The genomic Make-Up of a Hybrid Species – Analysis of the Invasive *Cottus* Lineage (Pisces, Teleostei) in the River Rhine system

Inaugural – Dissertation

Zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Kathryn Stemshorn
aus Köln

Köln, 2007

Berichterstatter:	**Prof. Dr. Diethard Tautz**
	Prof. Dr. Hartmut Arndt
Tag der letzten mündlichen Prüfung:	**11. Juni 2007**

List of Contents

Danksagung

First of all I want to thank my supervisor Prof. Dr. Diethard Tautz for the opportunity to enter the world of evolutionary genetics and furthermore for the great working and social atmosphere in his research group. I know that the latter should not be taken for granted!

Prof. Dr. Hartmut Arndt has kindly agreed to evaluate this thesis.

I thank Arne Nolte, who established the *Cottus* project and who put all the *Cottus* wisdom into me that he possibly could. He also introduced me to the molecular laboratory work in general and always was and is a great support in any respect.

Furthermore, I thank J. Freyhof, the discoverer of ‚prickled' sculpins in the River Sieg.

The sculpin project was possible only with benevolent support and permission from Dr. A. Mellin, T. Heilbronner, W. Fettweis, C. Bode, L. Jörgensen, L. Kroll, Dr. C. Köhler, W. Fricke, Dr. H. Arzbach and M. Kämmereit and with material or logistic support from Dr. F. Volckaert, Dr. B. Hänfling, Dr. D. Neely, Dr. L. Bervoets, Dr. G. Knaepkens, Dr. E. Winter, I. Steinmann, Dr. E. Korte, U. Weibel, C. Dümpelmann, Dr. W. Dönni, C. von Landwüst and Dr. A. Waterstraat as well as numerous local fischermen.

I want to thank all current and former members of the Tautz lab for the wonderful working environment, helpful discussions, open ears, chocolate, cake and champaign, and many good times in the lab and outside. Special thanks to Meike Thomas and Ruth Rottscheidt for shopping-lunch breaks, girl's talk and all of the other important things. Concerning all computer-related issues I want to thank Till Bayer and Chriz Voolstra who rescued me several times out of user-provoked crevasses. Till Bayer furthermore wrote some extremly helpful scripts which saved me lots of time handling and analyzing raw data. Many thanks to Susanne Kipp, Birgit Schmitz and Patrick Kück who kept me from dreaming of pipetting 96-well plates and thus probably dreamed of it themselves.

Finally I want to thank my parents, Renate and Folker Stemshorn, my siblings Anne and Jan and all of my friends for their support and encouragement and most of all for providing me with a safe runway from which I can take off to all kinds of scientific and other adventures and to which I always look forward to return to.

Zusammenfassung

Innerhalb der letzen Jahre wurde eine neue invasive Groppenlinie (*Cottus* Spezieskomplex) untersucht, die sich momentan im Unterlauf des Rheins ausbreitet. Mit Hilfe von molekularen Analysen konnte gezeigt werden, dass diese Linie durch Hybridisierung zwischen *Cottus perifretum* aus der Schelde und *Cottus rhenanus* aus dem Niederrheinsystem entstanden ist. Die Entstehung dieser Hybridlinie korreliert mit Anpassungen an einen neuen Lebensraum, die die Ausbreitung innerhalb von Flußhabitaten ermöglicht haben, die zuvor nicht von Groppen besiedelt waren. Daher stellt sich die Frage, ob das Hybridisierungsereignis die Invasion und die Anpassungen an solch eine neue Umgebung vereinfacht hat. Um mit der Beantwortung dieser Frage zu beginnen, sollte festgestellt werden, wie groß der Anteil der beiden Elternarten am Hybridgenom ist, und welche elterlichen Chromosomenfragmente in den Hybriden fixiert wurden. Um die Herkunft der unterschiedlichen Chromosomenstücke kartieren zu können, mussten zunächst einmal genomische Resourcen entwickelt werden. Als Basis wurde eine auf Mikrosatelliten basierende genetische Karte erstellt. Diese wurde mit physikalischen Karten von sequenzierten Fischgenomen verglichen und es konnte ein hoher Grad an konservierter Syntenie zwischen *Cottus* und *Tetraodon nigroviridis* und zwischen *Cottus* und *Gasterosteus aculeatus* festgestellt werden. Diese Genome konnten dann in der weiteren Analyse des Groppengenoms als Referenz benutzt werden. Weiterhin wurde eine Reihe von Markern entwickelt, die im Hinblick auf den Ursprung verschiedener Chromosomenfragmente in der Hybridlinie informativ sind. Mit Hilfe dieser Mittel war es möglich, das Hybridgenom zu kartieren und den jeweiligen Beitrag der beiden Elternarten zu bestimmen. Dabei wurden 25 genomische Fragmente entdeckt, die bezüglich ihrer elterlichen Herkunft fixiert sind. Diese Fixierung deutet darauf hin, dass diese genomischen Regionen Gene enthalten, die für die neuen Adaptationen in der Hybridspezies relevant sind.

Abstract

In the past years a new invasive lineage of sculpins (*Cottus* species complex) has been studied that is currently expanding in the Lower River Rhine. Molecular analysis showed that this lineage has originated through hybridization of *Cottus perifretum* from the River Scheldt and *Cottus rhenanus* from the Lower River Rhine system. The emergence of the hybrid lineage is correlated with new habitat adaptations that allow the expansion along river habitats that have previously not been used by *Cottus*. Thus the question arises, if the hybridization event facilitated the invasion of and the adaptation to such a new environment. To start tackling this question an estimate is required how much each of the parental species contributed to the hybrid genome and which chromosomal fragments became fixed. Several genomic resources had to be developed in order to map the ancestries of chromosomal fragments in the hybrid genome. As a basic genomic resource for *Cottus* a genetic map based on already established microsatellite markers was created. This map was compared with the physical maps of sequenced fish genomes and a high degree of conserved synteny between *Cottus* and *Tetraodon nigroviridis* and between *Cottus* and *Gasterosteus aculeatus* could be detected. These model fish genomes could then be used as a reference in the further analysis of the *Cottus* genome. Finally, a set of ancestry-informative markers was developed in order to determine the ancestries of chromosomal fragments in the hybrid lineage. These tools allowed to map the hybrid genome and to assess the contribution of each parental species to the hybrid lineage. 25 genomic fragments could be identified that were fixed for material from only one parental species and thus might harbor genes that are relevant for the specific adaptations in the hybrid species.

Declaration

The design of the whole project was developed together with Diethard Tautz. I conducted the major part of the practical laboratory work as well as the data analysis. In the different parts of this thesis I profited from the experience and previous work conducted on *Cottus* by a few colleagues whose input and contribution I acknowledge below.

Genetic map

Arne Nolte provided me with the mapping families for the preliminary genetic map. Furthermore he and Claudia Englbrecht developed the microsatellite markers, which are included in the genetic map. Arne Nolte introduced me into the laboratory methods and data analysis and conducted the sampling of prespawning adults for further mapping families. He also taught me the basics of raising and maintaining sculpins in the lab.

Genomic library construction

Arne Nolte provided me with the protocol for the library construction and introduced me into the basic techniques of cloning.

Development of ancestry informative markers

Tissue samples for DNA extractions for the parental DNA pools were provided by Arne Nolte. Furthermore, some of the microsatellite loci established by Arne Nolte and Claudia Englbrecht were employed for marker development.

Analysis of the hybrid lineage and an outgroup species

Tissue and DNA samples for the hybrid DNA pools were provided by Arne Nolte. Tissue samples from the outgroup species were provided by David Neely.

1 General Introduction

1.1 Hybridization: a neglected mechanism for animal speciation

Among zoologists hybridization is usually considered as a process opposing speciation. This paradigm is based on the observation, that hybrids between two species are often inviable or at least less fit and furthermore on the definition of species according to the biological species concept as reproductively isolated entities (Mayr et al. 1963). This definition does not allow hybridization to act as a creative evolutionary force. Considering however, that around 10 % of animal and 25 % of plant species are known to hybridize with at least one other species (Mallet 2005) the potential of this mechanism for speciation should not be neglected. Among plants hybridization has long been considered as a process, which can lead to the formation of new species and only recently examples of hybrid speciation are also emerging in the animal kingdom. The cyprinid fish *Gila seminude*, the 'swordtail' *Xiphophorus clemenciae* and the Colombian butterfly *Heliconius heurippa* all show signs of hybrid origin (DeMarais et al. 1992, Mavarez et al. 2006, Meyer et al. 2006). Hybrid species have furthermore been detected in the butterfly genus *Lycaeides* and among *Rhagoletis* fruitflies (Gompert et al. 2005, Schwarz et al. 2005). Moreover, Seehausen (2004) proposed that hybridization was one of the triggers for the explosive radiation in Lake Victoria cichlids. Just looking at current literature demonstrates that hybridization is gaining more attention as a mechanism that can lead to evolutionary novelties (Bullini 1994, Dowling et al. 1997, Barton 2001, Seehausen 2004, Mallet 2005, Mallet 2007). In plants there are already some well studied cases of hybrid speciation where even the genetic basis for the success of these hybrids is known (Rieseberg 2000). Such detailed analysis of hybrid speciation is only starting now in the animal kingdom but they will help to gain insights into the process of speciation and the creation of organismal diversity.

1.2 The evolutionary processes of hybrid speciation

Most cases of hybrid speciation studied so far concern polyploid hybridization. This hybridization mechanism seems to be more common in plants than in animals (Mallet 2007) and usually leads to a direct genetic isolation of the newly arisen hybrid population. Diploid or homoploid hybrid speciation however, the subject of this study, seems to be an unlikely event and harder to explain since the hybrid lineage has to establish itself in the face of ongoing gene flow with the parental species. The only well studied examples of homoploid hybrid speciation are the sunflowers species *Helianthus anomalus*, *Helianthus deserticola* and *Helianthus paradoxus* which are hybrids between *Helianthus annuus* and *Helianthus petiolaris*. All of these hybrid species exhibit favorably interacting (epistatic) gene combinations making them superior to the parents in extreme habitats (Rieseberg et al. 1996). This phenomenon has been described as transgressive segregation and it explains one possibility how a hybrid lineage can become established. The availability of an unoccupied habitat or ecological niche seems to be an important prerequisite for the establishment of a hybrid lineage such that direct competition with pure parental genotypes, which have

been evolutionary optimized for a given habitat, can be circumvented (Burke & Arnold 2001). Mallet (2007) described this situation with adaptive landscapes, where some adaptive peaks are occupied by the parental species and hybrids are found as 'hopeful monsters' mostly in the valleys and far from phenotypic optima. Some of these hybrids, however, might gain fitness or even extreme phenotypes due to their high genetic variance, allowing them to reach other adaptive peaks if these are available. Thus hybrid speciation would occur most easily through founder events of hybrid genotypes, that can potentially occupy a novel habitat which would then allow them to become ecologically or even geographically isolated from the parental species (Burke & Arnold 2001).

Another factor that aids homoploid hybrid speciation are chromosomal rearrangements, especially inversions (Livingston & Rieseberg 2003). Rearranged chromosomal fragments are protected from gene flow due to their lack of recombination. If such rearranged regions carry advantageous traits they could be fixed quickly in a hybrid population. Buerkle et al. (2000) modeled recombinational speciation events in which parental rearrangements were sorted in the hybrids, eventually leading to fit hybrid genotypes.

1.3 European sculpins (Pisces: Cottidae)

Sculpins (Scorpaeniformes, Cottidae, *Cottus*), are small benthic freshwater fishes usually inhabiting small, cold streams. They are distributed all over Europe, except for southern Spain, southern Italy, the northern part of Great Britain and Ireland. Further species of this genus are found in the whole northern hemisphere, but most species occur in North America, Siberia and Asia.

Since sculpins have never been of commercial value they were probably never artificially stocked leaving their distribution unaffected by humans. This is one of the reasons turning *Cottus* into a good model organism for studies of biogeography and natural patterns of differentiation (Hänfling & Brandl 1998, Englbrecht et al. 2000).

Phylogeographic analysis of European *Cottus* were conducted by Englbrecht et al. (2000) Schreiber et al. (1998), Hänfling et al. (2002) and Volckaert et al. (2002). Like several other freshwater species in Europe, sculpins retreated to glacial refugia during the last ice age. Following the ice age, recolonization started from the southern part of the Danube (Englbrecht et al. 2000). Englbrecht et al. (2000) could show, that several distinct haplotype lineages can be detected based on mitochondrial D-loop sequences: a western group with populations in the Seine, the Adour and the Lower Rhine which has been described recently as *Cottus perifretum* (Freyhof et al. 2005), an eastern group with populations in the upper and lower Danube, the Main and the Elbe (*Cottus gobio*), and a Lower Rhine group with populations in tributaries of the Middle and the Lower Rhine which has now been named *Cottus rhenanus* (Freyhof et al. 2005). The oldest phylogenetic lineage is the eastern group, which seems to be ancestral to the other lineages (Englbrecht et al. 2000, Kontula & Väinölä 2003). The oldest lineages probably split around 3 million years ago whereas *Cottus perifretum* and *Cottus rhenanus* diverged about 1 million years ago (Englbrecht et al. 2000, Hänfling et al. 2002).

An overlap between the well-differentiated *Cottus* lineages was noted by Englbrecht et al. (2000) in the River Rhine system. Different evolutionary haplotype lineages were detected, suggesting secondary contact between the divergent ancestral lineages and the possibility for hybridization.

1.4 A hybrid invasion of the Lower River Rhine

As mentioned above, sculpins are usually confined to well oxygenated cold headwater regions. Less than 20 years ago however, sculpins were discovered in the main channel of the Lower River Rhine (Schleuter 1991, Lelek & Köhler 1993), which presents a typical summer warm potamal habitat. At the same time sculpins were reported to be common in the Lower Rhine of the Netherlands (Cazemir 1988, van den Brink et al. 1990). Fish surveys indicated that sculpins were only found in few places before 1980 (De Nie 1997) whereas now they were abundant preferentially in large rivers, artificial canals and the Ijsselmeer. In 1992 fish abundance surveys in the Sieg detected sculpins with intense skin prickling which were found to expand upriver within the next ten years. The main channel of the Sieg had also not been inhabited by *Cottus* before even though *Cottus rhenanus*, the native Lower River Rhine species, has always been found in the tributaries to the Sieg. Molecular analysis based on mitochondrial haplotypes and diagnostic single nucleotide polymorphisms suggests that the invasive sculpins arose through hybridization between the western sculpin species *Cottus perifretum* and the native Lower River Rhine species *Cottus rhenanus*. Microsatellite analysis shows, that the invasive sculpins are genetically intermediate between the old lineages and that they form a distinct genetic group across their whole expansion range (Nolte et al. 2005b). Contact zones between the invasive sculpins and *Cottus rhenanus* have been well studied in the Sieg (Nolte et al. 2006). Where small streams disembogue into the main stream, stable narrow hybrid zones can be observed between *Cottus rhenanus* and the invasive hybrid lineage. The occurrence of a stable hybrid zone indicates, that the two lineages in contact present distinct entities, which do not merge (Nolte et al. 2006). Thus the invasive sculpins represent a homogenous hybrid lineage with obviously new adaptive potentials in terms of ecology. In contrast to their headwater inhabiting parental species they are found in summer warm and turbid waters in the main channel of the rivers Rhine, Sieg and Mosel (Nolte et al. 2005b) (Fig. 1.1). The question that arises is whether the hybridization event combined favorable parental traits such that the invasion of and the adaptation to this novel habitat became possible.

Morphologically the hybrid sculpins are more similar to *Cottus perifretum* in terms of body shape and skin prickling the latter being a character that is virtually absent in *Cottus rhenanus* (Nolte et al. 2005b). The function of skin prickling is not known, but since *Cottus perifretum* is found in the typical cold stream habitats and never invaded the main channel this character alone is probably not responsible for the invasive potential of the hybrid lineage.

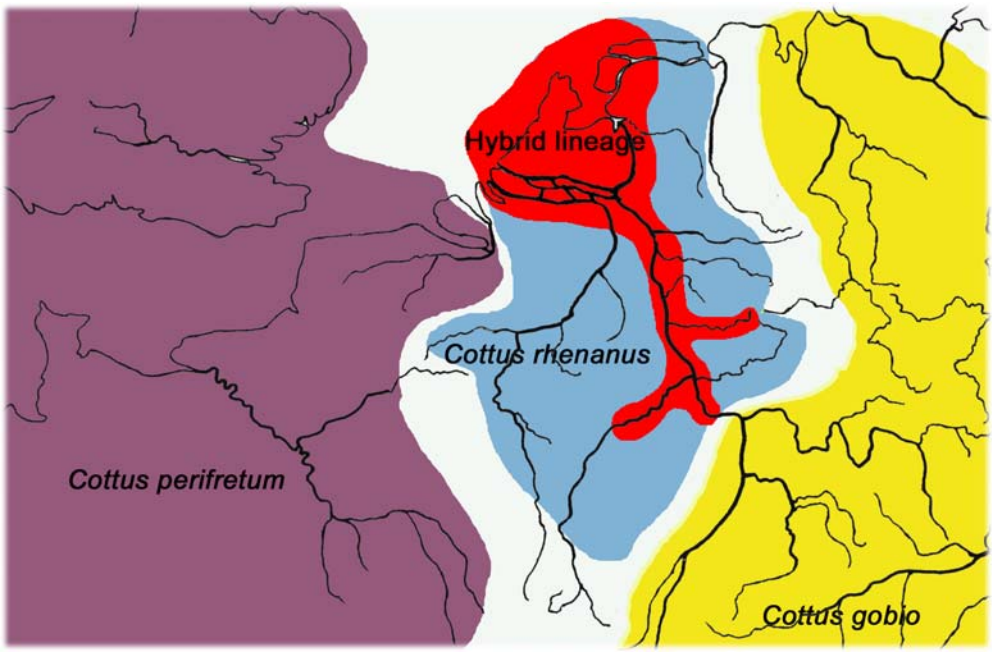

Figure 1.1 Distribution of *Cottus* lineages in and around the River Rhine system

1.5 Age of the hybrid lineage

The molecular analyses conducted so far allow inferences about the age of the hybrid lineage. Derived characters for the hybrid lineage could neither be found in the mitochondrial DNA nor in the first analysis of nuclear markers. The lack of unique characters is an indicator of recent origin.

This hypothesis is supported by the geographic history of the Rivers Rhine and Schelde. About 200 years ago, channels were build connecting the River Rhine with the Schelde system. The rocks used for the fortification of the channels presented suitable microhabitats for *Cottus* which might have allowed them to spread into the newly build waterways. This situation allowed for secondary contact between old phylogeographic lineages. Thus hybridization between *Cottus perifretum* and *Cottus rhenanus* only became possible quite recently in the Lower River Rhine area. A hybrid population between the two species probably existed for some time, before a uniform hybrid lineage arose, which had the potential to invade a new unoccupied habitat.

1.6 Mapping hybrid genomes

To reconstruct how processes of hybrid speciation have taken place it is necessary to explore the genetic architecture of hybrid species. This has only been done so far for the diploid hybrid sunflower species *Helianthus anomalus* (Rieseberg et al. 2003a, b), which is a hybrid between *H. annuus* and *H. petiolaris* that has emerged about 170,000 years ago. Rieseberg et al. (2003a) have used high-resolution genetic linkage maps from the hybrid lineage and were able to trace how the hybrid genome was assembled as a mosaic from different parental species. After linkage map generation, the ancestry of each mapped trait could be determined by surveying the

parental populations, which ultimately allows to trace the origins of whole genomic fragments.

To map hybrid genomes, an ancestry-informative marker system has to be developed. The markers have to be fixed for different alleles in the two parental species in order to be ancestry-informative in the hybrid lineage (Fig. 1.2). SNP (Single Nucleotide Polymorphisms) and indel (Insertion/Deletion polymorphisms) markers, which are specific for the two parental species, present a suitable marker system for this study. Microsatellite markers are not informative for this study since they harbor large genetic diversity with respect to allele frequencies between different stream populations of *Cottus*. Therefore one would have to know the exact source populations that contributed to the hybrid lineage in order to use this marker system.

Several populations of both parental species have to be screened in order to detect markers that are fixed for different states between the two species. To furthermore infer which of the two marker states is the ancestral and which is the derived one an outgroup species can be included into the analysis (Fig. 1.2). *Cottus ricei,* which is mainly found on the eastern slopes of the Rocky Mountains up to southwest Quebec and also in the Great Lakes, is employed as an outgroup species in this study.

After the establishment of ancestry-informative markers, different populations of the hybrid lineage have to be analyzed separately for these loci in order to first estimate the overall contribution of the parental species to the hybrid genome and afterwards to compare the homogeneity of these contributions in different populations. Ancestral alleles, which are detected in the hybrid lineage, could potentially have entered the hybrid genome from any lineage that retained the ancestral state. Therefore, only derived alleles are reliably indicative of the ancestry of a specific locus, while ancestral alleles give only indirect information.

