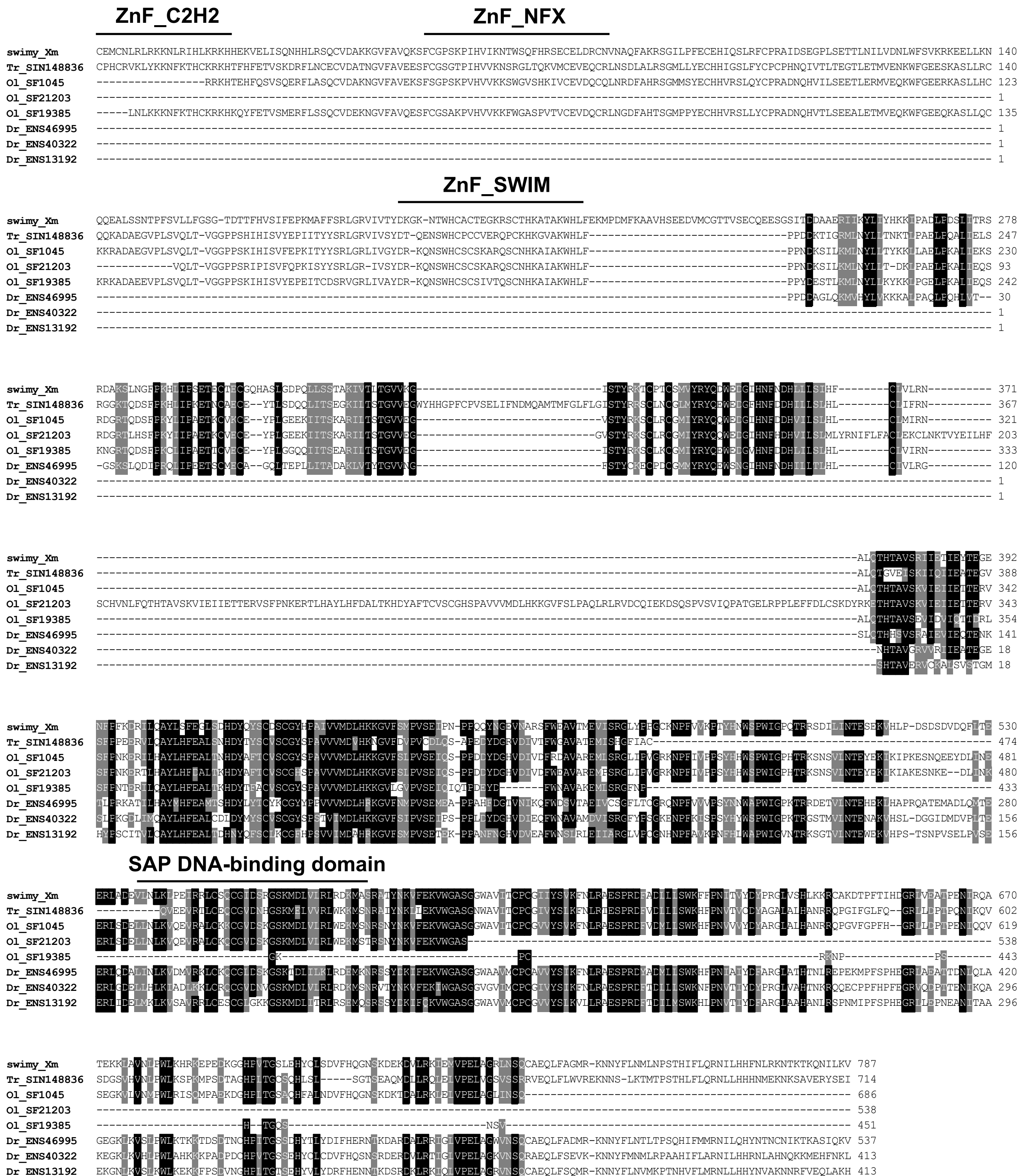


Fig. 27. Alignment of putative Swimy-like proteins from fish. Identical residues are shown in white type on black, and conservative substitutions in white type on gray. Swimy_Xm the protein sequence of Swimy in *Xiphophorus maculatus*. Tr, Ol and Dr, which are followed by accession numbers of matched nucleotide sequences, represent fugu, medaka and zebrafish, respectively.

4.3.2 *Fredi* encodes a novel helix-turn-helix protein

4.3.2.1 Expression pattern and location of *fredi* in *X. maculatus*

Fredi was discovered in collaboration with P. Bienert (F1 practical student) during the sequencing of BAC clone B14. It encodes a completely novel helix-turn-helix protein. In *X. maculatus* adult tissues, *fredi* is expressed predominantly in testis (Fig. 28). Its transcript was also detected in the PSM melanoma cell line.

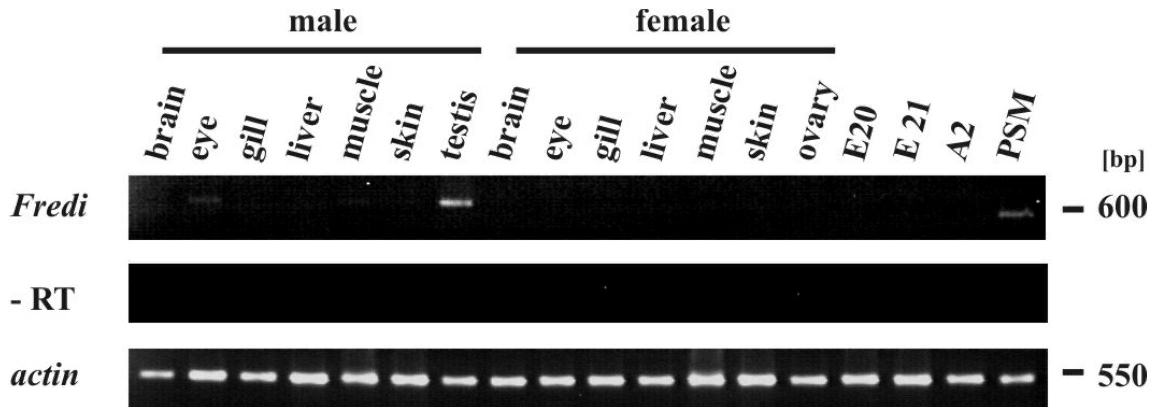


Fig. 28. . The expression of *fredi* in *X. maculatus.* The primer pair amplifies the putative exon 1 of *fredi*. E20 is stage 20 embryos of *X. maculatus* Jp163A. E21 is stage 21 embryos of hybrids (called BCn407) between *Xiphophorus maculatus* and *Xiphophorus hellerii*. A2 is the embryonic cell line of *Xiphophorus xiphidium*. PSM is a melanoma cell line from a hybrid between *Xiphophorus maculatus* and *Xiphophorus hellerii.* PCR program: 1* [3'00'' at 95°C], 35* [0'40'' at 95°C/ 0'40'' at 55 °C/ 1'00'' at 72°C], 1* [5'00'' at 72°C], 4°C.

Fredi is present at about 20 copies in the *X. maculatus* genome (Fig. 29B). The two Y-specific bands are located in the *ps-criptY* contig and in the *Xmrk* contig of the Y chromosome (Fig. 29A). *Fredi* has at least 2 copies on the X chromosome contig and 4 copies on the Y chromosome contigs (Fig. 41).

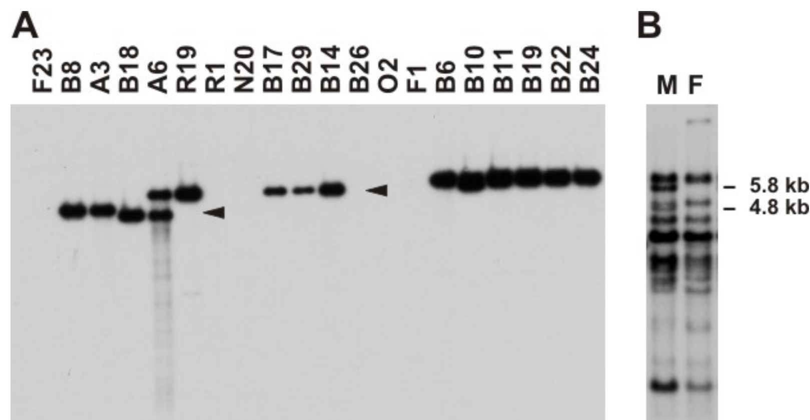


Fig. 29. Southern blot hybridization-based RFLP analysis of BAC clones containing the *fredi* gene. The BAC clones (A) and genomic DNA from male (M) and female (F) *X. maculatus* (B) were digested with *Hin*dIII and *Sst*I, and hybridized with probe generated from a RT-PCR product of adult testis (Fig. 28). The BAC clones (A) cover the *ps-criptY* contig and the *Xmrk* contigs of the X and Y chromosomes. The Y-specific bands are indicated by the black triangles (5.8 and 4.8 kb).

4.3.2.2 Is *fredi* a fish-specific gene?

A segment from BAC B11 (X-linked *Xmrk* contig) containing a *fredi* gene was sequenced in collaboration with C. Körting. Using its conceptual protein sequence (699 aa, called Fredix) as query to search public databases, sequences similar to *fredi* were found only in the fish lineage. All of them contain a HTH_7 domain (accession number: PF02796), which is a helix-turn-helix domain of recombinases involved in prokaryotic site-specific recombination (Argos *et al*., 1986; Garnier *et al*., 1987). There are three sequences similar to *fredi* in the genome of the medaka and two copies in the zebrafish. The amino-acid similarities between Fredi-like protein sequences from platyfish and medaka are 31-42%, and about 20% between platyfish and zebrafish. No EST sequence matching *swimy* could be identified by tblastx in the NCBI database. These suggest that *fredi* might be a ´´fish-specific`` gene.

An alignment of Fredi-like protein sequences from *Danio* and *Oryzias* was constructed, together with four copies of *fredi* assembled from the sex-determining region of *X. maculatus* (Fig. 30). Fugu sequences were not included because of alignment ambiguities. This alignment showed high sequence conservation in the HTH_7 domain and its downstream region (~55 aa in length).

```
Fredix                      ELLLDHIFDRLRDRLEHVLDRMPLDLDFLDFICTQELVFLNAMSCQIDVPSNILDALAHLHRLINLEKQSEE-QQTTVVH 79
FrediY-b18                  ------------------------------------------MSCQIDVPSNILDALAHLHRLINLEKQSEE-QQTTVVH 37
FrediY-b14                  -------------------------------------------------------------------------------- 1
FrediY-b17                  ELLLDHIFDRLRDRLEHVLDRMPLDLDFLDFICTQELVFLNAMSCQIDVPSNILDALAHLHRLINLEKQSEE-QQTTVVH 79
Fredi_Ol_SF27104            --LLQHIFERLRSRLEHVLQRLPLDLDFLDLMCTQEHVFLNALSSQLTVCPAILEELTHLQRLISSEKERTN-SCIAVVQ 77
Fredi_Ol_SF3459             --LLQHIFERLRSRLEHVLQRLPLDLDFLDFICTQEHVFLNALSSQLTVCPAILEELTHLQRLISTEKERTN-SCIAVVQ 77
Fredi_Ol_SF15321            --LLRHIFERLRSRLEHVLQRVPLDLDFLDFICTQEHVFLNALSSQLTVCPAILEELTYLQRLISTEKERTN-SCIAVVQ 77
Fredi_Dr_GENSCAN000000      ELLLNHLLQRLCNRLTLWLERIPLDIDFLEFTCAQELVIFSAISHQIPLAHEVYD---ALYQVHQMAVTSTNSDSHVIVC 77
Fredi_Dr_GENSCAN000000      QLLLDYVLQRLRSRLETAIGRLPLDMDFLEFICSQELVFLSAVSNSVTIPQDIMDGITELLGLIQLNLSNEEIRTPTNVL 80

                                       HTH_7 domain
                            ____________________________________________________
Fredix                      IEQGPAGRQRIVVSSDYLCSLLELHLPVPCIAKLLGISRRTVYRRMAESGLSDTGPRGPGVPTPDLDQCVRDIKSRQSHS 159
FrediY-b18                  IEQGPAGRQRIVVSSDYLCSLLELHLPVPCIAKLLGISRRTVYRRMAESDLSVRALYTTMS-DDDLDQCVRDIKSRQSHS 116
FrediY-b14                  -------------------------------------------------------------------------------- 1
FrediY-b17                  IEQGPAGRQRIVVSSDYLCSLLELHLPVPCIAKLLGISRRTVYRRMAESGLSVRALYTTMS-DDDLDQCVRDIKSRQSHS 158
Fredi_Ol_SF27104            WEQGPAGRQRMALSSDYLCSLLELNLPVTCIAKLLGMSRRTVYRRMNESDLSVKGLYSTMS-DEDLDQCVREIKSRQPNS 156
Fredi_Ol_SF3459             WEQGPAGRQRMALSSDYLCSLLELNLPVTCIAKLLGMSRRTVYRRMNESDLSVKGLYSTMS-DEDLDQCVREIKSRQPNS 156
Fredi_Ol_SF15321            WEQGPAGRQRMALSSDYLCSLLELNLPVTCIAKLLGMSRRTVYRRMNESDLSVKGLYSTMS-DEDLDQCVREIKSRQPNS 156
Fredi_Dr_GENSCAN000000      EEKGTNGRPRILLSEDSVSQLLSSGFPVTCIARLLGVSRFTIHRRMVEWGLSVKSTYSNLT-DAELDSLVAKIHTSNPNA 156
Fredi_Dr_GENSCAN000000      QESGSMGRPRLVISPDFITNLLDINCSVPTIASMIGVSTRTVFRRMSECNLSVRARYSTLT-DEELDDCVKELQRHIPHC 159

Fredix                      GYRMVKALLQARGLRVQYDRVRVSMHRVDTIGMASRMVQLGCIARRTYSVPGPQSLMHIDTNHKLIRYN 228
FrediY-b18                  GYRMVKALLQARGLRVQYDRVRVSMHRADTIGMASRMVQLGCIARRTYSVPGPQSLMHIDTNHKLIRYN 185
FrediY-b14                  ---MVKALLQARGLRVQYDRVRVSMHRVDTIGMASRMVQLGCIARRTYSVPGPQSLMHIDTNHKLIRYN 66
FrediY-b17                  GYRMVKALLQARGLRVQYDRVRVSMHRVDTIGMASRMVQLGCIARRTYSVPGPQSLMHIDTNHKLIR-- 225
Fredi_Ol_SF27104            GYRMTKALLQAQGHRVQYNRVRASMHRVDTSGVLSRMVHLGCIARRTYSVPGPMSLIHIDTNHKLIRYV 225
Fredi_Ol_SF3459             GYRMMKALLQAQGHRVQYNRVRASMHRVDTSGVLSRMVHLGCIARRTYSVPGPMSLIHIDTNHKLIRYV 225
Fredi_Ol_SF15321            GYRMMKALLQAQGHKVQYNRVRASMHRVDTSGVLSRMVHLGCIARRTYSVPGPMSLIHIDTNHKLRRYS 225
Fredi_Dr_GENSCAN000000      GYRMMMGLLRAQGHRVQWDRVRTSMHRVDTASIVSRMSQLGCVVRRTYSVPSPKSLMHIDTNHKLIRVR 225
Fredi_Dr_GENSCAN000000      GYRMMRGALKARGHLVQFERVRASLHRVDTLGVLSRMTQMGCVVRRTYSVSSPRSIMHIDTNHKLIRIE 228
```

Fig. 30. Sequence alignment of the N terminal region (~ 300 aa) of conceptual Fredi proteins from different fish species. Identical residues are shown in white type on black and conservative substitutions in white type on gray. The HTH_7 domain is overscored. Fredix is from a copy 3′ from *Xmrk* on the X chromosome, FrediY-b18 is from a copy in BAC B18 from the *Xmrk* contig on the Y chromosome, FrediY-b14 and b17 are from copies in the *ps-criptY* contig from BAC B14 and B17, respectively. Fredi_Ol SF27104, 3459, 15321 are predicted Fredi-like aa sequences from the medaka. Fredi_Dr_GENSCAN00000040878, 40229 are predicted Fredi-like aa sequences from the zebrafish.

4.3.3 A *DNA polymerase type B* related to organellar and viral genes is located in the sex determination region of *X. maculatus*

4.3.3.1 A *DNA polymerase type B* gene in *X. maculatus*

DNA-directed DNA polymerases can be classified, on the basis of sequence similarity, into at least four different groups: A, B, C, X (Delarue *et al.*, 1990; Jung *et al.*, 1987). B family polymerases can be divided into two groups: one is the DNA polymerase (containing pol2 domain); another is the DNA polymerase type B, organellar and viral (containing DNA_pol_B_2 domain). The latter has been found in plant and fungal mitochondria, and in virus. Some putative proteins, similar to DNA polymerase type B, have been deposited in *C.*

elegans, D. rerio and *F. rubripes* genome databases, but nothing is known concerning their function.

All members in the B family contain a characteristic DTDS motif, which is suggested to be involved in binding a magnesium ion (Argos, 1988), and possess many functional domains, including a 5´-3´elongation domain, a 3´-5´exonuclease domain (Wang *et al*., 1996), a DNA binding domain, as well as binding domains for both dNTP´s and pyrophosphate (Zhou *et al*., 1998).

A gene encoding a putative DNA polymerase type B was found in the *ps-criptY* contig by strategic sequencing. A predicted terminal exon of the *DNA polymerase type B* gene was assembled from genomic DNA of BAC clone F11. It encodes a putative protein of 1158aa in length, which contains a DNA_pol_B_2 domain and a DTDS motif (Fig. 31). The putative DNA polymerase type B-like protein of fugu has two exons. The exon number of the *DNA polymerase type B* gene in the *ps-criptY* contig of the platyfish is not clear (Fig. 31, white box with question mark) and additional exons have not been detected so far in ´´shotgun`` sequences from the *ps-criptY* contig.

4.3.3.2 The DNA_pol_B_2 domain of DNA polymerase type B

The 457 aa-long DNA_pol_B_2 domain is separated in *X. maculatus* into two parts 197 aa from the N terminal end through insertion of a 131 aa segment, compared to DNA_pol_B_2 domains in plant and fungal mitochondria and in viruses (Fig. 31). This insertion seems to be a good encoding sequence without stop codon. No repeats or inverted repeats were found at the boundaries of this insertion. A similar phenomenon was also found in the putative DNA polymerase type B like proteins of *C. elegans*, *D. rerio* and *F. rubripes.* On the other hand, such a difference might be well explained by an in-frame deletion in the organellar/viral gene lineage.

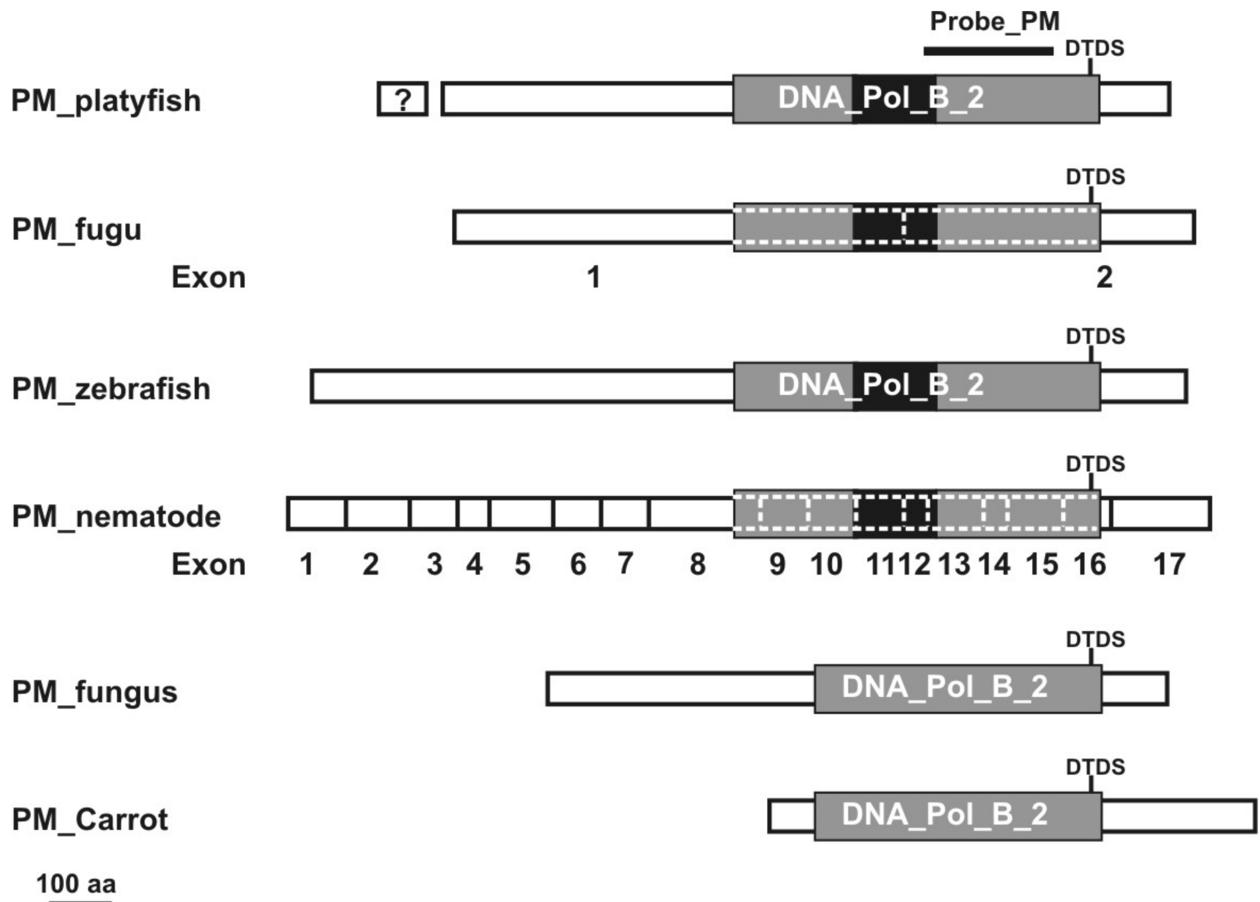


Fig. 31. Structure of some putative DNA polymerase type B proteins from fish and *C. elegans* as well as DNA polymerase type B of Carrot and fungal (Gelasinospora) mitochondria. The white box represents exons, the gray box is the DNA_pol_B_2 domain, the black box represents a 130-132 aa sequence, which is absent in plant and fungal DNA_pol_B_2 domain. DTDS is a family B polymerase motif. Accession number of PM_zebrafish is Q7ZYY7, PM_fugu is SINFRUP000000160091, PM_Carrot is BAC16364.1, PM_fungus is Gelasinospora _ S62752.

An alignment of the DNA_pol_B_2 domains of DNA polymerases type B was constructed for phylogenetic analysis. A representative tree, generated by distance analysis, is shown in Fig. 32. It is apparent that the DNA polymerases type B found in fishes are monophyletic. The remaining DNA polymerases type B fall into species-specific groups, with the *C. elegans*, viruses, fungus and plant genes grouping together, respectively.

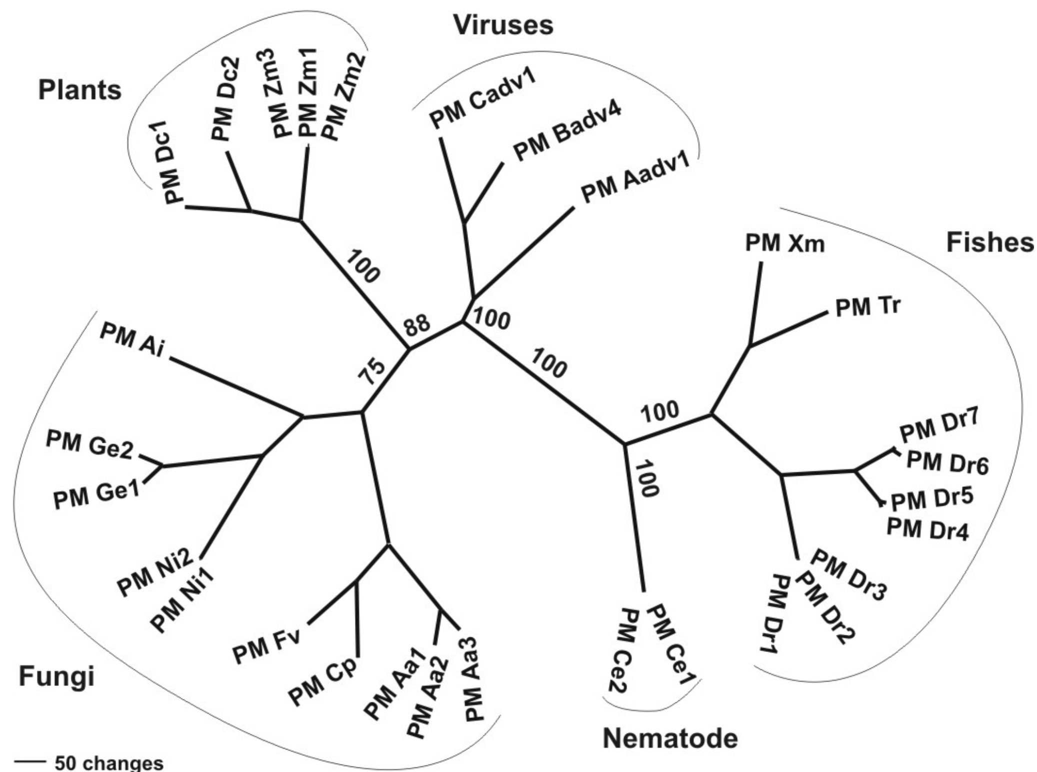


Fig. 32. Phylogenetic tree based on the DNA_pol_B_2 domains of the DNA polymerases type B. The tree was constructed by the maximum parsimony using PauP* 4.0 (Swofford, 2003) and is a consensus derived from a heuristic search of 1000 bootstrap replicates. The number indicates the percentage bootstrap support. The accession numbers of the sequences for analysis are: PM_Ge1 is PM_Gelasinospora_AAB41447.1, PM_Ge2 is S62752; PM_Ni1 is *Neurospora intermedia*_CAA36327.1, PM_Ni2 is S17909; PM_Ai is *Ascobolus immersus*_P22374; PM_Fv is *Flammulina velutipes*_BAB13496.1; PM_Cp is *Crinipellis perniciosa*_YP025890.1; PM_Aa1 is *Agrocybe aegerita*_AAG44763.1, PM_Aa2 is AAG44762.1, PM_Aa3 is AAC33727.1; PM_Cadv1 is canine adenovirus type 1 _Q65946, PM_Badv4 is Bovine adenovirus type 4_Q997I4, PM_Aadv1 is Avian Adenovirus gal 1_ Q7M5G2; PM_Zm1 is *Zea mays*_CAB57809.1, PM_ Zm2 is P10582, PM_ Zm3 is S07183; PM_Dc is *Daucus carota*_AAS15054.1; PM_Bn is *Brassica napus*_BAC16364.1; PM_Ce1 is *Caenorhabditis elegans*_C25F9.2, PM_Ce2 is NP507780.1; PM_Dr1 is *Danio rerio*_Q7ZYY7, PM_Dr2 is NP001004509.1, PM_Dr3 is ENSDARP00000041958, PM_Dr4 is ENSDARP00000034761, PM_Dr5 is ENSDARP00000006982, PM_Dr6 is ENSDARP00000034544, PM_Dr7 is ENSDARP00000036764; PM_Tr is SINFRUP00000160091. PM_ Xm is the putative DNA polymerases type B of *Xiphophorus maculatus* identified in this work.