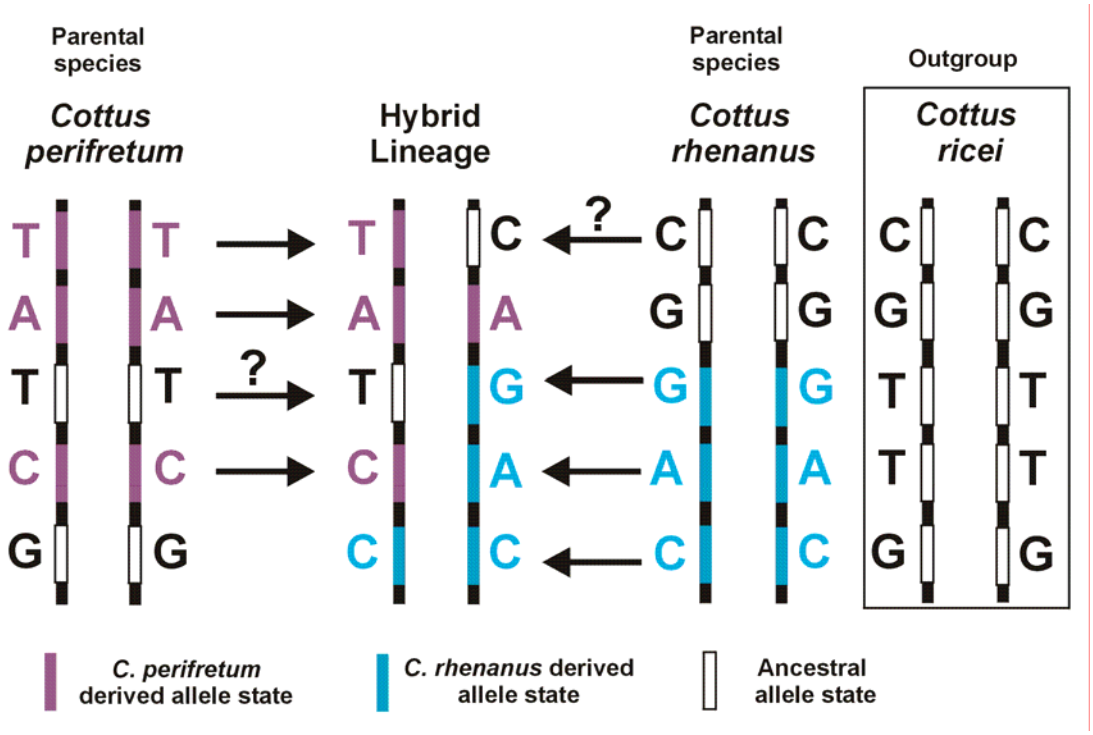

Figure 1.3 Mapping of the hybrid genome with ancestry-informative markers. Purple letters and bars indicate SNP alleles, which are derived for *C. perifretum* and blue letters and bars indicate derived *C. rhenanus* alleles. Alleles found in the outgroup species *C. ricei* are thought to present the ancestral allele state and are indicated by black letters and white bars. If both parental species posses derived alleles at one locus (i.e. both alleles differ from the one found in the outgroup), the ancestral state cannot be determined. Only derived allele states detected in the hybrid lineage are directly informative of the ancestry of this allele (indicated by a black arrow) whereas ancestral alleles are not reliably informative of ancestry (indicated by question marks over the arrows).

1.7 Employing the genomic resources from model organisms for the study of non-model species

Syntenic relationships offer the possibility to transfer genomic information available for model organisms to non-model organisms, which are genetically less well characterized (Schmid 2000, Gebhardt *et al*. 2003, Erickson *et al.* 2004). With a number of complete genome sequences becoming publicly available, the possibilities for comparative approaches are increasing. Studies range from basic comparisons of chromosome structure (Chowdhary *et al.* 1998) to the identification of synteny-defined candidate genes (Giampietro *et al.* 1999). Whole genome comparisons of different species reveal information about homologies, conserved regions, syntenic relationships, genome duplications or duplications of genomic fragments, and genome evolution in general. Comparisons like this are only possible for fully sequenced model organisms. However, comparisons of the genetic map of one organism with the physical map of another organism can also be very informative. Among plants this strategy has been employed to gain information about conserved synteny between the plant model *Arabidopsis thaliana* and different crop species (Dominguez *et al.* 2003, Gebhardt *et al*. 2003). One of the hopes is, that through comparative analysis, knowledge about the genetic make-up of non-model organisms can be gained without having to construct a physical map. Depending on the goal of the study, these approaches require high degrees of genome colinearity at the genetic level and at the

gene level (= microsynteny) (Schmid 2000), as well as sufficient similarity between the sequences to identify homologous regions. Consequently, the question arises of how closely related organisms should be for comparative analysis to be fruitful.

In this study, synteny information between the *Cottus* genome and the genomes of sequenced fish species will be employed, to infer the distribution of ancestry-informative markers over the *Cottus* genome and the gene content of marker regions.

1.8 Aim of the study

With this study I want to pave the way to show, that hybridization can act as a creative evolutionary force, which can lead to the formation of new lineages. The general phenomenon of hybridization coupled with new capacities for colonization has so far only been studied in plants. The *Cottus* case provides the opportunity to genetically characterize a hybrid lineage in animals for the first time. I also want to show how the available genomic resources of model organisms can be used to facilitate such an analysis in a non-model species.

2 Materials and Methods

2.1 Establishment of mapping families

For the preliminary genetic map crosses between the hybrid lineage and *Cottus rhenanus* were established. All populations used were taken from the River Sieg drainage.

To obtain crosses, mature prespawning adults were collected in the field in February 2002 and transferred to laboratory tanks. Fish were fed *ad libitum* with insect larvae. Spawning occurred readily in artificial shelters partially buried in sand at temperatures between 8-10°C. After spawning, only the guarding male was left with the egg clusters. After hatching of the larvae, the male was removed from the tank. Larvae were raised initially using live *Artemia* nauplii, and later with frozen chironomid larvae and mysiid shrimps until at least 3 cm in length. All animals were preserved in 70 % ethanol for future studies.

One cross involved a male from the population “Giertshagener Bach” (*Cottus rhenanus*; Stream Giertshagener Bach at Giertshagen, North Rhine-Westphalia, Germany; 50°45'N 7°36'E) and 2 females from the population “Wahnbach” (*Hybrid lineage*; Stream Wahnbach, Outlet into River Sieg at Seligenthal, North Rhine-Westphalia, Germany 50°48'N 7°16'E) resulting in two half-sib families (n= 24 and 63 progeny). A full-sib family was obtained from a female from “Ottersbach” (*Cottus rhenanus*; Stream Ottersbach at Eitorf, North Rhine-Westphalia, Germany; 50°47'N 7°26'E) and a male from "Wahnbach" (see above) and contains 78 progeny. Attempts to create an F2 generation intercross failed for unknown reasons. Note, however, that this is not due to general hybrid sterility as numerous F2 or backcross hybrids were found in natural hybrid zones (Nolte et al. 2006).

For a refinement of the genetic map pure hybrid and pure *Cottus rhenanus* families were established. Premature spawning adults were collected again in the field and set up in tanks as above. Larvae were not allowed to hatch, but instead DNA was extracted directly from the eggs. These families involve 5 pure hybrid families all coming from the Wahnbach (see above). The *Cottus rhenanus* families were established with parents from the Bröl for two families (Stream Bröl, North Rhine-Westphalia, Germany; 50°51`N 7°22`E), from the Derenbach for one family (Stream Derenbach, North Rhine-Westphalia, Germany; 50°47`N 7°20`E) and from the Ottersbach for two families (see above). Each analyzed family consists of the two parents and 94 randomly picked progeny in order to fit a 96-well format.

2.2 DNA-Extractions

DNA was extracted using a salt-extraction protocol. A few square millimeters of tissue are digested in 500 µl HOM buffer (80 mM EDTA, 100 mM Tris, and 0.5% SDS) and 5 µl Proteinase K (NEB 20 mg/ml) at 55°C over night. 500 µl of 4.5 M NaCl is added and the mixture is incubated for 10 min at 4°C. Subsequently 300 µl of

Chloroform are added, followed by centrifugation at 10.000 g for 10 min. 850 µl of the upper phase are transferred to a fresh tube and DNA is precipitated with 595 µl of pure Isopropanol (0.7 volume). The DNA is pelleted by centrifugation at 13.000 g for 10 min. Finally the pellet is washed two times with 500 µl 70 % Ethanol, dried and dissolved in TE-buffer (10 mM Tris, 0.1 mM EDTA)

This protocol was modified for the extraction of DNA from the *Cottus* eggs in order to be conducted in a 96-well plate. Per well one single egg is digested in 100 µl HOM buffer with 2 µl Proteinase K at 55°C and with shaking at 1300 rpm (Eppendorf, thermomixer comfort) over night. 100 µl 4.5 M NaCl are added and the mixture is incubated for 10 min at 4°C. Afterwards the plate is centrifuged for 30 min at 3220 g. About 100 µl of the supernatant are transferred to a new plate and precipitated with 100 µl of Isopropanol. The DNA is pelleted by centrifugation for 30 min at 3220 g. Afterwards the pellet is washed two times with 100 µl of 70 % Ethanol, dried and dissolved in TE-buffer.

2.3 Genotyping of microsatellite markers

Loci were taken from Englbrecht et *al.* (1999) and Nolte et al. (2005a). For the preliminary genetic map all individuals were genotyped for 171 microsatellite markers on a Megabace 1000 (Amersham Biosciences). For the refined map, the 10 pure mapping families were genotyped for all 49 microsatellite markers on linkage group 3 and genotyped on an ABI 3730 capillary sequencer (Applied Biosystems). PCR reactions were performed as multiplex; up to 8 fluorescently labeled (Fam, Hex, Tet for the Megabace and Fam, Hex, Ned for the ABI) primer pairs were combined and amplified using the Multiplex-PCR Kit (Quiagen) as described in Nolte et al. (2005a). The loci were combined in a way such that all fragments could be separated in a single lane without overlap and scored unambiguously.

2.4 Construction of a genetic map

Linkage distances and marker orderings were determined with the *Locusmap* software (Garbe and Da, 2003). The sex-averaged LOD-threshold was set to 3. The Haldane mapping function was used to convert recombination frequency to centiMorgan. Non-inheritance errors were checked again in the genotyping files and then classified as probable allele-drop-out errors, when the progeny was homozygous for a parental allele only found in one parent, or allele-mutation errors, when the progeny possessed an allele not present in one of the parents, which could be explained by a single step mutation of a parental allele. Graphics of the linkage groups were produced with the *MapChart* software (version 2.1; Voorrips 2002).

For the preliminary genetic map sex-averaged LOD-Scores ranged from 3.26-94.81 with an average of 21.21. The informative meiosis among the linked loci ranged from 62-330 with an average of 199.9. Identical inheritance was detected for 57 marker-pairs. 20 non-inheritance errors were detected, of which 16 concern a single locus and can be explained by allele drop out in the progeny. The remaining non-inheritance errors are spread over five different loci and can also mainly be explained by allele drop out except for one locus, were a mutation in one of the progeny alleles is the most probable explanation.

For the refined genetic map of linkage group 3 all hybrid families were analyzed together and all *Cottus rhenanus* families were analyzed together in order to

be able to compare the linkage maps between these two lineages. For the final comparison of linkage maps only markers, which could be integrated into the map in both families, were included in the linkage analysis. A composite map from both lineages was also created in order to include as many loci as possible in the map. For the loci included in the composite map sex-averaged LOD-Scores ranged from 3.14-357 with an average of 82.85. The informative meiosis ranged from 178-963 with an average of 652. Identical inheritance was detected for 18 marker-pairs. 19 non-inheritance errors were detected, of which 6 concern a single locus and can be explained by both allele dropouts in the progeny and by a single-step mutation. The remaining non-inheritance errors concern single loci and can mainly be explained by single-step mutations except for two cases, which can only be explained by allele drop-out.

2.5 Tests for Mendelian segregation

Tests for Mendelian segregation were performed for the mapping families employed in the preliminary linkage map construction using Pearson's chi-square test with an expected segregation ratio of 1:1 for all alleles (significance level $P < 0.05$). Every family was tested separately for every marker, which resulted in 513 pairwise comparisons of observed vs. expected allele numbers. Markers not following Mendelian segregation were checked for genotyping errors (see above).

2.6 Blast searches

BLAST searches (Altschul et al. 1990) were conducted against the *Tetraodon*, *Fugu*, *Danio* and *Gasterosteus* genomic sequences via the Ensembl Genome Browser (http://www.ensembl.org/). Similarity searches against the Medaka sequences were conducted via the Medaka Genome Project homepage (http://dolphin.lab.nig.ac.jp/medaka/index.php). The *Cottus* sequences of the microsatellite loci had an average length of about 500 bp (range from 119 – 1109 bp). Hits with e - values below 10^{-5} were considered as significant. The corresponding *Tetraodon* sequences were retrieved for sequence comparisons. Local alignments were produced with DIALIGN 2 (Morgenstern 1999) using the default settings.

For all loci included in the screen for ancestry-informative markers (see 2.7) BLAST searches (Altschul et al. 1990) were conducted only against the *Gasterosteus* genome.

2.7 Construction of a genomic library

For the development of ancestry-informative markers a genomic library was created. *Cottus* genomic DNA from two individuals of the hybrid lineage (Stream Wahnbach, see 2.1) was partially digested with *Mse*I. A digestion reaction of 800 µl was set up containing 160 µl of a mix of total genomic DNA (~400 ng/µl), 4 µl of MseI (NEB, 4000 U/ml), 80 µl NEBuffer 2 (NEB), 4 µl BSA (10 mg/ml) and 516 µl of H_2O. This reaction was split into 8 vials, each containing 100 µl of the digestion reaction. 4 reactions were incubated at 37°C for 5 minutes and the remaining reactions were incubated for 15 minutes at 37°C. Subsequently a range of 1000-1500 bp

fragments was eluted from a gel (0.8 % agarose) using the QIAquick gel extraction kit from Qiagen. This size range was chosen, since these fragments can be sequenced in one sequencing run. Furthermore it was known from a previous SNP screen, that one ancestry-informative SNP could be found about every 1000 kb. After extraction from the gel, fragments were end polished in a 50 µl reaction containing the eluted fragments, 10 µl 5x Phusion™ HF buffer (Phusion™ High-Fidelity PCR kit, Finnzymes), 1 µl 10 mM dNTPs and Phusion™ High-Fidelity Polymerase (Finnzymes). This reaction was incubated for 30 min at 72°C. Subsequently the end polished fragments were cleaned up again by a gel run. Afterwards fragments were ligated into pZeroII™ vector (Invitrogen) and cloned into electrocompetent Top10™ cells (Invitrogen). Plasmids were extracted via minipreps. Sequencing was conducted on an ABI 3730 capillary sequencer (Amersham Biosciences) with the universal primers SP6 (5'-ATTTAGGTGACACTATAG-3') and M13F-pUC(-40) (5'-GTTTTCCCAGTCACGAC-3') and for a part of the plasmids with PbsA (5' CTATGACCATGATTACGCCAAG-3') and PbsE (5' TAACGCCAGGGTTTTCCCAGT-3'). Forward and reverse sequences were assembled and edited with the program 'Seqman' (www.dnastar.com). A total of 960 plasmids has been isolated and sequenced.

2.8 Prescreen of *Cottus* genomic fragments for similarities to the *Gasterosteus* genome

For all sequenced plasmids BLAST searches (Altschul et al. 1990) were conducted against the genomic sequence of *Gasterosteus aculeatus* (see 2.6). Hits with e-values below 10^{-5} were considered significant. Only fragments yielding a significant hit were included in the screen for ancestry-informative markers (see 2.9). For loci, which yield a significant, hit the conserved synteny between the *Cottus* and the *Gasterosteus* genomes can be employed to roughly localize the fragments on the *Cottus* genetic map.

2.9 Development of ancestry-informative SNP and Indel markers

Primers for 563 genomic fragments have been designed with the program 'FAST-PCR' (Kalendar 2003) (Supplement 1). 122 of these loci are microsatellite loci from Nolte et al. (2005) and from Englbrecht et al. (1999), which are partially included in the linkage map. The remaining 441 fragments were taken from the genomic library.

To screen the fragments of the genomic library for ancestry informative SNPs and indels pooled DNA samples of each parental species were employed. For *Cottus perifretum* 5 individuals each from three different populations were pooled (numbers in brackets indicate sample points which are shown in Fig. 2.1): 'Zwanebeek' (66), 'Witte Nete' (65) and 'River Nete' (1). For *Cottus rhenanus* 5 individuals each were pooled from the populations 'Rur Düren Maas' (17), 'Flaumbach' (31) and 'Bröl bei Winterscheid' (24). Each DNA sample was adjusted to a concentration of 20 ng/µl. Even amounts of all samples were mixed and 1 µl (20 ng) of each DNA pool was used for amplification with the Quiagen Multiplex Kit and subsequent sequencing. Afterwards forward and reverse sequences from each parental pool were aligned using 'Seqman' (www.dnastar.com).

Loci, which contained fixed SNPs or indels in the parental species, were analyzed for the hybrid lineage. For this purpose pooled DNA samples were used. A total of three pools from three different populations was employed (numbers in brackets indicate the sample points which are shown in Fig. 2.1): one pool with 10 individuals from the population 'Ijsselmeer Enkhuizerzand' (6), one pool with 10 individuals from the Sieg (10) and one pool with 6 individuals from the population 'Mosel bei Koblenz' (15). Like the parental pools the samples were amplified with the Quiagen Multiplex Kit and sequenced afterwards. Pools were only sequenced in one direction depending on where the informative marker was found in the parental species.

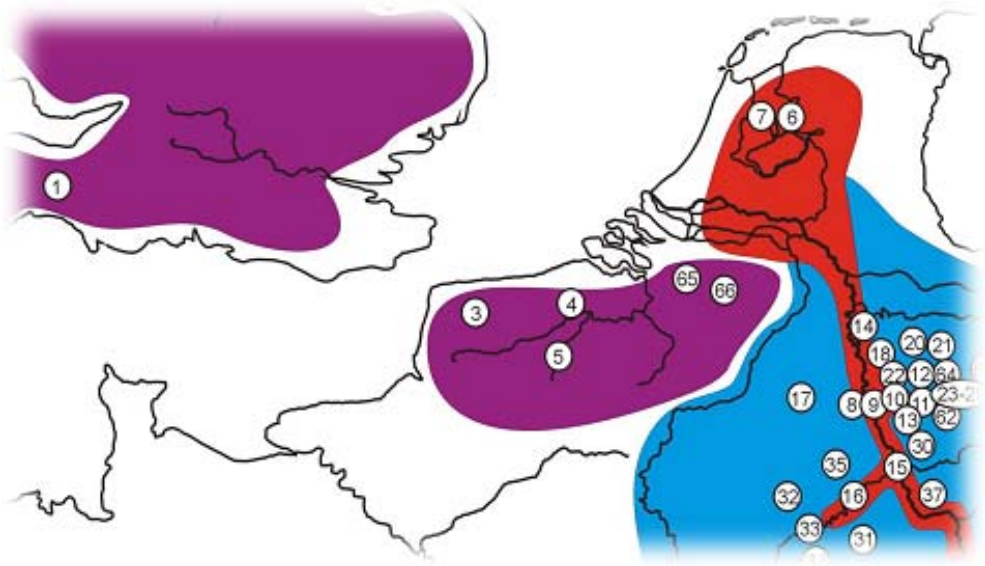

Figure 2.1 Map of the Rhine and the Scheldt area with the locations from which samples are available (this is a section of the map form Nolte el al. 2005b). The purple area represents the range of *C. perifretum*, the light blue area represents the range of *C. rhenanus* and the red area represents the distribution range of the invasive *Cottus*.

Furthermore outgroup species were analyzed for the informative marker loci. For several marker loci *Cottus aleuticus* (Kenia River, Soldatina, Alaska), *C. bairdii* (Brokenstraw Creek, Warren, Pennsylvania, USA) and *C. poecilous* (River Vistula, Poland) were used to generate outgroup sequences. For the majority of loci a pool of 5 DNA samples from *Cottus ricei* was amplified and sequenced. The hybrid and the outgroup sequences were aligned with the parental sequences using 'Seqman' (www.dnastar.com). The parental and ancestral allele states found in each hybrid population were recorded in respect to being present or not. Actual allele frequencies could not be estimated with the pooled samples.

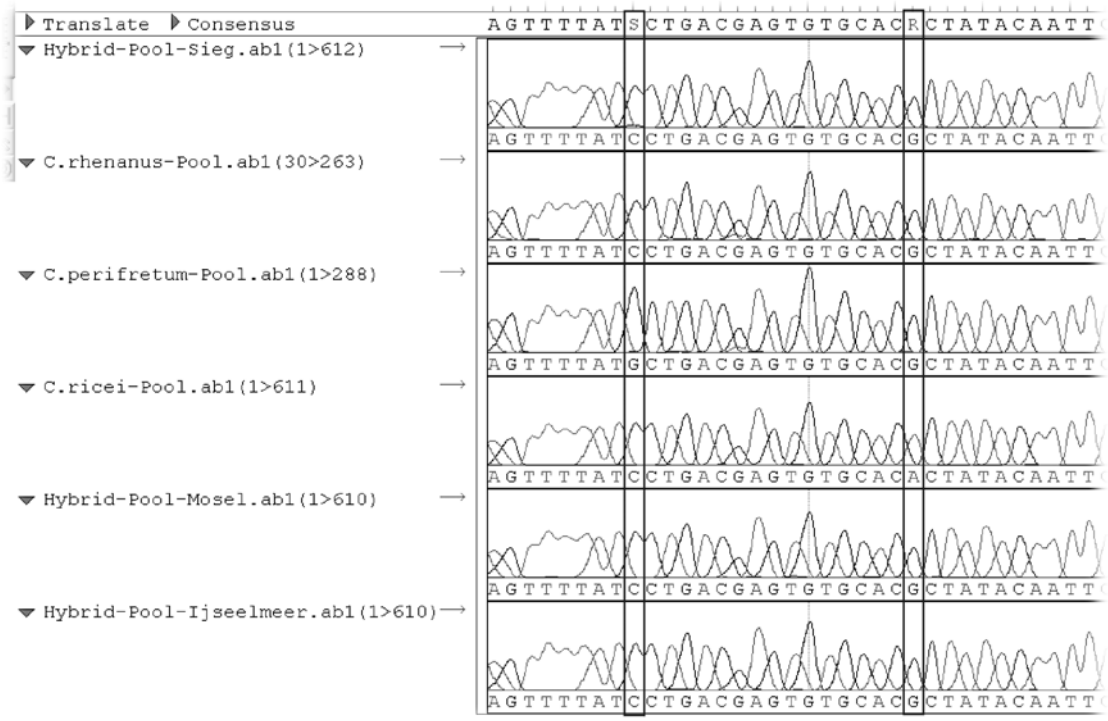

Figure 2.2 Alignment of sequences from the two parental, the three hybrid and the outgroup pool. Polymorphic sites are indicated by boxes. The first polymorphic site presents an ancestry-informative SNP with a derived allele for *C. perifretum*. *C. rhenanus* retained the ancestral state, which can be concluded from the comparison with *C. ricei* sequence. The ancestral allele is also fixed in all three hybrid populations. The second polymorphic site presents a private allele for the outgroup species.

2.10 Tests for parental allele contributions

Tests for parental allele contributions to loci with fixed and mixed ancestries in the hybrid lineage were conducted using Pearson's chi-square test with an expected contribution from both parental species of 1:1 (significance level $P < 0.05$). Pearson's chi-square test was also employed to test for differences in parental contributions to the three hybrid populations, again with the expectation of a 1:1 contribution (significance level $P < 0.05$).

2.11 Comparison of gene content of marker loci with fixed and mixed ancestries

Gene content of marker loci with fixed and mixed ancestries in the hybrid lineage was compared using Pearson's chi-square test. Marker loci were divided into four categories: 1. within coding regions, 2. within 10 kb upstream of coding regions, 3. within 10 kb downstream of coding regions and 4. no coding region. Fixed and mixed marker loci were compared with the assumption that the contribution of these two marker classes to each category is 1:1.

3 Results

3.1 A genetic map of *Cottus* based on microsatellite markers

Three mapping families consisting together of 170 individuals were genotyped for 171 microsatellite loci. 3.3% of the tests for Mendelian segregation distortion were significant at $P < 0.05$, indicating that the level of segregation distortion was within the limits that are expected by chance. 366 significant pairwise linkages (LOD > 3.0) were detected for 154 of these markers. The loci could be assembled into 20 linkage groups (Fig. 3.1). The lengths of the linkage groups ranged from 0-1681.7 cM with 2-49 markers per group. The longest linkage group is linkage group 3 with 1618.7 cM; the cumulative map length is 2738.1 cM. Given that the chromosome number in *Cottus* is 24 with no conspicuously large single chromosome (Vitturi & Rasotto 1990), it seems likely that linkage group 3 is artificial and will become fragmented when more mapping groups are included.

The published genome size of close relatives of *Cottus gobio* is slightly below 1 pg per cell (Hardie & Hebert 2003) and this value was also found for the *Cottus* lineages involved in this study in a first estimate (T. R. G pers. com., compare http://www.genomesize.com/). According to Dolezel *et al*. (2003) this can be converted into a genome size of about 1000 Mbp. One centimorgan would thus correspond to 0.36 Mbp.

A possible explanation for the apparent clustering in parts of linkage group 3 would be chromosomal rearrangements. The map is based on F1 crosses between the hybrid lineage and *C. rhenanus*, in which chromosomal variants do not segregate. Thus, mapping in first generation hybrids would integrate different signals that trace back to rearranged chromosomal fragments from the parental lineages. The resulting pattern corresponds to what is seen in linkage group 3, namely an inflated linkage group that would be assembled from multiple regions with a different architecture (Livingstone *et al*. 2000).

To address this question, new pure hybrid and *C. rhenanus* mapping families were established. Only markers that could be included into the linkage map in both, the *Cottus rhenanus* and the hybrid lineage were included in the analysis. Linkage analysis yielded a brake-up of linkage group 3 into 7 linkage groups in the *Cottus rhenanus* families and 6 linkage groups in the hybrid families (Fig. 3.2). One of these newly created linkage groups is still referred to as linkage group 3. The remaining linkage groups are added to the previous map (existing of 20 linkage groups) as linkage groups 21-26. Therefore the *Cottus* genetic map now exists of 26 linkage groups, which is more than would be expected from the haploid chromosome number (n = 24). The cumulative map lengths however is reduced to 1692.1 cM thus that 1 cM now corresponds to 0.53 Mbp.

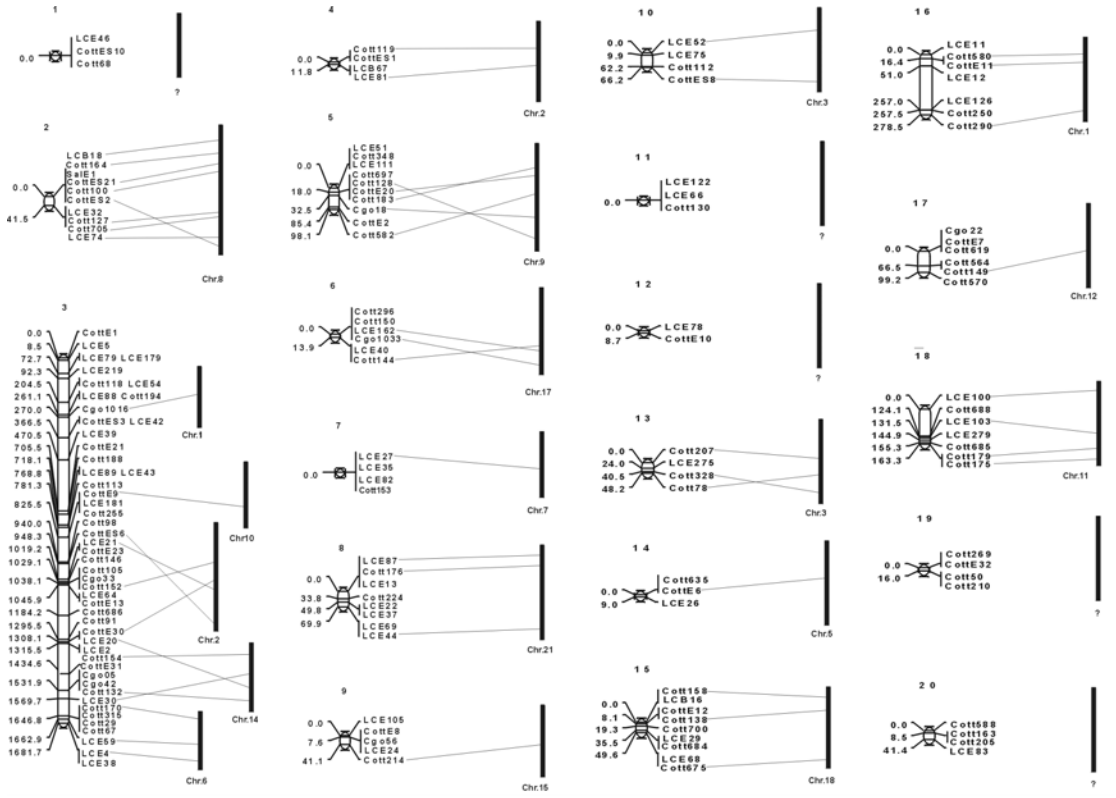

Figure 3.1 Comparison of the preliminary *Cottus* linkage groups with the chromosomes of *Tetraodon nigroviridis*. Significant BLAST hits and their relative position on the *Tetraodon* chromosomes are indicated by connecting lines between the *Cottus* locus and the *Tetraodon* chromosome. Locus names refer to Englbrecht et al. (1999) for all „Cgo“ labels and to Nolte et al. (2005) for the remainder.

Comparing the linkage groups of the hybrid and *Cottus rhenanus* families, three inconsistencies can be observed between the maps (Fig. 3.2): (1) Locus Cott146 is placed differently in linkage group 3 on the *C. rhenanus* and the linkage map of the invasive lineage, (2) a whole block including the loci CottE31, LCE59, Cott315 and Cott170 is placed within linkage group 3 of the invasive lineage, but is assigned as a single group in the *C. rhenanus* linkage map and locus LCE59 is found in different positions within this block, (3) locus Cott255 is found in different positions within linkage group 23. These differences could not be confirmed by comparing the linkage maps from the single families. One reason for this is probably that the loci are not equally informative in the different families.

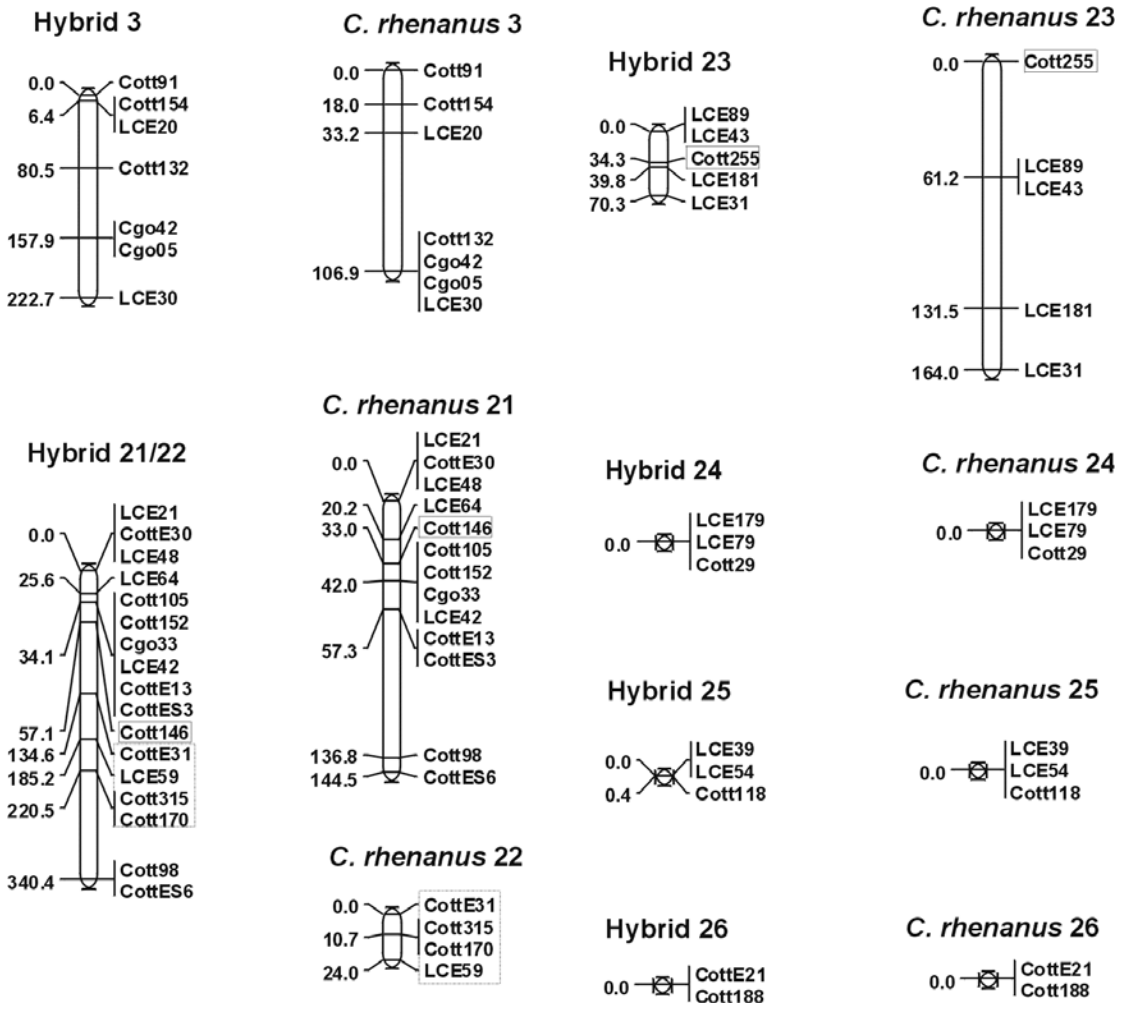

Figure 3.2 Subgroups of former linkage group 3, which have been established through the analysis of pure hybrid lineage and *Cottus rhenanus* families. Rearrangements between corresponding linkage groups are indicated with boxes.

3.2 Conserved synteny between the genomes of *Cottus* and model organisms

The flanking sequences of all typed microsatellite loci were used for similarity searches against the *Danio*, Medaka, *Fugu*, *Tetraodon* and *Gasterosteus* genomes. Using a significance threshold of $e < 10^{-5}$ 21 to 159 hits could be detected in the different genomes, most of which are even retained at a significance threshold of $e < 10^{-10}$ (Tab. 3.1).

Table 3.1 Number of BLAST matches of *Cottus* microsatellite flanking sequences in other fish genomes.

matches with	$e < 10^{-5}$ N out of 171	$e < 10^{-10}$ N out of 171
Danio	**21**	**11**
Medaka	**18**	**11**
Tetraodon	**77**	**64**
Fugu	**87**	**67**
Gasterosteus	**141**	**127**

The matches were usually due to blocks of very highly conserved sequences. For *Tetraodon* comparisons, these had a length of 19-120 bp (average 40 bp) with sequence similarities between 62-100% (average 92%).

Only about a third of the loci with matching flanking sequences showed a conservation of the microsatellite itself (i.e. at least 5 repeats of the respective sequence motif) in *Tetraodon*, confirming the expected high turnover of such sequences (Schlötterer 2000).

The total length of *Cottus* sequences analyzed in these BLAST searches was 86,530 bp. Given that 77 fragments yielded a significant hit with the *Tetraodon* genome sequence, one can estimate that at least one conserved block occurs about every 1100 bp. Thus, it should be possible to analyze even microsyntenic relationships throughout the genomes of these species.

An ordered map is available for the *Tetraodon* genome, which covers about 64% of the genome sequence (Jaillon *et al.* 2004). Comparisons of map positions of the *Cottus* markers with a hit in the *Tetraodon* sequence thus allow assessing large-scale synteny patterns. It can be observed that most markers from a single linkage group in *Cottus* yielded also hits on a single chromosome in *Tetraodon* (Figure 3.1). The major exception is *Cottus* linkage group 3 of the preliminary linkage map, which yields hits with five *Tetraodon* chromosomes. The observed syntenic relationships together with the sequence similarities between the *Cottus* and *Tetraodon* sequences suggest true homology of the associated regions.

Five *Cottus* linkage groups could not be associated with a *Tetraodon* chromosome so far. In some cases this was due to lack of significant hits with the respective markers (groups 12 and 19) and in other cases hits were only found on genomic fragments that are not yet anchored to a *Tetraodon* chromosome (groups 1, 11 and 20).

Given that *Tetraodon* has only 21 chromosomes (Grützner *et al.* 1999), a one to one syntenic relationship between all linkage groups cannot be expected. This is also reflected in the finding that *Cottus* linkage groups 10 and 13 map to a single *Tetraodon* chromosome (Figure 3.1). However, the general patterns are clearly comparable and suggest that large parts of the genomes will be alignable.

In July 2006 the annotated genome sequence of the three-spined stickleback *Gasterosteus aculeatus* became available (release 43.1b). Since this species is more closely related to *Cottus* than *Tetraodon* it seemed feasible to look for conserved synteny between the *Cottus* and the *Stickleback* genome.

Significant similarity hits were detected for 83 % of the *Cottus* loci. As shown in Fig. 3.3 most of the loci from a given *Cottus* linkage group yielded significant hits on single stickleback linkage groups, suggesting a very good correspondence of chromosomes. Exceptions are linkage groups 1, 9, 11, 19, 20 and 23. However, as *Tetraodon*, *Gasterosteus* has also only 21 chromosomes, compared to 24 in *Cottus.* Accordingly, a perfect association cannot be expected. Furthermore, some of the *Cottus* loci might not yet be integrated into the correct linkage group, which also explains hits from one *Cottus* linkage group on two *Gasterosteus* chromosomes. However, despite some unresolved associations between the linkage groups of the two genomes, a high degree of conserved synteny can be inferred.

With the help of the conserved synteny between the *Cottus* and the *Gasterosteus* genome, the subgroups of the former linkage group 3 can be confirmed. In the preliminary map, linkage group 3 yielded hits on 5 different *Tetraodon* chromosomes, which was taken as an indicator that this group actually resembles several, unresolved linkage groups. By combining the information from the 10

established mapping families (see 3.1), this group could be broken up into 6 linkage groups added to the previous map as linkage groups 3 and 21-26. Except for linkage group 23, which yields hits on *Gasterosteus* linkage groups II and X, all linkage groups are associated with only one stickleback chromosome (Figure 3.3). Linkage groups 25 and 26 are both associated with stickleback chromosome IV, which might be an indicator, that these two groups represent actually only one linkage group. This would bring the *Cottus* map closer to the 24 expected linkage groups. Even though the subgroups of linkage group 3 are supported by the syntenic relationships to the *Gasterosteus* genome, it has to be kept in mind, that these linkage groups were established by combining the information from both hybrid and *C. rhenanus* families. The question if rearrangements between the genomes of these two lineages exist has not yet been finally answered. Thus it cannot be excluded, that synteny relationships differ at some places between the *Gasterosteus* genome and the genomes of the hybrid lineage and *C. rhenanus* respectively.

The stickleback genome seems to be assembled to a higher degree than the *Tetraodon* genome, since a relatively lower number of *Cottus* loci (6 %) yielded hits on unassembled genomic fragments as compared to 30 % on the *Tetraodon* genome. Thus, the stickleback genome presents an even better genomic resource for the analysis of *Cottus* due to a higher percentage of significant BLAST hits and its higher degree of assembly.

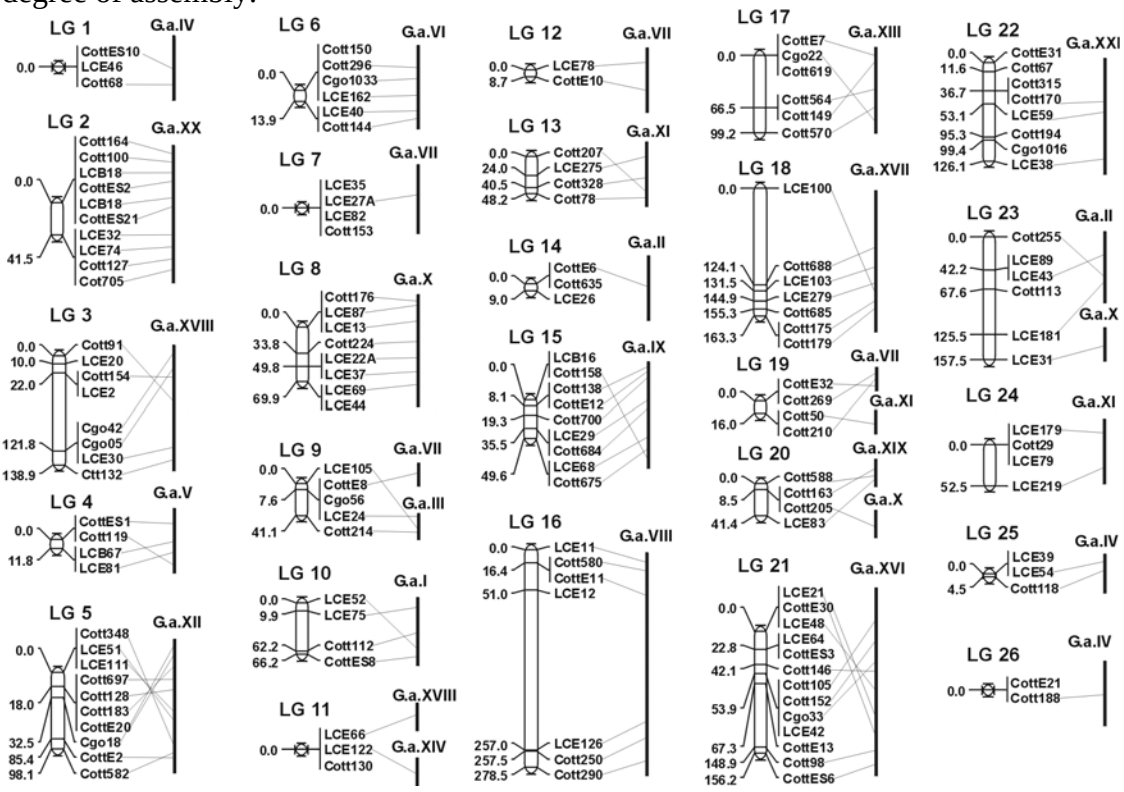

Figure 3.3 Comparison of the improved *Cottus* linkage groups with the chromosomes of *Gasterosteus aculeatus*. Significant BLAST hits and their relative position on the *Gasterosteus* chromosomes are indicated by connecting lines between the *Cottus* locus and the *Gasterosteus* chromosome.

The conserved synteny between the two genomes can be put to use to roughly integrate new markers into the genetic map of *Cottus*. If a BLAST search is conducted with a random genomic *Cottus* fragment against the genome of *Gasterosteus* and a

significant hit is yielded for example on *Gasterosteus* linkage group XX, then it can be inferred, that this fragment is localized on *Cottus* linkage group 2.

3.3 Development of ancestry-informative markers

For the SNP and indel screen a genomic library was constructed containing random 1-1.5 kb genomic fragments. Furthermore, flanking sequences of microsatellite loci (Nolte et al. 2005, Englbrecht et al. 1999), which had already been developed previously and are partially included in the linkage map, were analyzed for informative markers as well. Primers were developed for a total of 563 fragments potentially yielding PCR products in a range from 183-1368 bp with an average length of 690 bp. These fragments were amplified and sequenced for one pool of DNA for each parental species (*Cottus rhenanus* and *Cottus perifretum*). 427 loci (76%) could be amplified and sequenced for both parental pools. For the remaining ones, either the PCR or the sequencing reaction failed. In many instances, microsatellites prevented the production of a clear sequence read. When the individuals in the DNA pool are variable for the microsatellite, the sequence is not readable anymore beyond the microsatellite.

Sequences ranged in size from 48 to 1170 bp with an average of 427 bp. Of the sequenced loci 152 (36%) contained fixed SNPs or indels for the parental species. 21 loci (14%) contained indels and 26 loci contained more than one fixed marker. A total of 161 fixed SNPs were detected. If this is averaged over the entire length of sequenced fragments (205.828 bp), one SNP is found every 1300 bp whereas indels are only found with a frequency of one in every 9800 bp.

3.4 Analysis of the hybrid lineage and an outgroup species for ancestry-informative SNP markers

Pooled DNA samples from three different hybrid populations and one pool of DNA from *Cottus ricei* or DNA from other outgroup species was analyzed for all ancestry-informative loci. Sequences from all hybrid populations and the outgroup species could only be obtained for 108 (71 %) of the 152 SNP and indel loci. Of these 108 loci 14 contained ancestry-informative indels, whereas the remaining ones are SNP loci (Supplement 3). 3 of the indel loci furthermore contained a SNP, which gave the same signal as the indel in the hybrid and outgroup sequences.

Of the 108 ancestry-informative loci (Supplement 3) 7 contained polymorphic SNPs in the hybrid lineage with SNP alleles that were not found in the parental or the outgroup species. One of these loci (co311-m13) contained three such polymorphic SNPs. If the total amount of sequence (44.084 bp) obtained for the hybrid lineage is considered, one polymorphic SNP with a potentially private allele for the hybrid lineage is found about every 6300 bp.