4.3.3.3 Location and expression analysis of *DNA polymerase type B*

The *DNA polymerase type B* gene has ca. 10 copies in the haploid *X. maculatus* genome (Fig. 33B), only one of which in the *ps-criptY* contig (11 kb, but not a Y-specific band). No copy was found in the *Xmrk* contigs on the X and Y chromosomes (Fig. 33A). The RFLP analysis on genomic DNA indicated that this gene has a Y specific band (7 kb) in the genome, but this Y-specific band is located neither in the *ps-criptY* contig nor in the *Xmrk* contigs of *X. maculatus* (Fig. 33).

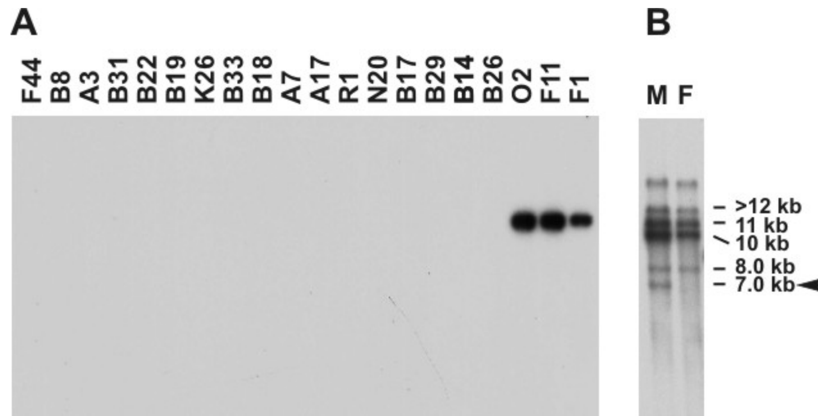


Fig. 33. Southern blot hybridization-based RFLP analysis of the *DNA polymerase type B* gene. BAC clones (A) and genomic DNA from male (M) and female (F) *X. maculatus* (B) were digested with *Hin*dIII and hybridized with the probe generated from the sequence encoding the C terminal region of the DNA_pol_B_2 domain called Probe_PM (Fig. 32). The BAC clones (A) cover the *ps-criptY* contig and the *Xmrk* contigs on the X and Y chromosome. The Y-specific band is indicated by a black triangle (7.0 kb).

The expression analysis of *DNA polymerase type B* indicated that it is expressed ubiquitously in male and female *X. maculatus*, although its expression was very weak in the ovary. Its transcripts were also detected in stage 20 embryos of *X. maculatus*, stage 21 embryo of hybrids (called BCn407) between *Xiphophorus maculatus* and *Xiphophorus hellerii*, in the embryonic cell line A2 and in the PSM melanoma cell line (Fig. 34).

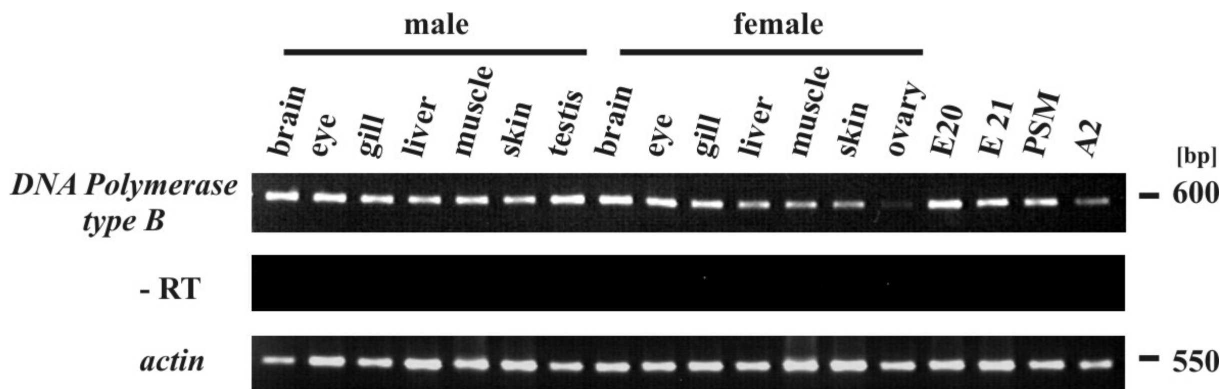


Fig. 34. RT-PCR-based expression analysis of *DNA polymerase type B* in *X. maculatus.* The primer pairs amplify the region encoding the DNA_pol_B_2 domain of DNA polymerase type B (Fig. 32, Probe_PM). E20 is the stage 20 embryo of *X. maculatus* Jp163A. E21 is stage 21 embryos of hybrids (called BCn407) between *Xiphophorus maculatus* and *Xiphophorus hellerii*. A2 is the embryonic cell line of *Xiphophorus xiphidium*. PSM is a melanoma cell line from a hybrid between *Xiphophorus maculatus* and *Xiphophorus hellerii.* PCR program: 1* [3'00'' at 95°C], 35* [0'40'' at 95°C/ 0'40'' at 55 °C/ 1'00'' at 72°C], 1* [5'00'' at 72°C], 4°C.

4.3.3.4 Distribution of *DNA polymerase type B* in ray-finned fishes

Using the probe_PM in Southern blot hybridization analysis, *DNA polymerase type B* was found to be present in all tested *Xiphophorus* and in some other poeciliids (except *P. latipunctata, P. gracilis and B. belizanus*) with a variable number of copies (Fig. 35A-B). The copy number per haploid genome of *Xiphophorus* was higher than in some other poeciliids. Moreover, RFLPs were observed between different populations of a same species (*X. maculatus* and *X. hellerii*) as well as between male and female within a same species (such as *X. maculatus, X. milleri, X. hellerii, P. mexicana* and *G. falcatus*). Interestingly, the *DNA polymerase type B* is also apparently located on the W chromosome of the *X. maculatus* Usumacinta strain like *swimy* (Fig. 35A, white triangle). With the same probe, no signal could be observed in the closely related family Fundulidae (tested on *Fundulus heteroclitus*) and in more divergent bony fishes (Fig. 35B-C). In zebrafish, a putative homologous protein (Q7ZYY7, E-value = 0) presents 59% similarity with the *X. maculatus DNA polymerase type B* exon product (NCBI, blastp). It is encoded by a single exon, has a length of 1392 aa and also contains the bipartite DNA_pol_B_2 domain like the DNA polymerase type B in *X. maculatus.* From the Ensembl zebrafish genome browser, more than 6 predicted proteins (E-value < e^{-100}) also contain the DNA_pol_B_2 domain, but they are encoded by a different number of exons.

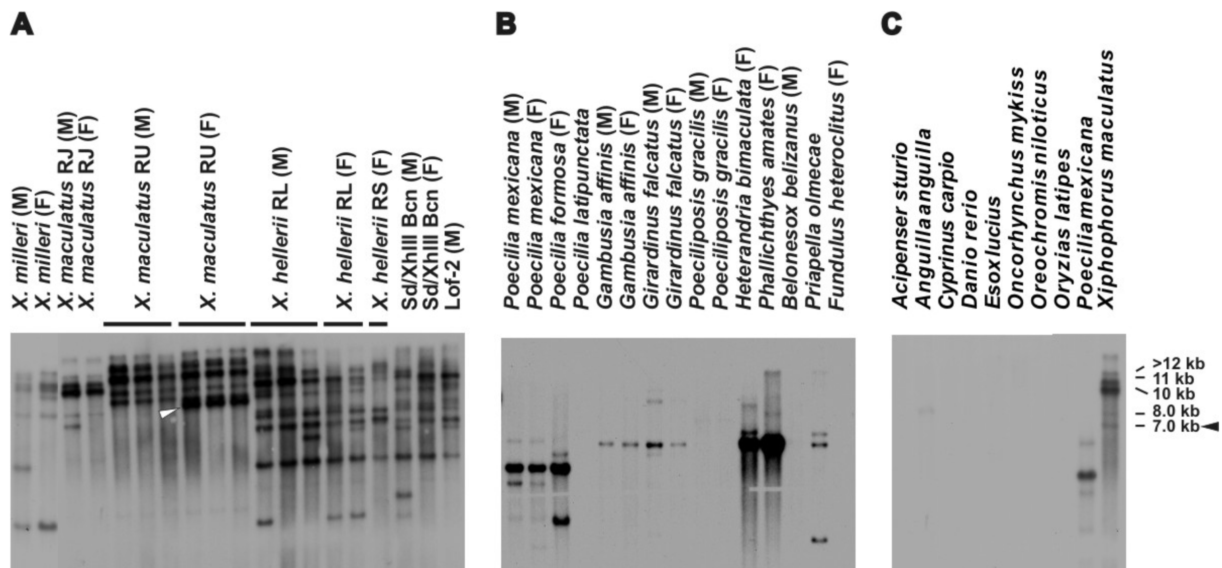


Fig. 35. Southern blot hybridization-based analysis of the distribution of *DNA polymerase type B* in *Xiphophorus* (A), poeciliids (B) and other bony fish species (C). The probe was the RT-PCR product encoding the DNA_pol_B_2 domain of DNA polymerase type B. Genomic DNAs were digested with *Hin*dIII. The tested fish are described in MATERIALS. The Y-specific band is indicated by a black triangle (7.0 kb) and the W specific band is indicated by a white triangle (8.5 kb). *Fundulus heteroclitus* belongs to Fundulidae, a family closely related to the Poeciliidae.

Surprisingly, in medaka (*Oryzias latipes*), no ´´shotgun`` sequences or scaffolds matched the *DNA polymerase type B* in the latest version of the medaka genome sequence (Jan. 2005). In *Takifugu rubripes* and *Tetraodon nigroviridis* genome, putative proteins similar to DNA polymerase type B were identified. So far *DNA polymerase type B* was not found in human, chimp, mouse, rat, chicken and fly (from Ensembl data, Jan. 2005).

4.3.4 Other gene candidates in the *ps-criptY* contig

A gene candidate (*gch1*) putatively encoding a GTP cyclohydrolase I was identified in the *ps-criptY* contig (Tab. 4). GTP cyclohydrolase I catalyses the conversion of GTP to dihydroneopterin triphosphate and initiates the pteridine pathway. *Tmpl* and *trail* are new genes identified in this work (Tab. 4). A single exon of *tmpl* was predicted from a 4229-nt assembled genomic sequence from the *ps-criptY* contig. *Tmpl* belongs to the tripartite motif protein family, which is involved in specific cell compartments by means of homo-multimerization (Reymond *et al*., 2001). The conceptual protein sequence (446 aa) of *tmpl* from the N- to the C-terminus contains ring-finger, B-box-type zinc finger, coiled coil and PRY domains, respectively. This is the typical domain order of the tripartite motif family. A gene candidate (*trail*), which is predicted to encode a tumor necrosis factor (TNF)-related apoptosis inducing ligand (Trail)-like protein, was also identified in the *ps-criptY* contig. Trail/Apo2L is one of several members of the tumour necrosis factor superfamily. It is involved in apoptosis through the interaction with an unusually complex receptor system, which comprises two death receptors and three decoy receptors in humans. Many cancer cell lines are sensitive to the Trail/Apo2L-induced apoptosis (Bouralexis *et al*., 2005; Tracey and Cerami, 1994). Only a partial sequence of *trail* could be assembled so far from *ps-criptY* contig shotgun sequences, but this sequence was predicted to encode a product with a TNF domain (SM00207). *Trail* was similar to some EST sequences from the whole bodies of the adult male and female stickleback (CD505120), from a pool of embryos and a lot of organs of the rainbow trout (BX299019) and from a pool of head and internal organs of the adult male zebrafish (C0807698, CN505671, CK026586 and CK030145). *Acr, igd1, igd2* and *fbx11*genes (Tab. 4) are present in both the *ps-criptY* contig and in the X-linked *Xmrk* contig. They were also partially assembled from *ps-criptY* contig shotgun sequences.

4.3.4.1 A multicopy *msh2* (pseudo) gene on the Y chromosome of *X. maculatus*

Mismatch repair is one of five major DNA repair pathways, the others being homologous recombination repair, non-homologous end joining, nucleotide excision repair, and base excision repair (Bernstein *et al*., 2002). The mismatch repair system recognizes and repairs mispaired or unpaired nucleotides that result from errors in DNA replication. Many proteins involved in the different repair processes also play a role in apoptosis when DNA damage is excessive, thereby helping to prevent carcinogenesis (Bernstein *et al*., 2002). The mismatch repair protein, Msh2 (MutS homologue 2), has a dual role in DNA repair and apoptosis and is required for the maintenance of genome integrity. Msh2 acts as a heterodimer with Msh6, which together function to recruit the Mlh (MutL homologue) – Pms (post-meiotic segregation) heterodimer, and to replace the mispaired base (Drotschmann *et al*., 2002). Germline mutations in the human *hMSH2* gene are very frequently detected in hereditary nonpolyposis colorectal cancer (HNPCC).

In order to investigate whether *msh2* is active or not in *X. maculatus*, its expression pattern was studied. The RT-RCR results showed that except in ovary, *msh2* is expressed in all tissues of males and females as well as in embryos, in an embryonic cell line A2 and in a melanoma cell line (Fig. 36).

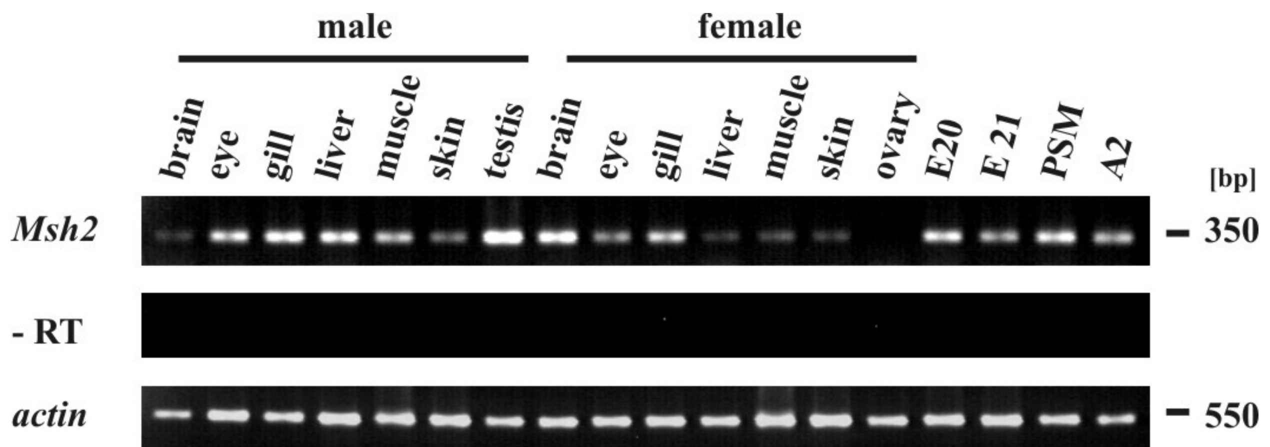


Fig. 36. RT-PCR-based expression analysis of *msh2* in *X. maculatus.* The primer pair flanks the exons 10-12 of *Msh2*. E20 are stage 20 embryos of *X. maculatus* Jp163A. E21 are stage 21 embryos of hybrids (named BCn407) between *Xiphophorus maculatus* and *Xiphophorus hellerii*. A2 is an embryonic cell line of *Xiphophorus xiphidium*. PSM is a melanoma cell line derived from a hybrid between *Xiphophorus maculatus* and *Xiphophorus hellerii.* PCR program: 1* [3’00’’ at 95°C], 35* [0’40’’ at 95°C/ 0’40’’ at 55 °C/ 1’00’’ at 72°C], 1* [5’00’’ at 72°C], 4°C.

Msh2 has 5-10 copies per haploid genome in *X. maculatus* and 4 Y-specific bands (Fig. 37B). About 20 BAC clones containing *msh2* were isolated from the *X. maculatus* BAC library and further analyzed. All size bands of *msh2* detected in *X. maculatus* genome could be found in these positive BAC clones (Fig. 37). There are two copies of *msh2* in the *Xmrk* contig on the X chromosome, one copy in the *Xmrk* contig on the Y chromosome and two copies in the *ps-criptY* contig (Froschauer, A., pers. commun. and Fig. 37). The BAC clone C8, which is positive in the BAC library screening using the RT-PCR product of *msh2* as a probe, contains the other two Y-specific bands of *msh2* (6.9 kb and 4.5 kb). Hence, C8 certainly belongs to the Y chromosome. However, no relationship was found so far between the BAC clone C8 and the current Y chromosomal contigs. The other signals came from autosomes (the BAC C1, 2, 4, 5 and 7 have been shown to belong to autosomes (Froschauer, 2003).Two *msh2* copies found in the *ps-criptY* contig might be truncated, because only partial sequences of them were obtained by strategic sequencing. This is consistent with the result that *msh2* copies in the *Xmrk* region of *X. maculatus* are truncated (Volff *et al.*, 2003).

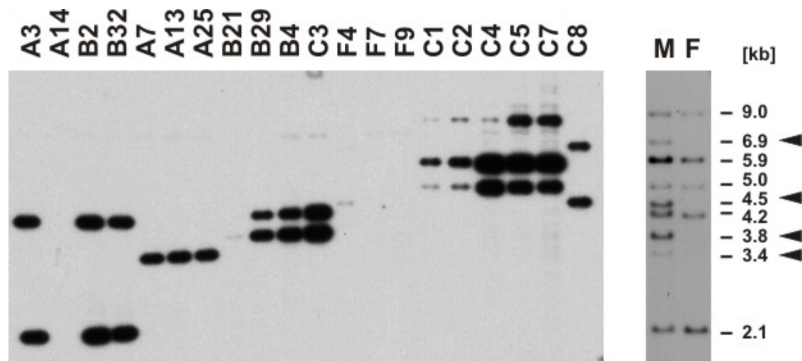


Fig. 37. Southern blot hybridization analysis of BAC clones containing *msh2.* The BAC clones (A) and genomic DNA from male (M) and female (F) *X. maculatus* (B) were digested with *Hin*dIII and hybridized with a *msh2* partial probe generated by RT-PCR of adult testis RNA. The BAC clones (A) are positive for the *msh2* probe in the BAC library screening. The Y-specific bands are indicated by black triangles (3.4 kb, 3.8 kb, 4.5 kb and 6.9 kb).

4.3.4.2 Sequence analysis of *ps-cript1* on the X chromosome and comparison with *ps-criptY*

There are three copies of *cript* in the *X. maculatus* genome. One of them is autosomal; the second one is linked to *Xmrk* at allelic position on the X and Y chromosomes; the third copy is Y-specific and located in the *ps-criptY* contig. In order to analyze these copies, a 5514 bp segment containing the copy called *ps-crip1* on the X chromosome was sequenced by ´´transposon tagging`` sequencing. The sequence analysis of *cript*, *ps-cript1* (on the X and Y

chromosomes) and *ps-criptY* with the NIX-Software (http://menu.hgmp.mrc.ac.uk/menu-bin/Nix/) showed that *ps-cript1* is truncated and only contains exons 1, 2 and a partial exon 3, together with the *Rex4*-LTR (a long-terminal repeat from an endogenous retrovirus) (Froschauer, 2003). Strikingly, this 5514-bp X-linked segment had 99.9% nucleotide identity with the corresponding sequence on the Y chromosome, suggesting recent recombination between the X and Y chromosomes in this region.

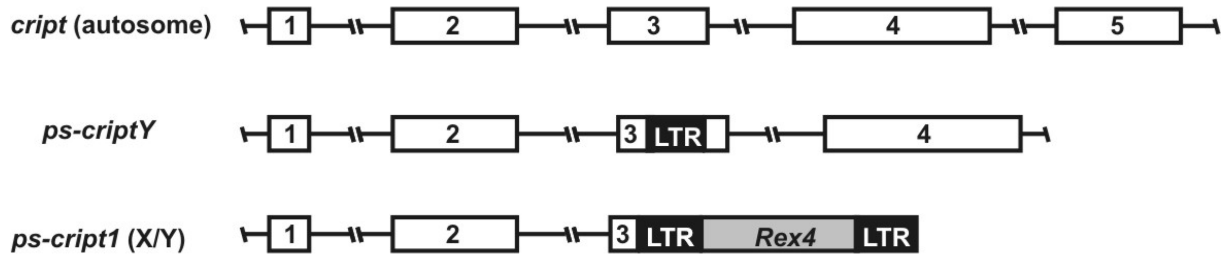
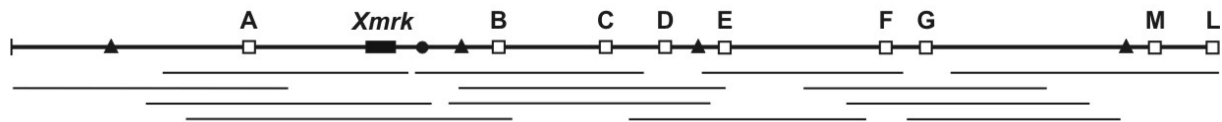


Fig. 38. Genomic structure and predicted exons of the *cript* sequences in *X. maculatus.* The white boxes are the *cript* exons; the black box is the Rex4-LTR (long terminal repeat); the gray box is the *Rex4* coding region. Exon numbers are shown in the boxes.

4.3.4.3 Sequence analysis of *mc4rP, mc4rQ* and *mc4rR*

As many as 11 melanocortin type 4 receptor genes (*mc4r*) have been identified in BAC contigs from both the X and Y chromosomes (Fig. 39). Three of them, called *mc4rP*, *mc4rQ* and *mc4rR*, are located in the *ps-criptY* contig (Fig. 39 and 41). Melanocortin receptors can bind the alpha melanocyte-stimulating hormone and are involved in the control of feeding and body weight (*mc4r*) in mammals (Huszar *et al.*, 1997; Vaisse *et al.*, 1998). In fish, the onset of sexual maturity determined the size of the body (Kallman, 1989), and some of these genes might correspond to the puberty (*P*) locus. All X- and Y-linked *mc4r* genes were completely sequenced through a collaborative effort of our group. *Mc4rP, mc4rQ* and *mc4rR* were accomplished by ´´transposon tagging`` sequencing in this work.

X chromosome (1 contig)



Y chromosome (4 contigs)

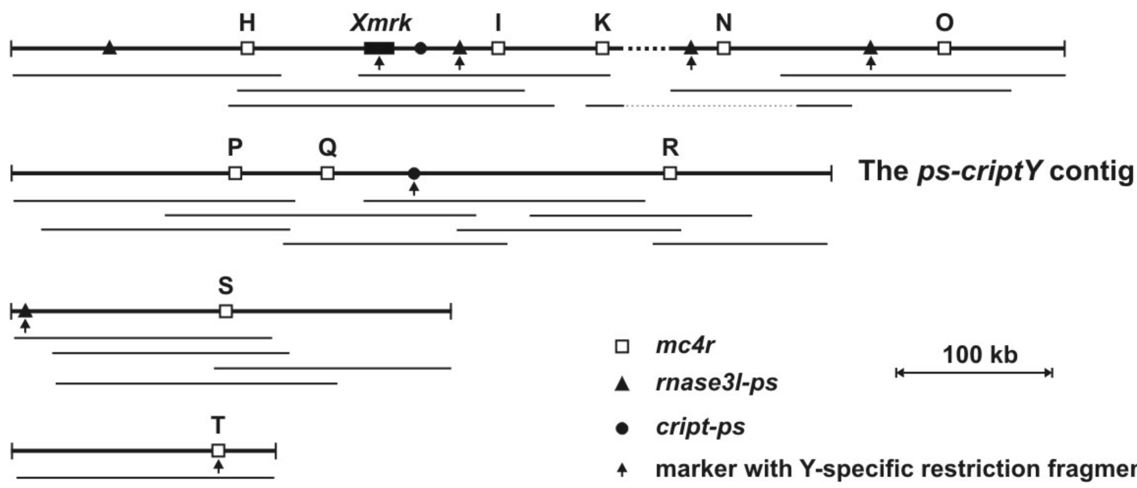
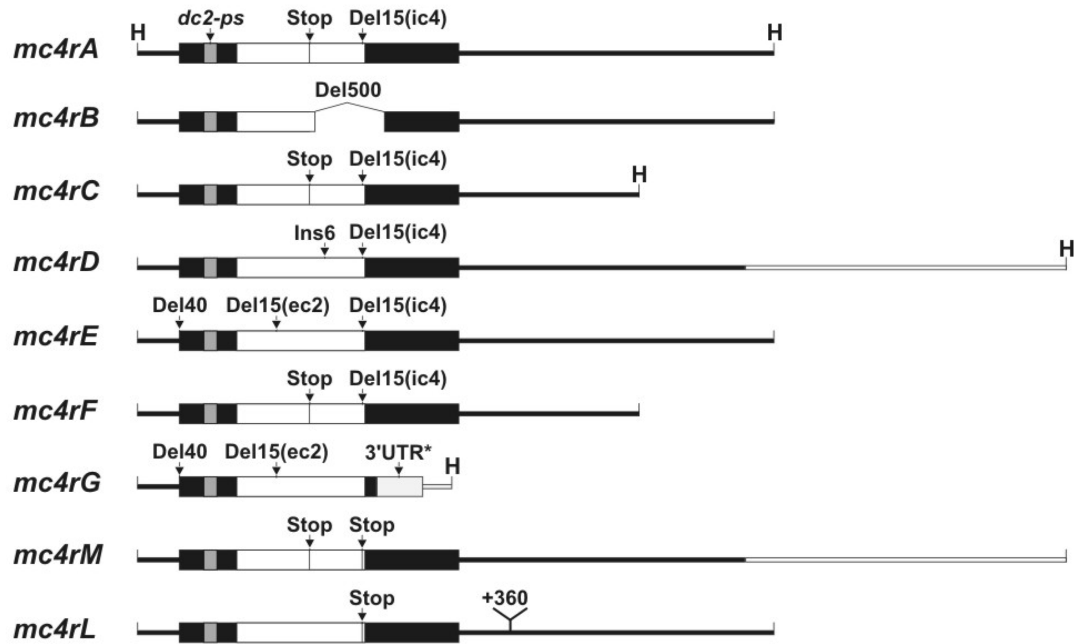


Fig. 39. Schematic locations of *mc4r* copies in different contigs on the X and Y chromosomes of *X. maculatus* (adapted from Froschauer *et al*., in preparation).

These three *mc4r* copies in the *ps-criptY* contig show more than 90% nucleotide identity to each other. In their 5’ untranslated region, all of three *mc4r* copies contain the last exon of an unrelated gene of unknown function called *dc2* in humans. Structure analysis (Fig. 40) revealed that *mc4rQ* has an open reading frame interrupted by a stop codon found after ca. Half of the coding sequence. Although it can not be excluded that truncated receptors might be functional, for example in a dominant-negative manner, the *mc4rQ* copy probably correspond to a pseudogene. Both *mc4rP* and *mc4rR* showed in-frame microdeletions in their coding sequences compared to wild-type receptor genes. These two copies present also an alternative 3’ untranslated region similar to one X-chromosomal copy (*mc4rG*). The phylogenic analysis showed that *mc4rP* and *mc4rG* belong to one group (Selz, 2003). *Mc4rP* and *mc4rR* copies are located in a region duplicated in the *ps-criptY* contig (Fig. 41, the red dotted line). This is another example of gene duplication in the *ps-criptY* contig.



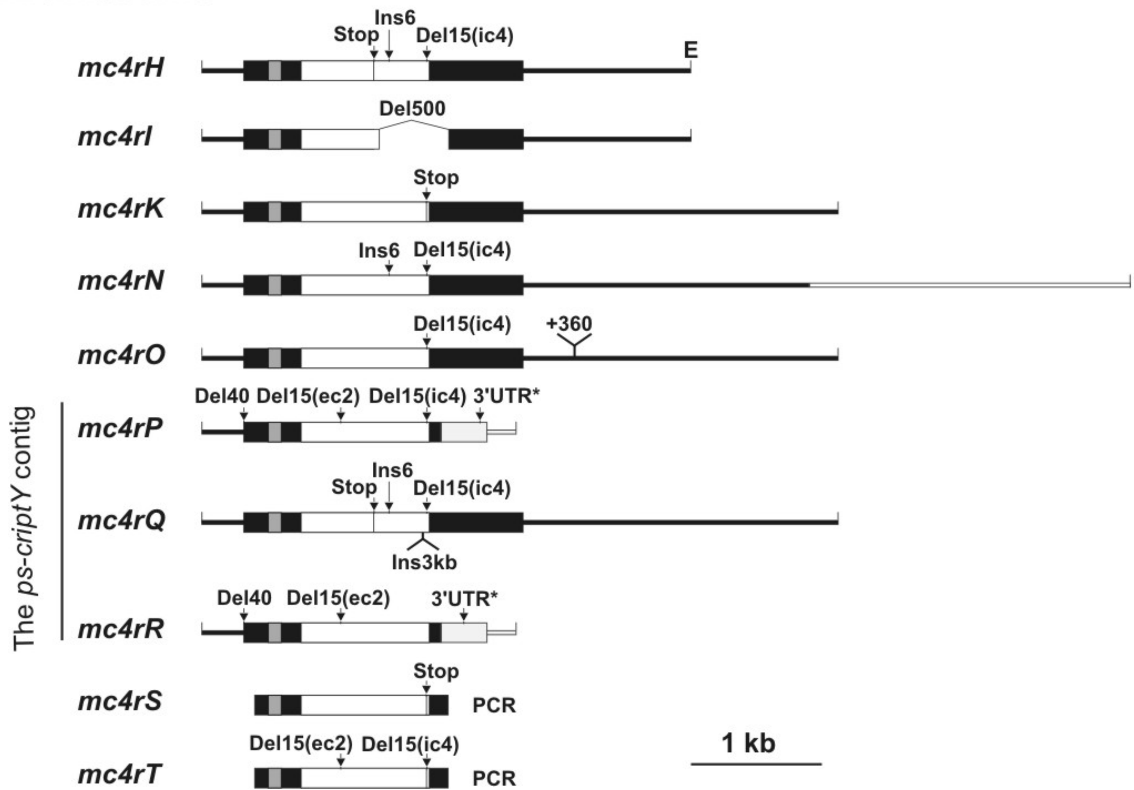


Fig. 40. Structure of sex-chromosomal *mc4r* genes of *X. maculatus.* The structural characteristics of the different *mc4r* copies are shown. The frame represents the mRNA of *mc4r*, the white frame its ORF (adapted from Froschauer *et al*., in preparation).

4.4 Structural analysis of the *ps-criptY* contig of *X. maculatus*

4.4.1 Genomic structure and instability of the *ps-criptY* contig

Numerous retroelements including an endogenous retrovirus (*Rex4*), several LTR (long terminal repeat) retrotransposons (*Jule*, *Rex7*, *Rex8* and *T-Rex*), non-LTR retrotransposons (*Rex1*, *Rex2*, *Rex3*, *Rex5*, *Rex6* and *L1-like*) and a LTR-like repeat (*XIR*), as well as various short repeats, have been found in the *SD*-linked *Xmrk* BAC contigs (Volff *et al*., 2003a; Volff *et al*., 2003b; Volff and Schartl, 2003; Volff *et al*., 2003). By sequencing and Southern blot hybridization, most of them were also identified in the *ps-criptY* contig (such as *T-Rex*; *Rex1*, *Rex2*, *Rex5, L1-like*, *XIR*). Besides these, numerous novel repeated elements were found in the *ps-criptY* contig in this work, including several non-LTR retrotransposons (*swimmer1* and *BABAR*), several DNA transposons (*Tc-1, helitrons*, AC-like transposons and putative elements encoding a transposase), as well as microsatellites.

The blast results of consensus sequences from the *ps-criptY* contig against the 5′ *Xmrk* and 3′ *Xmrk* contigs on the Y chromosome (Volff *et al*., 2003b) showed that there are about 29 kb of sequences in 34 contigs highly identical (> 95%) with the 5′ *Xmrk* contig and 10 kb of sequences in 21 contigs highly identical (> 95%) with the 3′ *Xmrk* contig, suggesting recent duplications between the *Xmrk* contig and the *ps-criptY* contig. Such intrachromosomal duplications could even be found within the *ps-criptY* contig, for example, a region containing *helitron, fredi, tnh2, etc.* (Fig. 41).

In BAC B17 DNA, 68.2 of 108.2 kb non-redundant sequences assembled contain *XIR* or other repeats. Complete sequencing of a 34 kb region including *ps-criptY* revealed a high concentration of the *XIR* repeat (Froschauer, A., unpublished data). These data indicate that the *ps-criptY* contig is subject to genomic instability.

Even though the great majority of markers located on the *ps-criptY* contig have counterparts on the X-linked *Xmrk* contig, major differences between the X and the Y chromosomes were observed in this region (Fig. 41). Particularly, some sequences of the *ps-criptY* contig were present at higher copy numbers on the Y than on the X chromosome, which include the Y-specific copy of *cript*, and a region containing gene candidates and elements duplicated in the *ps-criptY* region but not on the corresponding X chromosome contig (Fig. 41, frames). In

addition, a region containing *mc4r* as well as sequences homologous to the terminal fragments B17-FOR-*Dra*I and B26-REV-*Rsa*I is duplicated and inverted in this contig (Fig. 41, red arrows). Strikingly, some genes are so far only found in the *ps-criptY* contig but not found in the X contigs such as *swimy* and *polymerase type B* (Fig. 41). Taken together, these results strongly suggest that molecular differentiation between the X and Y chromosomes occurred in *X. maculatus* by duplications/deletions, inversions as well as through accumulation of transposable elements and repeats.

4.4.2 Synteny analysis of the *ps-criptY* contig

A synteny between the *ps-criptY* contig of the platyfish and human chromosome 2 was detected (2p21-22: *cript*, *gch1* and *msh2*). Furthermore, a synteny (*cript*, *gch1* and *msh2*) was also detected in scaffold 007971 of Tetraodon (*Tetraodon nigroviridis*), scaffold003107 of fugu (*Takifugu rubripes*) and scaffold Zv4_NA14529 of zebrafish (*Danio rerio*).

Another synteny (*mc4r*, *trail* and *igd1*) between the *ps-criptY* contig of the platyfish and scaffold 14565 of the pufferfish (*Tetraodon nigroviridis*) was also observed.

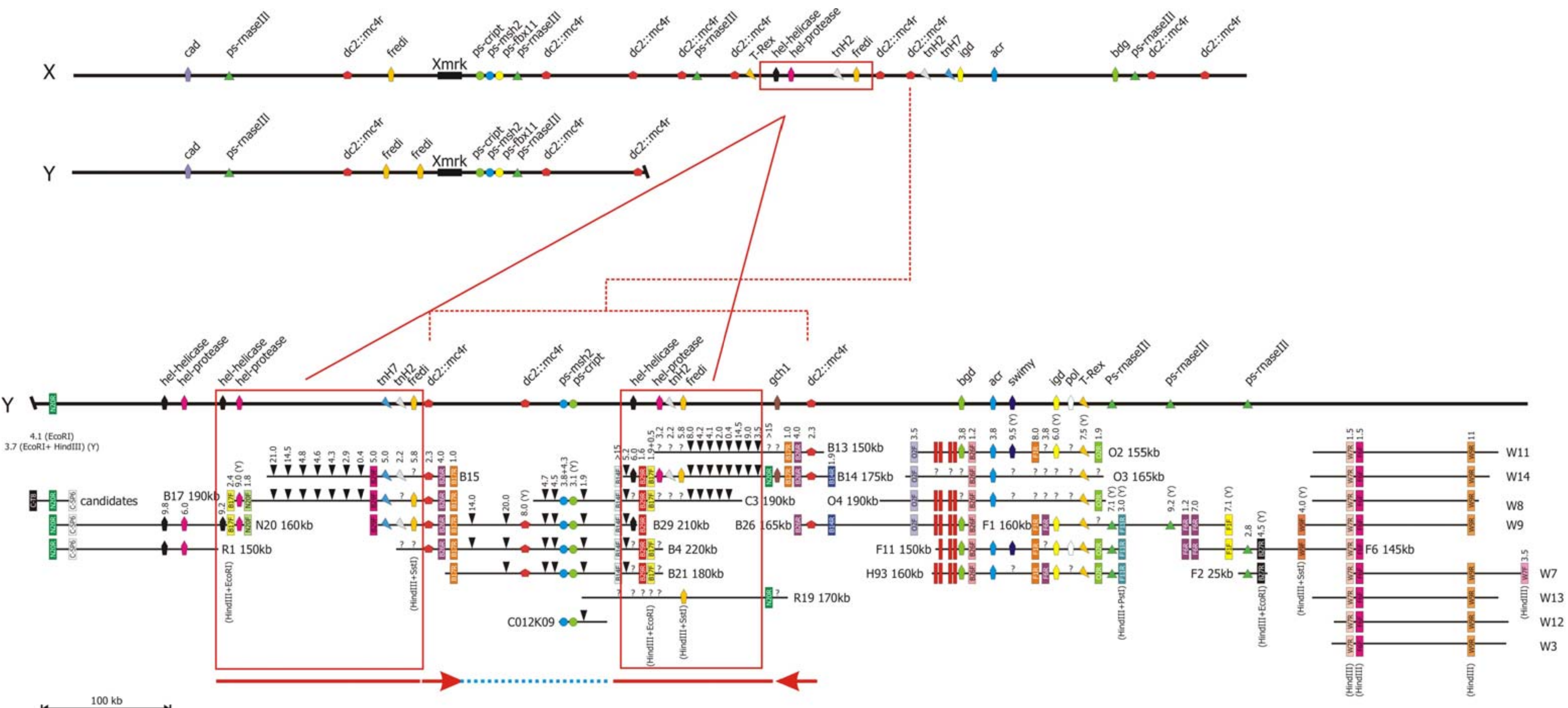


Fig. 41. The structural comparison of the *Xmrk* contigs on the X and Y chromosomes and the Y specific *ps-criptY* contig. The black solid lines represent the contigs on the X and Y chromosomes. The red solid lines and frames represent the region that is duplicated on the *ps-criptY* contig, but not on the X contig. The red arrows represent the region that is inverted on the *ps-criptY* contig. The blue dotted line represents the region containing *ps-criptY*.

4.5 *Helitron*, a rolling circle transposon, on the sex chromosomes of *X. maculatus*

Recently, a novel class of DNA transposons, called *helitron,* has been reported in eukaryotes (Kapitonov and Jurka, 2001). Such elements do not present any terminal inverted repeats and target site duplications. Autonomous *helitrons* have an open reading frame (ORF) which encodes a rolling circle replication initiator protein (Rep) and a helicase from the steroidogenic factor (SF1) superfamily, suggesting that *helitron* is transposed by some form of rolling circle replication (RCR) mechanism (Curcio and Derbyshire, 2003; Feschotte and Wessler, 2002; Kapitonov and Jurka, 2001).

Many evolutionarily diverse replicons can proliferate by RCR. These elements include circular single-stranded bacteriophages, bacterial or archaeal plasmids, some bacterial transposons (such as IS91, IS801 and IS1294) and the plant geminiviruses.

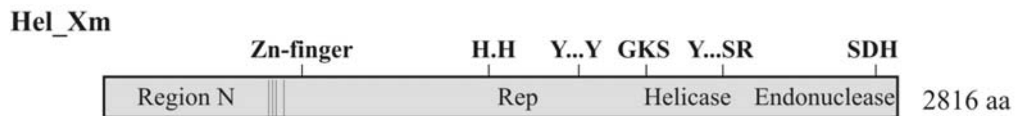
Autonomous *helitrons* and their derived nonautonomous counterparts have been described in the rot white fungus, plants, *C. elegans,* mosquito, fruit fly, zebrafish and pufferfish. Moreover, *helitrons* make up of about 2% of *Arabidopsis* and *C. elegans* genome (Kapitonov and Jurka, 2001). Most of the apparently autonomous *helitrons* encode Rep and helicase within a single ORF. The plant elements contain other genes in a separate ORF, including RPA (Replication protein A) genes (Kapitonov and Jurka, 2001). Interestingly, besides the Rep and helicase, the fish elements encode also an apurinic-apyrimidinic (AP) endonuclease (EN) within a single ORF (Poulter *et al*., 2003).

4.5.1 *Xiphophorus helitrons*

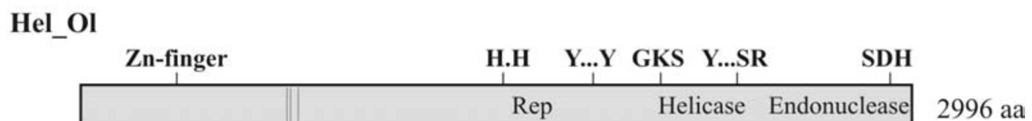
When the *ps-criptY* contig of *X. maculatus* was sequenced by ´´shotgun`` and analyzed, some sequences were found with significant matches to the *Danio* helitrons. A reconstructed consensus ORF of the *Xiphophorus helitrons* was assembled from these ´´shotgun`` sequences. Its deduced protein sequence is 2816 aa in length, 73.3% similar to one of the *Danio* helitrons (called Hel_Dr5) and has similar structure hallmark to the *Danio* helitrons*,* encoding Rep, helicase and AP endonuclease within a single ORF. It contains all conserved motifs of Rep, helicase and EN domains (Fig. 42A). The N terminal region of the protein contains repeats of an ~ 210 residue sequence. A NFX zinc-finger motif and a putative zinc-finger motif are at approximately 600-870 residues downstream of the N terminus. Like

helitrons in zebrafish, the ORF apparently lack introns. This is in contrast to the *helitrons* described from plants and *C. elegans* (Kapitonov and Jurka, 2001).

A. *Xiphophorus maculatus*



B. *Oryzias latipes*



1kb

Fig. 42. Reconstructed consensus *helitrons* of Xiphophorus and medaka fish. Shaded boxes indicate the open reading frames. The position of the putative zinc finger domains is indicated by Zn-finger. H.H and Y…Y represent motifs in the Rep domain; GKS and Y..SR are helicase domain motifs; SDH the motif of the AP endonuclease domain. Vertical lines indicate the repeats of an ~ 218 residue sequence. Region N represents the 325-758 residues downstream of the N terminal of *Xiphophorus* helitron.

Helitrons lack easily definable termini, compared to other DNA transposons. The reason why the *Arabidopsis* and *C. elegans helitrons* are accurately determined is that they often appear within well-defined repeats, hence defining the insertion and therefore the termini. These *helitrons* are integrated precisely into AT sequences and begin with conserved 5`TC and CTRR 3` nucleotides. But some elements such as *Danio helitrons* can not be analyzed by this way as they are not integrated into informative sequences (Poulter *et al*., 2003). The *Xiphophorus helitrons* termini could not be identified for the same reason.

Xiphophorus helitrons have less than 15 copies per haploid genome (Fig. 43B, 44). All *helitrons*-containing BAC clones (56 clones) from the *X. maculatus* BAC library have been analyzed. They are divided into at least ten families by the restriction fragment length polymorphism (RFLP) analysis using three probes generated from region N, Rep domain and EN domain (Fig. 43, 6). Additionally, two of 10 families are *helitrons* without delectable EN domain (Fig. 43C and Tab. 5).

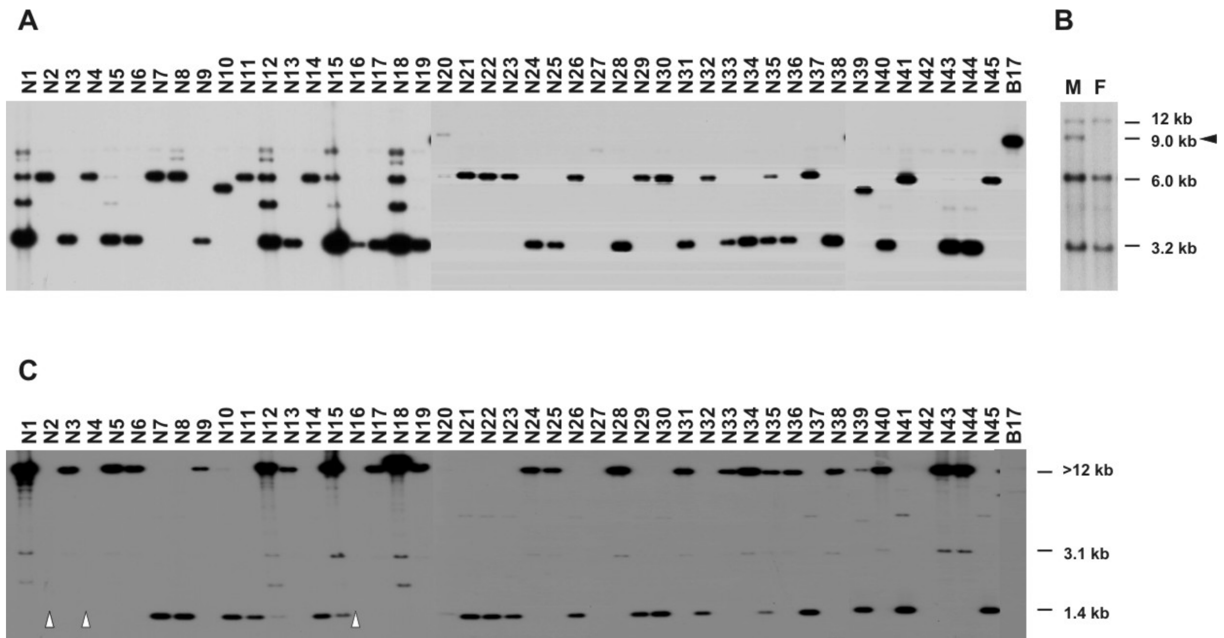


Fig. 43. Southern blot hybridization analysis of BAC clones (A and C) and genomic DNA from male (M) and female (F) *X. maculatus* (B) with the probe generated from helitron region N (A-B) and EN domain (C). The BAC clones in (A) are positive for the F17-FOR-*Dra*I in the BAC library screening. The Y-specific band is indicated by a black triangle (9.0 kb). White triangles show BAC clones without EN domain. All BAC clones and genomic DNA were digested with *Hin*dIII.

Table 5. The groups of *helitrons* in *X. maculatus*

Group	Protease	Helicase	Endonuclease
1	N1, 12, 15, 18	N1, 12, 15, 18	N1, 12, 15,18
2	N2, 4, 7, 8, 11, 14, 21, 22, 23, 26, 29, 30, 32, 37, 41, 45,	N2, 4, 7, 8, 11, 14, 21, 22, 23, 26, 29, 30, 32, 37, 41, 45,	N7, 8, 11, 14, 21, 22, 23, 26, 29, 30, 32, 37, 41, 45,
3	N10, 39	N10, 39	N10, 39
4	N3, 5, 6, 9, 13, 16, 17, 19, 24, 25, 28, 31, 33, 34, 35,36, 38, 40, 43, 44	N3, 5, 6, 9, 13, 16, 17, 19, 24, 25, 28, 31, 33, 34, 35, 36, 38, 40, 43, 44	N3, 5, 6, 9, 13, 17, 19, 24, 25, 28, 31, 33, 34, 35, 36, 38, 40, 43, 44
5	N20, B17, *etc.* (Y)	N20, B17, *etc.* (Y)	N20, B17, *etc.* (Y)
6	N20, R1, *etc.* (Y)	N20, R1, *etc.* (Y)	N20, R1, *etc.* (Y)
7	B14, B29, *etc.* (Y)	B14, B29, *etc.* (Y)	B14, B29, *etc.* (Y)
8	B11, B24, *etc.* (X)	B11, B24, *etc.* (X)	B11, B24, *etc.* (X)
9			N2, 4 (from group 2)
10			N16 (from group 4)

4.5.2 At least four *helitrons* of *X. maculatus* are located in the sex determining region

Thirteen of 56 *helitrons*-containing BAC clones are located in the sex determination region of *X. maculatus.* All Y-specific bands were found in BACs of the sex determination region (Fig. 44A, B). Except for BAC clones belonging to the *Xmrk* contig on the X chromosome, the remaining *helitrons*-containing BAC clones did not contain the 12 kb *Hin*dIII-band (Fig. 43A and 44A), suggesting that the 12 kb band is gonosome-specific.

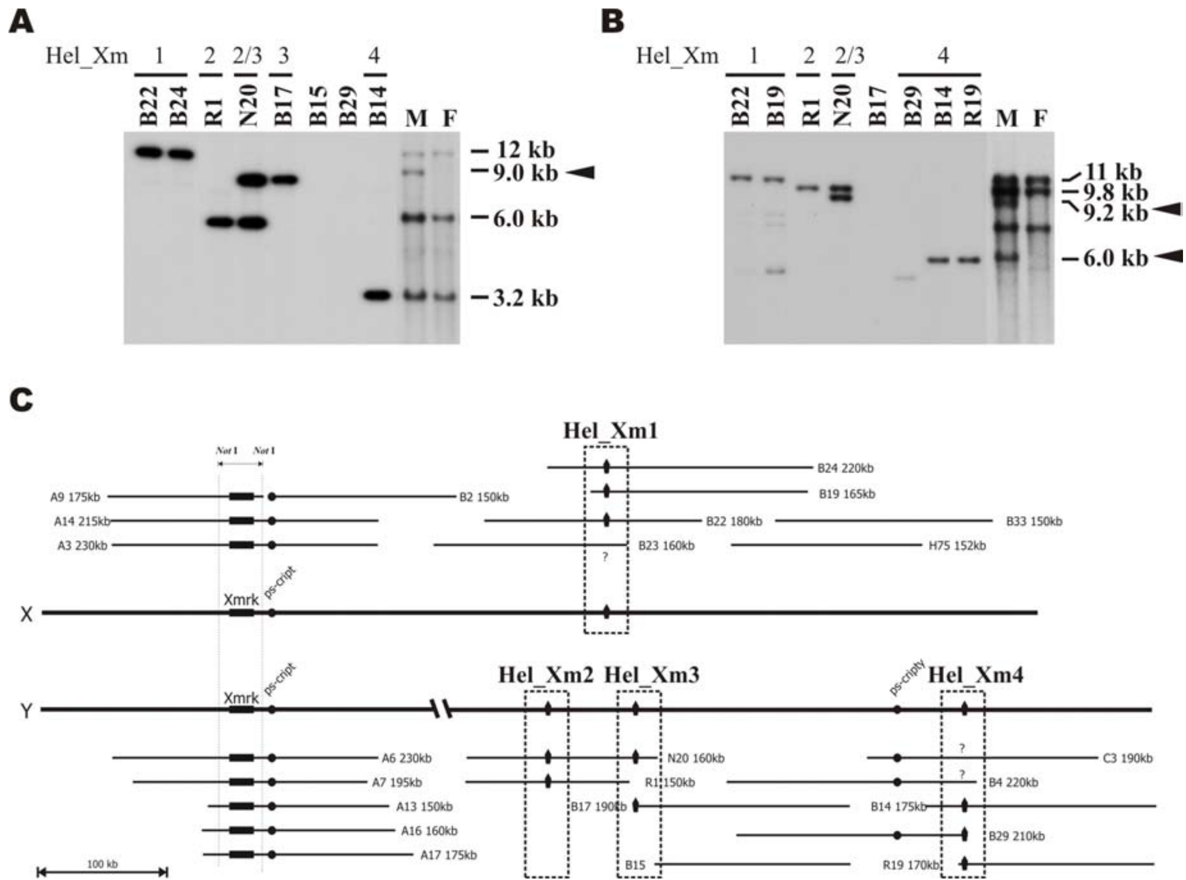


Fig. 44. *Helitrons* in the sex determination region of *X. maculatus*. *X. maculatus* BAC clones and genomic DNA were digested with *Hin*dIII (A) and *Eco*RI (B) and hybridized with the probes generated from region N (A) and helicase domain (B), respectively. Hel_Xm 1, 2, 3, 4 represent the different *helitrons* copies. (C) Partial contigs from the sex determination region of *X. maculatus* containing *helitrons.* The dotted-line frames indicate the locations of the *helitrons* in the sex determination region. *Xmrk*, *ps-cript* and *ps-criptY* are sex chromosome linked markers. The Y-specific bands are indicated by black triangles.