With the help of the outgroup sequence from *C. ricei* it was possible to determine for each locus which is the ancestral state (i.e. the one found in the outgroup) and which is the derived state. 62 of the analyzed loci contained derived states for *C. perifretum,* whereas 46 loci contained derived states for *C. rhenanus*. Only the derived allele states are clearly indicative of the ancestry of the specific allele. Ancestral allele states found in the hybrid lineage could potentially have entered the hybrid genome from any lineage, which retained the ancestral allele. If alleles are divided into the groups '*C. perifretum* derived', 'potentially *C. perifretum*

ancestral', '*C. rhenanus* derived' and 'potentially *C. rhenanus* ancestral' than it has to be taken into account, that the ratio of derived *C. rhenanus* markers to derived *C. perifretum* markers is 46 to 62. This means that at 46 of the marker loci one can expect derived *C. rhenanus* alleles or potentially ancestral *C. perifretum* alleles, whereas there are 62 loci at which one could find derived *C. perifretum* alleles or potentially ancestral *C. rhenanus* alleles. Therefore this factor has to be considered for any comparison using these four allele groups. If the 'derived' and 'potential ancestral' groups are combined for each species, this factor does not have to be considered anymore, since in this case every locus in the hybrid lineage can potentially contain one *C. perifretum* and one *C. rhenanus* allele.

84 loci (78%) showed mixed ancestries in the hybrid populations, which means that one derived parental state was present as well as one ancestral state. The remaining 24 loci (22%) showed fixed ancestries in all three hybrid populations, either for one of the derived parental or for one ancestral state. Of these fixed loci 8 contained only derived *C. perifretum* alleles, 1 contained only derived *C. rhenanus* alleles, 10 were fixed for ancestral alleles which might have been received from *C. perifretum* and 5 contained ancestral alleles that might have been received form *C. rhenanus* (Tab. 3.2).

Table 3.2 Ancestries of fixed marker loci in the hybrid populations

	***C. perifretum* derived**	***Ancetral potentially C. perifretum***	***C. rhenanus* derived**	***Ancestral potentially C. rhenanus***
Marker loci with fixed ancestries in the hybrid lineage	**8**	**10**	**1**	**5**

To estimate the parental contributions to the hybrid genome, each locus was scored for the presence of the 'derived *C. perifretum* state', the 'derived *C. rhenanus* state', the 'potentially ancestral *C. perifretum* state' or the 'potentially ancestral *C. rhenanus* state'. Each locus was analyzed as being representative of the whole hybrid lineage. This means each locus contains two states: either two times a fixed state or one derived state and one ancestral state. The occurrence of derived *C. rhenanus* states, derived *C. perifretum* states and ancestral states probably inherited from the one or the other parental lineage is added up over all loci (Tab. 3.2 and 3.3). Afterwards these numbers are corrected for the difference in derived markers for the two species (Tab. 3.3). This involves multiplying the number of 'derived *C. perifretum* states' and the number of 'potentially ancestral *C. rhenanus* states' by 0.75 (46/62).

Both diagrams in Fig. 3.4 indicate, that there is a difference in parental contributions to loci with fixed and mixed ancestries in the hybrid lineage. When derived and ancestral states are considered together (Tab. 3.2, Fig. 3.4 left graph) there is no significant difference in parental contributions to loci with mixed ancestries in the hybrid lineage, whereas a significantly higher contribution from *C. perifretum* than from *C. rhenanus* can be detected at loci with fixed ancestries. The same is true, when the allele states are split into 'derived' and 'potentially ancestral' categories. At loci with mixed ancestry, no difference in contribution can be observed between 'derived' and 'potentially ancestral alleles' from the two species (Tab.3.3, Fig. 3.4) whereas at loci with fixed ancestries the contribution from 'derived' and 'potentially ancestral' alleles from *C. rhenanus* both are significantly lower than the contributions from *C. perifretum*.

Table 3.2 Comparison of parental contributions to loci with mixed and fixed ancestries in the hybrid genome. Derived and potentially ancestral states are considered together for each parental species.

	C. perifretum states	*C. rhenanus states*	*P(same)*
Loci with mixed ancestry	**243**	**184**	**0.065665**
Loci with fixed ancestry	**91**	**28**	***8.83e^{-6}**

Table 3.3 Comparison of parental contributions to loci with fixed and with mixed ancestries in the hybrid genome.

	Derived C. perifretum states	*Derived C. rhenanus states*	*P(same)*	*Ancestral C. perifretum states*	*Ancestral C. rhenanus states*	*P(same)*
Loci with mixed ancestry	**147**	**74**		**134**	**149**	
Loci with mixed ancestry - corrected	**109**	**74**	**0.063324**	**109**	**110**	**0.39913**
Loci with fixed ancestry	**42**	**6**		**66**	**30**	
Loci with fixed ancestry - corrected	**31**	**6**	***6.03e^{-4}**	**66**	**22**	***2.54e^{-4}**

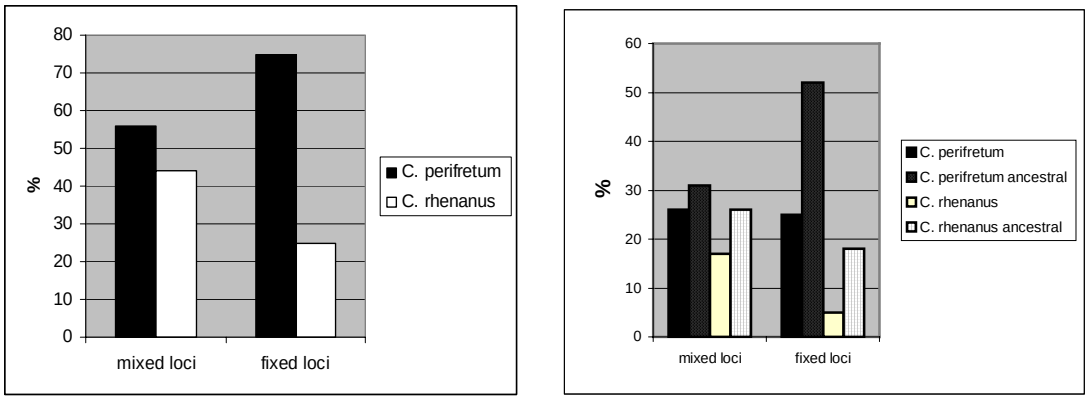

Figure 3.4 Comparison of the parental contributions to loci with fixed and with mixed ancestry in the hybrid genome.

This analysis was also conducted for the three hybrid populations respectively, but combining loci with mixed and with fixed ancestry in order to compare overall parental contributions to the three populations. It can be observed, that the contribution of *C. perifretum* declines from the Ijsselmeer, over the Sieg to the Mosel population and that the contribution of *C. rhenanus* rises from the Ijsselmeer over the Sieg to the Mosel population (Fig. 3.5, left diagram). These differences are not significant (Tab. 3.4) and they can only be caused by the loci with mixed ancestries, since the fixed loci all contain the same amount of the specific allele states in the three hybrid populations.

Table 3.3 Comparison of parental contributions to the three hybrid populations. Derived and 'potentially ancestral' states for each parental species are considered together.

	C. perifretum alleles	*C. rhenanus alleles*
Ijsselmeer	137	79
Mosel	126	90
Sieg	122	94
P (same)	0.89993	0.80368

Table 3.4 Comparison of parental contributions to the three hybrid populations.

	C. perifretum alleles	*C. rhenanus alleles*	*ancestral C. perifretum alleles*	*ancestral C.rhenanus alleles*
Ijsselmeer	67	26	70	53
Mosel	63	29	63	61
Sieg	59	29	63	65
P (same)	0.94162	0.97911	0.94418	0.8042

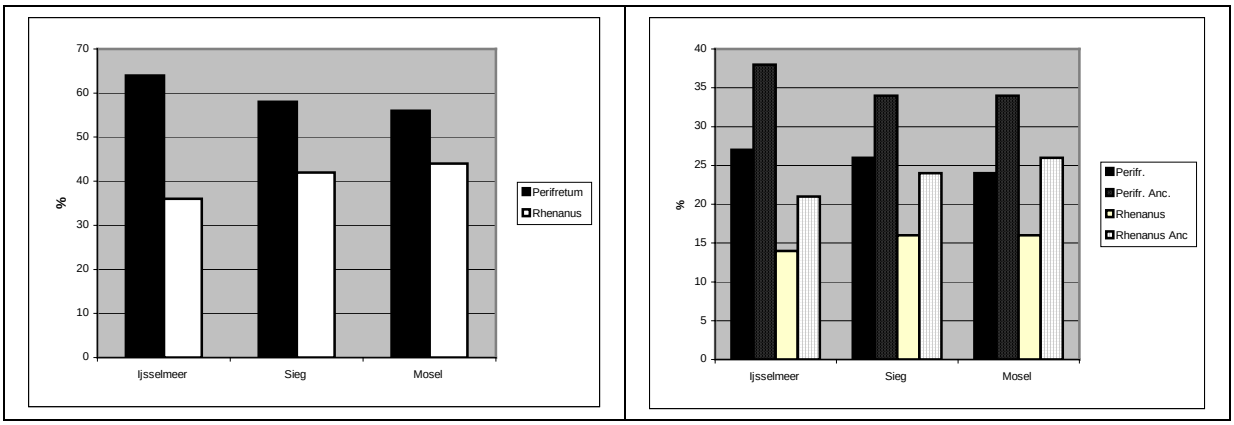

Figure 3.5 Comparison of parental contributions to the hybrid populations

When the comparison is conducted with the four allele classes, this trend of parental contributions is not so obvious anymore. Again no significant difference in parental contribution to the three hybrid populations can be observed for any of the four allele classes.

3.5 Localization of marker loci on the *Cottus* genetic map employing conserved synteny to the *Gasterosteus* genome

BLAST searches were conducted with all potential ancestry-informative loci against the genome sequence of *Gasterosteus aculeatus*. Since only loci, which yielded significant hits on the Gasterosteus genome, were included in the screen for ancestry-informative loci, all SNP and indel loci could be associated with one *Gasterosteus* linkage group. By employing the conserved synteny between the *Cottus* and the *Gasterosteus* genome, the approximate localization of each ancestry-informative locus could be inferred. According to the *Gasterosteus* chromosome, on which a significant hit was detected, marker loci were assigned to the associated *Cottus* linkage group (Figure 3.6). In some cases more than one *Cottus* linkage group is associated with one *Gasterosteus* chromosome (*Gasterosteus* linkage groups I, II, VII, and XI). In these cases it is not clear, if marker loci are actually detected on all of the associated *Cottus* linkage groups. For *Gasterosteus* linkage group XV, the corresponding *Cottus* linkage group is not known, because none of the loci included in the *Cottus* linkage map yielded a significant hit on this group.

Marker loci are not distributed evenly over the different linkage groups. The number of markers assigned to the different linkage groups ranges from 1-10 with an average number of 4.5 markers per *Cottus* linkage group. Marker spacings in Fig. 3.6 do not reflect actual distances between the markers but represent the order of markers on the linkage groups as inferred from the hit positions in the *Gasterosteus* genome.

Except for *Cottus* LG 22, which only contains derived marker states from *C. perifretum*, a mix of derived and ancestral states is found on all linkage groups.

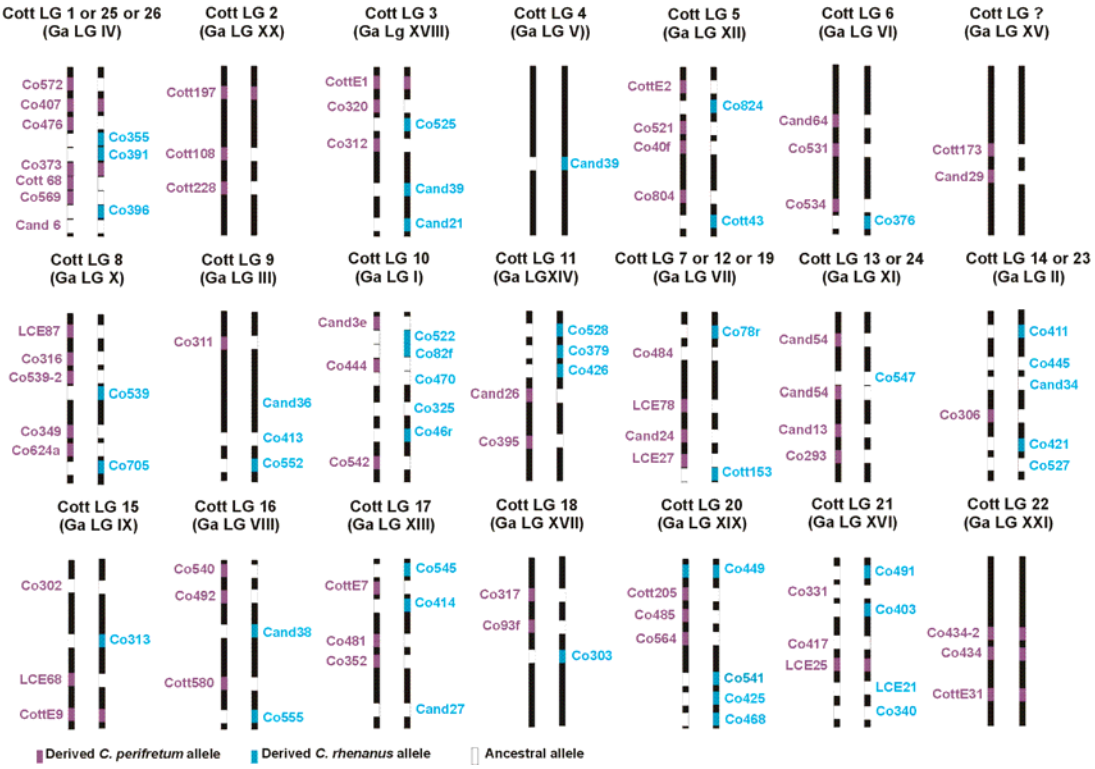

Figure 3.6 A map of the hybrid genome indicating the marker states found at each locus. The relative positions of the markers on the *Cottus* linkage group are inferred from the conserved synteny with the *Gasterosteus* genome.

3.6 Gene content of marker loci

According to the positions in the *Gasterosteus* genome, more than half of the 108 marker loci can be found within coding regions. Furthermore, 31 loci lie within 10 kb up- or downstream of coding regions (22 upstream and 9 downstream). Only 21 markers are not found within the vicinity of coding regions (Supplement 3).

When loci with fixed and mixed ancestries in the hybrid lineage are compared for gene content (Tab. 3.6) no significant difference for these two classes can be found for any of the comparisons. Therefore mixed and fixed marker loci contribute evenly to these four categories and marker loci with fixed ancestries are not preferentially found in the vicinity of or within coding regions.

Table 3.6 Comparison of gene content between loci with fixed and mixed ancestries in the hybrid genome.

	Fixed loci Total (n = 24)	Fixed loci %	Mixed loci Total (n = 84)	Mixed loci %	P(same)
Within coding region	11	46	45	54	0.92953
Within 10 kb upstream	7	29	15	18	0.99129
Within 10 kb downstream	2	8	7	8	0.79726
No gene	4	17	17	20	0.9492

4 Discussion

4.1 A genetic map of *Cottus*

Constructing genetic maps based on F1 crosses is not a common approach, but could not be circumvented in this study, due to the relatively long generation time of *Cottus*. However, a basic genetic map could be established, which is supported by the colinearity with the genomes of *Tetraodon* and *Gasterosteus*. The preliminary genetic map contained one inflated linkage group, which raised the idea that a genomic rearrangement might have occurred between the hybrid lineage and one of the parental species, *Cottus rhenanus*. This idea evoked further analysis of this linkage group, since a rearrangement could have been of great importance for the divergence process of the hybrid lineage. Rearrangements and especially inversions are thought to be able to play an important role in the process of sympatric or parapatric speciation, since they can protect the rearranged regions from gene flow (Livingston & Rieseberg 2003). Especially in the case of hybrid speciation were the newly emerging lineage is found in sympatry with the parental species, chromosomal rearrangements can contribute to isolation, especially when they act synergistically with isolation genes (Rieseberg 2001). Lai et al. (2005) could show for the three hybrid sunflower species *Helianthus anomalus*, *H. deserticola* and *H. paradoxus* that karyotypic rearrangements are found in these species, resulting from the sorting of parental chromosomal rearrangements and from *de novo* rearrangements. The majority of pollen viability QTL occurred on rearranged chromosomes and mapped close to rearrangement breakpoints.

New mapping families were therefore established for the hybrid lineage and for *Cottus rhenanus* to resolve the question of a possible inversion or rearrangement. A combined linkage analysis of all families from one lineage indicated differences between the two maps, which might have been caused by rearrangements including small-scale inversions and one insertion, but these differences could not be validated by the analysis of the single families. But independent of these remaining uncertainties, these results suggest that a large inversion, which could protect a considerable part of a hybrid chromosome from gene flow, can be excluded, at least for linkage groups 3 and 21-26.

To finally solve the question if rearrangements exist between the hybrid lineage and the parental species, the establishment of F2 generations or backcrosses will be necessary in order to obtain reliable linkage maps. Furthermore, mapping families of *Cottus perifretum* are needed as well, in order to determine if rearrangements can be detected between the hybrid lineage and one of the parental species, or maybe even between the parental species. From the currently available data however, large-scale chromosomal rearrangements between any of these lineages cannot be expected.

4.2 The implications of conserved synteny between *Cottus* and model organisms

The tackling of specific evolutionary questions often requires working with non-model organisms. However, when it comes to understanding the genetic basis of an evolutionarily interesting trait, the limited genetic options in non-model organisms may prohibit even standard approaches that are commonly used in model organisms.

In order to conduct genetic analysis, like the mapping of a hybrid genome, a linkage map has to be constructed. In non-model organisms, it will often only be possible to obtain an F1 cross for mapping, which limits the map resolution. It is therefore of special interest to assess in how far completed genome projects can aid such efforts in non-model organisms. Studies in plants have already been conducted to evaluate whether microsyntenic relationships exist between model and non-model plant species. Colinearity can generally be observed at the level of genes within flowering plant families and could aid fine-mapping and map-based cloning experiments (Schmid 2000). The results shown here suggest that the same may also hold for teleost fishes.

Microsatellite markers provide both a system for polymorphism analysis and a system for anchoring the locus via the sequences that flank the microsatellite repeat. However, since microsatellites normally reside within non-coding regions, it is often thought that they can only be matched with relatively closely related species. Interestingly, Rico et al. (1996) had already found that a given microsatellite locus can be amplified across a large range of fish taxa. Here I found that almost half of the flanking sequences from *Cottus* yield a significant match with *Fugu* and *Tetraodon* and 84% yield significant hits on the *Gasterosteus* genome.
Intriguingly, the matches occur with highly conserved short stretches of unknown function. Given the large number of hits that were detected, it would seem that the density of such conserved non-coding regions is very high in these fish genomes. While it is generally interesting to speculate about the functional role of these sequences (Gaffney and Keightley 2004), they also turn out as potentially highly useful tools for linking genome information between diverse fish species.

Given the known partially conserved synthenies even between mammal and fish genomes (Grützner *et al.* 2002; Jaillon *et al.* 2004), it is not surprising that evidence for highly conserved synteny between the fish genomes themselves could be found. Already a simple map construction strategy as the one employed for the *Cottus* genetic map, in conjunction with an only partially annotated genome such as *Tetraodon*, already yields clearly comparable chromosomal parts and this picture becomes even more convincing when the *Cottus* genetic map is compared with the more closely related and better annotated *Gasterosteus* genome. Nevertheless, intrachromosomal rearrangements have to be considered, which do not allow direct transfer of all positional information from the *Tetraodon* or the *Gasterosteus* to the *Cottus* genome. The comparison of the genetic maps already suggests that inversions or transpositions exist between the *Cottus* and the *Tetraodon* and *Gasterosteus* genome respectively. However, it is not clear how this picture will change when a more reliable map of the *Cottus* genome becomes available. Still, because of the apparent high density of conserved sequence elements, it will be possible to trace microsynthenic relationships, even if the whole chromosome segment is rearranged, or fused to another chromosome.

Figure 4.1 shows a sketch of the phylogenetic relationships between the major fish lineages. *Fugu*, *Tetraodon*, *Gasterosteus* and *Cottus* belong to the Acantopterygii

(Nelson 1994), which include also medaka (*Oryzias latipes*) as a further genome for which full sequence information will soon be available. The interrelationships within the Acantopterygii are still under debate, but both Nelson (1994) and Miya (2002) agree that *Cottus* (Scorpaeniformes) is more closely related to the Tetraodontiforms (*Takifugu rubripes, Tetraodon nigroviridis*) and to *Gasterosteus* than to the Atheriniforms (*Oryzias latipes*). The other major model fish, *Danio rerio*, belongs to the Ostariophysi. Given that we find about a quarter of the *Cottus*/*Tetraodon* matches even in *Danio*, it would seem that it will be straight forward to link genetic markers that are found in any of these teleost fish species to known genome information of one of the model organisms.

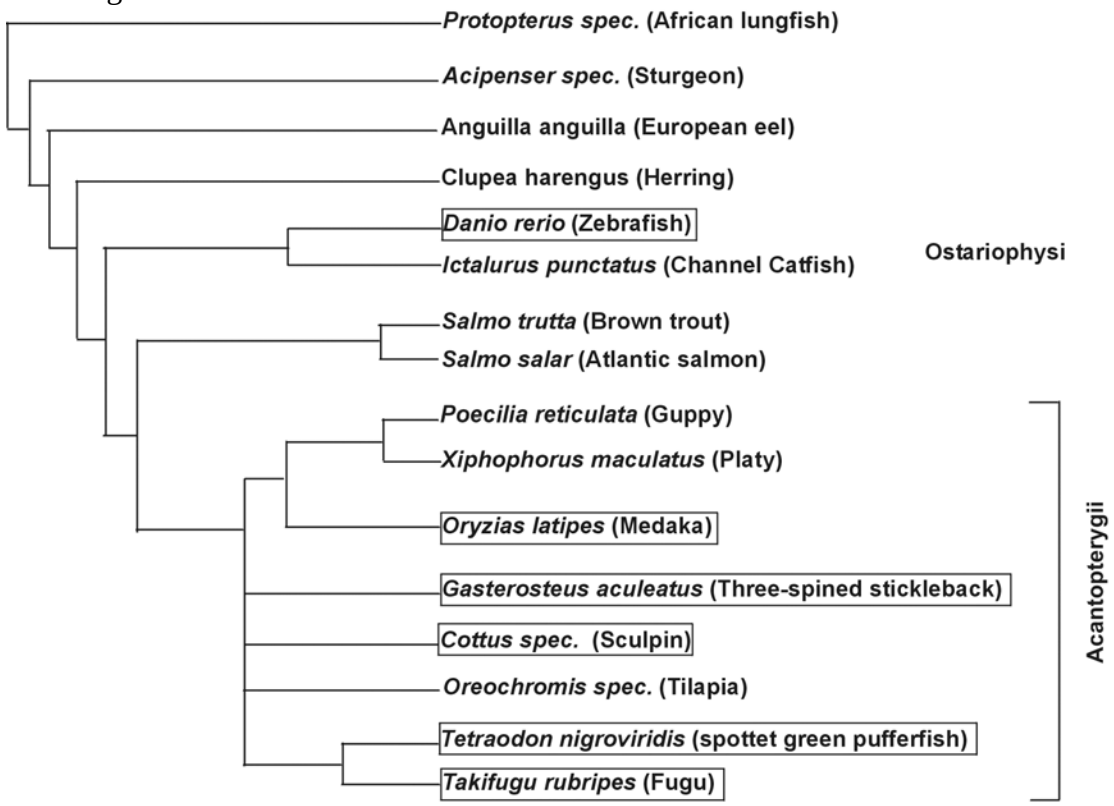

Figure 4.1 Schematic cladogram illustrating the relative phylogenetic positions of model fish species such as *Danio, Orizias, Cottus, Tetraodon, Gasterosteus* and *Fugu* among other teleost fishes of special interest. Based on Nelson et al. (1994) and Miya et al. (2004).