Four groups of *Xiphophorus helitrons* are located in the sex determination region. Interestingly, *Xiphophorus helitrons* have more copies on the Y chromosome than on the X chromosome. Three copies of *helitrons* on the Y contigs and one copy on the X contigs were found so far. In the *ps-criptY* contig, these three copies are located within a region of approximatively 400 kb (Fig. 44). In order to compare these four copies with the consensus nucleotide sequence of *Xiphophorus helitrons*, their region N, Rep domain, helicase domain and EN domain from the corresponding BAC clones (Fig. 44C) were amplified and sequenced. After comparison of sequences of the PCR products of these four copies with the consensus nucleotide sequence of *Xiphophorus helitrons*, no in-frame stop codon, insertion or deletion was found in all the PCR products tested. All of their nucleotide identities were above 95%, and those of the region N PCR products are shown in Tab. 6.

Table 6. The nucleotide identities of the region N between four *Xiphophorus helitrons* located in the sex determination region.

Nucleotide identity (%)	Hel_Xm1	Hel_Xm2	Hel_Xm3	Hel_Xm4
Hel_Xm1	-	-	-	-
Hel_Xm2	96.55	-	-	-
Hel_Xm3	96.91	97.64	-	-
Hel_Xm4	96.18	98.36	96.91	-

4.5.3 Transcripts of *Xiphophorus helitrons*

In order to investigate whether *X. maculatus* contains autonomous *helitrons*, the presence of transcripts was analyzed. The RT-PCR results showed that *helitrons* transcripts exist in every tissue of male *X. maculatus* and some tissues of the female (such as brain, eye and gill, *etc.*), corresponding to the higher copy number on the Y chromosome. *Xiphophorus helitron* transcripts were also detected in stage 21 embryos, in an embryonic cell line A2 and in a melanoma cell line (Fig. 45).

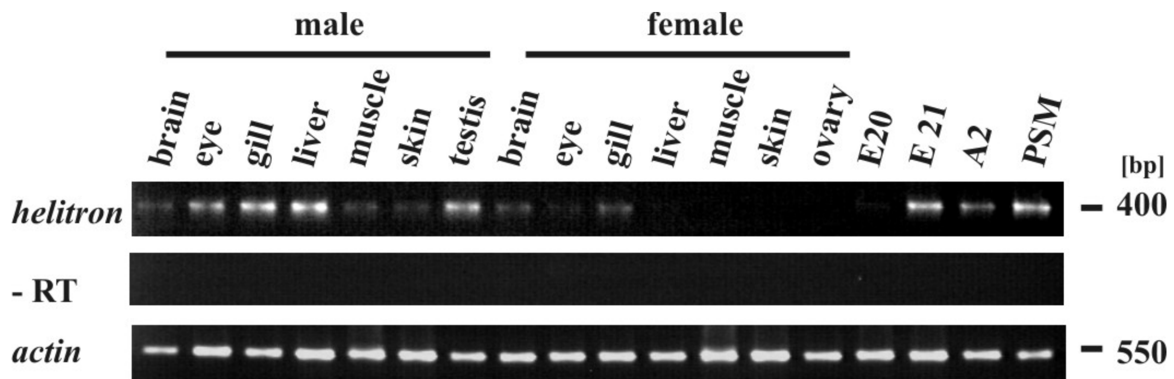


Fig. 45. RT-PCR-based expression analysis of *helitron* in *X. maculatus.* The primer pair amplifies the region encoding the helicase domain. E20 is stage 20 embryos of *X. maculatus* Jp163A. E21 is stage 21 embryos of hybrids (BCn407) of *Xiphophorus maculatus* and *Xiphophorus hellerii*. A2 is the embryonic cell line of *Xiphophorus xiphidium*. PSM is a melanoma cell line derived from a hybrid of *Xiphophorus maculatus* and *Xiphophorus hellerii* (Wakamatsu, 1981).

Over 10 RT-PCR products of region N, Rep domain and helicase domain were cloned from a cDNA library of the PSM melanoma cell line and sequenced. No in-frame stop codon was found in these RT-PCR products, compared with the consensus ORF of *Xiphophorus helitrons.* Their nucleotide identities to each other are all above 97%, and most of them are above 99%. But two deletions, causing frameshift, were found in two of 12 tested transcripts from the helicase domain as well as one deletion was found in one of 13 tested transcripts from the region N. Sequence comparison was also performed between these RT-PCR products and the four copies in the sex determination region, indicating that the Hel_Xm2 shows the highest level of nucleotide identity (> 98%) to the RT-PCR products of *helitrons*.

4.5.4 *Helitron* distribution in the genus *Xiphophorus*, in the family Poeciliidae and in other ray-finned fishes

To investigate *helitron* distribution in fish, the genomic DNA of other strains of *X. maculatus*, backcross offspring Sd/XhIII BCn, some other poeciliids and some distantly related bony fishes were hybridized with the probe generated from region N. *Helitrons* were found in all tested *Xiphophorus* and most poeciliids. Interestingly, different copy numbers of the *helitron* were observed between males and females in all tested *Xiphophorus*, *Poecilia mexicana, Gambusia affinis* and *Poeciliposis gracilis*. Copy number polymorphism was also observed in the different strains of *X. maculatus* and *X. hellerii* as well as between individuals of *X. hellerii* and of the Rio Usumacinta strain of *X. maculatus.* Two gonosome-specific bands (12 kb and 9 kb) were lost in the *X. maculatus*/*X. hellerii* backcross offsprings (Sd/XhIII BCn, over 20 times backcrossing) (Fig. 46).

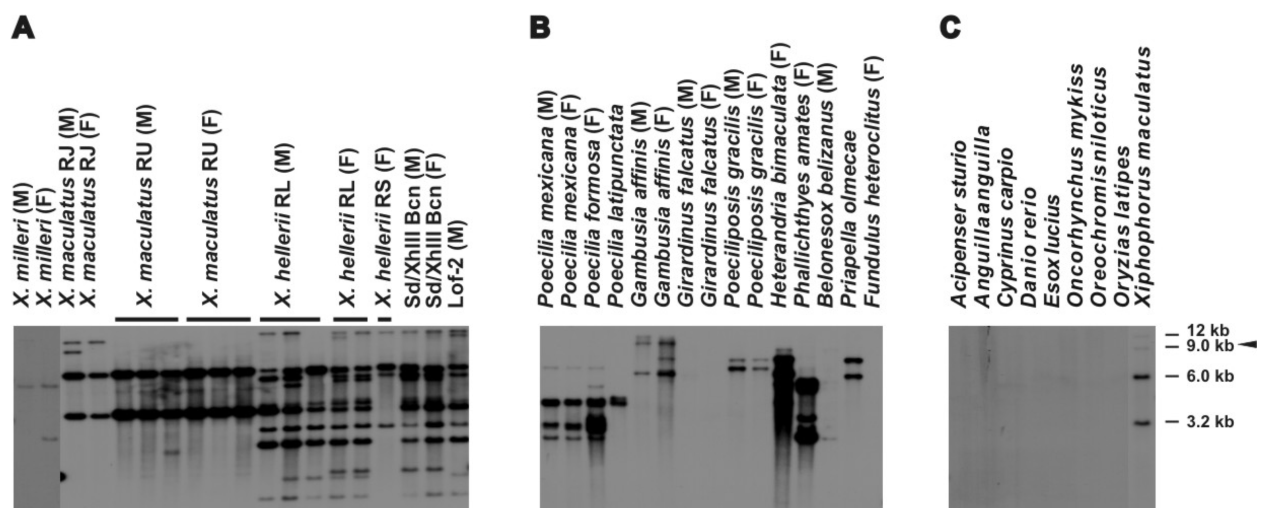


Fig. 46. Distribution of *helitrons* in Xiphophorus (A), poeciliids (B) and other bony fishes (C). The probe was generated from region N of *helitron*. Genomic DNAs were digested with *Hin*dIII. The tested fishes are described in MATERIALS. The Y-specific band is indicated by a black triangle (9.0 kb). *Fundulus heteroclitus* belongs to Fundulidae, a family closely related to the Poeciliidae.

With the same probe, no signal could be observed in the closely related family Fundulidae (such as *Fundulus heteroclitus*) and more divergent fish species (Fig. 46C). The genomic DNA of the same distantly related bony fish species was blotted again with another probe generated from the helicase domain. Again, no signal was observed (data not shown). But it has been reported that *helitrons* exist in the zebrafish genome (Poulter *et al*., 2003). *Helitrons* in the madaka fish will be described in the following section.

4.5.5 The medaka (*Oryzias latipes*) *helitron*

The *Xiphophorus helitron* protein sequence was used as a query to search with significant similarity in the medaka (*Oryzias latipes*) database. Over hundred short ´´shotgun`` sequences (1-2kb) were identified. A reconstructed consensus ORF of medaka *helitrons* was assembled. The consensus sequence obtained is 2996 aa in length and presents 71.9% similarity with Hel_Dr5 and 68.0% similarity with the consensus sequence of *Xiphophorus* helitron at the protein level. The medaka *helitron* has structure hallmarks similar to the fish *helitrons,* encoding in this order Rep, helicase and AP endonuclease within a single ORF (Fig. 42). It contains all conserved motifs of Rep, helicase and EN domains. The N terminal region of the protein also contains repeats of an ~ 380 residue sequence. A putative zinc-finger motif is at approximately 350 residues from the N terminus. The ORF also apparently lack introns, like *helitrons* in zebrafish and platyfish. After the scaffolds of medaka genome were released, many short scaffolds (< 8 kb) matched with the consensus sequence of the medaka *helitron.* No scaffold was found to contain the intact ORF of medaka *helitrons* in this version so far (March 2005). Additionally, more than five long scaffolds (> 10 kb) were found to only match the N terminal region (about 1400 residues) of *helitron.* These data suggest that some *helitrons* might be truncated in medaka genome.

4.5.6 Phylogenetic analysis of *helitrons*

All *helitrons* found in eukaryotes contain the Rep and helicase domains. An alignment was constructed for phylogenetic analysis of the helicase domain of *helitrons* in plants, fungi, invertebrates and vertebrates.

A representative phylogenetic tree is shown in Fig. 47. *Helitrons* in the fungus *Phanerochaete chrysosporium* were used as outgroups. It is clear that fish *helitrons* are monophyletic, including the two consensus *helitrons* found in this work. Strikingly, two of the *Danio helitrons* are more similar to the *Xiphophorus*, *Oryzias* and *Sphoeroides* elements than to the other four *Danio* elements. The remaining *helitrons* form species-specific groups, with *helitrons* of plants, nematode, fly and fungi grouping together, respectively. Unexpectedly, *C. elegans* and *anotheles helitrons* fall to a well-supported group with plant *helitrons* (Fig. 47).

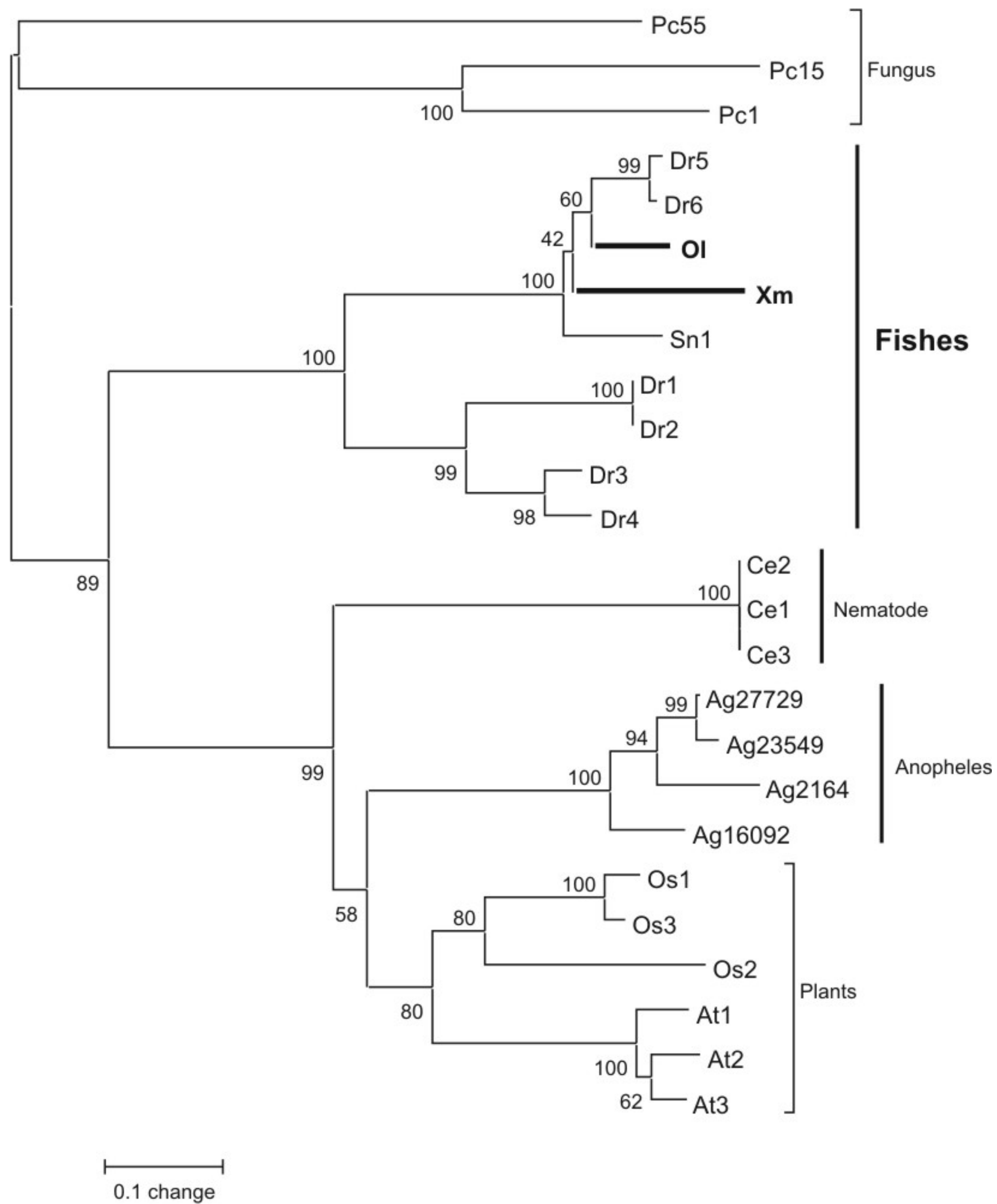


Fig. 47. Phylogenetic tree based on the alignment of the helicase domains of *helitrons*. The tree was constructed by the neighbor-joining method using MEGA 2.1 (Kumar *et al*., 2001) and is a consensus derived from a heuristic search of 1050 bootstrap replicates. The numbers indicate the percentage bootstrap support. Pc is *Phanerochaete chrysosporium* (the white rot fungus), Dr is *Danio rerio* (zebrafish), Xm is *Xiphophorus maculatus* (platyfish), Ol is *Oryzias latipes* (medaka), Sn is *Sphoeroides nephelus* (pufferfish), Ce *is Caenorhabditis elegans*, Ag is *Anopheles gambiae* (anopheles), Os is *Oriza sativa* (rice), At is *Arabidopsis thaliana* (mouse-ear cress). *Helitrons* of Os, At and Ce were reported by Kapitonov and Jurka (2001). *Helitrons* of Dr, Sn, Pc and Ag were described by Poulter (2003). Hel_Xm and Hel_Ol are the consensus ORFs of the *helitron* analyzed in this work.

5. Discussion

5.1 Platyfish *X. maculatus* as a model to study early sex chromosomes differentiation

5.1.1 Why choosing *X. maculatus* to study the mechanism of the sex determination?

No information about sex determination, sex markers and sex chromosomes is available so far from the classic fish models such as the zebrafish, *Danio rerio* and the pufferfish *Takifugu rubripes*. The medaka has been studied from the viewpoint of sex determination and sex chromosome evolution since the early days of experimental genetics (Wittbrodt and Schartl, 2002). Its genome project is close to complete. The *dmrt1bY* is a strong candidate for the master *SD* gene in the medaka. However, it is apparently not the universal master *SD* gene in fish. Hence, alternative fish models need to be analyzed.

In fish, the amazing diversity of sex determination system and their frequent switching during evolution provide hope for identifying different molecular mechanisms in divergent species, particularly for those in which a positional cloning strategy is possible. The platyfish, *Xiphophorus maculatus* is one of them (Volff and Schartl, 2001). *X. maculatus* is a freshwater fish and well known as a model for cancer study. The platyfish is viviparous, meaning that functional *in vivo* analysis by classical embryological methods is almost impossible in this fish. No Xiphophorus genome is being sequenced. Hence, *X. maculatus* is not the first model chosen for the isolation and characterization of a sex-determining gene. But this fish has its own advantages. First, *X. maculatus* has been identified to have three types of sex chromosomes: X, Y and W. Second, its sex chromosomes can be distinguished by some markers (such as *XIR*), although they are homomorphic. Third, a genetic map is available for *Xiphophorus* (Kazianis *et al*., 2004b; Morizot *et al*., 1991; Walter *et al*., 2004). The most important is the availability of loci and markers intimately linked to the major *SD* gene. These loci and markers have been well studied and most of them are involved in the development of normal and malignant pigment cells. A putative genetic map around the *SD* locus has been established by analysis of genetic hybrids and mutants. The *SD* locus is flanked by the X*mrk*

oncogene and its progenitor *egfrb*, which are delimiting a region of about 1 Mb (Gutbrod and Schartl, 1999). These made *X. maculatus* a suitable model to isolate a master *SD* gene in fish.

5.1.2 *X. maculatus* is suitable for study on early sex chromosomes differentiation

It is widely accepted that dimorphic sex chromosomes evolved from cytogenetically indistinguishable pairs of autosomes (Ohno, 1967). However, the exact mechanisms through which the homomorphic pair diverged to produce the gene-rich X chromosome and the small, gene-poor Y chromosome in human are not fully understood. In order to analyze the dimorphic sex chromosome evolution (e.g. mammalian sex chromosome evolution), it is necessary to examine species that appear to be at a very early stage of the process. In mammalian, however, even the most primitive mammals have clearly differentiated sex chromosomes. In fly, some models like the neo-X and neo-Y chromosomes found in certain *Drosophila* species have been used to analyze the evolution of sex chromosomes (Charlesworth, 2002; Steinemann and Steinemann, 2001). But meiosis in *Drosophila* differs from mammals in that no recombination occurs in males. Fish have a great diversity of sex determination mechanisms. Sex chromosomes are found in approximately 10% of examined fish species, and most of them are homomorphic, with the others showing clearly heteromorphy (Sola *et al*., 1990; Traut and Winking, 2001). Several evidences indicate that the sex chromosomes of the platyfish *X. maculatus* are at the early stage of differentiation. YY males are viable and fertile, and recombination can occur all along the length of sex chromosomes (and WW females are produced in the laboratory and are viable as well). In addition, the sex chromosomes can be distinguished by FISH with *XIR* as a probe (Nanda *et al*., 2000a). *Oryzias latipes* and *Oreochromis niloticus* has been used to analyze the early differentiation of sex chromosomes (Griffin *et al*., 2002; Schartl, 2004a; Volff and Schartl, 2002). But the platyfish has three forms of homomorphic sex chromosomes (X, Y and W) and can be utilized to examine not only the early sex chromosome differentiation but also the relationship between the XX/XY and ZZ/ZW systems. Hence, *X. maculatus* is a competitive model for the study on initial differentiation of different types of sex chromosomes.

5.1.3 BAC contigs close to the *SD* locus were set up to isolate the master *SD* gene in *X. maculatus* by positional cloning

The *SD* locus of *X. maculatus* is located between the *Xmrk* oncogene and its progenitor *egfrb*, which delimit a region of about 0.6 cM (centiMorgans). This should allow the positional cloning of the master *SD* gene in *X. maculatus*. After the construction of a BAC library of *X. maculatus*, X- and Y-linked *Xmrk* and *egfrb* contigs were established using *Xmrk* and *egfrb* as starting points, respectively (Froschauer *et al*., 2002). In addition, the Y-specific *ps-criptY* contig was also set up using *ps-criptY* as starting point, which is close to the *SD* locus. This *ps-criptY* contig has been extended over 1000 kb and molecularly analyzed in this work. Duplications, inversions and the accumulation of repeats and transposable elements are found in the *ps-criptY* contig, compared to the corresponding contig on the X chromosome (details see chapter 5.3). These events might correspond to the initiation of differentiation between both types of sex chromosomes.
It is no doubt that the master *SD* gene can be found in *X. maculatus* by positional cloning. Among a handful of putative genes predicted from the analysis of BAC contigs in the sex determination region, *swimy* is exclusively expressed in some spermatogonia and appears to be a good candidate for the *SD* gene, although direct evidence is lacking so far. The identification of the *SD* gene in *X. maculatus* must be beneficial to reveal the mechanism of early sex chromosomes differentiation and gonosomes evolution in fish.

5.2 Where is the *ps-criptY* BAC contig?

The *ps-criptY* BAC contig is a Y-specific contig. *Ps-criptY*, which is the starting point for the *ps-criptY* BAC contig, is a Y-specific *cript* pseudogene and is very close to the *SD* locus of *X. maculatus*. The *ps-criptY* BAC contig has been extended over 1000 kb in this work. Thus, BAC contigs covering a total of 2,500 kb of the X and 3,500 kb of the Y chromosome have been assembled. To the best of my knowledge, these are some of the largest contigs ever assembled for the sex chromosomes of a non-mammalian vertebrate species.
The great majority of markers located on the *ps-criptY* contig have copies on the X-linked *Xmrk* contig, suggesting that the X-linked *Xmrk* contig is the corresponding region of the *ps-criptY* contig on the X chromosome. Hence, the *ps-criptY* contig is putatively located between the *Xmrk* gene and its counterpart *egfrb*. However, this gene pair delimit a region of about 0.6 cM (centiMorgans) covering the *SD* locus (Gutbrod and Schartl, 1999). The question then

arose why does this long *ps-criptY* contig not overlap either with the Y-linked *Xmrk* contig or with the Y-linked *egfrb* contig?

Structural analysis and comparison of the *ps-criptY* contig with X- and Y-linked *Xmrk* contigs may provide answer to this question. This region contains genes and elements duplicated in the *ps-criptY* contig but not in the X-linked *Xmrk* contig. An inversion in this contig was also observed (Fig. 41, arrows). These events probably decreased or suppressed the recombination between the *ps-criptY* region and its corresponding region on the X chromosome, leading to the formation of the Y-specific contig and to an underestimation of the physical distance between the *Xmrk* gene and its counterpart *egfrb* on the *Xiphophorus* Y chromosome. A similar phenomenon is also observed in linkage group 19 of the three-spined stickleback in females and males: the genetic distances between markers adjacent to the sex locus are much shorter in males than in females (Peichel *et al*., 2004). This only reflects the suppression of recombination in the sex determination region and is not the actual genetic distance. An inversion of a huge region containing the Y-linked *Xmrk* contig and *ps-criptY* contig might probably occur on the Y chromosome, but this inversion is not consistent with a putative gene order established by the analysis of sex chromosome crossovers (Gutbrod and Schartl, 1999). In any case, the *ps-criptY* contig is improbably located in the 5′ region of *Xmrk* gene, because this 5′ region is hotspot of X/Y cross-over (Gutbrod and Schartl, 1999). In addition, no recombination was observed so far between the *SD* locus and *Xmrk* or *ps-criptY.* Hence, the *ps-criptY* contig might be located in the 3′ region of *Xmrk* and close to the master *SD* gene. Recently, Southern blot hybridization analysis of a new Y-specific marker (called fahrrad) suggested that the BAC W clones of the *ps-criptY* contig might overlap with the *Xmrk* contig on the Y chromosome (Schultheis, C., pers. commun.).

5.3 Genomic plasticity in the sex determination region and initial differentiation of sex chromosomes in vertebrates

5.3.1 Genomic plasticity in the sex determination region

The genome instability in the sex determination region of *X. maculatus* has been well described (Froschauer *et al*., 2001; Volff *et al*., 2003a). Molecular dissection of the *ps-criptY* contig provided more evidence of this genome fluidity in the sex determination region.

The great majority of markers located on the *ps-criptY* contig have counterparts on the X-linked *Xmrk* contig. However, considerable differences are evident between the X and the Y chromosomes in this region (Fig. 41). Particularly, some sequences of the *ps-criptY* contig are present at higher copy numbers on the Y than on the X chromosome, which include the Y-specific copy of *cript*, and a region containing gene candidates and elements duplicated in the *ps-criptY* region but not on the corresponding X chromosome contig. The presence of genes with more copies on the Y than on the X chromosome is compatible with a model in which the copy number and consequently the level of expression of a master *SD* gene determine the nature of a gonosome (Volff and Schartl, 2001).

The more than 90% nucleotide identity between all *mc4r* copies in the sex determination region indicates that they are the result of rather recent events of gene duplication. In their 5' untranslated region, all platyfish *mc4r* genes contain the last exon of an unrelated gene of unknown function called *dc2* in humans (a complete version of *dc2* was also identified in the platyfish). This sequence is not present in the unique *mc4r* gene of other fish species and vertebrates. Hence, the 5' region of the original *mc4r* gene underwent genomic rearrangements before serial duplications in the sex-determining region. Intrachromosomal duplications of the *ps-criptY* and the Y-linked *Xmrk* contig were also observed by their high nucleotide identity (> 95%), besides *Xmrk* and its counterpart *egfrb*. The above 95% nucleotide identity among the four copies of *helitron* each other indicates also recent transposition or intrachromosomal duplication.

Strikingly, some genes are only found in the *ps-criptY* region but not in the constructed X chromosome contig, such as *swimy* and *polymerase* type B (Fig. 41). An inversion in this contig was also observed (Fig. 41, arrows). Gene deletion has been described in the lof-2 mutant (Schartl *et al*., 1999).

Approximatively 16 gene candidates have been located in the *ps-criptY* contig. However, most of them are truncated, resulting from a relatively high frequency of gene rearrangement. This is similar to the gene truncation observed in the Y-linked *Xmrk* contig (Volff *et al*., 2003b). Numerous truncated genes are just the reflection of the genomic plasticity in the sex determination region.

Finally, partial sequencing of BAC clones B17, B29 and B14 together with the complete sequencing of a 34 kb region including *ps-criptY* revealed a high concentration of the *XIR* repeat. Hence, the *ps-criptY* region might correspond to the Y-specific *XIR*-rich region previously detected by fluorescent *in situ* hybridization (FISH) (Nanda *et al*., 2000b).

Taken together, these results strongly suggest that molecular differentiation between the X and Y chromosomes occurred by duplications, deletions, inversions as well as through the accumulation of repeats and transposable elements. Since such initial differential events should occur in a region around the master sex-determining gene, the *SD* gene might be located in or very close to the *ps-criptY* contig (or alternatively on the corresponding X region in the case of a female-inducing gene).

In sex-determination regions of other fishes such as medaka fish (*Oryzias latipes*) and threespine stickleback (*Gasterosteus aculeatus*), local duplication and accumulation of transposable elements and species- and Y-specific repeats are also present (Nanda *et al*., 2002; Peichel *et al*., 2004). It must be noted that the gonosomes of *X. maculatus*, *O. latipes* and *G. aculeatus* are all homomorphic. In snakes, a female specific repeated DNA sequences was firstly identified as a minor satellite from an Indian snakes (the Banded Krait, hence called *Bkm*), containing of 12-26 tandem repeats of GATA and GACA sequences. It is distributed throughout the snake genome but accumulated on the W chromosome, which is from homomorphic to heteromorphic to Z chromosome in higher snakes (Jones and Singh, 1985). In the more divergent human Y chromosome, approximatively 44% of the euchromatin in the male-specific region consists of ampliconic regions containing Y–specific repeats (Skaletsky *et al*., 2003). Even in plants (*Silene latifolia*, *Marchantia polymorpha*, and *Carica papaya*) and in *Drosophila miranda*, They show patterns similar to the *Xiphophorus* Y chromosome with duplications and accumulation of repetitive DNA (Bachtrog, 2003; Filatov *et al*., 2000; Liu *et al*., 2004; Okada *et al*., 2001; Steinemann and Steinemann, 1998). Apparently, the accumulation of repetitive sequences preferentially occurs on the sex chromosome that will be degenerated or has been degenerated. It is proposed to correspond to the initial event of sex chromosome differentiation (Griffin *et al*., 2002; Nanda *et al*., 2000a).

In sum, sex chromosomes are preferred sites for duplications, deletions, inversions and accumulation of repetitive sequences. Whether the genomic plasticity of sex-determining regions is directly implicated in the variability of sex determination systems in fish is still an open question.

5.3.2 The initial differentiation of sex chromosomes in vertebrates

It is commonly accepted that sex chromosomes are derived from undifferentiated pairs of ancestral autosomes, which contained or acquired a master *SD* gene supposed to be an autosomal gene. The sex chromosomes became differentiated first via suppression of

recombination leading to the eventual functional and structural degeneration of the Y (XY system) or W (ZW system) chromosome. An accepted model of the process of dimorphic sex chromosomes evolution has been established, based on many cytogenetic observations and theoretical consideration (Ayling and Griffin, 2002; Vallender and Lahn, 2004).

The evolution of sex chromosomes in vertebrates has been extensively reviewed (Ayling and Griffin, 2002; Schartl, 2004b; Vallender and Lahn, 2004). Here only the initial differentiation of sex chromosomes in vertebrates is discussed. As the view mentioned above (Chapter 5.1.2), the species with homomorphic sex chromosomes are good models to examine the initial differentiation of sex chromosomes. Most of fish with identified sex chromosomes possess cytogenetically indistinguishable sex chromosomes. Some of them have been extensively studied for sex determination, such as medaka, platyfish, tilapia and threespine stickleback.

The medaka is the sole species of teleost fish for which the major *SD* gene is characterized. In medaka, the master *SD* gene *dmrt1bY* (or *DMY*) shares phylogenetic affinities to the autosomal *DMRT1* and has been formed by recent duplication of this gene (Matsuda *et al*., 2002; Nanda *et al*., 2002). This occurred approximatively 10 million years ago (Kondo *et al*., 2004). The initial medaka Y chromosome has been formed through the insertion of a segment duplicated from an autosome (Schartl, 2004a; Volff and Schartl, 2002). This process created a chromosomal region of 260 kb on the Y that is absent on the X. Thus pairing and crossing-over events are inhibited here leading to the isolation of the *SD* gene. This indicates that gene duplication and then subsequently transposition of duplicated region can modify the control of sex determination.

A standard model of sex chromosome evolution postulates that Y chromosome evolution starts either from a single male specific locus, which makes it different from the otherwise homologous X, or from mutations at two loci, one controlling male and the other female formation, thereby establishing a proto-X/proto-Y chromosome system (Charlesworth, 1996). The medaka apparently follows the former model of sex chromosome evolution. The platyfish might follow the latter one. The first exon of X-linked copies of *fredi*, which is regarded as candidate for the *SD* gene, has been disrupted by a miniature inverted-repeat transposable element (MITE) (Schultheis, C., pers. commun.). Such an insertion did not happen in the Y-linked copies. This means that an insertion into the X chromosome copies makes it different from the other homologous on the Y chromosome. Subsequent suppression of recombination around this insertion keeps this difference between the gonosomes. Additionally, one part of

the sex determination region of the platyfish might have been formed by trans-chromosomal duplication, which was inferred by syntenic conservation (detail see chapter 5.6). Finally, the molecular differentiation of the X and Y chromosomes of *X. maculatus* occurred by duplications, deletions, inversions as well as through the accumulation of repeats and transposable elements. Such events might correspond to an initiation of differentiation between both types of gonosomes.

In the tilapia (*Oreochromis niloticus*), the initial differentiation of gonosomes has been proposed to correspond to the stochastic accumulation of heterochromatin (Griffin *et al*., 2002). This is consistent with the accumulation of repeats and transposable elements in the sex determination region of many species.

In the homomorphic sex chromosomes (ZW system) of primitive birds, ostrich and emu, a deletion on the W chromosome has been identified (Ogawa *et al*., 1998; Shetty *et al*., 2002). The accumulation of a repeat sequence (called *Bkm*) has also been observed in homomorphic sex chromosomes (Z and W chromosomes) of some boid snake (Jones and Singh, 1985). These data suggest that the initial differentiation of gonosomes between the XY system and the ZW system might be similar.

The large variety of sex determination systems that exist even in closely related species of lower vertebrates could suggest that there are permanently new top *SD* gene emerging (Schartl, 2004b). Thus it is necessary to study different models of sex chromosomes evolution. On the other hands, such swapping of top *SD* genes, that sometimes seems to happen stochastically, may be a good illustration of the surprising evolutionary dynamism in sex determination system and sex chromosomes of the lower vertebrates.

5.4 *Swimy* and *fredi* are candidates for sex determination or sex differentiation genes in *X. maculatus*

5.4.1 *Swimy* is exclusively expressed in some spermatogonia of *X. maculatus*

Swimy is the first gene in *X. maculatus* that is exclusively expressed in some spermatogonia of adult testis. It has at least 7 copies per haploid genome. Apart from its Y–specific band located in the *ps-criptY* contig, the other copies are not found in X/Y-linked *Xmrk* contigs.

Moreover, no stop codon or frameshift is found in the coding sequence of the *ps-criptY-swimy* copy, which has a very high nucleotide identity with the cDNA sequence. Putative promoter and polyA-tail could be predicted by NIX analysis at 408nt away from the upstream and about 2500 nt away downstream of the ORF of the genomic sequence, respectively. These data suggest that *swimy* of the *ps-criptY* contig was not truncated and is very likely functional.
Additionally, *swimy* displays two W-specific bands in the Usumacinta strain of *X. maculatus*. It is a new W-specific molecular marker found on the Y chromosome in *X. maculatus*, along with *polymerase type B* (see Chapter 5.6).
Swimy encodes a DNA binding protein containing four Zn-finger domains and one SAP DNA binding domain. No known protein has been identified so far to contain the same components and display the same organization as *swimy*. All Swimy-like sequences in fish contain the SAP DNA binding domain, which is found in diverse nuclear proteins involved in chromosomal organization. The C terminal region (~ 210 aa) of Swimy does not contain any known motif or domain and is putatively Leu-rich. A short leucine rich hydrophobic motif (LXXLL) was reported that could interact with ligand binding domain of the androgen receptor (AR), and this interaction is necessary for AR regulation of transcription, which is mediated by steroid receptor coactivators (SRCs) (Estebanez-Perpina *et al*., 2004). Thus, Swimy might be a novel hormone receptor coactivator, and forms a new protein family along with fish homologies.
Swimy is exclusively expressed in *X. maculatus* spermatogonia, which were derived from primordial germ cells (PGCs). In the zebrafish, PGCs are critical for the survival of gonads and the development of females (Slanchev *et al*., 2005). This is not consistent with results in mice where gonads are present but in a smaller size, and mice do not invariably develop into males in the absence of PGCs (Beck *et al*., 1998). Hence, it is necessary to examine whether PGCs have the similar role in other fish species.
Several gonad- and germ cell-specific genes have been identified in the zebrafish (Lv *et al*., 2005) and in the medaka (Shinomiya *et al*., 2000). However, none of them has similarity with *swimy*, and *swimy* does not match any EST sequences available in the database of NCBI. The role of *swimy* in PGCs or germ cells is under scrutiny.
Although some similar sequences of *swimy* were found in the medaka, zebrafish and fugu, *swimy* was not found outside of the fish lineage. It suggests that this gene evolves extremely rapidly, as observed for upstream sex-determining genes in several organisms. Particularly, it

is exclusively expressed in spermatogonia. Thus, *swimy* is a very promising candidate for a sex determination or sex differentiation gene in *X. maculatus*.

5.4.2 *Fredi* in *X. maculatus*

Fredi is expressed predominantly in testis of the adult fish and encodes a novel helix-turn-helix protein, suggesting that it works as a DNA binding protein. Its expression in the tail of embryonic fish (Froschauer, A., unpublished results) suggests that *fredi* might be involved in the development of the male swordtail (*X. hellerii*). The swordtail is the distinctive characteristic of male fish in the related species *X. hellerii*.

Fredi has four copies in the Y-linked contigs and two copies in the X-linked contigs. Interestingly, the two copies of *fredi* in the X-linked contigs are interrupted by a DNA transposable element, MITE (Schultheis, C., unpublished data), leading to a new hypothesis for the initiation of sex chromosomes (detail see Chapter 5.3.2). In addition, *fredi* was not found outside of the fish lineage and does not match any EST sequences in the public database. Hence, *fredi* is also a putative candidate for a sex determination or sex differentiation gene in *X. maculatus*.

5.5 Gene candidates for pigmentation in the sex determination region of *X. maculatus*

Approximatively 30 gene candidates have been identified in the sex determination region of *X. maculatus*. Most of them are pseudogenes, including *cript*, *msh2*, *acr*, *igd*, *fbx11*, *dc2*, *etc.* However, some of them are putative candidate genes for loci, involved in pigmentation and located in the sex determination region.

A gene candidate putatively encoding a GTP cyclohydrolase I was identified in the *ps-criptY* contig (Tab. 4). This gene is also linked to *cript* in the genome of fugu. During the embryonic development of zebrafish, neural crest precursor cells segregate into neural elements, ectomesenchymal cells and pigment cells. The latter then diversify into melanophores, iridophores and xanthophores. The differentiation of neural cells, melanophores, and xanthophores is closely coupled with the onset of pteridine synthesis. GTP cyclohydrolase I catalyses the conversion of GTP to dihydroneopterin triphosphate and initiates the pteridine pathway. GTP cyclohydrolase I is expressed in both the melanophore and xanthophore lineages during early development in zebrafish (Ziegler, 2003). Hence, the GTP

cyclohydrolase I gene of the platyfish is an excellent candidate for the red-yellow (*RY*) pigmentation locus, or alternatively, for the macromelanophore-determining locus *Mdl* (macromelanophores are large melanocytes being precursors of melanoma cells in *Xiphophorus*).

As many as 11 *mc4r* genes were identified in BAC contigs from both the X and Y chromosomes. Three of them are from the *ps-criptY* contig (such as *mc4rP*, *mc4rQ* and *mc4rR*). Melanocortin receptors can bind the alpha melanocyte-stimulating hormone (alpha-MSH, beta-MSH, and gamma-MSH) and adrenocorticotropic hormone (ACTH), and are involved in mammals in the control of feeding and body weight (*mc4r*) and the differentiation of melanocytes (*mc1r*). In fish, the onset of sexual maturity determined the size of the body (Kallman, 1989). Consequently, *mc4r* genes are candidates for the puberty locus *P*.

In addition, *mc4r* genes were found to be expressed in skin in goldfish (Cerda-Reverter *et al*., 2003) and platyfish (Selz, Y., pers. commun.). Particularly, Northern blot analysis showed that *mc4r* mRNA is strongly overrepresented in melanoma compared to healthy tissues. This suggests that the expression pattern of one of several *mc4r* copies has been modified in the platyfish. A higher expression was found in testis compared to ovary. Interestingly, the *mc4r* copy preferentially expressed in testis was different from those expressed in other organs and tissues. *Mc4r* copies with alternative 3' untranslated region (such as *mc4rG* in the X-linked *Xmrk* contig) are preferentially expressed in melanoma tissues and cell lines, suggesting that they might correspond to *Mdl* (Selz, 2003).

5.6 Conserved syntenies between the sex determination region of *X. maculatus*, human and other fishes

A synteny between the sex-determining region of the platyfish and human chromosome 2 has been detected (2p21-22: *msh2*, *cript* and *fbx11*; 2q32-37: *tmeff*, *acr* and *bdg*). Very interestingly, XY sex reversal (feminization) has been observed in patients with interstitial deletion in 2q31 (Slavotinek *et al*., 1999) and deletion of 2q32-qter (Fryssira, 1988). This indicates the presence of a so far unknown male-determining gene in the region of the human chromosome 2 syntenic to the sex-determining region of the platyfish. Hence, identification of *SD* in *X. maculatus* might lead to the discovery of a novel sex-determining gene in humans.

The conserved linkage between *msh2*, *gch1* and *cript* was also observed in a single scaffold of Tetraodon (*Tetraodon nigroviridis*), fugu (*Takifugu rubripes*) and zebrafish (*Danio rerio*).

Thus the linkage of *msh2*, *gch1* and *cript* in the sex determination region of *X. maculatus* is speculated to come from a transchromosomal duplication of autosome, then the *msh2* and *cript* genes were truncated by the genomic plasticity affecting the sex determination region.
Other syntenies between the sex determination region of the *X. maculatus* and single scaffolds of pufferfish were detected, such as a synteny of *mc4r*, *trail* and *igd1* in Tetraodon (this work) and a synteny of *EZH2*, *Cullin-1*, *PHC*, *Claudin* and *TNIK* in scaffold654 of fugu (Groh, J./ Froschauer, A., unpublished results). It is no doubt that molecular analysis of these syntenies will be beneficial to reveal the evolution of sex chromosomes in *X. maculatus*.

5.7 *DNA polymerase type B* in fish

5.7.1 *DNA polymerase type B* domain in fish

DNA carries the biological information that instructs cells how to exist in an ordered fashion: accurate replication is thus one of the most important events in the cell life cycle. This function is performed by DNA-directed DNA-polymerases, which add nucleotide triphosphate (dNTP) residues to the 5′-end of the growing DNA chain, using a complementary DNA as template. Small RNA molecules are generally used as the primer for the chain elongation, although the terminal proteins may also be used. DNA-directed DNA polymerases can be classified, on the basis of sequence similarity, into at least four different groups: A, B, C, X (Delarue *et al*., 1990; Jung *et al*., 1987).
DNA polymerase type B, organellar and viral (containing DNA_pol_B_2 domain) belongs to DNA-directed DNA polymerase family B and has been found in plant and fungal mitochondria, as well as in virus. Some putative proteins, similar to DNA polymerase type B, are also deposited in *C. elegans, D. rerio* and *F. rubripes* genome databases. In this work, *DNA polymerase type B* was identified in the sex determination region of *X. maculatus*. Surprisingly, in medaka (*Oryzias latipes*), which is more closely related to platyfish than zebrafish, no ´´shotgun`` sequences or scaffolds were found to match the *DNA polymerase type B* so far in the latest version of medaka genome data (up to Jan. 2005).
Strikingly, all putative DNA polymerases type B described in fishes and *C. elegans* have a 130-132 aa-adding sequence in the middle of DNA_pol_B_2 domain. This addition is not a transposable element or a repeat. Because all DNA polymerases type B in fishes and *C. elegans* share a common ancestor, this addition might have occurred at least before

divergence between *C. elegans* and vertebrates. The functional role of this addition to the DNA_pol_B_2 domain remains to be determined. Alternatively, there might be an in-frame deletion of the 130-132 aa sequence in the organellar and viral gene lineage. All of those putative DNA polymerases type B in fishes and *C. elegans* contain DNA_pol_B_2 domain and DTDS motif, suggesting that they are involved in DNA replication like the other members of the DNA-directed DNA polymerase family B.

5.7.2 *DNA polymerase type B* is also a novel W-specific marker in *X. maculatus* Usumacinta, along with *swimy*

DNA polymerase type B is located in the *ps-criptY* contig on a non-Y-specific band, suggesting a relationship between this Y specific contig and the corresponding region on the X chromosome. However, this gene displays a Y specific band in *X. maculatus* genome. Hence, *DNA polymerase type B* must have another copy on the Y chromosome outside the current contigs of *X. maculatus*.

Interestingly, *DNA polymerase type B* also displays a W-specific band in the *X. maculatus* Usumacinta strain, like *swimy*. Both of them are novel W-specific markers of *X. maculatus* also found in the Y chromosome. These markers hint relationships between the sex determination region of the Y and W chromosomes in WY/YY sex determination system (female heterogamety). Hence, further molecular analysis of *DNA polymerase type B* and *swimy* might shed light on the relationship between the XY and ZW systems. Nevertheless, more male and female individuals have to be analyzed to confirm this linkage.

5.8 DNA transposon *helitrons* in *X. maculatus*

The distribution of *helitrons* is widespread. Although not found in human and mouse genome, *helitrons* have been found so far in plants, arthropods, basidiomycete fungi, vertebrates and some invertebrate deuterostomes by computational analysis (Tab. 7). The experimental result that *helitron* was recently inserted into intron 11 of *sh2* gene in the maize mutant *sh2-7527* indicated that *helitron* is a transposable element (Lal *et al*., 2003). In this work, at least four *helitron* copies have been found in the sex determination region of *X. maculatus* by strategic sequencing, and a *helitron* sequence has been reconstructed in the medaka fish.

All known *helitrons* share a common ancestor and the element phylogeny is often similar to the host species phylogeny (Poulter *et al*., 2003). After adding the *Xiphophorus helitron* and medaka *helitron,* the phylogenic analysis results show that they fall into a well supported group with other fish *helitrons* (Fig. 47). Four zebrafish *helitrons* copies form a subgroup within the fish *helitrons* group, the remaining two zebrafish *helitrons* (Hel_Dr5 and Hel_Dr6) fall into another subgroup together with *helitrons* of the other three fishes. These suggest that the duplication and subsequent divergence of *helitron* might have occurred before divergence between the zebrafish and these three analyzed fishes (such as the platyfish, pufferfish and medaka).

Table 7. Distribution of *helitrons* in different species (up to January 2005)

Species		References
Fungi	*Phanerochaete chrysosporium* (the white rot fungus)	Poulter *et al*. (2003)
Plants	*Arabidopsis thaliana* (mouse-ear cress)	Kapitonov and Jurka (2001)
	Oriza sativa (rice)	Kapitonov and Jurka (2001)
	Zea mays (maize)	Lal *et al.* (2003)
Nematode	*Caenorhabditis elegans*	Kapitonov and Jurka (2001)
Insects	*Anopheles gambiae* (mosquito)	Poulter *et al*. (2003)
	Drosophila melanogaster (fruit fly)	Kapitonov and Jurka (2003)
Vertebrates	*Danio rerio* (zebrafish)	Poulter *et al*. (2003)
	Sphoeroides nephelus (pufferfish)	Poulter *et al*. (2003)
	Xiphophorus maculatus (platyfish)	This work
	Oryzias latipes (medaka)	This work

No in-frame stop codon, insertion or deletion has been found so far in the four copies of *Xiphophorus helitrons* in the sex determination region. Their nucleotide identities to each other are very high (> 95%). These suggest that these four elements are less divergent and very young in the genome. Less than 20% of *Xiphophorus helitrons* transcripts contain deletions or insertions, suggesting that many of them might be recently active or still active. This is in agreement with the results of computational analysis for zebrafish *helitrons* (Poulter *et al*., 2003).

Zebrafish was the first fish reported to contain *helitrons*. But there is no signal in the zebrafish genome, when hybridized with the *X. maculatus helitron*. Medaka fish has the same phenomenon. Hence, the conclusion could not be made that the other bony fishes do not

contain *helitrons,* although they have no signal on the Southern blots with the *Xiphophorus helitron* as a probe. This is in contrast to the retrotransposon *Rex3* described in bony fishes (Volff *et al*., 1999). This indicates that *helitron* diverged more rapidly than some retrotransposons in the fish lineage.

The backcross hybrids (Sd/XhIII BCn) are the offsprings that the female hybrids (from the crossing between the female *X. maculatus* and male *X. hellerii*) backcrossed with the male *X. hellerii* over 20 times (Schartl *et al*., 1999). So it is easy to understand that the Y-specific band (9kb) of *Xiphophorus helitrons* was lost after one backcrossing. I presume that the loss of another sex chromosome specific band (12kb) result either from the deletion or from the recombination between the X chromosomes of the female in backcross hybrids.

Helitrons have a number of features that suggest that they might be of considerable evolutionary importance. As one suggestion given by Kapitonov and Jurka (2001), *helitrons* might function as ´´powerful tools of evolution`` by virtue of their apparent ability to recruit the host genes and multiply and modify them in the host genome. It will be interesting to determine how significant this type of genome modification is in eukaryotes. Another suggestion from Kapitonov and Jurka (2001) is that *helitrons* probably represent the missing evolutionary link between prokaryotic RC elements and plant geminiviruses. Geminiviruses might have evolved from plant *helitrons* transposons. In a word, *helitrons* are DNA transposable elements that keep the RCR mechanism early only found in prokaryotes and can now be added to a growing list of entities known to reside in eukaryotic genome.

5.9 Perspectives

In order to isolate the master *SD* gene in *X. maculatus* by positional cloning, an over 1000 kb Y-specific *ps-criptY* contig has been constructed and partially sequenced, together with about 2Mb large contigs of both X and Y chromosomes established and analyzed in our group. The molecular analysis of these contigs shows that the Y chromosome has undergone gene rearrangement before putative suppression of recombination between the gonosomes. Several markers of the *ps-criptY* contig have copies on the X-linked *Xmrk* contig, suggesting a relationship between the *ps-criptY* contig and the X-linked *Xmrk* contig. However, major differences between the X and the Y chromosomes were observed in this region, such as a region duplicated on the Y but not on the X chromosome. Additionally, the Y-specific accumulation of repeats was also observed in the *ps-criptY* contig. Some putative candidates for the *SD* gene were isolated and analyzed, such as *swimy* and *fredi*. Both of the *ps-criptY*

and Y-linked *Xmrk* are quite close to the *SD* locus. But these contigs are not connected to each other so far. Hence, it is necessary to reach overlap by continuous extension of these two contigs. The finding of the master *SD* gene might be beneficial the study on speciation, since the master *SD* gene may correspond to a speciation gene (Volff, 2005).

The examination of the expressional time point of *swimy*, the further functional analysis of *swimy* and study of similar sequences in the medaka and zebrafish should provide more information about *swimy* as a candidate for the master *SD* gene in *X. maculatus*. Complete sequencing and comparison of the corresponding sex determination region on the X and Y chromosomes will provide new information on the evolution of sex chromosomes.

It is no doubt that the master *SD* gene in *X. maculatus* can be identified by positional cloning. The further analysis of the contigs mentioned above will shed new light on the molecular mechanism of sex determination and the evolution of sex chromosomes in fish.

6. Appendix

6.1 Tables

Table A1. BAC clones that were identified or tested by Southern blot hybridization. The clones with bold font were picked up by Dr. A. Froschauer, the others have been picked up in this work.