For the future research on the hybrid *Cottus* lineage, the available genomes of other model fish species are a valuable resource, which might speed up the search for candidate genes responsible for the success of the hybrid lineage in a novel habitat.

4.3 Corresponding signals from genetics and morphology

Loci in the hybrid genome, which are fixed for one ancestry, contain an excess of *C. perifretum* material. This reflects the morphological similarity between *C. perifretum* and the hybrid lineage. Furthermore two of the loci where we find fixed *Cottus perifretum* states correspond to trait loci, which have been identified with the help of an admixture mapping approach (Nolte Phd thesis). This study analyzed the correlation between morphological and ecological characters and the occurrence of specific microsatellite alleles in order to detect quantitative trait loci (QTL) responsible for diagnostically different morphological characters.

Two of the microsatellite loci (LCE21 and CottE9) from this QTL analysis were employed in the screen for ancestry-informative markers and yielded fixed derived states for *Cottus perifretum*. In the admixture mapping approach, these two loci were significantly associated with skin prickling, a morphological trait which is found in *Cottus perifretum* and in the hybrid lineage, but not in *Cottus rhenanus*. Three other loci which, were fixed for ancestral states potentially received from *Cottus perifretum* (co413, co547 and co340), could be found in close proximity to skin prickling-associated loci. This proximity is only inferred from the synteny relationships with *Gasterosteus* and thus needs further support. According to the hits on the stickleback genome, the fixed loci are between 80 and 700 kb apart from the skin prickling loci (Tab. 4.1). This corresponds to a distance of maximally 1 cM in the *Cottus* genome (1 cM = 590 kb). According to Briscoe et al. (1994) and Collons-Schramm et al. (2003) admixture linkage disequilibrium extends over 5-20 cM for the time frame appropriate for the *Cottus* system. This implies, that the fixed loci in the hybrid lineage might well be associated with the potential skin prickling QTL. It is not surprising, to find fixed *C. perifretum* regions to be associated with prickling loci, since this morphological character is found in the hybrid lineage. This finding turns these regions into interesting candidates for further research. If genes that underlie skin prickling are really found in these regions than it would be of great interest to find out if this morphological character became fixed due to a selective advantage for the hybrid lineage or if it became fixed by chance.

The admixture-mapping analysis suggests, that in many instances loci associated with one specific trait seem to be physically linked. In many instances, skin-prickling loci were grouped into regions with a distance of less than 20 cM between significant markers, implying genomic cohesion of genetic factors that determine *C. perifretum* morphology. This suggests that one should observe large chromosomal blocks with fixed *C. perifretum* ancestry in the candidate regions associated with skin prickling. For locus LCE21 one can indeed observe, that the two neighboring loci (co340 which also lie in close proximity to a prickling locus and LCE25) are also fixed for *C. perifretum* ancestry (LCE25) or for ancestral alleles potentially derived from *C. perifretum* (co340). To validate this hypothesis, however, a finer mapping of the hybrid genome is needed in order to define blocks with different ancestries more precisely.

Table 4.1 Fixed ancestry-informative loci, which might correspond to regions identified in an admixture-mapping analysis (Nolte Phd thesis)

Ancestry-informative Locus	*Fixed ancestry*	*Admixture mapping Locus*	*Associated trait*	*Hit on Gasterosteus Linkage group*	*Position on Gasterosteus Linkage group (bp)*	*Distance inferred from synteny*
LCE21	*C. perifretum*	**LCE21**	Prickling	XVI	12.922.306	-
Co340	Potentially *C. perifretum*	**Cott146**	Prickling	XVI	14.075.930/ 14.611.755	540 kb
CottE9	*C. perifretum*	**CottE9**	Prickling, Habitat	IX	18.334.973	-
Co547	Potentially *C. perifretum*	**Cott78**	Prickling	XI	10.104.235/ 10.793.926	690 kb
Co413	Potentially *C. perifretum*	**Cgo56**	Prickling	III	14.662.844/ 14.742.507	80 kb
Cott197	*C. perifretum*	**CottES21**	Habitat	XX	4.905.894/ 4.387.625	510 kb

Another character studied in the admixture mapping analysis was habitat association of *Cottus rhenanus* and the hybrid lineage. CottE9 is one of the loci

associated with this trait. One other locus fixed for *C. perifretum* ancestry (Cott197) might also lie in close proximity to a habitat-associated locus.

Potential candidate regions for habitat association are of even more interest because if hybridization is really responsible for the adaptation to the novel environment than one should not only find loci derived from *C. perifretum* to be associated with this trait but also loci with *C. rhenanus* ancestry.

4.4 Hints for ongoing gene flow between parts of the hybrid genome and the parental lineages

A comparison of the three hybrid populations for parental contributions revealed, that the contribution of *C. perifretum* is highest in the Ijsselmeer population and declines slightly over the Sieg to the Mosel population. The opposite trend is observed for the *C. rhenanus* ancestry. These differences can only be caused by different parental contributions to the loci with mixed ancestries, since all of the hybrid populations show the same parental states at the fixed loci. This finding can be explained by a scenario, in which the Sieg and the Mosel population collected more *C. rhenanus* material on the way up the River Rhine and/or the Ijsselmeer population, on the other hand, still has some influx from *C. perifretum*. Does this contradict our idea, that the hybrid lineage is a separate entity? The answer to this question depends on what is to be called ‘distinct’. Mallet (2007) proposes to define species as genotypic clusters that remain distinct even when hybridization and gene flow occur. This implies that gene flow is allowed for some, but not for all parts of the genome. A similar concept has been proposed by Wu (2001) in which he states that genes or a set of interacting genes are the unit of adaptation and not the whole genome. In his view speciation starts with a few differential adaptations between two populations or races. In a next step more differential adaptations and a certain degree of reproductive isolation are acquired (for example through epistatic interactions of differentially adapted genes with other genes) and populations can still fuse or diverge further. At the next level the divergent populations are beyond the point of fusion, but still share a portion of their genomes via gene flow. Only in the final step complete reproductive isolation is achieved.

If the trend of different parental contributions in the three hybrid populations is indeed an indicator for ongoing gene flow in some parts of the hybrid genome, than on the other hand a few adaptively important regions might be sufficient for the maintenance of the integrity of the hybrid lineage. This idea needs further support first of all through the estimation and comparison of actual allele frequencies between different hybrid populations and furthermore through the analysis of gene flow across contact zones between the hybrid lineage and both parental species.

4.5 Speculations about the hybridization scenario

In the hybrid lineage 9 polymorphic SNP loci could be detected (Tab. 4.2) where one allele could not be explained by allele states from one of the parental species or from the outgroup species. One explanation for this finding is, that these alleles were not sampled in the parental or the outgroup species, due to the relatively small amount of pooled samples. If this is the case, these alleles must have risen in frequency in the hybrid lineage, since they could be detected readily in the pooled samples of 6 to 10 individuals. Another explanation is that these alleles are only found

in the actual source populations that gave rise to the hybrid lineage. These source populations are not known to date. Furthermore other lineages than the proposed parental species could have contributed to the hybrid gene pool. This scenario cannot be excluded, since also all of the ancestral alleles found in the hybrid lineage could have come from any lineage, which retained the ancestral allele state. One hint however, that this is not the case, comes from the comparison of contributions from the different allele categories to the hybrid lineage. The same signal was obtained, no matter if allele states were grouped into 'derived *C. perifretum'*, 'derived *C. rhenanus'*, 'potentially ancestral *C. perifretum'* and 'potentially ancestral *C. rhenanus'* or if 'derived' and 'potentially ancestral' states were considered together for each species. If the ancestral alleles have been introduced into the hybrid genome from any other than the proposed parental species, than one should expect a different signal, when alleles are grouped into four categories. Therefore it seems unlikely, that other lineages than the proposed ones contributed a considerable amount of genetic material to the hybrid lineage.

A final explanation for the polymorphic SNP loci are private SNP alleles of the hybrid lineage itself. This would not be surprising but rather expected given the considerable amount of fixed SNPs that are found between the two parental species. If the average rate of fixed SNPs of 1 in about every 1300 bp (fixed SNPs between *Cottus perifretum* and *Cottus rhenanus*) is taken to estimate the overall amount of nucleotide divergence between the parental genomes this leads to an estimate of 0.078 %. This would be the divergence rate per million years since these two species are thought to have diverged 1 million years ago. This rate is somewhat higher than the nucleotide substitution rate of 0.02 – 0.05 % per million years observed in flanking sequences of microsatellite markers of diverse fish species by Rico et al. (1996). There is probably no reason to believe, that the nucleotide substitution rate is considerably lower in the hybrid lineage. Therefore the lack of fixed derived states allows speculations about the emergence and the age of this lineage. If there would have been a founder event, including only a small amount of fit hybrid genotypes, genetic drift alone should have led to the fixation of private alleles. Since such alleles cannot be observed it has to be suggested, that the hybrid lineage emerged from a considerably large hybrid population and furthermore, that the hybrid lineage is very young. The latter suggestion is supported by the recent occurrence of the invasive lineage in the River Rhine less than 20 years ago and furthermore by the history of the Rivers Rhine and Schelde as already mentioned in the introduction.

Table 4.2 Single nucleotide polymorphisms found within the parental species and the hybrid lineage. Parental and shared parental polymorphisms have been detected within the 427 fragments screened for ancestry-informative loci (205.828 bp of total sequence) whereas the polymorphic loci in the hybrid lineage were detected within the 108 ancestry-informative loci (44.084 bp of total sequence).

Polymorphisms in:	C. perifretum (205.828 bp)	C. rhenanus (205.828 bp)	Shared between C. perifretum & C. rhenanus (205.828 bp)	Hybrid lineage (44.084 bp)
Total	117	99	23	7
Per bp	1/1700	1/2100	1/8900	1/6300

Further speculations about the hybridization scenario can be made by comparing the amount of loci with mixed and fixed ancestries. Only 22 % of the analyzed loci are fixed, demonstrating, that most parts of the hybrid genome are not yet stabilized in terms of chromosomal block size. Ungerer et al. (1998) estimated the time span for the formation of the hybrid sunflower species *H. anomalus* based on a junction clock, which can be established due to the recombinant nature of hybrids that

leads to an accumulation of junctions following the hybridization event. This junction clock stops once the hybrid genome becomes stabilized and parental species blocks become homozygous (Rieseberg et al. 2000). At this point the distribution of junctions provides an estimate of the speed of hybrid speciation. For *H. anomalus* only about 60 generations were sufficient to create the chromosomal block sizes observed in the hybrid species today. However, differences in the arrangements of chromosomal blocks were observed between different haplotypes suggesting that some polymorphism for genomic composition may have been maintained or alternatively, that drift led to the fixation of slight differences in genomic composition among geographically isolated populations of *H. anomalus*. Ungerer et al. (1998) suggests, that the major part of the hybrid sunflower genome became stabilized, before a population expansion. This does not seem to be the case in the hybrid *Cottus* lineage. Presumably, only some parts of the hybrid genome became stabilized due to selection before the hybrid lineage expanded into the new habitat. The remaining parts of the genome can still recombine to smaller block sizes and either remain polymorphic or eventually become stabilized by drift. The maintenance of a certain amount of polymorphism is rather the rule than the exception looking at the number of ancestral polymorphisms (1 in about 8900 bp), which are still found in the parental species (Tab. 4.2) that diverged 1 million years ago. Stabilization of some genomic regions by drift could lead to differences in genomic composition between geographically distant hybrid populations. It remains to be analyzed in more detail if such differential fixation events already contributed to the modest differences in parental contributions, which can be observed between the three hybrid populations.

4.6 What does it take to be a hybrid species?

In his review on hybrid speciation Mallet (2007) states, that in contrast to polyploid hybridization it is hard to define homoploid hybrid species. This is mainly due to the fact that an even contribution of both parental genomes cannot be expected, if backcrossing has been involved in the speciation process. In these cases it becomes hard to distinguish introgression from hybrid speciation. He suggests to restrict the term ‘ hybrid species’ to “cases where hybrid allelic combinations contribute to the spread and maintenance of stabilized hybrid lineages generally recognized as species”. In which respects does the hybrid *Cottus* lineage fit this definition? First of all, only parts of the hybrid genome have become stabilized so far. However, one should expect that these stabilized parts contributed to the ‘spread and maintenance’ of this lineage. As mentioned earlier, the lack of fixed derived marker states might be an indicator for a rather large hybrid source population. If this conclusion is right, the fixed genomic regions in the hybrid genome must have been under selection in order to become fixed. This argument becomes even more plausible, when the young age of the hybrid lineage is considered. The stabilized genomic regions contain material from both parental species, thus exhibiting ‘hybrid allelic combinations’. Yet it remains to be demonstrated, that traits from both parental lineages actually formed an adaptively advantageous combination in the hybrid lineage.

5 Conclusions

With the help of an ancestry-informative marker system and by employing the conserved synteny between the *Cottus* and *Gasterosteus* genomes, it was possible to map the genome of the hybrid *Cottus* lineage. It could be shown that the hybrid genome received genetic material from both of its proposed parental species *Cottus perifretum* and *Cottus rhenanus*. The three hybrid populations studied do not exhibit significant differences in parental contribution, indicating that the hybrid lineage is a distinct entity. However, a slight difference in parental contributions can be observed at loci, which harbor alleles from both parental species, which could either be an indicator of ongoing gene flow between parts of the hybrid genome and the parental genomes or a sign of differential fixation of parental chromosomal blocks by drift.

A large part of the hybrid genome is not yet stabilized in terms of parental block size, yet the fixation of other parts of the hybrid genome is an indicator that the specific regions have been under selection and might thus be adaptively important. It remains to be shown that the combination of advantageous genetic material from the parental species allowed the hybrid lineage to successfully invade a novel habitat.

6 Literature

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipmann, D.J. (1990) BLAST: ‚Basic local alignment search tool'. J. Mol. Biol. 215: 403-410.

Arnold, M. L. Bulger, M. R. Burke, J. M. Hempel, A. L. Williams, J. H. (1999) Natural Hybridization: how low can you go and still be important? Ecology 80, 2, 371–381.

Barton, N. H. (2001) The role of hybridization in evolution. Molecular Ecology 10, 551-568.

Brink van den F. W. B., Velde van der, Cazemir, W. G. (1990) The faunistic composition of the freshwater section of the River Rhine in the Netherlands: present state and changes since 1900. Limnologie aktuell 1, 191-207.

Brisco, D., Stephens, J.C. & O'Brien, S. J. (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. Journal of Heredity 85, 59-63.

Burke, J.M. & Arnold, M.L. (2001) Genetics and the fitness of hybrids. Annu. Rev. Genet. 35, 31-52.

Cazemir, W. G. 1988. Fish and their environment in large European river ecosystems. The Dutch part of the River Rhine. Sciences de l´eau 7, 97-116.

Chowdhary, B.P., Raudsepp, T., Frönicke, L. & Scherthan, H. (1998). Emerging patterns of comparative genome organization in some mammalian species as revealed by zoo-FISH. Genome Res. 8: 577-589.

Collins-Schramm, H. E., Chima, B. Operario, D. J., Lindsey, A., Criswell, L. A. & Seldin, M. F. (2003) Markers informative for ancestry demonstrate consistend megabace-length in African American populations. Human Genetics 113, 211-219.

De Nie, H. W. (1997) Atlas van de Nederlandse Zoetwatervissen. Media Publishing, Doetinchen, 151 pp.

Dolezel, J., Batos, J., Vogelmayr, H. & Greilhuber, J. (2003) Nuclear DNA content and genome size of trout and human. Cytometry 51A: 127-128.

Dominguez, I., Graziano, E., Gebhard, C., Barakat, A., Berry, S., Arus, P., Delseny, M. & Barnes, S. (2003) Plant genome archeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species. Plant Biotechnol. J. 1: 91-99.

Dowling, T. E. & Secor, C. L. (1997) The role of hybridization in the evolutionary diversification of animals. Ann Rev Ecol Syst, 593-619.

Englbrecht, C.C., Largiader, C.R., Hänfling, B. & Tautz, D. (1999) Isolation and characterization of polymorphic microsatellite loci in the European bullhead Cottus gobio L. (Osteichtyes). Mol. Ecol. 8: 1966-1968.

Englbrecht, C.C., Freyhof, J., Nolte, A., Rassmann, K., Schliewen, U. & Tautz, D. (2000) Phylogeography of the bullhead Cottus gobio (Pisces: Teleostei: Cottidae) suggest a pre-Pleistocene origin of the major central European populations. Mol. Ecol. 9: 709-722.

Erickson, D.L., Fenster, C.B., Stenoien, H.K. & Prices, D. (2004) Quantitative trait locus analysis and the study of evolutionary process. Mol. Ecol. 13: 2505-2522.

Freyhof, J., Kottelat, M. & Nolte, A. (2005) Taxonomic diversity of European Cottus with description of eight new species (Teleostei: Cottidae). Ichthyol. Explor. Freshwaters 16, 107-172.

Gaffney, D.J. & Keightley, P.D. (2004) Unexpected conserved non-coding DNA blocks in mammals. Trends in Genetics 20: 332-337.

Garbe, J. & Da, Y. 2003. Locusmap user manual, Version 1.1 (Department of Animal Science, University of Minnesota).

Gebhardt, C., Walkemeier, B., Henselewski, H., Barakat, A., Delseny, M. And Stüber, K. (2003) Comparative mapping between potato (Solanum tuberosum) and Arabisopsis thaliana reveals structurally conserved domains and ancient dupliations in the potato genome. The Plant J. 34: 529-541.

Giampietro, P.F., Raggio, C.L. & Blank, R.D. (1999) Synteny-defined candidate genes for congenital and idiopathic scoliosis. Am. J. Med. Gen. 83: 164-177.

Gompert, Z., Fordyce, J. A., Forister, M. Shapiro, A. M. & Nice, C. C. (2006) Homoploid hybrid speciation in an extreme habitat. Science 314, 1923-1925.

Grützner, F., Lütjens, G., Rovira, C., Barnes, D.W., Ropers, H. & Haaf, T. (1999) Classical and molecular cytogenetics of the pufferfish *Tetraodon nigroviridis*. Chromosome Research 7, 655-662.

Grützner, F., Crollius, H.R., Lutjens, G., Jaillon, O., Weissenbach, J., Ropers, H.H. & Haaf, T. (2002) Four-hundred million years of conserved synteny of human Xp and Xq genes on three Tetraodon chromosomes. Genome Research 12, 316-322.

Hänfling, B. & R. Brandl (1998) Genetic differentiation of the bullhead Cottus gobio L. across watersheds in Central Europe: evidence for two taxa. Heredity 80, 100-117.

Hänfling, B., Hellemans, B., Volkaert, F.A.M. & Carvalho, G.R. (2002) Late glacial history of cold-adapted freshwater fish Cottus gobio, revealed by microsatellites. Molecular Ecology 11, 1717-1729.

Hardie, D.C. & Hebert, P.D.N. (2003) The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. Genome 46: 683-706.

Jaillon, O., Aury, J., Brunet, F., Petit, J., Stange-Thomann, N., Mauceli, E., Boneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., de Beradinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J., Gouzy, J., Parra, G., Guigo, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schlachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J. & Crollius, H.R. (2004) Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431: 946-957.

Kalendar, R. (2003) Fast-PCR v. 2.9 (for Windows). http://www.biocenter.helsinke.fi/bi/bare-1_html/oligos.htm.

Kontula, T., Väinölä, R. (2001) Postglacial colonization of Northern Europe by distinct phylogeographi lineages of the bullhead, Cottus gobio. Molecular Ecology 10, 1983-2002.

Lai, Z., Nakazato, T., Salmaso, M., Burke, J. M., Tang, S., Knapp, S. J. & Rieseberg, L. H. (2005) Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. Genetics 171, 291-303.

Lelek, A. & Köhler, C. (1993) Erfassung der Fischfauna des deutschen Niederrheinabschnittes mit besonderer Brücksichtigung der Rheinsohle im Hauptstrom. Studie für die Rheinfischereigenossenschaft, Bonn, 27 pp.

Livingstone, K. D., Churchill, G. and Jahn, M. K. (2000) Linkage Mapping in Populations With Karyotypig Rearrangements. The Journal of Heredity 91, 423-428.

Livingstone, K. & Rieseberg, L. (2003) Chromosomal evolution and speciation: a recombination-based approach. New Phytologist 161, 107–112.

Mallet, J. (2005) Hybridization as an invasion of the genome. Trends in Ecology and Evolution 20, 229-237.

Mallet, J. (2007) Hybrid speciation. Nature 446, 279-283.

Mavàrez, J. et al. (2006) Speciation by hybridization in *Heliconius* butterflies. Nature 441, 868-871.

Mayr, E. (1963) Animal Species and Evolution. Harvard University Press, Cambridge, Massachusetts.

Meyer, A., Salzburger, W. & Schartl, M. (2006) Hybrid origin of a swordtail species (Teleostei: Xiphophorus clemenciae) driven by sexual selection. Molecular Ecology 15, 721-730.

Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M. & Nishida, M. (2002) Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phyl. Evol. 26: 121-138.

Morgenstern, B. (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15: 211-218.

Nelson, J.S. (1994) Fishes of the world, third ed. Wiley, New York.

Nolte, A., Stemshorn, K. and Tautz, D. (2005a) Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Molecular Ecology Notes 5, 628-636.

Nolte, A. W., Freyhof, J., Stemshorn, K. C. and Tautz, D. (2005b) An invasive lineage of sculpins, *Cottus* sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. Proceedings of the Royal Society B 272, 2379-2387.

Nolte, A. W. (2005c) Evolutionary genetic analysis of an invasive population of sculpins in the Lower Rhine. Phd Thesis.

Nolte, A. W., Freyhof, J. and Tautz, D. (2006). When invaders meet locally adapted types: rapid moulding of hybrid zones between two species of sculpins (*Cottus*, pisces) in the Rhine system. Molecular Ecology 15, 1983-1993.

Rico, C., Rico, I. & Hewitt, G. (1996) 470 million years of conservation of microsatellite loci among fish species. Pro. R. Soc. Lond. B 263: 549-557.

Rieseberg, L. H., Sivervo, B., Linder, C. R., Ungerer, M. C. & Arias, D. M. (1996) Role of Gene Interactions in Hybrid Speciation: Evidence from Ancient and Experimental Hybrids. Science 272, 741-745.

Rieseberg, L. H., Baird, S. J. E and Gardner, K. A. (2000) Hybridization, introgression, and linkage evolution. Plant Molecular Biology 42, 205-224.

Rieseberg, L. H. (2001) Chromosomal rearrangements and speciation. Trends in Ecology and Evolution 16, 351-358.

Rieseberg, L. H., Widmer, A. Arntz, A. M. Burke, J. M. (2003a) The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. Philosophical Transactions of the Royal Society of London B 58, 1141-1147.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J. L., Schwarzbach, A. E., Donovan, L. A. & Lexer, C. (2003b)

Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301, 1211-1216.

Schleuter, M. (1991) Nachweis der Groppe (*Cottus gobio*) im Niederrhein. Fischökologie 4, 1-6.

Schlötterer, C. (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365-371.

Seehausen, O. (2004) Hybridization and adaptive radiation. Trends in Ecology and Evolution 19 (4), 198-207

Schwarz, D., Matta, B. M., Shakir-botteri, N. L. & McPheron, B. A. (2005) Host shift to an invasive plant triggers rapid animal hybrid speciation. Nature 436, 546-549.

Schmid, R. (2000) Synteny: recent advances and future prospects. Cur. Op. Plant Biol. 3, 97-102.

Schreiber, A., Engelhorn, R. & Riffel, M. (1998) Auswirkungen der Fußgeschichte von Rhein, Neckar, Doubs und Donau auf die Populationsgenetik von Groppen (Cottus gobio) und Bachneunauge (Lampetra planeri) in Südwestdeutschland. Verhandlungen der Gesellschaft für Ichthyologie 1, 185-102.

Ungerer, M.C., Baird, S.J.E., Pan, J. & Rieseberg, L.H. (1998) Rapid hybrid speciation in wild sunflowers. Proceedings of the National Academy of Science 95, 11757-11762.

Vitturi, R. & Rasotto, B. (1990) Karyotype analysis of *Cottus gobio* L. (Pisces, Cottidae). Cytobios 62: 81-86.

Voigt, C. Hofer, B. (1909) Die Süsswasserfische von Mittel-Europa. Engelmann, Leipzig, xxiv+558 pp.

Volckaert, F.A.M., Hänfling, B., Hellemans, B. & Carvalho, G.R. (2002) Timing of populatione dynamics of bullhead Cottus gobio (Teleostei: Cottidae) during the Pleistocene. Journal of Evolutionary Biology 15, 930-944.

Voorrips, R.E. (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93: 77-78.

7 Supplement

- Supplement 1 Primer list of loci included in the screen for ancestry-informative markers
- Supplement 2 Count of allele states in the hybrid lineage
- Supplement 3 Gene content of marker loci

Supplement 1 Primer pairs for all loci included in the screen for ancestry-informative markers.

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
Cand10	TAATGCATTGCATCACCCACTGCAGA	CATTTTTTTCAAGACTGTCTGGCATTGG	912
Cand28	CATTTGATGCGTGGAATTCTGCA	TGATGGATTAAGCGGCGCGTGATGCT	924
Cott100	TCCTTTTCATGCCATTTTCC	AGGGACGTTTCCCAGTGTC	354
Cott119	TGCTTGTGAACCGAGTCTTG	ACCCAGGTCAGGCAGAGAG	499
Cott132	ACAATCAGGGAAAGTCTGGG	ATGGAGCCATGAAAGAGCAC	315
Cott138	TYTTCAGCAGCTTTATCCCG	CGTGAACGACACTCTGATCG	459
Cott144	CCCAACTTGCTAAAAATGGC	CCAGGGTGTTGGTTACAAGG	378
Cott149	CACAACAGCCATCACTGGAC	TGGCAAATGCACAGCTAAAG	358
Cott152	CTACGGCTTGAGATTGGTCC	CGATCATCTCACTGCAGAATC	318
Cott153	AGCGGCTTCTAATCCAAATG	AGGTGTGGACCGAGATGAAC	359
Cott154	AGTTTGGGTCGCACAATACC	ATGTTGTCCAGGTGCTTTCG	340
Cott158	AGCTGATGACACAGACACGG	CTTTGGCTGAAAGACGAACC	365
Cott164	ATGGCCAGACAGACAAGAGC	ACTAATGCCTGATGCAACCC	624
Cott170	ACATGGTGCATAATGTTGCC	CTTGCTCACTTCTGCGTCTG	322
Cott179	AACGATGGCATTTCAAGGTC	GCTCTGAATGAAACGGAAGG	472
Cott183	TTGTTGTGCTTGAGTGGGAG	GCCATGACATCATTGTCACC	499
Cott184	GAAACACACATAATAGAAAACGGG	ACACACACACACACACACGG	351
Cott207	ATCATGAAGTCCTTGTCGGG	ATGAAGGAGTTTCATTGGGC	311
Cott214	CAACGACAGAGGCTTTTGG	TAAATCCCATCTCCCTCGTG	306
Cott272	TGTTGTTGATGTTGATCGGG	AGAGGAGAAGGCTACCTGGC	347
Cott293	GAGAGAGAGAGTCAGGTGAGGC	GCGATTTAGACTCCTGTGGG	311
Cott323	CCCCATGATGAGAGAAGAGG	TTTGAGTGTCCTGAAAAGCG	361
Cott328	TGGGACACAGATGTTTAGCG	ACTTGTGTTTGTGTGGGCTG	421
Cott580	CTCTCACACGCACACTTTCTG	CACACAAACACAGTGCCCTC	364
Cott582	TGAGTCGAGGTGAAAGTCCC	CTGGGGATGAAGGTGATGTC	441
Cott675	AAAGAGGCAGGCTGTTTGTG	CTTCCTTTCCTCCTTCACCC	334
Cott78	AGGATCAGACGGGTATGTGC	CTTCCTCAGATGGCCGTTAC	682
Cott108	TAAACATGCCCCCGTGTAAC	ACCAACTGTCACCGTCATTG	375
Cott722	TCTTGAGATCTTTCTGAGCATCAC	AGACCTCCATTAGGCAGCAC	367
Cgo1034f	GCTGGATTTACCACAGCCAC	TTGCTGCGGTTTATTGTTTG	510
Cgo1017f	AAACCCACACTCCACCTCTG	GATGTCAGGGAGGCTGAAAG	350
LCE81-SNP	TTATGTTATTTGTATTTGTTTCGGG	ACAATCTCGACAGTTCAATG	271
CottE30-SNP	GCAGCTCAGTAGAAAGCGGA	TGAATGTGGAAAGTGATTAGAACC	294
Cott697-SNP	AGCCAAGCGACCATCAATAG	CCCCCGACAGCTCAGATATT	316
Cott570-SNP	TGAACAGAAAAGTAGATTTGTG	GCAACTAAAGCGAGACCACC	326
LCE51-SNP	ATAAGCGCCAGTCTGAAAGG	CTCTCGCATGAGGTTAGCAA	328
Cott688-SNP	ACAGAATCTGCTCGACATGC	GTACCCCTGGTGGTCTGACA	330
CottE23-SNP	TTGCCAAGTGAGCAGCTTTA	CGTGTGAACATTCGTGCTCT	336
LCE52-SNP	CAATACTGGCAAAAGTGACACA	TGATATCGAATCCAGACGAGG	340
Cott210-SNP	AGCAAATAGTTCACCCAGCG	GTGCTCAAAGACAGTCACGC	358
Cott313-SNP	GGTTGAGCTCCAGTGTGTGA	TGTCCTGCTCTTGCTCAGTG	358
Cott684-SNP	TTGATACACTGACTGCAATGAACT	CAGTGAAAGGCGAACACAGT	361
Cott300-SNP	GCTGTAGACTTTATGAGCAGCG	TCTTCTGATGCGCTCTTTTCT	367
LCE100-SNP	TGTGCTAAAGGAGATGACCAGA	TCCCCCTATCGTGGATGTGTT	370
LCE69-SNP	CGTTTTCTCTCACAATCCAGG	CCCCCTCCTTTTAATAAATCA	372
Cott50-SNP	GAGATGATGTCATCCCTCTTGT	TCACCTCGGTGAGTCCACA	373
Cott173-SNP	GCAGCCTTGTGTTGATCGTA	AAGAATGAACCCTGTGTGGG	388
Cott163-SNP	CAATCACTGCATCCCATTTG	ATAGGGCTTGTGTCTGAGCG	389
CottES19-SNP	GGCACTTGAACACCATCAAA	AAAAATCCTCCCATCCAAAGA	389
Cott564-SNP	AAGTGGGTCATACTGGGACG	GGTTAGAAATGTTGGCAGGC	394
LCB13-SNP	TTGTGACACATTGATACACCCA	GGGCTCAAATGTTCTACCGG	400
Cott222-SNP	AGCTTTTCCCCTTTCTGCTC	GCAAAGATGATGGATGGAAGA	402
LCE279-SNP	GCTCAACTTCAAATGAGCCA	CAATGCAGGTGTTTTAGGCA	405

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
Cott213-SNP	TANATGGGTTTGCCATGGAT	CGATGTGTGCTGATGCAAAT	411
LCE219-SNP	GAGTTGTTTCACTGCGCAAA	GCATCTGCACGCATTCTCTA	414
Cott175-SNP	TTTGTGTCTGTGTGAAGGGTG	AAAGCTGGCAGTTTGGTCTG	421
LCE80-SNP	CCCCTCGAGGTCACGGTAT	CACATCCTAGCATTCCTGCTT	422
LCE13-SNP	TTGTGTTTCTGTAGTGGGGCT	AGCACCCATGCCTTTTTATG	424
LCE54-SNP	CAACACACTGCTTCCCACTG	ATCCAGGATCCCTGCAAAGT	425
Cott98-SNP	GGTTCATCCTATCCATGAACAAA	GAACTGCAGGACACAGCAAA	425
Cott687-SNP	CCCTAATCTGTGTCAAAATCACA	ACTGGGCAGGAAACAATGAC	427
LCE126-SNP	CAGCTGGCACATGACTTCAC	AGGTTCTCCTGTACCCCCTC	432
Cott250-SNP	ATCTTTGTTTTAGGCGCAGC	GCAAATCGTGCAATTGAAATC	433
Cott29-SNP	TGTTTATGCGCAGACAGAGG	TGACAGAGTTTAGACTTGCCCT	437
CottE10-SNP	GAGTCCTGAGTAACAGCAGCA	GGCACTGGTAATGAACTGCTC	437
Cott205-SNP	CAAATGTGCAAATATGGCTGA	CACAAACGAATTGCTGCTGT	440
LCE29-SNP	ATATTGGAAGGGGAGGCAAA	TCCTCTTTCATCACATGCACA	449
Cott91-SNP	TCTAACTTGTGGCCTGGTGA	GATTGAGTGTGTGGCTGCAT	455
Cott228-SNP	TTTTGCCCTTTGTCTCTTTCA	TTATTTTCGGGGTAAACGACC	456
LCE55-SNP	AAAAAGTACTCCCATAAGTCGGC	GTGAGGAATATCTCTGCCCG	457
CottE31-SNP	GACGTAACCACCCGACCAC	AGTCAGGACCAGTCGCACTC	458
CottE32-SNP	GCCGGAAGAAAACTTGACAG	GCTCACCGTTGCTGTGTCTA	461
CottE7-SNP	GAGGAAGACTCGAGAGGAATGA	CTTGCTCCTCCCAGAATGTG	466
Cott68-SNP	TCACCCCTTTACGTTTCTAGATATT	CAGGCCCCTCAATTGAATC	466
LCE122-SNP	TTTGAAAATGCTGCCCACTT	CACACCGATAGATAGAGCGACA	466
Cott146-SNP	CAACCAGCAAAAGGCAAGAT	GCGGCTTGGACTTGGTATT	469
LCE181-SNP	GCTCTTGCTTAAGACTCGCC	CTTCCTTCTGGTCACCTCCA	472
LCE25-SNP	CTGAGCCGGTGACGTCCT	CGAATCAAATAATCAGGCTTATCC	475
LCE111-SNP	TGCCTCTGATGCTGATTCAT	GGGTGATTCTGTTTAGGCCA	483
Cott188-SNP	GTACAGCTTCTTCCCGGGTT	CCCTACGATGGGAGGTGTG	489
CottE8-SNP	CAACGAAATGCAGTTTAGCATC	AAAATCGCGTCAGCTTTTGT	505
LCE39-SNP	GTGGAAGGTGGATGAGCAAA	TTTCTGTTGGCAGTACAAAGTCT	508
LCB67-SNP	TGTTCTGCAGCTCAGAGTCG	ACACACAGACTAATGACAGG	517
CottE12-SNP	TCAGACGTGTTGTTTGTTTGC	AAAAGTGGAATGAGAAAGAGAGAGA	517
Cott348-SNP	GGAAAGGCTGCAGACTCAAG	CAAAAATGACAATGCAGAGCA	519
Cott197-SNP	GGAACCAGGATTAGGTCCTC	CAGCAGGAAAAATAAAACACGA	526
LCE43-SNP	ACAACGTCAGGGAAATTTCCACG	GGATCAATGCGAGGGTAAAA	532
LCE38-SNP	TGCATGGTTTAGATGTTTCCTTT	ATGGTCATTAACCAGTGGGC	543
Cott255-SNP	TGAAGATAACGTGTCTGCCTG	GGTACTGCTTCTGCAAACTGC	549
Cott584-SNP	TTTTTGCTATCATTACACAGGCA	TCTGAGGTTCATCCGGTGTC	554
Cott118-SNP	TCTCTGTGCCACTGGTCTCC	ATGAGAGTGGGTCATCTCGC	555
LCE48-SNP	CCCTCAGGTCACGGTATCAT	GCAGATCAGCTTCATACATTTTT	573
CottE2-SNP	CCAGAGATAAAAAGGACGGG	TTTCCTCCTCCTCCTCCCT	583
Cott708-SNP	TGAGTCCTGAGTAATCCAATAATTC	GTTTGTTTTGTTAGTGCCGGA	589
CottES10-SNP	CCTCGAAGGTCGACGGTAT	TAGAACTAGTGGATCCCCCG	590
LCE76-SNP	TGGTTTCATAGCCATTTGGG	TGCTTTTGGGAGATAAACATGA	591
Cott43-SNP	TGTGTAGGAGATGCAGTAGGGA	ATGCCTGACTGAATTGTGGG	594
LCE83-SNP	ACAACCGGCGGATCCTTT	AAACAACTGTTTGCAGAAGCAT	595
LCE79-SNP	TTCTCCTTTTTGTTTTGAGAACG	TTTCTTACTAATCTTGTTTGGGCTG	599
Cott128-SNP	AAGCATGTTTTGTTTCTGTTTTGA	AAAGCACTAAAAGTTGAGAAAGCA	601
Cott221-SNP	GGAACTTCACACCGCCACTA	TCAAATATCCAAATGATGATTGC	604
Cott296-SNP	CAACTGCTGCTCCATGTTTATC	TTGCTAAGCGCAGACAGAGA	611
LCB12-SNP	TCGAGTGAGGTAATGATAGCTGA	TTTGGTGAGTATTTGTGGATCA	614
LCE22-SNP	AGCGAAATAAATGGAAACCG	GCTGATATCGAATTGCATCAAA	625
CottES1-SNP	GCGGCCGCTCTAGAAACTA	CCTCGAGGTCGACGGTATC	635
LCE105-SNP	TCAGAAGGATTTCAATCGGG	GCGGTAATGTATCCCTGCAC	642
CottE1-SNP	CATGGTCATGACAGAGCTGC	AAATGTACAATTTTGCTTCCCTG	645
Cott700-SNP	GGGAGATACTCTTTACAGTGGGC	TGGAATTCGTCATGTAACCG	645
CottE16-	CCCCTTCACCTCCGTCAG	GTCACACCAGCCAGTGGAG	661

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
SNP			
LCE42-SNP	AAACAGATGGCGGAGATCAG	AGGAGTACGAGCCGAGCC	663
LCE89-SNP	AGCTGATATCGAATTGACTCAAAGT	GGATAGTTGTTTGAGTTAACAGGCA	693
Cott105-SNP	CGATCGTGTATCCCTTCACC	GAAGGAATGAAGTGAACAGTGAAA	696
LCE75-SNP	GGAGACAAATGTTAAATGTAAATGG	AAGGCACATGACATTTTGCTC	697
LCE68-SNP	CTGCACTTAGTCCCTTTGACC	GCAAACAATCCGGGTATACAT	715
LCE275-SNP	CAGCGATGTATGTCTTCAGTCAA	GATGTGCGTCCCTTTTCAGT	776
LCE78-SNP	TGAAAAGTCTCGGGAAGCTG	AGAGGAACGGCGTATGTCAA	811
LCE40-SNP	CCCTCAGGTCACGGTATGAT	ACAAGAGCCACAAACAGGGTG	812
LCE66-SNP	TCGCCTCAGAAGAGGTTTGT	TTCTCATGCAGAGACCTGACA	817
LCE37-SNP	TGCTTTCGGTTCGTATTTGTT	CCCTCCCATGCAGATACTGT	827
LCE32-SNP	ACTCAGATGTGCTTGTGGTTTGA	TCGAATTTCATTTATCTGCTTCA	829
LCE74-SNP	TCATGACCCCTTTAAGTAACTGC	CTCCACGTCCTTCATTACACC	833
LCE11-SNP	CCTGGAAACTGGAAGCTCTG	AAAATGCAATACCTCTCTCTGTGA	865
LCB4-SNP	ACGAACCCACAGAGTCAGGA	ATCAGGCTCAGAAACGGATG	944
LCE31-SNP	ATGACTGTTCAAGGTCCGACA	TCAACATCTTGAATGTGCCC	952
SNP-Cand13	TGGTCTATGTATACCTGTCAGCTTG	GATCCAGATCAGAAATTGGACC	786
SNP-Cand19	ACACATTCACCCCCTCAAAA	TGGAAACATAATGTGGTGGAA	1216
SNP-Cand6	TTCTCTCAGAAAGCCATAGTTTGA	TGTCCTCCCTTTGACGTGAC	1108
cand1	TTGTGTTTGCATGTCAGCAGAG	GAAAACCGTGCTGCCGATAAGC	838
cand11	TGACTAATCTGAGTGCGTGTC	ATTGGGCCCTCTAGATGCATGC	1245
cand12a	AGGCACATAAAAGACCTCCAC	GCTGAGCATAACAACCATCCCA	645
cand12e	AAGATTGAAGGGCATTTCCCT	CACGACACTCGAAACGCCGCTG	637
cand15	TAAGCTTCACACACAGATGCCTGG	GCAAATTCAGCCATAACGCCT	902
cand16	CCATACAGGTGAATACAGTGATCC	TGTACTGCCTCAAAAGCTACACAG	739
cand2	GTATCATGACTGACATAGCCGGCA	CGAGTCAGAATTGGACTCCCGTCG	803
cand20	AGCAGAAATGTCATGCTTTGC	CCTGCAATCATATGGAATGACCCA	1105
cand21	GCAATTTGGATACCCCGGCGAGTG	GACGCCAGGAATGGGAAGTGCACC	613
cand22	AAAGAAACGCTCACTTCGACTC	TGAGTTGCTTAAGTTCTCCATGG	1072
cand24	TATTTAACCAGGTCGGCCCTG	GAAGAAGCAGTTACACGGATCTT	873
cand26	TAACCCAGCTGGAGCAATCATCG	ACGATTGCAAAATGTCCATCG	924
cand27	TATCAAAAATGGAGCGGGCTCTA	TCAGAGTTGCCAACAATGACAGC	538
cand27e	TTACGTCAAATTGAGGACTGGAG	ACCTCCATGAGCACGCACACAC	558
cand29	CCCATTCAAAGTGATGCAAACAGC	ACAATGCACACTATTTGGTCGTCG	1164
cand30	TGACATTGAGATTCTTGACCCAG	GCCAGCACTCTCAAGCAGCACG	518
cand32	GTTGCTTTGGATAAAAGCGTCAG	GTATGGCGAGTCACTATGGGCAC	1216
cand33	AACATAGTCAACCTCAGTGCCCT	CATATCGATGTGAGACAGCTGAG	850
cand34	ATCAACCCAACTATGCTCATGG	TAACGAGTGTAACTTGTGCCCA	1008
cand35	GTCTTTCATTGATGGCTCGTGAG	ATCGTGACCATAATGTCCTGTTGC	1112
cand36	ATGAACTACCCACCCCACTGG	TGCTTTGGTGAAAACCAATGCCA	1176
cand38	CTTTATCAACACAGCAGGTGGT	CTCATTGCCAGTGGTCCAGGGA	1134
cand3a	TTGACCTTCTGAGTCAGAGGCAGG	CGTTCAAACATTCCCGCAGAG	529
cand3e	TCTCCAGCATGAGGATGGGACC	TCATTTAAGGAGCCGGCATGAT	608
cand40	TCTTTGAGTTAGGGCTGGGCGGTA	GGGTATTTTTCCAAGTAGGCCA	1181
cand42	GTCGCCCCATCTGTTGCTGAGC	TAGCTTCCATCAGTAGACAGTGTG	1242
cand43	ATCAGCACAGCGCCGGCCATTCTC	TAATGCGCAATCTGACTCAGTG	1313
cand44	ATATGTCGTTGCTGTCGTTGCTGG	TGCATCATGAACACAGCACAT	671
cand45	AATAGCCAATCTTCTCGCTGATGG	CAGATATTGGGACAATCTGGTCAC	513
cand46	CTATCAGGTGTGATGTGAAACAGC	CAACAAACTGTGACGTTAAGGCA	1328
cand47	CAGCTCATCACCTATGGATGAGTG	TCTAGTTGGAACAACATGTGCCCA	1079
cand48	TATGTGTGTGTGTGTCCGGTAGAG	ATGGGTCCAAAAGCAGGACGA	962
cand52	GAGGAGCCCTGATGATGCCGT	ATGAGTCCTGAGTAACATCTCCAG	997
cand53	GAAGTTTCAGTTGATTACCGGCT	GGCCTCGAGGTGTTCTCGGGTCTT	834
cand54	AACTGGCTTTGTTGTGGTCTCCGA	TACCCCCTTGGGGCAACTCAG	930
cand56	ACAATGAAACCAGGTTCCAGC	AGTCTGGGTGACCTTTTGTGCA	803
cand59	TCTTGCAGATGAATGAGTCCTG	TGAGTCCTGAGTAAGCAAGTTCTC	986
cand61	TGAGTCCTGAGTAACCTCAAGC	GTAAGGTCTGATCTAATTGGCTGC	1038
cand62	GGAATTCTGCAGATGATGAGTCC	CAGTGGTGATGGATATGAGTCCTG	1027
cand64	ATGAATAACCACACACACTCGGCT	TCCTGAGTAACTGTAAAGCAGTGC	915
cand67	GCCAGTGTGATGGATGATGAGTCC	TGAGTCCTGAGTAAGGTCCGT	912