Code	**BAC number**	**Probes for identification**	**Hybridization**	**size [kb]**	**Position**
B4	068 C04	*Xmcr/ Xcript/ Dc2/helitron*	positive	220	*ps-cript*Y contig
B13	065 C10	*Xmcr/ Dc2*	positive	145	*ps-cript*Y contig
B14	041 B06	*Xmcr/ Dc2/helitron*	positive	170	*ps-cript*Y contig
B15	088 J08	*Xmcr/ Tnh7/ Dc2/Tnh2*	positive		*ps-cript*Y contig
B17	082 N08	*Xmcr/ Tnh7/ Dc2/helitron/fredi*	positive	190	*ps-cript*Y contig
B21	104 G16	*Xmcr/ Xcript/ Dc2*	positive	180	*ps-cript*Y contig
B26	130 D13	*Xmcr/ Dc2/bdg/O2-FOR*	positive	165	*ps-cript*Y contig
B29	085 A12	*Xmcr/ Xcript/ Tnh7/ helitron/fredi*	positive	220	*ps-cript*Y contig
C3	103 D02	*Xcript/Helitron*/B14-FOR	positive	190	*ps-cript*Y contig
F1	031 A22	*RNase3l/pol/igd1/T-Rex*	positive	160	*ps-cript*Y contig
F11	123 N02	*RNase3l/swimy/pol/igd1/T-Rex/acr/bdg*	positive	140	*ps-cript*Y contig
F2	065 F20	*RNase3l/* B6-REV	positive	25	*ps-cript*Y contig
F6	020 H13	*RNase3l/* F1-FOR/W9-FOR/W1-FOR	positive	145	*ps-cript*Y contig
H93	115 P05	*RNase3l/igd1/T-Rex/acr/bdg*	positive	160	*ps-cript*Y contig
N1	118 F21	B17-FOR	positive		
N2	112 F18	B17-FOR	positive		
N3	094 F06	B17-FOR	positive		
N4	116 I14	*B17-FOR*	positive		
N5	104 K08	B17-FOR	positive		
N6	092 M05	B17-FOR	positive		
N7	128 H07	B17-FOR	positive		
N8	103 P18	B17-FOR	positive		
N9	076 A06	B17-FOR	positive		
N10	130 B03	B17-FOR	positive		
N11	109 P04	B17-FOR	positive		
N12	120 M17	B17-FOR	positive		
N13	085 H01	B17-FOR	positive		
N14	120 G24	B17-FOR	positive		
N15	078 B03	B17-FOR	positive		
N16	105 O21	B17-FOR	positive		
N17	093 O07	B17-FOR	positive		
N18	117 O05	B17-FOR	positive		
N19	111 M06	B17-FOR	positive		
N20	123 K08	B17-FOR/*helitron*	positive	160	*ps-cript*Y contig
N21	099 N18	B17-FOR	positive		
N22	101 L09	B17-FOR	positive		
N23	111 D24	B17-FOR	positive		
N24	083 B04	B17-FOR	positive		
N25	125 B13	B17-FOR	positive		
N26	083 C22	B17-FOR	positive		
N27	099 G13	B17-FOR	negative		
N28	058 P18	B17-FOR	positive		
N29	070 N13	B17-FOR	positive		

Code	**BAC number**	**Probes for identification**	**Hybridization**	**size [kb]**	**Position**
N30	022 K12	B17-FOR	positive		
N31	058 E10	B17-FOR	positive		
N32	002 E12	B17-FOR	positive		
N33	072 P24	B17-FOR	positive		
N34	055 N02	B17-FOR	positive		
N35	018 L18	B17-FOR	positive		
N36	061 M23	B17-FOR	positive		
N37	001 J14	B17-FOR	positive		
N38	060 C17	B17-FOR	positive		
N39	003 P09	B17-FOR	positive		
N40	009 P05	B17-FOR	positive		
N41	017 H14	B17-FOR	positive		
N42	057 F14	B17-FOR	negative		
N43	027 E04	B17-FOR	positive		
N44	057 C13	B17-FOR	positive		
N45	057 C23	B17-FOR	positive		
O1	063 C22	B26-FOR/*acr*/F11-FOR	negative		
O2	009 E03	B26-FOR/*acr*/F11-FOR	positive	155	*ps-cript*Y contig
O3	101 N19	B26-FOR/B33-REV/F11-FOR	positive	165	*ps-cript*Y contig
O4	083 M05	B26-FOR/*acr*/F11-FOR	positive	190	*ps-cript*Y contig
R1	068 J15	N20-REV/*helitron*	positive	150	*ps-cript*Y contig
R2	066 F09	N20-REV	positive		
R3	057 D19	N20-REV	positive		
R4	055 A11	N20-REV	positive		
R5	037 E09	N20-REV	positive		
R6	022 E13	N20-REV	negative		
R7	013 E15	N20-REV	negative		
R8	044 B06	N20-REV	negative		
R9	032 H03	N20-REV	negative		
R10	005 K03	N20-REV	negative		
R11	051 J21	N20-REV	negative		
R12	is N40	N20-REV	negative		
R13	012 C08	N20-REV	negative		
R14	088 E01	N20-REV	negative		
R15	110 O09	N20-REV	positive		
R16	084 H24	N20-REV	negative		
R17	114 H13	N20-REV	negative		
R18	121 B04	N20-REV	positive		
R19	099 A19	N20-REV/*helitron*/*fredi*	positive	170	*ps-cript*Y contig
R20	081 O10	N20-REV	negative		
R21	086 I03	N20-REV	negative		
R22	086 D05	N20-REV	negative		
R23	078 C04	N20-REV	negative		
R24	108 B11	N20-REV	negative		
S1	034 G20	F11-REV	positive		
S2	055 P04	F11-REV	positive		
S3	054 P05	F11-REV	positive		
S4	011 M05	F11-REV	negative		
S5	030 E02	F11-REV	positive		
S6	122 E15	F11-REV	positive		
S7	084 L10	F11-REV	positive		
S8	is F8	F11-REV	positive		

Code	**BAC number**	**Probes for identification**	**Hybridization**	**size [kb]**	**Position**
S9	119 C19	F11-REV	positive		
T1	068 I20	F1-FOR	positive		
T2	037 J02	F1-FOR	negative		
T3	is P2	F1-FOR	negative		
T4	008 B12	F1-FOR	positive		
T5	003 O07	F1-FOR	negative		
T6	021 L08	F1-FOR	negative		
T7	045 N01	F1-FOR	negative		
T8	045 I04	F1-FOR	negative		
T9	084 P23	F1-FOR	negative		
T10	108 F20	F1-FOR	negative		
T11	is P6	F1-FOR	positive		
T12	102 K02	F1-FOR	positive		
W1	050 O14	F6-FOR/W9-REV/W1-REV/W1-FOR	positive		
W2	052 C15	F6-FOR/W9-REV/W1-REV	positive		
W3	070 D09	F6-FOR/W9-REV/W1-REV/W7-REV	positive		
W4	056 E03	F6-FOR/W9-REV/W1-REV	positive		*ps-cript*Y contig
W5	055 L05	F6-FOR/W9-REV/W1-REV	positive		
W6	018 M20	F6-FOR	negative		
W7	024 M18	F6-FOR/W9-REV/W1-REV	positive		*ps-cript*Y contig
W8	072 M13	F6-FOR/W9-REV/W1-FOR/W7-REV	positive		*ps-cript*Y contig
W9	005 C12	F6-FOR/W1-FOR/W1-REV/W7-REV	positive		*ps-cript*Y contig
W10	023 B08	F6-FOR	negative		
W11	097 M23	F6-FOR/W9-REV/W1-REV/W1-FOR	positive		*ps-cript*Y contig
W12	091 J20	F6-FOR/W9-REV/W1-REV/W7-REV	positive		*ps-cript*Y contig
W13	097 B11	F6-FOR/W9-REV/W1-REV/W1-FOR	positive		*ps-cript*Y contig
W14	090 F10	F6-FOR/W9-REV/W1-REV/W1-FOR	positive		*ps-cript*Y contig

Table A2. Oligonucleotide primers for sequencing and PCR.

Name	Sequences of 5´-3´	Name	Sequences of 5´-3´
Act1	GTAGGTGATGAAGCCCAGAGC	N20t163-R	TCCATTTCTTTTTCTGTTCC
Act2	AGGGAGCTCGTAGCTCTTCTC	N20t169-R	AGATCTTCAACCAGCTGCAC
CEQ-FOR (M13)	GTAAAACGACGGCCAGT	N20t006rev-F	TAGGCGTGTGAGACAGAAGG
CEQ-REV (M13)	CACACAGGAAACAGCTATGAC	N20t103rev-F	CATTACACCCTCATGATTAC
B29-H61-F1	TTAGGTAGACCAGAGTTTGC	Tnh7-F1	ATGATCGAGAGGGTGTTGG
B29-H61-R1	AAAGGATGTTAGCTGCCATC	Tnh7-R1	TTGATAAGTAGCTGGGTCTG
B29-H61-F2	GCAGTTAGCAAGTGCTAACG	B17t18-F1	ATAAGTGGCAGTACTTGTGC
B29-H61-R2	CACAGTAATCTACAATATGG	B17t18-R1	GAAGCACTAAGTGCATAGGC
B29-H61-F3	GCATAGCTCTGCTATATAGC	HTH-F1	TCTAAGAGTTGCAGGACACC
B29-H61-F4	TGTTTACTGACACAACCATG	HTH-R1	CTGACAATTCTTTTGACATG
B29-H61-R4	GCAACACCATTAAGGCAAAG	Fredi-F1	ATGTCTTGTCAGATAGATGTTC
B29-H61-91F	AATTACCAATGTCTCCTTGG	Fredi-R1	AACCAAACTCCTCCACTGAACC
B29-H61-91R	GTATTGGTCTCTGGTAACTC	MSH2-F1	AAGACGGTGAAGCTGGAGCC
B29-H61-91R2	GTTTCATAGACTCAAGTGTGC	MSH2-R1	CAGCTCAGCGACGTGGATCC
B29-H61-188F1	GAGAAGGTGACTTTAACTGC	MSH2-F2	GSAACAACAAGAARTTCAC
B29-H61-28R1	ACTGCACCTGCTTCATGTCG	MSH2-R2	AAARCTCTGGTCRCACACAC
M02-B104-F1	CTTTCCAAAGTCCTGGAAGC	mcrR1	TCTTGTTCATCGATATGGTG
M02-B104-R1	GGGTGCGTTGTGTTTGGTGC	mcrR2	GGAACTGATCTTCTATTCTG
M02-B104-R1a	GGATGTAATGGTTGATAAGAC	mcrP1	GTTCATCGATATGGTGTCAG
M02-B104-F2	GGATAAGTGGTTCACTGATC	F11t252-F1	ATGGAACCACTTTCCTCCTG
M02-B104-R2	ATTTACAATCCTCAACTTCG	F11t252-R1	GACTTTTCACATGGTGTCCAG
M02-B104-R2a	ATTTACAATCCTCACTTCGG	F11t109-F1	ATAAAAGGGCCTCCTTCACC
M02-B104-R2b	TAACTTTAATGCAGTGCAGG	F11t109-R1	GATTAGCGTGGTCGTGTGTG
M02-B104-14F	TCAGATCTCAAGAGTCTCAG	F11t134-F	GGAATTGCCGAAACAAGAC
Rep-F1	GGCTCTCCTCATTGCCATTG	Swimy-F1	CTGAGGAATGCATTGCAGGT
Rep-R1	ATGCATCAACACAATACTGG	Swimy-R1	AAGACCTACCTGGCATGCTG
Rep-F2	GGAATACACATGTGGTGTTG	Swimy-F2	GGGTGCCTCTGGTAAGTGTG
Rep-R2	TTAGCCAGGCACATATCAGG	Swimy-R2	TCCCCTGGGGTAGTCATACA
Helicase-F1	GTTGTTATTGGGTAAAGGAG	Swimy-F3	ACATGAATGGGCAGGTGAGT
Helicase-R1	TCTTACACGACTTATAGCC	Swimy-R3	TGCCACACAGGAATCAGAAC
Helicase-F2	CCTGAGGTTGAATTGGAACG	Swimy-F4	TGGGCTTGCATTAATCTTCC
Helicase-R2	CGACCATGAACATACGATAAG	Swimy-F5	TAATCTCTGCCCATGGCTGT
Helicase-F3	AAGCAACAACAATCCACACC	Swimy-R5	TGTCCCTTCTGTCAATGGTG
Helicase-R3	CATAGTCCTGTGCCCCAATG	Swimy-F6	GTCAGTGCAGTGTCCAGGTC
Endonuclease-F1	ACCGGGACAAGCTTATGTTG	Swimy-R6	AAGCTGTTCTCCAGGAGGTG
Endonuclease-R1	GGCACAACTACTGCCTCCAC	Swimy-R7	GTCCTGCAATACTGGGACAAA
Endonuclease-F2	AGGATACCTCCAAATGGTTG	Swimy-F8a	GCAGTGGGACAGACACAACTT
Endonuclease-R2	AGCTGTCTTAATGCAAGAAACG	Swimy-F8b	CAGGAGGAAAGTGGTTCCAT
Endonuclease -F3	GGCATGGGCATGTACAGTAC	Swimy-F8c	TTGGTGACCCTCAGTTGTTG
Helitron-R1	CAGACGGGAGAAATGTGGTG	Swimy-F8d	CACTTGTCCCTGTGGCATTA
Helitron-R2	AGCAAATGGCCTTTTTGACC	Swimy-F8e	CAACGATCACTTGCTGCTGT
Helitron-R3	TACTCACGCCATCAGGACAC	Swimy-R8a	GCAGCAAGTGATCGTTGAAG
Helitron-R4	ACTGGCCATTTTCGCACAGC	Swimy-R8b	GCAGATGGAGCACAAGATCC
Helitron-F4	GCGAAAATGGCCAGTTAGTG	Swimy-R8c	GCCCCGCAACTTTTAGAATG
Helitron-R5	GCAGCTGATCCATGATGTTG	Swimy-R10	CTGGACACCATGTGAAAAGTC
Helitron-F5	GCACAACATCATGGATCAGC	Swimy-R11	TGGTAGGCCACATCTGAAGA

Name	Sequences of 5´-3´	Name	Sequences of 5´-3´
Hel-Mk-Rep-F1	ACGAGGCTCACCACATCTTC	Swimy-F10	CTGCAAAGAAACATCCTCCA
Hel-Mk-Rep-R1	CCACAATGCTTGTCATCCAC	Swimy-F11	TGGGTAATGCTCAACTTTTGTG
Hel-Mk-Rep-F2	GAGTGCTCTCGGAAAGTTGTG	Swimy-F15a	CCCTGTTCTGAGGGATAATGAA
Hel-Mk-Rep-R2	TTTACCAGTCCCTGCACCTC	Swimy-F15b	TCTTCAGATGTGGCCTACCA
Hel-Mk-Helicase-F1	GAAAATGGTGTGTGGAAAAGG	Swimy-R15a	AGGCTTTTGAGTCCAGGTGA
Hel-Mk-Helicase-R1	CGTCTCAGACCACCTCTTCC	Swimy-R15b	TCTCCTCTGGTGAGGACCTG
Hel-Mk-EN-F1	GACATTGGAAGACAACGGAAG	Polymerase-F1	CAATTCAGTGGCTAGAGTGG
Hel-Mk-EN-R1	ACGACAAAGAACCCCATCATG	Polymerase-R1	AGGCCTCGTGGAGGATAAAC
Hel-Mk-F1	AATATGCGTCGCACCAAAGT	Polymerase -R2	TTGACTCTCTCGCTGCACTC
Hel-Mk-R1	TGCCAAACCATTCTGCATTC	Polymerase -F2	TCGTTTTAAGAAACGCGACA
N20t032-F	GCTGGTGTTGCTTTTCTCTG	Polymerase -F3	CGGGTGAGTCTGTGCATTAT
N20t213-R	CGATGGTTGATCACAATCTC	TNF-F1	GAGGTGGATTATCAGGAAAATTGC
N20t072-F	ATGTGTTAAAACCCGGATGG	TNF -F2	AGCAGCTTCACCTGCTCATC
N20t140-R	TGTAAAATGCCCACGCACTC	TNF -R1	AAACTTCCCGTGACATGAGC
N20t071-F	AGGTGGACAAAATGGAGTGC	TNF -R2	ACGTACTGCAGCAGGGGT

6.2 Protein sequences of some gene candidates and Alignment of *helitrons*

>Swimy_Xm, 939 bases, C4B checksum.

ZnF_C2H2

MYINAFLKFYSFLFVTAEFHIFRCGLPCRDTTHFHCPYCFSTIDRRDRFVNHLVNHRMTFHMVSSRAAAQPPVHLPATDIQAT

ZnF_C2H2

QAPVRPAATDIQATQAPVRPAATDIQATQAPVRSEDMITCEMCNLRLRKKNLRIHLKRKHHEKVELISQNHHLRSQCVDAKKG

ZnF_NFX

VFAVQKSFCGPSKPIHVIKNTWSQFHRSECELDRCNVNAQFAKRSGILPFECEHIQSLRFCPRAIDSEGPLSETTLNILVDNL

ZnF_SWIM

WFSVKRKEELLKNQQEALSSNTPFSVLLFGSGTDTTFHVSIFEPKMAFFSRLGRVIVTYDKGKNTWHCACTEGKRSCTHKATA
KWHLFEKMPDMFKAAVHSEEDVMCGTTVSECQEESGSITDDAAERIIKYLIYHKKIPADLPDSLITRSRDAKSLNGFPKHLIP
SETECTECGQHASLGDPQLLSSTAKIVTLTGVVKGISTYRKTCPTCSMVYRYQDWEDGIHNFNDHLLLSIHFCLVLRNALQTH
TAVSRIIETIEYTEGENFPFKDRILQAYLSFEGLSDHDYQYSCDSCGYHPAIVVMDLHKKGVFSMPVSEIPNPPQQYNGEVNA

RSFWEAVTMEVISRGLYPPGCKNPFVVKPTYHNWSPWIGPQTRRSDILINTESEKVHLPDSDSDVDQFLTEERLADEVLNLKL

SAP DNA-bind domain

PEIRRLCSQCGIDSRGSKMDLVLRLRDKMASRATYNKVFEKVWGASGGWAVITCPCGIIYSVKFNLRAESPRDFADLLLSWKF
FPNITVYDYPRGLVSHLKKRCAKDTPFTIHDGRLVEATPENIRQATEKKLAVNLPWLKHRKEPEDKGGHPVTGSLEHYCLSDV
FHQGNSKDEKDVLRKIEMVPELAGRLNSQCAEQLFAGMRKNNYFLNMLNPSTHIFLQRNILHHFNLRKNTKTKQNILKVAGPH
ANLVLDKNGCIIMGNAQLLCYRCGSL

>Partial-polymerase type B_Xm, 1158 aa, bold letters are the additional region in animals.

RCTVRLVWGNISHHINISVTDNDAILPAFEEFLDELVQSNAELPSESNLELILQIVGNPTGGSKRKAERTLECELINKKRRHL
YIVENTGNKLCFATSLAHVAHPEFTDKQALEQGRKWQHQAGLTDQTAVTFSDVAKFENILQRKIVVFHRTSKDRALSKFETDF
QNRSNPCFLLLHNNHFYGIRNLAGFTGSRYICKFCYGGHNNSNTHHCQGYCGVCCGYSCRQLQYNPVHCDDCNRICRNSTCFD
RHKEPRNRPHAEMRVSDCETIKFCSTCKRLYRVPMNKEKTLHICESKCVICGEKNLPPGTDVTLNDHQCYIQTSTIDDKLHDK

→ **DNA_Pol_B_2 domain**

LVFYDFETFVDQSGVHKPFLVCSKTVKGVEWHAYGLDCAQQFLLHFRRPMFKGHTFIAHNARGFDSYLLLNSMVQLGIKPFLI
MQGGKVLCFTDPDYKLKFIDSLSFLTMKLSAMPKALGFHDRSKGYFPHEFSAEEHLKYVGVFPPLDSYGVKLMNPDERQKITD
WYGEASKGIFDFEKESLHYCKNDVDILFQGCVKFREEFFKETNVDPFKNITIASACMQVFVTNFLPEKSLAIPSAVDY**RRGSK**
TFSNASIQWLEWKMDSENLHIEHALNSGERKIGPYFVDGFAVISGLATIFSFNGCLYHACPRCFKQTEVCPLRKVPFEQIYAA
TVERSKILQAVYGVRVETVWEHEWDEMKKSDPGVIRFLEKFDAPEPLVPRNALYGGRTCALKLRFTAGPGESVHYVDFTSLYP
YVNATCEYPLGHPTLIYKDFDDPVNYFGFIRATVYPPRGLFFPVLPYKTSRGKLVFTLCRTCADINNQSGICTHEDEARSLTG
VWVSAEFQKALQCGYRLGKITEVWHFERSSSSIFKGYIHTFLKGKQEASGYPPEAMDQESRLKYVRDYQINQGIQLDAGKIEV
NPAKRQVAKLCLNSFWGKFAQRNDLSQTSFVSDPDEYFNFFFSGKYVVKYFHFINPETCLIQWNYSKRCIIRPNKSNNIFIAA

DNA_Pol_B_2 domain ←

FTTAYARLKLFSCLERVQDKILYIDTDSLIYVVKDGESPLELGNYLGDLTDELGGDTIQEFVAAGPKSYAYQTKNQKKVVIRV
KGITQTYECSERVNFDSIRELVGGYLEGSRHGVIKTPQHTIKRDKKGFVLRNATFLKRFQVVYDKRRLFPDGSTLPFGY

>Partial msh2 from liver cDNA, 293 bases, 0 checksum.
MQQQEVHHAGCSEDGVRFTNSKLSSLNEEYTRNREEYEEAQNAIVKEIINIASGYVDPLQTLSDVIAQLDAVASFAVASVSAP
VPYVRPXLLDDGRRRLELLQARHPCMETDADTAFIPNDISFVQGEKSFYIITGPNMGGKSTFIRQVGLIALMAQIGCFVPCEK
AELSVIDSVLARVGAGDSQVKGVSTFMSEMLETAAILRSATENSLIIIDELGRGTSTYDGFGLAWAISEHISSKINCFCLFAT
HFHELTALAAQQPAVHNLHVTALTTHNTLTMLYRVKPGVCDQSF

>Alignment of *Helitrons* in the platyfish and the medaka. Identical residues are shown in white type on black and conservative substitutions in black type on gray.

```
Hel_platyfish  -----------------------------------------------------MSCSVVAAIKHHISPVCTWKTSDVDEV 27
Hel_medaka     MPRKKGFNMRRTKVCKVQEEGKTSGCVCQEQIQKPLISGMKVRLTXSSDSTTSLTCCVVAAINHVVAPVCTWRDADVDNI 80

Hel_platyfish  GVEGRKLAEYVARERPNRGPKPELCKLVEDLTIFGRKWKVVIGN-IMFGTFGFEQDGELYEILKKYLLINGMCMFSLHGA 106
Hel_medaka     RNEGVQLVCSFNGSDGCXNNSXKLFKVSEKLSVFGRIWDVAINEPVYSGDSKSEDFSLLYEMMQERLMEEDLCLLDLQGF 160

Hel_platyfish  TCLVIQHGLYFVVVDFGTRNSQGLASQSGTSVVVFNTCLNDLMIHLLNLRESLNAIEYGISAISVQEMYSN--------- 177
Hel_medaka     VCAVIHHKEYFVVVDCGARDSSGMFSNIGTSVAVFNSCFNDLMLHIWNLKKSLNAEWFGISSMFVKEASWDGCNNTFGLG 240

Hel_platyfish  ---------------------------------------VETCAAVYADRQATDACHSASSHVASLS------------G 206
Hel_medaka     AEDARVNQSVLGKRQTGFIKEHVQELNDKIVVGDKCKVLMECKXTLKIGKKGXNLKKKSSSVILNEXNCNRFDTERCVRG 320

Hel_platyfish  SFHQGDVQFKYAGVQCTAISAVALTKHTLDSVFSWNADILDDVVVLGDQLYTFLRDNNLISGGSQLLCVPDLPKKMHVDG 286
Hel_medaka     SFHQGDRQLQNRGVQCMAIALVSXAKHTXRSVFEWEATDLDHVVFLGDQLYSSVCENKIIPGNLKFLSVPDLPMDSEIDG 400

Hel_platyfish  QSFEYAYGDYVAGAIDTLDPELLESGVHTSLWDGLSKMCAKYETSFITISGSTCALISSNGRYAVVDSHARNTDGMVHXX 366
Hel_medaka     KLFKFEYGEYVSGIVNVSNGKFVDSGEVITLKKGLEHMLVKYKTCFLTIGSSTCAVISENGHYAVVDSHARSGTGMADVN 480

Hel_platyfish  GKSVVLYFNTLDDVFVYIXRFSXQLNVTPKLFEISGVDIVQTGSSKIASEHX--------------------AAVPGSSX 426
Hel_medaka     GRSIIMYLPSLNDLYNYICCLSISLAGKEMAFEIASVAATDNDHSEKNKQSVLEKDCKTGLNTESQFGFGCSLDVQTSSL 560

Hel_platyfish  VFDHTFTGVDMQDDSCG-EHSSQLDMSDMDDDVVVTGVQSX-VLYFNPVSEEIARSLCGKLNVEYQRANXVS-CVVGELG 503
Hel_medaka     PLPESNRKRKLLTGQCESKKTKWYDANEASLDVEFIGEVSSKKLEFYPVDFDVTQALCKQLNVHCENTACANPLEVXMLG 640

Hel_platyfish  VPCLTEKIVGDGNCFFRALSQAISGTQKNHRKIRLAVVKQLQRNSHTYDSILRSEYSSISQYIAVSRMQYVGSWATEVEI 583
Hel_medaka     IPCQTDHIVADGNCFFRAVSKVISGTEKNHIKLRRAVVTHLQKNVGNYSSLLRSENMSVDEYLNVSRMRFSGSWATEVEI 720

Hel_platyfish  KAXADYFGVXIFTFCDDKWLEYX-SLSSVSNHALYLQNISGNHYETVTCVKQPRSQTCYGYCMNSDLS-GEYKTRQVTX- 660
Hel_medaka     QATADYLGVSVYTXHNERWLKYACQRKAISKQGIYLQNINECHYEVVTCVXXFQQQRCFGXCKSMEXNXLGHNMRQXKRX 800

Hel_platyfish  --EQNIARIKXVKTETSIEDECVGIL------QDNQNTSL------------FTPISEXTAKTLCXNLKIDFEEHXFQKX 720
Hel_medaka     SEMLSEGDCKSVKHXRSEVGDCNXVLXSALVNNDHQNSEIXQSNNLVENAFSFYPLSFESAVLLSKHIDLDFVKQYEDIP 880

Hel_platyfish  ILRGPLGHVCKTKNIINDGNSFFRAVAYALSASEKNHRKIRLAVISHMTKNTEECKKYLAKNFASVTEYVNQSQMKYIGH 800
Hel_medaka     TNNVTLGSICKTEVILNDGNSFFRAISRVISGSENCHRRVRLAVVKFMLKNSDIHMKLLXXGYATVSEYIXKSRMKYVGS 960

Hel_platyfish  CATEIEFKSTANMLGLDIYVFNGTQWTKYNSNSSHLTNEAIYLQNCD-EHFDVVICAKQGDKDFCFGLCEEN-VSLKKRH 878
Hel_medaka     FVSEIDIQATSNFLGVEIFVHNGSIWRKYNPVNKILGNQAIYLRECXXKYYEVVVCVHCSGKXGCANWCKSNSICSEKPL 1040

Hel_platyfish  IRTRSPQVPKNERQAVLSG--------------KANDSFSTYLRQKKNQRYTTKYRTKMVYRQKIKLNKTNTYKKNLLYK 944
Hel_medaka     LRCRSFQPHNVSCDSVGESSFSKYLKKRRSFYHKMQYQIIELHKQKKDMSKRKYASDSLHRSRALELSKINYXENAIYR 1120

Hel_platyfish  EQKKKWIRNKYRQDQTYQKKLRQVSISKYKEDKCHREKVKQISIAKYNTDQSHREKVKLISIAKYNTDQSHREKVKQISI 1024
Hel_medaka     MCVKERSSAKYRTDVLHRDRVKNLSKNKYKCDXLHKHNVKERSKRKYQSNVTHRENVKALSSRKYQSNVTHRENVKALSS 1200
```

```
Hel_platyfish  AKYKENKSHREKVKHFSRKQYLN-PQHKIHIISNVKLKRQEIKMKSKEFDFVVAQFFDKVKEGPNFVCCVCFRLLFKHQV 1103
Hel_medaka     RKYQSNVTHRENVKALSSRKYRQDPHFRKSVLDGMKLKRQLNKKQLEEFDVVMKQFLDKVKDGPGFVCCVCHRLLFEHQV 1280

Hel_platyfish  LNCHKDSYRKTKEMSLITDKCISDDYVHICNNGCISPCN-LKTCRNKLYICYCCHHKISKGQMPPESSCNNLTVDDIPPQ 1182
Hel_medaka     LHCKKELYSN-AGLGLIADQCITSGYLHICNANCSKNCEWMKTSRSKLWICYTCHAKIKKGKVPAESVKNNLELEPIPAE 1359

Hel_platyfish  LACLNTLEQHLIALHIPFMKMLALPKGGQNGVHGPVTCVPANVAETCSMLPRSNMEGFLLPVKLKRKLTYKGHYDYQYVD 1262
Hel_medaka     LACLNQLEQHLIALHIPFMKLLALPKGGQNGVHGPITCVPANTEQTTYLLPRMSIEGSLVPVKLKRKLTYKGHYEYQYVD 1439

Hel_platyfish  SMHVQEALWYLKHCNFHYKNVEFNESWINEFXQEDNNSVLNNTSVNEENEDISIGENDDDLLHDRQQHCMFQDTCLMPVD 1342
Hel_medaka     TERVKQALQFLQTHNYYNNVKFNNDWINGFCKETSDVVEKDVECTANVDDSNVLG--DELLHDRQQHCMFQDTCLMPVD 1517

Hel_platyfish  IGQEALDLYVDNVLNVAPGENNNPIKLLSDCTNEAKCFPVLFPSGFNTYHEKRQYRLTLSRYFNNRLLHADGRFARNVEY 1422
Hel_medaka     IGQEALDQYFDHILNVAPAEGNSPVKLLSDITNEAKCFPVLFPKGGPTFHDGRQHPLTLSRYFNNRILNADCRFAQNVEY 1597

Hel_platyfish  IFFAQYMSELEQVXSKVSIALRKGTSRTPQN---MSEVLRDEQSIXKLLXFDDGYRFLKPIRGTPAFWQTAQRDLLACVR 1499
Hel_medaka     IFFAQYMTELEKVVSNVSIAVRKGKYNFKNKRVVTGELLXSDDSIKRMLEFDDGYRFLKPIRGTPAFWQAAQKDLMACVR 1677

Hel_platyfish  MLGKPTWFASFSSADMRWTNLLYSILKQEGRTQTLEQLQWAEKCELLRRNPVTAARMFDFRWHVFVREVLMSPAXPIGKI 1579
Hel_medaka     QLGIPTWFCSFSSADMRWKNLLSCILKHEGRTETVEQLEWADRCELLRRNPVTAARMFDFRWHCFLKEVLMSPACPIGKI 1757

                                   Rep donain
Hel_platyfish  EDYYYRVEFQQRGSPHCHCLFWISGAPILDKNTDEEVIAFIDKYVTCEIPSEEDALSEVVTSVQQHSKRHSKTCKKKKTV 1659
Hel_medaka     KDYFYRVEFQQRGSPHLHCLFWIEDAPVIDRNTDAEVIAFIDRYVTCELPSNIDLL-EIVTSVQKHSETHAKTCKKKGTV 1836

Hel_platyfish  CRFNFPRPVSCRTFICRGEKYQDPVKTCTCNLDKTDGSADCECLXKNKTRPEQMDSDVAXNILTKIKNAISXDNCPYNTV 1739
Hel_medaka     CRFNFPRPPSARTFISR-STSAEETQSCQCVLSQKE---SCNCKENAQTX--QLKKKEASALIAHIKAALSTDGPTYDNV 1910

Hel_platyfish  EEXFEGLXMNQXVFETAYKRFSRNTHVVLKRQINEIWINQYSRPLLKAWDANIDIQYCVDAYACCVYIVSYMSKSEREIG 1819
Hel_medaka     EDLFESVGLNQSTFETAYNCCARKTHVVLKRQLKEVWINQYSKPLLKCWNANMDIQFVVDAYACVVYIISYISKSEKEMG 1990

Hel_platyfish  LLLGNAQREAAKEGNVSAKXALKRLGSVYLHNRDVXAQEAVYRLTNMHLKECSRKVVFIPTGDNIVKMSLPISVLKQKAT 1899
Hel_medaka     LLLQCAQKEASRDGNASAKESLKTLGSTYLHNRDVSAQESVYRLTNAHLKECSRKVVFLPVGENTVKMSLPLKVLKEKAS 2070

Hel_platyfish  SQDLTPEDMWMTGIVDRYKNRPNDDVFPDMCLAKFASEYRVLPKNEKCRNPVKLNKNFGFVVKRTRTKPAVVRYARFSET 1979
Hel_medaka     KHSVSTDDMWMTSIVDRYKNRPKSDDFPSMCLAVFASEFRVLSKNESAAKRIKLNNNTGYVVKRTRTQPAVVRYVKFSET 2150

Hel_platyfish  KEPERFYQSIMQLFLPYRFDSELKPAHCETFGDFYXTGVISFVDGTRHSVKXVVDLNRSEFEVESDHFEAVDN-VTGDVM 2058
Hel_medaka     KQPELFYQSNLQLFLPYYSDSQLKPQGFETYQDFFHKGTIRLGNSGLQLVKSVVTLNRSKFEIDGDKLVNAQQGIDNAGV 2230

                                                                 Helicase domain
Hel_platyfish  LEDAWAELCPEVELERLECVELQR-ERQIENSDPEQIPDL--SLQCKEFSVFEKRKXTRTEGLALIRSLNEKQFSVFYQI 2135
Hel_medaka     VEDXWCELCPEQELERLECQELLQNQKSVDDEQEFDIPDLRVSAQHSHFFQLNKTKMPRSDGLALIRSLNEIQMDIFYQV 2310

Hel_platyfish  RQWCLAKVNGKNPEPLHIFITGGAGTGKSHLIKAIEYESKRLLSTVCQSPDNTCVLLTAPTGIAAYNLEATIHTTFSIG 2215
Hel_medaka     RKWCVEKVLGKKPEPFHLFITGGAGTGKTHLIKAIKYESNRLLSPLCRHVDDICVLLTAPTGIAAFNVQASTIHNTFCIG 2390
```

Helicase domain

```
Hel_platyfish  KDVRLPYTPLSEEKLNSLRVKYCDLQLLIIDEISMVDHNLLSYVHGRLRQIKQTVKS--YGNXSIIVVGDMYQLPPVKGK 2293
Hel_medaka     KDVRLPYTPLGEEKLNSLRAIYSSLQILIIDEISMVDHNLLAYIHGRLRQIKQTGNSALFGNVCIIAVGDFFQLPPVRGK 2470

Hel_platyfish  PLYSDGVATNIWSXLFKIVELTEIVRQKDAVFSQLLNRMRTHSKGTPILADDLQILKRCETGEVSSALHIFATNKQVNEH 2373
Hel_medaka     PLYVENVGMDLWCGVFXIVCLETIVRQKDQEFAKLLGRVRTRCKGTEMLKSDIQQLKSRETGEGSSALHIFPTNGQVNQH 2550

Hel_platyfish  NIHRLYETCPEFVSIGAQDYVNDKKTGKLRLLEGNHAKASNTCLSEVLLLGKGARVMLCKNVDVXDGLVNGVCGTVTXIL 2453
Hel_medaka     NLTHLHSSCPXYVTIKAQDYVNCKETGRQVLLSGNHSNANNTCLAXRLLLGKNARVMLCKNVDVADGLVNGACGIVTDIV 2630

Hel_platyfish  IPEKDKFPNVVYVKFDXXRVGMQKRKSCHYASSDLAGSTXIGPEEERATVKGGMRRQFPLRLAWACTVHKVQGLTVDEAV 2533
Hel_medaka     YRNSD-FPQIVFVNFDNKDIGRQRKKQYENHPGLVSGSVAIVPEEERLTGRGGLRRQFPLKLAWACTVHKVQGLTIDNVV 2709
```

Endonuclease domain

```
Hel_platyfish  VSFSKIFAPGQAYVAISRVRSVLGLTIQDFNEKKIFCKDEXLVSLQSMTPLLSXXIQLDRFNTSVFTVFLMNVQSLNRHV 2613
Hel_medaka     VSLKKTFAPGQAYVALSRVTSLSGLIIQDFKEQAIYCNKTVSEYVANMPQYLINKENLTSS-SCDFSIFLMNVQNLXWHV 2788

Hel_platyfish  KDLXCYTEHWKPKCIAITETWVS-STHTDTVKIDGYSFTNRPRCLSYTSRHPELIALQDQQHGGVGIYCXDDVEFEILQQ 2692
Hel_medaka     KDLALCTKQLHANCIAVTETWLPNAPFTDALNISGYNFYSMPRSLSYSGVDPRLNVIKGLKHGGVGMYTVENILFEHVEI 2868

Hel_platyfish  PELNLECLVYRFCSFNMVLGVIYRPPLYPLSLFKNNLGQLLDWLEKQSDTIALXGDFNDXILKSSIITKFVCDKGYLQMV 2772
Hel_medaka     QNVDIECLVHRFITSNIVVAVLYRPPSYAISLFRENLKKLVDLLEPLGDTIAIMGDFNENLVNSKTITDFMAHKGFIQLV 2948

Hel_platyfish  VEATTEKDTLIDHVYVKSKTYKVEAVVVPTYFSDHEGIMCGFSL---- 2816
Hel_medaka     SENTTLKNTIIDHVYLKTTNYKVECTVVQTYFSDHDGVLCRFTHVSDC 2996
```

7. References

Adam, D., Dimitrijevic, N. and Schartl, M. (1993). Tumor suppression in Xiphophorus by an accidentally acquired promoter. *Science* **259**, 816-9.

Aida, T. (1921). On the inheritance of color in a freshwater fish, *Aplocheilus latipes* Temminck and Schlegel, with special reference to sex-linked inheritance. *Genetics* **6**, 554-73.

Anand, R. (1986). Pulsed field gel electrophoresis: a technique for fractionating large DNA molecules. *Trends Genet.* **2**, 278-83.

Anders, A. and Anand, F. (1963). Genetisch bedingte XX- und XY-Weibchen beim wilden *platypoecilus maculatus*. *Z. Verebungsl.* **94**, 1-18.

Anders, A., Anders, F. and Rase, S. (1969). XY females caused by X-irradiation. *Experientia* **25**, 871.

Anders, F. (1991). Contributions of the Gordon-Kosswig melanoma system to the present concept of neoplasia. *Pigment Cell Res.* **4**, 7-29.

Argos, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Res.* **16**, 9909- 9916.

Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayama, S., Pierson, L. S. *et al.* (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5**, 433-40.

Ayling, L.-J. and Griffin, D. K. (2002). The evolution of sex chromoosmes. *Cytogenet Genome Res.* **99**, 125-40.

Bachtrog, D. (2003). Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila. *Nat. Genet.* **34**, 215.

Baroiller, J. F., Guiguen, Y. and Fostier, A. (1999). Endocrine and environmental aspects of sex differentiation in fish. *Cell Mol. Life Sci.* **55**, 910-31.

Beck, A. R. P., Miller, I. J., Anderson, P. and Streuli, M. (1998). RNA-binding protein TIAR is essential for primordial germ cell development. *PNAS* **95**, 2331-36.

Bellamy, A. W. and Queal, M. L. (1951). Heterosomal inheritance and sex determination in Platypoecilus maculatus. *Genetics* **36**, 93-107.

Berg, D. E. and Howe, M. M. (1989). Mobile DNA. Wahington, D.C.: American Society for Microbiology.

Bernstein, C., Bernstein, H., Payne, C. M. and Garewal, H. (2002). DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutat. Res.* **511**, 145-78.

Berta, P., Hawkins, J., Sinclair, A., Taylor, A., Griffiths, B., Goodfellow, P. and Fellous, M. (1990). Genetic evidence equating SRY and the testis-determining factor. *Nature* **348**, 448-50.

Bouralexis, S., Findlay, D. and Evdokiou, A. (2005). Death to the bad guys: targeting cancer via Apo2L/TRAIL. *Apoptosis* **10**, 35-51.

Brunner, B., Hornung, U., Shan, Z., Nanda, I., Kondo, M., Zend-Ajusch, E., Haaf, T., Ropers, H. H., Shima, A., Schmid, M. *et al.* (2001). Genomic organization and expression of the doublesex-related gene cluster in vertebrates and detection of putative regulatory regions for DMRT1. *Genomics* **77**, 8-17.

Bullock, W. O., Fernandez, J. M. and Short, J. M. (1989). Xl1-blue: A high efficiency plasmid transforming recA Escherichia coli strain with b-galactosidase selection. *Bio. Techniques* **5**, 376-9.

Burgoyne, P. (1982). Genetic homology and crossing over in the X and Y chromosomes of Mammals. *Hum. Genet.* **61**, 85-90.

Burgtorf, C., Welzel, K., Hasenbank, R., Zehetner, G., Weis, S. and Lehrach, H. (1998). Gridded genomic libraries of different chordate species: a reference library system for basic and comparative genetic studies of chordate genomes. *Genomics* **52**, 230-32.

Capel, B. (1998). Sex in the 90s: *SRY* and the switch to the male pathway. *Annu. Rev. Physiol.* **60**, 497-523.

Cerda-Reverter, J. M., Ringholm, A., Schiöth, H. B. and Peter, E. R. (2003). Molecular Cloning, Pharmacological Characterization, and Brain Mapping of the Melanocortin 4 Receptor in the Goldfish: Involvement in the Control of Food Intake. *Endocrinology* **144**, 2336-49.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Curr. Bio.* **6**, 149-162.

Charlesworth, B. (2002). The evolution of chromosomal sex determination. *Novartis Found Symp.* **244**, 207-19.

Charlesworth, B. (2003). The organization and evolution of the huamn Y chromosome. *Geno. Biol.* **4**, 226-8.

Craig, N. (1995). Unity in transposition reactions. *Science* **270**, 253-254.

Curcio, M. J. and Derbyshire, K. M. (2003). The outs and ins of transposition: from mu to kangaroo. *Nat. Rev. Mol. Cell Biol.* **4**, 865-877.

Delarue, M., Poch, O., Todro, N., Moras, D. and Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* **3**, 461-467.

Devlin, R. H. and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *aquaculture* **208**, 191-364.

Drotschmann, K., Hall, M. C., Shcherbakova, P. V., Wang, H., Erie, D. A., Brownewell, F. R., Kool, E. T. and Kunkel, T. A. (2002). DNA binding properties of the yeast Msh2-Msh6 and Mlh1-Pms1 heterodimers. *Biol. Chem.* **383**, 969-975.

Dzwillo, M. and Zander, K. (1967). Geschlechtsbestimmung und Geschlechtsumstimmung bei Zahnkarpfen (Pisces). *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut (Mitt hamb zool Mus Inst)* **64**, 147-62.

Estebanez-Perpina, E., Moore, J. M. R., Mar, E., Rodrigues, E. D., Nguyen, P., Baxter, J. D., Buehrer, B. M., Webb, P., Fletterick, R. J. and Guy, R. K. (2004). The molecular mechanisms of coactivator utilization in ligand dependent transactivation by the androgen receptor. *J. Biol. Chem.* **10**, 1074.

Ferrari, S., Harley, V., Pontiggia, A., Goodfellow, P., Lovell-Badge, R. and Bianchi, M. (1992). SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.* **11**, 4497-506.

Feschotte, C. and Wessler, S. R. (2002). Mariner-like transposases are widespread and diverse in flowering plants. *Proc. Natl. Acad. Sci. USA* **99**, 280-5.

Filatov, D. A., Moneger, F., Negrutiu, I. and Charlesworth, D. (2000). Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**, 388-90.

Fischer, J., Maier, H., Viell, P. and Altenbuchner, J. (1996). The use of an improved transposon mutagenesis system for DNA sequencing leads to the characterization of a new insertion sequence of Streptomyces lividans 66. *Gene* **180**, 81-9.

Fornzler, D., Altschmied, J., Nanda, I., Kolb, R., Baudler, B., Schmid, M. and Schartl, M. (1996). The Xmrk oncogene promoter is derived from a novel amplified locus of unusual organization. *Genome Res.* **6**, 102-113.

Foster, J., Brennan, F., Hampikian, G., Goodfellow, P., Sinclair, A., Lovell-Badge, R., Selwood, L., Renfree, M., Cooper, D. and Graves, J. A. (1992). Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. *Nature* **359**, 531-3.

Fredga, K. (1988). Aberrant chromosomal sex determining mechanisms in mammals, with special reference to species with XY females. *Philos. Trans. R Soc. Lond B Biol. Sci.* **322**, 83-95.

Froschauer, A., Koerting, C., Bernhardt, W., Nanda, I., Schmid, M., Schartl, M. and Volff, J. N. (2001). Genomic plasticity and melanoma formation in the fish Xiphophorus. *Mar. Biotechnol.* **3**, 72-80.

Froschauer, A., Korting, C., Katagiri, T., Aoki, T., Asakawa, S., Shimizu, N., Schartl, M. and Volff, J. N. (2002). Construction and initial analysis of bacterial artificial chromosome (BAC) contigs from the sex-determining region of the platyfish Xiphophorus maculatus. *Gene* **295**, 247-54.

Froschauer, A. (2003). Identifizierung und Molekulare Analyse Xmrk-gekoppelter Gene in der geschlechtsbestimmenden Region des Genoms von Xiphophorus maculatus. PhD dissertation, Julius-Maximilians-Universitaet Wuerzburg.

Froschauer A.,Y. Selz, C. Schultheis, C. Körting, Q. Zhou, W. Bernhardt, R. Hanel, M. Schartl and J.-N. Volff. Multicopy melanocortin receptor gene at the melanoma locus on the sex chromosomes of the platyfish *Xiphophorus maculatus*, manuscript in preparation.

Fryssira. (1988). 46, XY, del(2)(q32-qter) in a female with multiple congenital abnormalities. *Eur. J. Hum. Genet. [Suppl 1]* **6**, 88.

Garnier, T., Saurin, W. and Cole, S. T. (1987). Molecular characterization of the resolvase gene, res, carried by a multicopy plasmid from Clostridium perfringens: common evolutionary origin for prokaryotic site-specific recombinases. *Mol. Microbiol.* **1**, 371-76.

Gomez, A., Volff, J. N., Hornung, U., Scharl, M. and Wellbrock, C. (2004). Identification of a second *egfr* egene in *Xiphophorus* uncovers an expansion of the epidermal growth factor receptor family in fish. *Mol. Biol. Evol.* **21**, 266-75.

Gordon, M. (1927). The genetics of viviparous top-minnow Platypoecilus: the inheritance of two kinds of melanophores. *Genetics* **12**, 253-83.

Graves, J. A. M. and Shetty, S. (2001). Sex from W to Z: Evolution of vertebrate sex chromosomes and sex determining genes. *J. Exp. Zool.* **290**, 449-62.

Graves, J. A. M. (2002). The rise and fall of SRY. *Trends Genet.* **18**, 259-64.

Griffin, D. K., Harvey, S. C., Campos-Ramos, R., Ayling, L.-J., Bromage, N. R., Masabanda, J. S. and Penman, D. J. (2002). Early origins of the X and Y chromosomes: Lessons from tilapia. *Cytogenet. Genome Res.* **99**, 157-63.

Gruetzner, F., Rens, W., Tsend Ayush, E., El Mogharbel, N., O brien, P. C. M., Jones, R. C., Ferguson Smith, M. A. and Graves, J. A. M. (2004). In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* **432**, 913-917.

Gutbrod, H. and Schartl, M. (1999). Intragenic sex-chromosomal crossovers of Xmrk oncogene alleles affect pigment pattern formation and the severity of melanoma in Xiphophorus. *Genetics* **151**, 773-83.

Hanahan, D. (1995). Techniques for transformation of *E. coli.* In *DNA cloning*, (ed. D. M. Glover), pp. 109-35. Oxford: IRL Press.

Harley, V., Jackson, D., Hextall, P., Hawkins, J., Berkovitz, G., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P. (1992). DNA binding activity of recombinant SRY from normal males and XY females. *Science* **255**, 453-6.

Harrington, R. W. (1963). Twenty-four-hour rhythms of internal self-fertilization and of oviposition by hermaphrodites of *Rivulus marmoratus*. *Physiol. Zool.* **36**, 325-41.

Heery, D. M. G. F. and Powell, R. (1990). A simple method for subcloning DNA fragments from gel slices. *Trends Genet.* **6**, 173.

Hornung, U., Nanda, I., Kondo, M., Shimizu, A., Asakawa, S., Volff, J. N., Winkler, C., Shan, Z., Haaf, T., Shima, A. *et al.* (2004). *DMRT* genes and sex determination in Medaka. In *Chromosomes Today*, vol. 14 (ed. M. Schmid and I. Nanda), pp. 27-37. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Huszar, D., Lynch, C. A., Fairchild-Huntress, V., Dunmore, J. H., Fang, Q., Berkemeier, L. R., Gu, W., Kesterson, R. A., Boston, B. A., Cone, R. D. et al. (1997). Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* 88, 131-41.

Izsvak, Z., Ivics, Z. and Hackett, D. A. (1997). Repetitive elements and their genetic application in zebrafish. *Biochem. Cell Biol.* **75**, 507-23.

Izsvak, Z. and Ivics, Z. (2003). Sleeping Beuty transposition: Biology and Applications for molecular therapy. *Mol. Ther.* **9**, 147-56.

Jackson, J. A. and Fink, G. R. (1981). Gene conversion between duplicated genetic elements in yeast. *Nature* **292**, 306-311.

Jackson, I. J. (1999). The mahogany mouse mutation: further links between pigmentation, obesity and the immune system. *Trends Genet.* **15**, 429-31.

Jager, R., Anvret, M., Hall, K. and Scherer, G. (1990). A human XY female with a frame shift mutation in the candidate testis-determining gene SRY. *Nature* **348**, 452-4.

Jones, K. W. and Singh, L. (1985). Snakes and the evolution of sex chromosomes. *Trends Genet.* **1**, 55-61.

Jordan, S. A. and Jackson, I. J. (1998). Melanocortin receptors and antagonists regulate pigmentation and body weight. *Bioessays* **20**, 603-6.

Jung, G., Leavitt, M. C., Hsieh, J.-C. and Ito, J. (1987). Bacteriophage PRD1 DNA polymerase: Evolution of DNA polymerases. *Proc. Natl. Acad. Sci. USA* **84**, 8287- 8291.

Just, W., Rau, W., Vogel, W., Akhverdian, M., Fredga, K., Graves, J. A. and Lyapunova, E. (1995). Absence of *Sry* in species of the vole *Ellobius*. *Nature Genet.* **11**, 117-8.

Kallman, K. D. and Schreibman, M. P. (1973). A sex-linked gene controlling gonadotrop differentiation and its significance in determining the age of sexual maturation and size of the platyfish, Xiphophorus maculatus. *Gen. Comp. Endocrinol.* **21**, 287-304.

Kallman, K. D. (1975). The platyfish, Xiphophorus maculatus. In *Handbook of genetics*, (ed. R. C. King), pp. 81-132. New York: Plenum.

Kallman, K. D. (1984). A new look at sex determination in poeciliid fishes. In *Evolutionary Genetics of Fishes*, (ed. B. J. Turner), pp. 95-171. New York: Plenum Publishing Corporation.

Kallman, K. D. and Bao, I. Y. (1987). Female heterogamety in the swordtail, Xiphophorus alvarezi Rosen (Pisces, Poeciliidae), with comments on a natural polymorphism affecting sword coloration. *J. Exp. Zool.* **243**, 93-102.

Kallman, K. D. (1989). Genetic control of size at maturity in *Xiphophorus*. In *Ecology and evolution in livebearing fishes (Poeciliade)*, (ed. G. K. Meffee and F. F. Snelson), pp. 163-84. Englewood Cliffs, NJ: Prentice-Hall.

Kapitonov, V. V. and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. *PNAS* **98**, 8714 - 8719.

Kapitonov, V. V. and Jurka, J. (2003). Molecular paleontology of transposable elements in the Drosophila melanogaster genome. *Proc. Natl. Acad. Sci. USA* **100**, 6569-74.

Kazianis, S., Gutbrod, H., Nairn, R. S., Mcentire, B. B. and Della Coletta, L., *et al.* (1998). Localization of a CDKN2 gene in linkage group V of Xiphophorus fishes defines it as a candidate for the Diff tumor suppressor. *Genes Chromosomes Cancer* **22**, 210-220.

Kazianis, S., Morizot, D. C., Vielkind, J. R., Woolcock, B., Wigler, M. and Lucito, R. *et al.* (2004a). Sex-determination in platyfishes and swordtails, submitted.

Kazianis, S., Nairn, R. S., Walter, R. B., Johnston, D. A., Kumar, J., Trono, D., Della-Coletta, L., Gimenez-Conti, I., Rains, J. D., Williams, E. L. *et al.* (2004b). The Genetic Map of Xiphophorus Fishes Represented by 24 Multipoint Linkage Groups. *Zebrafish* **1**, 287-304.

Kirchner, J. M., Ivanova, V., Samson, A., Noskov, V. N., Volff, J.-N., Resnick, M. A. and Walter, R. B. (2001). Transformation-Associated Recombination (TAR) Cloning of Tumor-Inducing Xmrk2 Gene from Xiphophorus maculatus. *Marine Biotechnology* **3**, 168-76.

Knower, K. C., Kelly, S. and Harley, V. R. (2003). Turning on the male-SRY, SOX9 and sex determination in mammals. *Cytogenet. Genome Res.* **101**, 185-98.

Kobayashi, T., Matsuda, M., Kajiura-Kobayashi, H., Suzuki, A., Saito, N., Nakamoto, M., Shibata, N. and Nagahama, Y. (2004). Two DM domain genes, DMY and DMRT1, involved in testicular differentiation and development in the medaka, *Oryzias latipes*. *Dev. Dyn.* **231**, 518-26.

Kondo, M., Nagao, E., Mitani, H. and Shima, A. (2001). Differents in recombination frequencies during female and male meioses of the sex chromosomes of the medaka, *Oryzias latipes*. *Genet. Res.* **78**, 23-30.

Kondo, M., Froschauer, A., Kitano, A., Nanda, I., Hornung, U., Volff, J. N., Asakawa, S., Mitani, H., Naruse, K., Tanaka, M. *et al.* (2002). Molecular cloning and characterization of DMRT genes from the medaka Oryzias latipes and the platyfish Xiphophorus maculatus. *Gene* **295**, 213-22.

Kondo, M., Nanda, I., Hornung, U., Asakawa, S., Shimizu, N., Mitani, H., Schmid, M., Shima, A. and Schartl, M. (2003). Absence of the Candidate Male Sex-Determining Gene dmrt1b(Y) of Medaka from Other Fish Species. *Curr. Biol.* **13**, 416-20.

Kondo, M., Nanda, I., Hornung, U., Schmid, M. and Scharl, M. (2004). Evolutionary origin of the medaka Y chromosome. *Curr. Biol.* **14**, 1662-69.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. *Nature* **351**, 117-21.

Kosswig, C. (1928). Ueber Bastarde der Teleostier Platypoecilus und Xiphophorus. *Z. indukt. Abstammungs Vererbungsl.* **47**, 150-8.

Kuhn, C., Vielkind, U. and Anders, F. (1979). Cell cultures derived from embryos and melanoma of poeciliid fish. *In Vitro* **15**, 537-44.