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
cand69	TGAGTCCTGAGTAATGGAGCAGG	GAGTAACGTATTTGTGCTTGTGG	962
cand70	TGAGTCCTGAGTAATCAGAGCAG	ACACAGCGTTATGTTTGAGCCCTG	929
cand9	AGGAGCCTCTTTATCTGCGTTGG	GACGGATGTACTGGCTGCCCA	1032
cont39	TGTGTTGCAGATGTATCGGATGAC	TTCTCCAACAATAACAGCAGCAG	1005
cand13	TCAAGTGTGTCTGTAGGGGAC	CCTAACAGTGATGAGACCTCTG	786
cand19	ACCCCCTCAAAAAGCCCACGGGT	CATGACGTAACATGAGGTGTCTT	1216
cand6	GCAAACCACATCTTCTGCAAGC	CCCTTTGACGTGACACACAGAA	1108
co385	TGCAATGGTCATTATGCTGAGTG	GAGAGACAATGCCCGTCTACC	615
co387	CGTATGTGGCTGTAATGTTGTGC	TAAATGTGACCCCCAGCAATGTG	988
co388	TTGTTCATCAGGGTGACGGCCA	TTTGAGAGACTGGTGTTGACCCAC	1080
co389	TTTCATCTGCGAGAGTGAACTGG	GAAGCAGTGAAACACCGTCTC	926
co391-m13	TGCAGGTTGTGTGTGTAATGCGA	AACTGTACCCAGCATGCAGTTCAA	715
co392-m13	GCTCGAAAAAACGGACCGCGTT	CATACGTGGACTTGTCACGCT	640
co392-sp6	TACCCGGGGATCCCACAGGGAT	ATCGTCGTGTATTCCGGACAG	646
co394-m13	TAATGGCCGTGATATGAAGCCGT	GACACAATGTAACTACATGGTGC	537
co395	TAAAGCAGGATATCGGCTCAG	TGTTAGGTAGCAATCCATGACTC	1126
co396	AAACTTCCCTGACAACAAGCA	GATAAGGAAGTGCCGCCATGTC	809
co397	CATAAAAGTGTGAGAGTGGCCCGT	CCTGCCCCAGTGTCGGATCAT	1294
co398-m13	TGGTCACGACAAGGCACACGT	CACACACACACACTAAGGTGATGC	583
co398-sp6	AGAATAAAGCAACTTCGCCCA	GCCGGAGAAATGAACGGACTAGC	486
co399	TAAGCAGGTCGAGCACCCCAC	GCAGACAGGGCCGTGCGATATGTG	1002
co400	ACTTCACAGATTACCTCCGGCA	AACCTGGAAGCTCATTTTGTGCCA	935
co402	TAGCACATACTTAGGTGAGGTGC	TCTACATCAAAAGCACGATGGCA	1120
co403	GGTCTCCTCAAATATGCACCAA	TAAAAACTGGGCCCCAGCTTGTC	1084
co404	GTGTGTTTGGGAAATTAGCTGCA	TCAAAAGCCACACACAGTCCCA	1211
co405-sp6	AATCCAGTTCACTCGAGCGCT	ATCCGCTGCCCTGATGCAGACACG	417
co406	TGTGTAATTCAGACAGGAGCTC	TGTTACAGATGGTCACTTGAACG	1083
co407-sp6	AAACACAGACACTCTCTGAGC	TAGTAGATGTTGGCGGGGCTCTGC	594
co408	TAACAGACGAGGAGTCAAACGCT	TAATGAGCCCTGGCACTGCTT	983
co409	CCGCTACAAATAAGTCGGTGTC	GATTTGAAATGAGACCCCATCAC	829
co410	AGCCTAATGGATGAGAACTGC	GCAAGTGTTATGCTGGGCGCGTA	870
co411	TGGAGAGTCTAAGAACATCGGGTG	TCATTCAGTGGTAACAACCAGC	1096
co412-m13	ACACTTCCAAATGAGGGGGCA	CTGTCAAGATGAAGCTCACGCT	472
co412-sp6	TCCAAAGTGACTTACACAGCA	CAGAGGAGTAATCAGATCCCCGT	367
co413	AAGTTTCAACGGACACATGCA	TTACAGCACTAAGTGGTTCAGAG	918
co414	TAATCTCGCTGAGTCATCCAGAGG	CCAAATAATTCCGGGTGGTCGA	1094
co416-sp6	GAAGGAAACTTGTCTCCCGTGC	GTTACCGCTGAAAGCCTCTCG	515
co417	GATCATCTGTTTGTCCCGACAG	CCTATGACGATGTAATGTCTCCAC	655
co418-m13	TGGCGTGGTAAAACCGGGACAT	CATTATGCAACAGGAACAGTGGGT	709
co419	CTGTAATGCGCTATACAGGGAGG	AGTGAAAAGGCAACGCTACTC	1100
co421	AGCCTGAAGGTCGTCCAGGTG	ACTCATTGCCCAACCAAAAACG	1304
co422	TCTTGAGTAAAGTGCCACTGTG	ATCACCACCTTGTCCCTGACGGA	1286
co423	ATCACTTGTAGTTTACAGCCCTG	GATCTCAGTCATTACTGTGCCA	1238
co424-m13	ATATTAATTACGTGGCGCCGTCAG	AGACAGCTGTACCAATGTCTCCAC	646
co424-sp6	AATAAGCTTACCGTCTCATGCCT	TCATCCAGGCCGTCAGTCCAA	578
co425	AACCCATAAAGCAACTGCTCTTCC	AGTCCTCAGGTAGTTGTCAAGGCT	1002
co426	TTGATAACGGTGCTGCAATGG	CTCAATTAGAGCGTTCAACACAGC	1034
co427	AACCTCGTCCAATATCGGTGC	AACGCGTGTGATATTTTGCCCTG	1115
co428	TAAGTGTGCATCTGGCCGAACAG	AATGAGATGTTCTTCAGGTCGTC	1156
co429	GGTCACACAAATATTCCGAACC	CACTCGGACTCCTGACAACGT	864
co431	AATCCTTTGCAGTCAATGACAGC	GAAAGAACGCACTGGTGAGCT	891
co432	AGCACAGAGGTTTTCACTTCTGG	GTATCTGCTCACATCGATAGCCGA	980
co433	AATATGCGCGGAGCCCTTTCAA	AACCCAACCTGACCTCCACTGAGG	839
co434-m13	GCGGTTTACATCATTCAGATGCA	GAAGTGATGACCAAACTGGCCT	708
co434-sp6	GTCGAACACAAATCACTCTGTCG	AGTACAACATCTGGTTGCCCCGCT	432
co435-m13	TGATCTTAAGGCTCAAGTTGGGA	ATACACTTTGGAAACCCCGCA	465
co436-m13	TAACAACAACTGGATGTCGCCA	GCCCGATCAGTGTTTCAAGTCGAC	662
co438	TAAGAGTCCCAGGACCCACAC	CCAGTGATGAGGGGACATGTCTGG	918
co439	AGGACAAAGTGCACGTTGGCCA	CAACTGAGCAAATTCACTCGTG	991
co440-m13	TCACTAAACTTGGAGACCTGG	TGGATTTCCTCAATGGCCGGA	727

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
co441	CAGGAAGTAGGAAAGCACCCCGGT	AGACTTGCTACTGTACTGACAGTG	1124
co442-m13	TAATATACAGAGGTTAGCGCGTG	TTAGCCTTGTGTCGCCATGCA	603
co443	CAAAATGGAGTACTCTGCATGAC	AGGTAACTTGTCCTAACACAGTCC	945
co444	AAATGCACAATGCTGTAGCAC	TAAGGAGTAAGCCCTCACGCA	1259
co445	TAACCAACACCACGGTGACTG	TAACTGCGAAATCCAGCAGTG	1026
co448	TAAGTCTGATCGGCAGCAGCCA	TCCTCGAACTATGCAAATGAGC	899
co449	TAAACCGGTCCCACAGAAGCCA	AACACAGACTCGGTGTCTTTGGCT	1293
co450-m13	GAAAATGATCCTTTGTAGCCCGCT	CAGTACGACTGTAAAACAGTGAGC	493
co450-sp6	CAGCGAGTTTCATGACGATCAC	TGTTTGCTAGTGTAGAGTGTTGG	325
co452	AGCATGCTTGCGCACACACAC	CATTTCAAAGGAATCGCTTGTCC	790
co453-sp6	CAGGAACCAGGAACATAGCGGCCT	TCGCTTTGGATAAAAGTGTCAGC	355
co454	CGCCGGTGCTTATAGTCCAGGA	TATGCAAAGTTCCAGCGGATTACG	915
co456	GCTGAAAGACACAGGAGCATCAT	ACGCAGTATCGACTATCGGTATCG	1050
co457	GATGAAACTGGCCTTTGCGGT	AGCATAAGGCTATGTGCAGGT	945
co459	AAATATGCTCTGTCCGTGGCA	AGGTCAGTTGGTGTTACACACTCC	959
co460	GGAAAAGTGAATCTTCCACTCAC	TCCGTTGTTTGATGTAGCGAC	1141
co462	TAAATCGTCTGGACGCCGCAA	GTTTTTACAACTCTGTGGGGAGGT	997
co463	ACATCTGTTTCCTGTGCAGGGT	ATTCAGGGCGAAGCAGAACTC	1279
co464	AGGGTACCTGTTCAATGGCGT	AAATAATCCACTCTAGGCAGGCCT	1006
co466-m13	AGGCTTACCTCAATGTGACTACG	TGGAGAATAACGTCAACGGGCCA	653
co466-sp6	ATTTGCCAGCTGTACAGTGTCAC	GGGTCGCCTGGAATCAGTTTGTGC	455
co467	ACTAGCTCTGCGTTGGCGGAA	TAAATGGGTGTTGAGTGCGAC	965
co468	TTTCATTAGGTAGGAGGCAAGCCA	AATTGCCTTGTAAATGGCTGC	960
co470	ATGGTCTGTTGAAGCATTACCCT	GCTCTGTGAGAACATCTCCGCCCT	852
co471	TTGTGTTTACGAGAGTGTTGCGA	AGGGTAGCCACTCTCACACTGCT	1235
co472	AGAATATGTGCACCTCTTAGGCCT	AGTGCTCTTGTTAGTGGCTGAC	1055
co474	AGATGCATTCCTGAGATTCAGCAC	GAGACTGAAGATACATGTTCGTGG	910
co476-m13	ATGACTTGAGGGCCTGTCAGC	CATTACTGCGCACCTCAAGAGTCG	607
co476-sp6	GAAATTTATTGGCCAGCCGCTCTC	GGTGGAATGCATAATGTCATGACC	503
co477-sp6	TCCTGCAAAAACAGGACACACGGT	GTGGCCTCTAAGTGAGTGCTG	601
co478-m13	ACAATACACTGCTGTATCCCGTG	CTTCACTGAACTGATACTGGGGGT	472
co479-m13	CAGCTTGATACAATCTGCTTCG	TAACTTTAGTCAGGATGGTGGGCA	660
co479-sp6	CTGCTGAGAGTGAAAGCACAACTT	TTCATATAGGAAGGTATCCGGGCA	432
co480	AATCTTCAACCCAGCATTGGT	CAGTTATATTGGCCAGCACAGAT	846
co481	TACATTTGCAGAGAGCAAAGCCCT	ATACTACACTGTTCCATGAGCCGT	897
co483-m13	GTGCTTACAGTTACAGTCGGCCCT	AATGAGACAGAACGGCTTCAT	431
co484	TAAGGGATGAGAGAACCACGATCC	TGTCATGACCCGGGCCAGGAA	959
co485	ACTGGGTTTGTCTGAACTCTGCA	AGCGTATCTTTTGAACTGGGAC	973
co487-m13	GTTTAGGATGTTTGTTGGCCGAA	CTAAAATCAACGCTGAGCTCC	428
co487-sp6	GTTACAACTCTGGCAATACAGC	CTTTGTTCTCTAAGCAGGTAGCA	441
co489-m13	AACCTAACTGGGTCACTCGCA	AGTGGTGGGACATTCACTCGTT	664
co489-sp6	GCAGAACTGCTCAGAATTCGCT	TGTATTGACGAAGTGATGAGGTC	602
co491	ATCTATTCACACACAGCAGGAACC	TAACTGAAGTCGCCGTGCCGAC	913
co492	TTTATGGGCTTTGAGGCTAAGAC	TAATGAAGGACGCCACTTGCT	1093
co494	GGTGACAAATAGGTTGTCGGTC	CAGCGCTGCATCAAAGGGCAG	601
co495-m13	TTCAAAAACAGAACTGACGGTGC	ACGTAACGTCCCTCCAGCAGC	492
co495-sp6	TAAGAGTTGAGTGTCCTGATGTG	CGTAGCCTGGGATCAAACCACC	594
co497	GATTGACAAACACTCGGTGTTTCC	TCTGCACTGGGTTAATGCAGTCAG	1258
co498	TAGGCAAATGGAAAGTCCGCGCT	TGGTACCTACAGTTGAAACAGTCG	901
co499	CATCAAAGCCACATATGGACTCCA	CGATCAGATTAGACCCAAACACGA	417
co502	TTTGATGAGACACAGCAGAGTGTC	TTACAGAGACCTTTTCAGAGTGG	977
co503	TGAGTGTATGAACTATGGCTGTG	GCTGTAACTGAGCATTAGGGA	856
co505	CTAGGAACCACAAGTAGCCCCGCA	GCACTTACTTCCATTACGCGTGGA	852
co506	AATATAAGCAGGCATACGCTCTCC	AACATTAGGACTGCATGTCCA	1057
co507	TGAGGAGTGTCAACAAATCCACGA	AGGAGCCTGGAGCCATAGCAG	852
co511	ATCTGGTCAATATCAGCATCCAC	TGACCAGTAAACTGTAGTGCTG	806
co512	CCCTTGACGAACTCCACAGAG	GAGATACCCAAGTACATTCTGCCA	1118
co514	TAAACGTGGGTATTTGGATGCAG	TGAGAAAAAACTGACTGGCCTCAC	943
co516	GGTTAAGTCCCGTCACGAGTCCT	TAAGTCCTCAGGCCCGCAGGCAA	702
co518-m13	AATTCGCGTCATTCGCTACTGG	CAACAATAACGTGCAACCGGT	496

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
co518-sp6	CAGATATCATGGTGGGACCGT	TTCGTCCTCCAAAGTGGGGCT	487
co519	TGCTAATGAGACCACAGAAGC	TGGGGTTATTTGCATTGCATGG	789
co520-sp6	ACACATACAAGTGAAGCTCGT	GTCAGCTGCAGTTAGCCTTGAAGC	432
co521	CTTCTCTGAGACCGCCAGCTG	TACAAAATGAGGCATCAAGCTCC	754
co522-sp6	CTCCAACTCTGAAAAGCAGAGTC	CAAAATAGCATCACAATCCCGAG	522
co523	AAGGATCCTGTCACGGACCAC	CGGCGGAATCCTAGTCAGACGGA	775
co524-sp6	TCGGTTCAACACAAATGCGGCT	ATGGAGTCCAGATTTCCTGGTG	464
co525-m13	TAACAGCCCTGAAACTGGCCGT	ATTCCACTGTTTGGAGTCCCAGC	678
co526	ACACTGCTCGTTAACCCGTTTGAG	ACCTATTTGCATCAGCAGATTGG	883
co527	GTCCTCAGACATGTTAGCGGCGCT	CACCCATCATTCAACCAGGAA	922
co528	TGTCCAATATGCTCGGTTGAA	TACCCGTAGAATAGGTGGCGGA	1302
co529	CCCTTCATTTGGTGGGAGGTA	TAATGATAGTGTGGCAATGGCTC	1012
co531-m13	TAAACAGACACCATCAGGACC	CATGCATGACGAAACATTTGCTGC	741
co532	TCCCTCTAGTGATCGGACGGA	CTATCAGTCAGTGACATCAGTGG	1187
co532-sp6	TGCCGAGCAAACAAACGAGCTGC	GCGCGTATCATGTATCACGTGAC	566
co533-m13	CCCTGCAGAATTACCCAAAAGTG	GCGTTTAGAATTTGACCCTGC	521
co533-sp6	TAAACTACAGAGACGCCGCACAA	GGATTCCAAAACTCTGATCGCA	413
co534	CGCATGTACTCACTTGTGACG	TGGAATGTACCCATGATTCACGGA	878
co535	TGATTGGCTTAGAGACACTGTG	GGTTATTTATCGGCTGAACACCT	865
co536	TGTATTGTGTCAGAATGTGGTCG	CTGGATACAAGGGCCGTTTAGC	1213
co537	TGTCTGATTTGACAGGGCGCA	GCTGAATGTCATTGACTTGCTG	1095
co538	AGGAGCGCCTCTGGGTTCAGTC	GTAGCTCACCTGGGCGAGCAT	994
co539-sp6	TGTCGGCGTGGAGCTGCTGTT	AGATATACTCAGCGTATGCCTGC	406
co539-m13	ACTTGCTAATGGCATCACTCAGG	GCATGATGGTCACCACGCGCT	737
co540	CAGTGTTCATGAGAGCAACAC	CAAACCACGTCTACCAGTAGGA	987
co541	ACAAGATGTTGGTACCTAGATGC	TAATGTGCAACACAAGGTTGGGCA	1030
co542	GGGGGATAGGTTGTTGTCCTC	GGCGTCGCCTTTAAAGCACCA	521
co543	GGTGTGGAGGCTGGAACCTCAG	GCAACATAACACATGGAGATGCGT	1013
co544	TGATTCAGAAACAGTGGCCTG	TAGACTAAGCCTGATTTGCAGC	894
co545	TTCATGCTTAAACACGTCAGAGG	ATTCTGCCGTTAAAAGTGCCTG	841
co546-m13	TCTTAATCCGCCAGCGAAAACAA	TTTGGGATTTGGGCGGCGGGTCAA	434
co546-sp6	AATGCTTTCCCATCGTAACCAGC	GGGTCATCAGATAGAGGACAATGC	592
co547-m13	GTTGAATGTTTGTGGGCTACTG	GATTGTTATCTGGACTGAAGCCAC	605
co549-m13	TAAAAGGAGCGACTGTGGCTCAG	CTCAGAATGTAAAGGGGTACTCAC	470
co550-m13	AGTCTGTCAATGTATCCATGCGT	GGTAGTTGGAAACACACTCCAT	687
co550-sp6	ACTAAACACATCAGCTTGGAGG	TTGTATGTGAGCTGTTGCAATGG	433
co551-sp6	GCGTCTAACGTACCTCCCGCCGTA	CAAGGACTGGAGAATGTTGTCCCT	364
co552	TTAGTATCAAGGCTGCTGACATGC	TGTTTGAGAATCCACAGAGTGC	669
co553	ACACGGGTGAACTACGTTGTCC	TAGTTGCATTACACCACTGGGA	911
co555	CTGGTCTGGTTGGAATGCTCC	GATCTGATCTGCAGCTGCCAC	835
co556	GCCTATTGACTGGAATCAAGC	ATGAGTCGTGATACTTACCAGC	1077
co557-m13	GACGAGTGGAACCCCCAGAAGC	AGAGGACAGCGGTGTTCCCATTCC	408
co560	AAAGCCTCCTAGTTAGCAGATTCC	CATTTTACAGAGCAACTCCGTC	993
co561-sp6	ATGAGTATTGCATACCTGCATCC	ACAGGGTGTTTATCTTGCTGCCGT	620
co564-m13	TGTCCGATTTCCAAAAGGGTCTGG	TGAATGTAAGCGTTACAGGAGCT	446
co564-sp6	AATACGCTAAATGCCAACAGC	GTCTCTTCGTCTGCGTAGACACG	480
co566	AGGTCTGCATCTAGTCCTCATGAC	ACCATGAAGATGTGTCTCCGGT	978
co568-m13	ATTCCCCAGACCTAGCTACGCA	GAAGGATAGGCTCTGGTTCCT	704
co568-sp6	AGGCACCGGAGACCGGATTCCA	TCGGAGGTTCTGGTGGCGGTCCAT	483
co569	CACACACAATAAAGCTTGCTAGC	TCTGTGTCGCAGCGAGGCACAT	426
co569-m13	CTGGGAATTGAGCAGCGCCCAC	ACAGCCCTGGATTTCGACTCAC	355
co570	ATCCGATCAATCTGTGAGTTGTGC	GATTCCGTGTGGGAGCGACAC	799
co571	AGCCCACCTTAAGTTAGCCTGAG	CTTCTATGGATACACACGGGAAGG	1031
co572-m13	TTAATTACTAGCTGACCGCAGAG	TCCAGACTTAAATGTGGGTCTC	498
co572-sp6	GGGCATATAGAAGTACGTCCTC	GTGTACACTTGACGGCAGCAC	529
co574-m13	TCTGTTTGCAGCATGGCATGG	ACACACTCGGCCAGACACCTC	579
co575	TTAGAAAGACAACATCACCCACG	TACCTCCTGAGAGCGAATCAA	1029
cand28	TGATGGATTAAGCGGCGCGTGAT	GATTCACAGGTCCAGCATGAA	498
cand8e	TGTTGGCAAATCCTAAACCCA	CTGGCCACCGTCAAGGTTGTG	484
cand58e	CTTCTATACAACCAGAGGAGTCG	CTGGTTGAGATTCTGAGCCGATGG	440