Kumar, S., Tamura, K., Jakobsen, I. B. and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244-45.

Lahn, B. T., Pearson, N. M. and Jegalian, K. (2001). The human Y chromosome, in the light of evolution. *Nat. Rev. Genet.* **2**, 207-16.

Lal, S. K., Giroux, M. J., Brendel, V., Vallejos, C. E. and Hannah, L. C. (2003). The Maize Genome Contains a Helitron Insertion. *Plant Cell* **15**, 381 - 391.

Lee, B. Y., Penman, D. J. and Kocher, T. D. (2003). Identification of a sex-determining region in Niletilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Anim. Genet.* **34**, 379-383.

Lee, B. Y., Hulata, G. and Kocher, T. D. (2004). Two unlinked loci controlling the sex of blue tilipia (*Oreochromis aureus*). *Heredity* **92**, 543-49.

Liu, Z., Moore, P. H., Ma, H., Ackerman, C. M., Ragiba, M., Yu, Q., Pearl, H. M., Kim, M. S., Charlton, J. W., Stiles, J. I. *et al.* (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* **427**, 348-52.

Lv, D., Song, P., Chen, Y., Gong, W. and Mo, S. (2005). Cloning and characterization of full length of a novel zebrafish gene Zsrg abundantly expressed in the germline stem cells. *Biochem. Biophys. Res. Commun.* **329**, 632-7.

Makarova, K., Aravind, L. and and Koonin, E. (2002). SWIM, a novel Zn-chelating domain present in bacteria, archaea and eukaryotes. *Trends Biochem. Sci.* **27**, 384-6.

Marin, I. and Baker, B. S. (1998). The evolutionary dynamics of sex determination. *Science* **281**, 1990-4.

Matsuda, M., Matsuda, C., Hamaguchi, S. and Sakaizumi, M. (1998). Identification of the sex chromosomes of medaka, *Oryzias latipes*, by fluorescence *in situ* hybridization. *Cytogenet. Cell Genet.* **82**, 257-62.

Matsuda, M., Sotoyama, S., Hamaguchi, S. and Sakaizumi, M. (1999). Male-specific recombination frequency in the sex chromosomes of the medaka, *Oryzias latipes*. *Genet. Res.* **73**, 225-31.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C. E., Shibata, N., Asakawa, S., Shimizu, N. *et al.* (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417**, 559-63.

McClintock, B. (1956a). Controlling elements and the gene. *Cold Spring Harbor Symp. Quant. Biol.* **21**, 197-216.

McClintock, B. (1956b). Intranuclear systems controlling gene action mutation. *Brookhaven Symp. Biol.* **8**, 58-74.

Meyer, M. K. and Schartl, M. (2002). Xiphophorus mayae, a new species of swordtail from Guatemala (Teleostei: Poeciliidae). *Ichthyol. Explor. Freshwaters* **13**, 59-66.

Morizot, D. C., Slaugenhaupt, S. A., Kallman, K. D. and Chakravarti, A. (1991). Genetic linkage map of fishes of the genus *Xiphophorus* (Teleostei: Poeciliidae). *Genetics* **127**, 399-410.

Nairn, R. S., Kazianis, S., Mcentire, B. B., Della Coletta, L., Walter, R. B. and *et al.* (1996). A CDKN2-like polymorphism in Xiphophorus LG V is associated with UV-B-induced melanoma formation in platyfish-swordtail hybrids. *Proc. Natl. Acad. Sci.* **93:**, 13042-13047.

Nanda, I., Volff, J. N., Weis, S., K. rting, C., Froschauer, A., Schmid, M. and Schartl, M. (2000a). Amplification of a long terminal repeat-like element on the Y chromosome of the platyfish, Xiphophorus maculatus. *Chromosoma* **109**, 173-80.

Nanda, I., Zend-Ajusch, E., Shan, Z., Grutzner, F., Schartl, M., Burt, D. W., Koehler, M., Fowler, V. M., Goodwin, G., Schneider, W. J. *et al.* (2000b). Conserved synteny between the chicken Z sex chromosome and human chromosome 9 includes the male regulatory gene DMRT1: a comparative (re)view on avian sex determination. *Cytogenet. Cell Genet.* **89**, 67-78.

Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A. *et al.* (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc. Natl. Acad. Sci. USA* **99**, 11778-83.

Nanda, I., Hornung, U., Kondo, M., Schmid, M. and Schartl, M. (2003). Common spontaneous sex-reversed XX males of the medaka *Oryzias latipes*. *Genetics* **163**, 245-51.

Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R. *et al.* (2002). A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics* **154**, 1773-84.

Niethammer, M., Valtschanoff, J. G., Kapoor, T. M., Allison, D. W., Weinberg, T. M., Craig, A. M. and Sheng, M. (1998). CRIPT, a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90. *Neuron* **20**, 693-707.

Ogawa, A., Murata, K. and Mizuno, S. (1998). The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. *PNAS* **95**, 4415-18.

Ohno, S. (1967). Sex chromosomes and sex-linked genes. Berlin: Spring.

Okada, S., Sone, T., Fujisawa, M., S. Nakayama, Takenaka, M., Ishizaki, K., Kono, K., Shimizu-Ueda, T., Hanajiri, T., Yamoto, K. T. *et al.* (2001). The Y chromosome in the liverwort Marchantia polymorpha has accumulated unique repeat sequences harboring a male-specific gene. *Proc. Natl. Acad. Sci. USA* **98**, 9454-59.

Passafaro, M., Sala, C., Niethammer, M. and Sheng, M. (1999). Microtubule binding by CRIPT and its potential role in the synaptic clustering of PSD-95. *Nat. Neurosci.* **2**, 1063-9.

Peichel, C. L., Ross, J. A., Matson, C. K., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Mori, S., Schluter, D. and Kingsley, D. M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr. Biol.* **14**, 1416-24.

Peters, G. (1964). Vergleichende Untersuchungen an drei Subspecies von Xiphophorus helleri Heckel (Pisces). *Zeitschrift fuer Zoologische Systematik und Evolutionsforschung* **2**, 185-271.

Poulter, R. T. M. and Butler, M. (1998). A retrotransposon family from the pufferfish (fugu) *Fugu rubripes*. *Gene* **215**, 241-249.

Poulter, R. T. M., Butler, M. and Ormandy, J. (1999). A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the *CR1* family of non-LTR retrotransposons. *Gene* **227**, 169-79.

Poulter, R. T. M., Goodwin, T., J. D. and Butler, M. I. (2003). Vertebrate helentrons and other novel Helitrons. *Gene* **313**, 201-12.

Raymond, C. S., Kettlewell, J., Hirsch, B., Bardwell, V. and Zarkower, D. (1999). Expression of Dmrt1 in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Dev. Biol.* **215**, 208-220.

Rens, W., Gruetzner, F., O'Brien, P. C. M., Fairclough, H., Graves, J. A. M. and Ferguson-Smith, A. M. (2004). From The Cover: Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. *PNAS* **101**, 16257-61.

Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., Riganelli, D., Zanaria, E., Messali, S., Cainarca, S. *et al.* (2001). The tripartite motif family identifies cell compartments. *EMBO J.* **20**, 2140-2151.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C. and Markham, A. F. (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887-90.

Roushdy, J., Michel, J., Petry, H., Anders, A. and Anders, F. (1999). Paragenetic suppressors of suppressor genes - a new class of oncodeterminants. *J. Cancer Res. Clin. Oncol.* **125**, 123-33.

Rozen, S., Skaletsky, H., Marszalek, J., Minx, P., Cordum, H., Waterston, R., Wilson, R. and Page, D. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873-6.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular Cloning. A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sambrook, J. and Russell, D. W. (2001). Molecular Cloning. A laboratory manual. 3nd edition. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-67.

Schartl, M. (1988). A sex chromosomal restriction-fragment-length marker linked to melanoma- determining *Tu* loci in *Xiphophorus*. *Genetics* **119**, 679-85.

Schartl, A. and Schartl, M. (1990). Genes and Cancer: Molecular biology of the Melanoma Oncogene of *Xiphophorus*. *BFE.* **7**, 461-6.

Schartl, M. (1990). Homology of melanoma-inducing loci in the genus Xiphophorus. *Genetics* **126**, 1083-91.

Schartl, M. (1995). Platyfish and swordtails: a genetic system for the analysis of molecular mechanisms in tumor formation. *Trends Genet.* **11**, 185-9.

Schartl, M. and Wellbrock, C. (1998). Ploygenetic inheritance of melanoma in *Xiphophorus*. In *Human polygenic diseases*, (ed. T. A. Dragani), pp. 167-87. Chur, Switzerland: Harwood Academic Publishers.

Schartl, M., Hornung, U., Gutbrod, H., N., V. J. and Wittbrodt, J. (1999). Melanoma loss-of-function mutants in Xiphophorus caused by Xmrk-oncogene deletion and gene disruption by a transposable element. *Genetics* **153**, 1385-94.

Schartl, M. (2004a). A comparative view on sex determination in medaka. *Mech. Dev.* **121**, 639-45.

Schartl, M. (2004b). Sex chromosome evolution in non-mammalian vertebrates. *Curr. Opin. Genet. Dev.* **14**, 634-41.

Schreibman, M. P., Scharl, M., Kallman, K. D. and Magluilo-Cepriano, L. (1994). Molecular approaches to study the genetic regulation of the fish reproductive system. In *Perspective in comparative endocrinology*, (ed. R. E. Davey R. E. Peter and S. S. Tobe), pp. 343-51. Ottawa: Natl. Res. Council of Canada.

Schultheis, C. (2003). Molekulare Analyse der Geschlechtsbestimmungsregion der Gonosomen des platyfisches Xiphophorus maculatus. Diploma-thesis, Julius-Maximilians-Universitaet Wuerzburg.

Schwartz, D. C. and Cantor, C. R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67-75.

Selz, Y. (2003). Funktionelle Untersuchungen zur Rolle der Melanocortin-Rezeptoren in der Melanombildung bei dem Fisch Xiphophorus. Diploma-thesis, Julius-Maximilians-Universitaet Wuezburg.

Shetty, S., Kirby, P., Zarkower, D. and Graves, J. A. (2002). DMRT1 in a ratite bird: evidence for a role in sex determination and discovery of a putative regulatory element. *Cytogenet. Genome Res.* **99**, 245-51.

Shinomiya, A., Tanaka, M., Kobayashi, T., Nagahama, Y. and Hamaguchi, S. (2000). The vasa-like gene, olvas, identifies the migration path of primordial germ cells during embryonic body formation stage in the medaka, Oryzias latipes. *Dev. Growth Differ.* **42**, 317-26.

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A. M., Lovell-Badge, R. and Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240-4.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P., Cordum, H., Hillier, L., Brown, L., Repping, S., Pyntikova, T., Ali, J., Bieri, T. *et al.* (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825-37.

Slanchev, K., Stebler, J., Cueva-Mendez, G. d. l. and Raz, E. (2005). Development without germ cells: The role of the germ line in zebrafish sex differentiation. *PNAS*, 10.1073/pnas.0407475102.

Slavotinek, A., Schwarz, C., Getty, J., Stecko, O., Goodman, F. and and Kingston, H. (1999). Two cases with interstitial deletions of chromosome 2 and sex reversal in one. *Am J. Med. Genet.* **86**, 75-81.

Sola, L., Monaco, P. J. and Rasch, E. M. (1990). C-bands, Ag-NOR polymorphisms and sex chromosomes in three populations of Poecilia latipinna. *Cytogenet. Cell Genet.* **53**, 148-54.

Southern, E. M. (1975). Dection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-17.

Southern, E. M., Anand, R., Brown, W. R. and Fletcher, D. S. (1987). A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Res.* **15**, 5925-43.

Steinemann, M. and Steinemann, S. (1998). Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of Drosophila miranda a model for sex chromosome evolution. *Genetica* **102**, 409-20.

Steinemann, S. and Steinemann, M. (2001). Biased distribution of repetitive elements: a landmark for neo-Y chromosome evolution in Drosophila miranda. *Cytogenet. Cell Genet.* **93**, 228-33.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. and Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell* **33**, 25-35.

Thuring, R. W., Sanders, J. P. and Borst, P. (1975). A freeze-squeeze method for recovering long DNA from agarose gels. *Anal. Biochem.* **66**, 213-20.

Tracey, K. J. and Cerami, A. (1994). TUMOR NECROSIS FACTOR: A pleiotropic cytokine and therapeutic target. *Annu. Rev. Immunol.* **45**, 491-503.

Traut, W. and Winking, H. (2001). Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Res.* **9**, 659-72.

Vaisse, C., Clement, K., Guy-Grand, B. and Froguel, P. (1998). A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat. Genet.* **20**, 113-14.

Vallender, E. J. and Lahn, B. T. (2004). How mammalian sex chromosomes acquired their peculiar gene content. *Bioessays* **26**, 159-69.

Veith, A. M., Froschauer, A., Koerting, C., Nanda, I., Hanel, R., Schmid, M., Schartl, M. and Volff, J. N. (2003). Cloning of the dmrt1 gene of Xiphophorus maculatus: dmY/dmrt1Y is not the master sex-determining gene in the platyfish. *Gene* **317**, 59-66.

Volff, J. N., Koerting, C., Sweeney, K. and Schartl, M. (1999). The non-LTR retrotransposon Rex3 from the fish Xiphophorus is widespread among teleosts. *Mol. Biol. Evol.* **16**, 1427-38.

Volff, J. N., Koerting, C. and Schartl, M. (2000). Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. *Mol. Biol. Evol.* **17**, 1673-84.

Volff, J. N. and Schartl, M. (2001). Variability of genetic sex determination in poeciliid fishes. *Genetica* **111**, 101-110.

Volff, J. N., Koerting, C., Altschmied, J., Duschl, J., Sweeney, K., Wichert, K., Froschauer, A. and Schartl, M. (2001a). Jule from the fish Xiphophorus is the first complete vertebrate Ty3/Gypsy retrotransposon from the Mag family. *Mol. Biol. Evol.* **18**, 101-11.

Volff, J. N., Koerting, C., Froschauer, A., Sweeney, K. and Schartl, M. (2001b). Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J. Mol. Evol.* **52**, 351-60.

Volff, J. N., Koerting, C., Meyer, A. and Schartl, M. (2001c). Evolution and discontinuous distribution of rex3 retrotransposons in fish. *Mol. Biol. Evol.* **18**, 427-31.

Volff, J. N. and Schartl, M. (2002). Sex determination and sex chromosome evolution in the medaka, *Oryzias latipes*, and the platyfish, *Xiphophorus maculatus*. *Cytogenet. Genome Res.* **99**, 170-77.

Volff, J. N. and Schartl, M. (2003). Evolution of signal transduction by gene and genome duplication in fish. *J. Struct. Funct. Genomics* **3**, 139-50.

Volff, J. N. Kondo, M. and Schartl, M. (2003). Medaka dmY/dmrt1Y is not the universal primary sex-determining gene in fish. *Trends Genet.* **19**, 196-9.

Volff, J. N., Froschauer, A., Korting, C., Bernhardt, W., Nanda, I., Schultheis, C., Schmid, M. and Schartl, M. (2003a). Genome fluidity in the fish *Xiphophorus*-steps toward a great future. In *Aquatic Genomics - Steps Toward a Great Future*, (ed. N. Shimizu T. Aoki F. Hirono and F. Takashima), pp. 185-194. Tokyo: Springer.

Volff, J. N., Koerting, C., Froschauer, A., Zhou, Q., Wilde, B., Schultheis, C., Selz, Y., Sweeney, K., Duschl, J., Wichert, K. *et al.* (2003b). The Xmrk oncogene can escape non-functionalization in a highly unstable subtelomeric region of the genome of the fish Xiphophorus. *Genomics* **82**, 470-479.

Volff, J. N., Zarkower, D., Bardwell, V. J. and Schartl, M. (2003c). Evolutionary dynamics of the DM domain gene family in metazoans. *J. Mol. Evol.* **57**, 241-49.

Volff, J. N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity* **94**, 280-94.

Wada, H., Shimada, A., Fukamachi, S., Naruse, K. and Shima, A. (1998). Sex-linked inheritance of the lf locus in the medaka fish (Oryzias latipes). *Zool. Sci.* **15**, 123-26.

Wakamatsu, Y. (1981). Establishment of a cell line from the platyfish-swordtail hybrid melanoma. *Cancer Res.* **41**, 679-80.

Walter, R. B., Rains, J. D., Russell, J. E., Guerra, T. M., Daniels, C., Johnston, D. A., Kumar, J., Wheeler, A., Kelnar, K., Khanolkar, V. A. *et al.* (2004). A Microsatellite Genetic Linkage Map for Xiphophorus. *Genetics* **168**, 363-72.

Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H. and Steitz, T. A. (1996). Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. *Biochemistry* **35**, 8110- 8119.

Weis, S. and Schartl, M. (1998). The macromelanophore locus and the melanoma oncogene Xmrk are separate genetic entities in the genome of Xiphophorus. *Genetics* **149**, 1909-20.

Wilkins, A. (1995). Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* **17**, 71-7.

Winkler, C., Hornung, U., Kondo, M., Neuner, C., Duschl, J., Shima, A. and Schartl, M. (2004). Developmentally regulated and non-sex-specific expression of autosomal dmrt genes in embryos of the Medaka fish (*Oryzias latipes*). *Mech. Dev.* **121**, 997-1005.

Wittbrodt, J., Adam, D., Malitschek, B., Maeueler, W., Raulf, F., Telling, A., Robertson, S. M. and Schartl, M. (1989). Novel putative receptor tyrosine kinase encoded by the melanoma- inducing Tu locus in Xiphophorus. *Nature* **341**, 415-21.

Wittbrodt, J. S. A. and Schartl, M. (2002). Medaka - a model organism from the far East. *Nat. Rev. Genet.* **3**, 53-64.

Woram, R. A., Gharbi, K., Sakamoto, T., Hoyheim, B., Holm, L. E., Naish, K., McGowan, C., Ferguson, M. M., Phillips, R. B., Stein, J. *et al.* (2003). Comparative Genome Analysis of the Primary Sex-Determining Locus in Salmonid Fishes. *Genome Res.* **13**, 272-80.

Yamamoto. (1975). Medaka (*Killfish*): Biology and strains. Tokyo: Keigaku Publishing Co.

Zarkower, D. (2001). Establishing sexual dimorphism: conservation amidst diversity? *Nature Review Genetics* **2**, 175-85.

Zhou, M., Mao, C., Rodriguez, A. C., Kiefer, J. R., Kucera, R. B. and Beese, L. S. (1998). Crystallization and preliminary diffraction analysis of a hyperthermostable DNA polymerase from a Thermococcus archaeon. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 994- 995.

Ziegler, I. (2003). The pteridine pathway in zebrafish: regulation and specification during the determination of neural crest cell-fate. *Pigment Cell Res.* **16**, 172-82.

PUBLICATIONS

Volff J.-N., C. Körting, A. Froschauer, Q. Zhou, B. Wilde, C. Schultheis, Y. Selz, K. Sweeney, J. Duschl, K. Wichert, J. Altschmied and M. Schartl (2003) The *Xmrk* oncogene can escape non-functionalization in a highly unstable subtelomeric region of the genome of the fish *Xiphophorus.* **Genomics 82**: 470-479 (cover page).

Froschauer A.,Y. Selz, C. Schultheis, C. Körting, Q. Zhou, W. Bernhardt, R. Hanel, M. Schartl and J.-N. Volff. Multicopy melanocortin receptor gene at the melanoma locus on the sex chromosomes of the platyfish *Xiphophorus maculatus*. Manuscript in preparation.

Zhou Q., A. Froschauer, C. Schultheis, M. Schartl and J.-N. Volff. Multicopy *DNA polymerase type B* genes in fish genomes. Manuscript in preparation.

SCIENTIFIC COMMUNICATIONS

Zhou Q., A. Froschauer, C. Körting, C. Schultheis, M. Vucic, A.-M. Veith, M. Mösl, I. Nanda, M, Schmid, M. Schartl and J.-N. Volff (2003) Molecular analysis of the sex-determining region of the Y chromosome in the platyfish *Xiphophorus maculatus*. Jahrestagung der Deutschen Gesellschaft für Genetik, Kassel, Germany (poster performed by QZ). **Genes, Chromosomes, Genomes 9**: 61.

Zhou Q., A. Froschauer, C. Körting, C. Schultheis, M. Wenning, P. Bienert, M. Mösl, M. Schartl and J.-N. Volff (2003) Construction and initial analysis of a bacterial artificial chromosome (BAC) contig of the sex determination region of the Y chromosome of the platyfish *Xiphophorus maculatus*. First european conference of poeciliid biologists, Zürich, Switzerland (poster performed by QZ).

Zhou Q, A.-M. Veith, C. Körting, C. Schultheis, Y. Selz, M. Vucic, R. Hanel, P. Bienert, M. Naciri, A. Froschauer und J.-N. Volff (2003): Positionelle Klonierung des Gens für Geschlechtsbestimmung im Platyfisch *Xiphophorus*. *2. BioFuture Meeting, Berlin, Germany, January 2003* (poster performed by JNV).

Zhou Q, C. Schultheis, A. Froschauer, A.-M. Veith, C. Körting, Y. Selz, J. Brandt, M. Wenning. A. Dettai and J.-N. Volff (2005): Positional cloning of the sex-determining gene in the platyfish *Xiphophorus*. *4. Biofuture Meeting, Berlin, Germany, January 2005* (poster performed by JNV).

A. Froschauer, C. Körting, A.-M. Veith, W. Bernhardt, Q. Zhou, R. Hanel und J.-N. Volff (2002): Positionelle Klonierung des Gens für Geschlechtsbestimmung im Platyfish *Xiphophorus. 1. BioFuture Meeting, Berlin, Germany, 28.-29.01.2002* (poster performed by JNV).

A. Froschauer, A.-M. Veith, C. Schultheis, Q. Zhou, M.Wenning, Y. Selz, J. Brandt, C. Körting and J.-N. Volff (2004): Positional cloning of the sex determining gene in the platyfish *Xiphophorus*. *3. BioFuture Meeting, Berlin, Germany, January 2004* (poster performed by JNV).

Selz Y., A. Froschauer, Q. Zhou, C. Körting, R. Hanel, W. Bernhardt, C. Schultheis, M. Schartl and J.-N. Volff (2002) Serial duplications of a melanocortin receptor gene at the melanoma locus of the platyfish *Xiphophorus maculatus*. 6th signal transduction society meeting “signal transduction – receptors, mediators and genes”, Weimar, Germany (poster performed by YS).

Selz Y., A. Froschauer, Q. Zhou, C. Körting, C. Schultheis, R. Hanel, W. Bernhardt, M. Schartl and J.-N. Volff (2003) Multicopy melanocortin receptor gene at the melanoma locus in the platyfish *Xiphophorus maculatus*. 12th international AEK cancer congress, Würzburg, Germany (poster performed by YS).

Froschauer A., C. Schultheis, Q. Zhou, C. Körting, M. Vucic, Y. Selz, R. Hanel, M. Naciri, A.-M. Veith, I. Nanda, M. Schmid, M. Schartl and J.-N. Volff (2003) Genomic instability in the *Xmrk* oncogene region of the melanoma fish model *Xiphophorus*. 12th International AEK cancer congress, Würzburg, Germany (poster performed by AF).

Schultheis C., A. Froschauer, C. Körting, Q. Zhou, M. Vucic, A.-M. Veith, M. Schartl and J.-N. Volff (2003) Conserved synteny between the sex-determining region of the platyfish *Xiphophorus maculatus* and a region of human chromosome 2 involved in sex determination. Jahrestagung der Deutschen Gesellschaft für Genetik, Kassel, Germany (poster performed by CS). **Genes, Chromosomes, Genomes 9**: 57.

Selz Y., A. Froschauer, Q. Zhou, C. Körting, R. Hanel, W. Bernhardt, C. Schultheis, M. Schartl and J.-N. Volff (2003) Serial duplications of the melanocortin receptor gene *mc4r* in the platyfish *Xiphophorus maculatus* with regard to melanoma formation and sexual maturity. Jahrestagung der Deutschen Gesellschaft für Genetik, Kassel, Germany (poster performed by YS). **Genes, Chromosomes, Genomes 9**: 60.

Schultheis C., A.-M. Veith, A. Froschauer, C. Körting, Q. Zhou, M. Wenning, K. Schmitter, M. Schartl and J.-N. Volff (2003) Variation at the top: the search for novel sex determination genes with the platyfish Xiphophorus maculatus as a model organism. First European conference of poeciliid biologists, Zürich, Switzerland (poster performed by CS).

Lebenslauf

Persönliche Angaben:

Name: Qingchun Zhou
Geburtsdatum: 22. 04. 1966
Geburtsort: Yichang, Provinz Hubei, V. R. China
Familienstand: verheiratet
Staatsangehörigkeit: chinesisch

Schulbildung:

Sept.1973 - Juli 1978 Grundschule in Yichang, Provinz Hubei, V. R. China
Sept.1978 - Juli 1984 Mittelschule in Yichang, Provinz Hubei, V. R. China

Hochschulstudium:

Sept.1984 - Juli 1988 Studium an der Huazhong Pädagogischen Universität, Wuhan
Studienfach: Biologie
Abschluß: Bachelor of science

Sept.1992 - Juli 1995 Studium an der Huazhong Pädagogischen Universität, Wuhan
Studienfach: Entomologie
Abschluß: Master of science

Thema: „ The establishment of a new bioassay for pesticide to Evaluate insecticidal toxicity to *Heliothis Armigera* cell line “

Berufstätigkeit:

Juli 1988 - Dez.1993 Assistent an der Huazhong Pädagogischen-Universität, Wuhan
Juli 1993 - Dez.1998 Dozent an der Huazhong Pädagogischen-Universität, Wuhan
Seit Dez.1998 ausserordentlicher Professor an der Huazhong Pädagogischen-Universität, Wuhan

Promotionsstudium:

Sept. 2001 – Juni 2005 Promotionsstudium am Lehrstuhl für Physiologishe Chemie, Universität Würzburg, Deutschland

Thema: „Molekulare Analyse der geschlechtsbestimmenden Region des Y Chromosoms von *Xiphophorus maculatus*”

Erklärung

Die vorliegende Arbeit wurde von mir selbständig und nur unter der Verwendung der angegebenen Quellen und Hilfsmittel angefertigt.

Diese Dissertation hat weder in gleicher noch in ähnlicher Form in einem anderen Prüfungsverfahren vorgelegen.

Ich habe früher außer den mit dem Zulassungsgesuch urkundlich vorgelegten Graden keine weiteren akademischen Grade erworben oder zu erwerben versucht.

Würzburg, March 2005 Qingchun Zhou