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
co290	GGAGTGCTCTGTAGACTTTGTGGT	GTGCATAAGTGACCAAGCCTT	576
co291-m13	AGGCCTGGCTAGCTCAGTCGGT	TTGACAGCATTGGAATGGAAGGCT	205
co292	CCACATGAAAGCACCGACCTTGCA	GGATCGGCGGTTCGATCCCTGGTT	666
co293-m13	CCCAATGCAGCCTGACAGCGA	ATGGGCATCGTTTCCAGCTCAT	528
co293-sp6	AGACATCATGTGGCATTCAGCTGC	GGGGGGTTCATCCAGCAAACCCCT	411
co294	AACGCCGAGGTCTGTCCTGGA	GTGACATCCACCGTAGTCCCA	874
co297	GCAGAATTGTGTATGCTCTAGC	ATACATCAGCTGTGTGCCGTCTG	857
co299	CTTTACGAGCAGCGCAGCATGC	GTTTGGTGTTGGCGTTCAGTA	793
co300	CCAAATGTGTCCTCAGCTGACAGG	GCTAACGACGTGCATGACTAGCCA	420
co302-sp6	TCAGGATGGACACCGGAGACGTGC	GGCGGAGCATCCTAAACAGCAGAG	302
co303-m13	ACCCAATACTCTGGTGGCCGAGCA	TCAGTTCACAGTAAACACTCTGG	400
co303-sp6	AAGCAGCCTTAACAACAGCCTGTC	GTGACTGTGGCTCTCCCACAGCTT	374
co304-sp6	CAAAAACACACTTGGACCACTGCT	GCCTGGTTGTAGAGAATCCTGTCC	426
co306	CACAGCTAATCCTGGCGGGCTGAG	AAGCTGTTTGCATTGATGGCCTG	882
co307	CCATTGTTGTGGTGTACAGAGCCA	GCTCTCTGGAGAACGATGACAG	1077
co308-sp6	AGCCATTTGCATCAAGCATCGCA	TACGAGTCAGATAGCAAAGTGGT	292
co309	TTACGTACTGAGCAATGCTGC	AGTTGACATCACATTTGCGTGG	875
co311	CTCTGAGCGTAGGATCAGAGGGTC	ACACTCAGACTTTGAGTCGCGCT	764
co312	CACAGTGTTATGGGTGTTTGCTGG	TCAGTATGTCTAAGCCCAGAGGCA	1136
co314-m13	TGTAGCTGGAGTCAGGACTTCGTT	GTGTTGGGAGCCAAACAGAGGCAT	573
co315	GATGCAACACATTCTCACGCCGA	CGGTCCTCCATTTGAACGGGA	1067
co316	CATATTGGCCTACAAGGCAGCT	ACTATGTGCCTCCGTGTTACGAGC	634
co317-m13	TCGAGCAATTCACAGGACAGGCTT	GTTATTGACCCAGAAGTCTGACC	457
co318	GGTGATTTCAACAGACGAGTCTT	GTCCTTGTAATGTTGGTGCCGA	773
co319-m13	TTGCAAACATCCAAGATGGCGACG	TTGCGTAACTCAAGCTCAGCAA	452
co319-sp6	GCATATGACTTCACAGCAGGCTGC	ATGTTCTGGATCAGACCAGACTT	448
co320	ATGCATTCACACCTGCGAGCTGC	AGGTGGCCAGCTTAATCCCCCA	864
co321	CATTTGTAGGAGACGGTCTTGGCT	AATATCTGAGCACCGGCATGCTGC	865
co322	AGAACGTCTCTGATCGGTGATGCT	TATAGCTGATTTGAGGGCCCAA	913
co323-m13	CATGAGCCCAGAGACATGCACGT	CCACCAAGGCATAGTAACACCAGC	591
co323-sp6	TTGGTTTCTATGAGGCTGCATGGA	GCAGCTTCATTTAGGCTGCGAA	540
co324	TCCCAGTGAGCTAATGCAGGTC	GCTTCAGATTCAGGGTCCTGG	994
co325	AACAGCTCCTGTTGGACCACGT	TCTACAAAGTGTCCCATCAACAC	1044
co326	CACAGCTGTTGCTTACGGGAA	GAGATTAAACGCTCCTCAGTGTGC	880
co327-m13	TCAGACGGCCTGTATGGCAGCCA	TTTTGGCACGATTGTGAACAGACC	479
co329	TTCTCTGAGCAGAGCCTGAACGCA	GTATTCAAAAGCAGAACTGCGTGC	693
co330	AGAATCTGCCATTTCCAGCAGAGC	AGGAAGTGGCCCCGACATGGTC	897
co331-m13	GTCATCCTGCTAGTAAGCACTGAC	AATTGATCAGACATCCCTCTGTG	655
co331-sp6	AACCTTTTCTGCAAGATGCAGTGG	CCATAACCAGATGTGGGTGACTG	388
co335	TGTAACCGCCGATGCACAGCTG	TAATATTCGCTGCGGTGACAGAA	818
co336-sp6	CAGACATCAGGAGCATATGGCGCT	GGAAGAGGTGCTCTATTGAGCTGG	444
co337-m13	CTTTCTACACATGTAAGAGCGGTG	TGTACATCAGTTGCAAGTCGGTGC	492
co337-sp6	GGCATTGCCTTTTGGGGACGCA	CCCGTGTTTGACATAGCACATGAG	521
co338	TCGAGCAGATTTGTTTGCTGAG	CCTGCAGACTGATTAGCCAATGAA	1020
co340	TCAGCGCACGCTTACCGAGAATCG	GATCCAAATGCAGGACAGGGCTGC	857
co341-sp6	TTACAAACAGGGGTCAGGCCCCTC	TTGGTGCTGGCCGGCGTTTGAG	574
co342	CCCCCGAACATAGCAAGATCCGCA	TAAAATGCAGCGCCCCCTGGTG	925
co343	TATCGGAACTCGAGACCTCAGCTG	CACGACTCGGCATAAACTGCACCA	853
co344-m13	GAAACGTCTGGCGGCGCTGTT	CTAAACAGATACCGGGAACCTGTC	655
co344-sp6	TAATACCCAGAAAAGCGCCGTTCC	GGTCGCTGGGTTCCAACCATCACG	651
co345-sp6	CACACGCTATCATAGGCGCGCA	GTACATTCACACGCAGCCCACTGC	304
co346-m13	ATGCACAGGTACTCTCAGTTGTGC	TTTATGGCCGGAAGGTCACCTGCA	555
co346-sp6	ACAAGTCCCATCATCGTATGACG	CAGCACAACAAGTAGGGCCTTCTT	494
co348-m13	ATTGCTCAAGACACCAACGATGTC	TGTCATGGCATTACTACACAGG	483
co349-m13	GCACATTCATCATGGCAGTTTGGA	TCTCTAGGACAGGCATGTGCTACG	499
co352-m13	GATCAGTCGGTGTCTCCGTGTGAC	TGCATCATGGTAGTGAAGGTGAGC	504
co352-sp6	AGATCCTGATATCTGTGCTACAGC	TTTGATTGTCAGGGGGTCTGTA	463
co353	AGCGTCCAAAGCATGCGTTTGCAC	GCATTTTGCTGTACTGCCTGAGG	1067
co354	GGCCAACGCATCTCAGCTGCA	GGAGTCATGCCAACACTCGCTG	1190
co355	CAATCCCACTGAAAGGTCCAGCA	CGGCTTCGTAGCATCCAATGGCAG	1051

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
co357-m13	ACAACAGAGCACCCAGGGGTCC	GTATGTGAGAAGGCAGTTCCT	386
co357-sp6	TGTAAAGGCTGCTGGTGCGCTCAC	ACGAGGTGAACCTGCAGGGAGTG	542
co358	GCTACACTCGCTACGGCACAGG	ACACGTGTCGGTTCGAGACCCA	797
co359	AGACACTTGATGCCTCACTGGGTG	GGGAATCCATCCACAGTATGCCAC	770
co360	TCTAATGAAAGGCTGCAGCTCACC	ATCAACACAGGTTTCCAGAGCCTG	415
co361	GCAGCAGCTGTGGTGCTCACAC	CCTGCAGAGACAATGCAGCCT	859
co362	AGTGCATGGAGGAGCGTCGGA	GCGGCACATTTACAGATCTGCGAT	1022
co363	TACACGAGCGCTGTCATCCCGAGG	GTAACTGAAGTCACATGAGGCTG	1025
co364	AGTCCTCCATGTTCTGCCGGACGA	AAACTGAACAATGTTGGCCAC	1091
co365	TGACCACATTGGTGGGCAGGAAGC	CCGCTCCGCTGAATGAAGGCTG	961
co366	TTCTGTCGAAAGAGCTGGACAGC	ATGTTCGCCCATGTTCGGGGCAGC	733
co368-sp6	TGAGAGCTGTTACAACCAGTCAGC	TGGAGCAGGACAAGTTGGGCTGC	469
co369-m13	TGAGGCAGCTTTGTCACCACGGAT	TGCAAATTATGACACTCGGGAC	469
co369-sp6	ATCAGTTCCAACATAGGACGCT	ATCCAGAGATGTCTGTGCCCA	671
co372-sp6	CCAAACATCCACACGGCCGGTA	GCATGTATCAGTGGGACCGAGTCC	448
co373-13	AGCCAAGCTCCCTCACATGGGGAT	CACCACGGAGAGCAGCCATGAGCT	509
co373-sp6	AAATCACCCAAACACGCGCCTG	GCTGTAGCGACTGAGGGAACGGGT	470
co374-sp6	GGTATGCAGCCCTGTAGGCCAT	TCGCAGAAACATTCGGGGGGTT	660
co375	TTCGCTAGTCGAGCGCGAGCATGG	TGGTTGAGTTTAGGTGTGCAGAC	865
co376	TTCTGGATCAACTAGAGCACTGGT	CTCTGTTGAGGCGCTGTCGATGAG	1029
co378	ATCCAAACACTATGGTCGCCGCAG	TAATCAGCTGTGAAGGGTTCAC	927
co379	TTCTCCAGTGGGCCTGTCGCAAGC	CACTTAGAACAAGTGTGGAGGACC	953
co380	TATAGCTGGGTATCATCGGCAT	CCGTTCCAACGAGGCTGCGCAA	847
co381	TATGAGCGTACACTTGATCGAGG	TAAACTCGACAGGCCCGTCGT	864
co382-m13	GGGAGGACCACCTCTGACCTTCAG	TCCTAGTTTAGCCAACAGAGAGC	464
co382-sp6	TCCTGTTCAGTAGCGGCTGCGGAG	AAGTGTAGCGCTCGGCCGTTCTGG	595
co383-m13	GATACGGCAGTCTACTTCAGCTGC	TCATGAAGTACTGGTTCCCCA	579
co383-sp6	GTGCCTGATGTTATGACCCAGAGC	ACGAGAGGGACAACTTCGGGGGTC	599
co17r	CCGCTGAAAAACATCCCCACAGCA	CCTTAGTCAGGGAGGTAACCCA	391
co18r	GCACCTCCGTTATAGGCTCAGTG	AGCTGTTAAGTTCCAGTGTCAG	399
co23r	CTGACACAAGCAGCTAGCCCTG	GTCGCTTTGGATAAAAGCGTCAA	381
co25f	TCAGAGAGGCTGTCACACTGCGT	ATACCTATGTAAACACACCCGACG	414
co29f	AAGCCATCGAGTCCGTGCTCAG	CAAATGTCACAGTAGGATCGTGTG	448
co30r	CTCCAACACAAAATCGCACTGCAG	GTCATGCTATGTGGTGGTCAT	437
co34f	CCCAGCAGAGTTACGCCATGCA	GTCGACTCACATTCTCGGGGT	374
co34r	GGAATATCCCATCACCTGTCTCGA	TTGTATGTGTTGCCATGCAAGGAC	378
co37r	GGTGTGTGTTTCACTGCCAGAACG	CGCTGAGTGCTGAAGCGACCTGTC	485
co38r	TAACCTTGTGCACGTCTGGAA	TTAGGAGGCAATTAGTGCCATCG	408
co3f	AATGAAAACACGGCCCCTTCGCT	AGAAGTCCAGCGCGCTATTCCA	400
co3r	GAAGGCGTTGAACGCTTCTGTT	TGATAAACATGCAGACAGGCAGTC	454
co40f	AAGGTCAAGCTTCCAGACCACC	CGTGGATCAATnCAAGCAGATCAC	430
co44r	ATCTGATCGCCAATCAGCACAAGC	GAATCATACTGTCTCTGTTGCAGG	412
co45r	TCAGCATTCAAGTGTGAGCCCA	CGTATTCATCAAGAGCTTGAGCCT	480
co46r	GCCGTCTAAGTCCTGGTCATGGAA	CTTCCCGCAGAAGGTTCGGCACGA	372
co56	TGGCCTGGGGCAACAGCCTCAT	TTATTGTGAAGTAGCGTGTGGAG	926
co70r	CGATGAACCGTCACAATCTGC	ACAATGTGGGAACCACATCTC	441
co71f	CGCAAAACGTCCCGTTGGGCTG	GTACATTTGGAGAGGTCGCTG	393
co73f	TGTGAATCGACACGAACGCCGCA	AGCTTAAGCAGAACGGTCCGA	425
co78r	TGATCCCCAGACACTGGAGCT	TCTCTCTTGGTGCGGGTGGAGCGA	437
co80r	GATCAACTGTCAGAGTGAACGTT	TCAAGATACACCGATGACGCTGTT	435
co81	CATTGCATGAAGTACTACGTGCAC	TGGAACACTGTGACGTCATCAA	607
co82f	AGGGTGAAGGTGTCCACTGCCCCT	AACACCATGAACCACACGGAC	380
co87r	GAAATGTTGCCATGGCTCACCGTA	CCAACTACGACCCGCCAGCAA	413
co88f	CATGGAACAGAATTCTCACCGGGT	AGCGTTGTATCTGTTGTACAGGGT	405
co88r	ACTAAAGTCCCACTGTGGCCTGAG	ACTGTGAATCTCTTTGGAGGCT	435
co89f	GCAGCCAACATGGCCCTCTGAA	CCCTGATTTCAGACAGCCCAA	447
co89r	GGGGATCCACCAGAGCACCCTCTT	TGGCTGTCTGAAATCAGGGAGCT	429
co93f	CAACACATCTGGGCATACCGGT	TGGCGAGGGACTCGAGTGGACGTT	481
co93r	GTCGCTCCGTGTCAAATGCGGGAT	ATCACTCAAACGTCCACTCGA	361
co96f	TAGCAGTGTATTGGGCTACACACC	CGCCGTGGATGGAACCCACAA	396

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
co9f	CATTTGCAGGGAAGCTGGTGCAC	AACGGTTTTCCATATGTGAGCCA	416
co9r	ATACACAGGATCTGCTGCAGTTGC	ACGCCGGTTCAGTTTCAATGCACC	402
co107-sp6	GTGATGAACTTCAACCTGCAGC	CCCCAGACGAGCAGACGTCATGCT	222
co137	TTATACTTCAAGCCAGGCCGATGG	TAAAACAAGACCGCAGGGTTTGG	622
co225	CATCCATGTTGAGGGTCGCGGGGA	AAATTATGAGGCAGCCAATGC	1000
co264	TGAATTCAGCCGTCACGGGCCA	TGCTAGCTAGTGCCACTTCAT	996
co26f	AATTTGTCGACACCAACACAC	AGCAGTTCAAAGGTGATGCTT	183
co26r	TGCGAAAGCAGAAGCTTGGCCAC	GTATTGGGGCTCTACAGTTGCA	415
co281-m13	CACTTAGATGTTGACGAGCTGGTT	CAATTAGAAACACAGTCCAGCCAG	515
co303-m13	ACCCAATACTCTGGTGGCCGAGCA	TCAGTTCACAGTAAACACTCTGG	400
co303-sp6	GCGATAGTCCAAGAAGCAGCCT	ATTTAGTGACTGTGGCTCTCC	392
co405-sp6	TCCTGTGAGCGAAACCCAAACTGC	ATCCGCTGCCCTGATGCAGAC	376
co422	TCTTGAGTAAAGTGCCACTGTG	AAATCCCAGCTGACCTCTGAC	1368
co56	TGGCCTGGGGCAACAGCCTCAT	TTATTGTGAAGTAGCGTGTGGAG	926
co940	GTAACGGCCGCCATATGTGCTGGA	GCAGTGCATCTACACGGTGCT	958
co927	GGTTCTTTCCAGCCGTGGCAAAGG	TCAGAGAAGACCACTGCCCGAGAG	712
co960r	GCACACGTTTGGTTTGGTGGCTG	ACTTCAGACCCAGTAAATGGCACC	503
co865r	TGTTGCCCTGATGAGGCCCACAGG	GGCTGTTGACCAGAAACACGCAT	593
co868	CAGTCAATCAGATGCAGCACTA	ACTACTCTCAGTACCATGGAGACG	664
co871f	CATCAGGTCTCGTAGGATGCCAGC	AGAGAACTATTGGTTCCAGCCT	620
co871r	CATGCACTTTGACAGGGAACTGCT	CCCCATTATTGAAACAGCAAGC	609
co872f	CGTACAGTCCCTGAAGTTGAGCAC	AGCTGTTTCAGGATCTGTTTGTGG	470
co873	GACTTTGGCTGACGTGAGGCATCC	TGGGAATCCAGAAGTCAATGCTG	652
co874	TCCGACATTCTTGAGGTCATGGGA	TGTACATATACATGCCAGGCCGTG	1167
co875f	CTTGGTAAAATCAGAGCGTGGCT	GACAACCAGCTATAACTAGGTTCG	481
co876f	CGACAAAGTCCCCCTGTAGTGG	TTAGGTTAAGGCCCAAATGCAC	619
co880	CAGAAAGTCGCATTCTCACCTGGA	AAAGTAGTCCCTGTCTCCCACAGG	861
co881r	GTTGTGCACTACACCAACAGCGT	TGTCTGCGTGCAGTGGCTGCTGCA	457
co882	GCAGAATCTGACCCGTGGTGACCT	ATGCACACCTAAATTCGACGA	676
co883	GAGTGCACACAGCCGGTCAACCTC	TGAAAAGTTGCATTGGGGCATTCC	991
co887r	CCGATAATGCTGAGGCACGTTGTC	AAAATCAGGGCCATGTTGTCG	491
co888	TGTGTGCACTCAGCCCTGGCA	ATCTCCCAACAGCACTCAGTG	1039
co889	GGCAGGGATATCTCCGTGGCTT	CAATAAGTGCATTCCACCTTGAG	854
co892f	AAGTCATGTCCAGTGTCATGTGCA	TAATAAGGAAGCGTACACTGGTC	536
co894	TCAAACGAGAGGGTGCCGTGACC	CCAATTGAGATCCACTCAAGGT	733
co895f	AGCCCGGAACGCTTCACTGTGG	AACGTGTGGGCTTCATACCACGT	584
co896f	AGTCTGATACGGGATCTGGTGTCC	AGAGTCTAAGAGTGTCCCTCGT	471
co896r	GTGCGTCCAACTGAGTGCTAGC	GGTATTCGGTATCGGGGCATCC	412
co897r	CCAGGAACCTCGGTGTGACTCTGG	TGTTGAGTGTTTAGCTCGCAG	501
co901r	AGCTAACGTAATGCTGCGCTC	GGAAACAATCAGGGCGGGGCTTGC	576
co902f	GACCAATCATGTGCGCCCTTGG	AATTGTAATCAGCGCCCCTGAA	505
co904	GCCATATCAGCAAAGACTTCCACG	CCTGTTGATGATTGTGTGGTGG	923
co905f	CAGACTGACCAGAGCTGCTCAC	ATAACTCCACTTGAAGACCACAGC	443
co577	CCAGCTGTTTGAGTGAGGCAGC	GTTCTCTCCGTTTCAGAGACCTG	672
co609e	GAATAAGTGTCGTGTTGCTCGT	TGGGTGTGTGTGTGTGTGTGTACG	562
co612a	AGCAGGCATCTGTTTTGAGTCACG	GCCAACAGCTACATGCCCGGCT	605
co612e	AGACCGTTTCACTCGGCGCCT	GTAATTAGTGGAGGACGTGGAGAA	564
co618	CTATCTCTGGACCAACCAGAAAGC	AACACCACATTCACTTCACTGGCT	983
co623a	AAATGTCATCGAACCTCTGCGA	GTAAAGAATGAAGTCGCTCTCCT	375
co624a	GAATTCCAGCCACTAGAGCCAGTC	AATATGCATTCCCCCAGCCTG	527
co635	GTCTATGTCAAGCTGAAGGCCT	TCAGATTCTGCAGCAAATCCA	1240
co651	AAATGTCAGCAAGTCCTCACTACG	AACGTGTGGTTAACGTTCGGA	762
co671	GCCATGTTAAGACAGTGACTGC	ACAATGCCATCAGCACTGCAA	1123
co677	TGGATTAAACCTGCACTGTAGC	TCCGGCAGTCCGTGTCGTTGTG	642
co684e	GTCATCAGTGCCTTACAAGCAAGC	ATATTTGTTCCTGGCTGGATCAC	502
co694	GTGGCTGTAAGCCAACCCGAC	TGCCACTTTTACTCAAGCAAGGGA	366
co703e	TGGTAGGACTTTGGACTCGAC	GAGCAGCTGAGAGGTTGCTCC	564
co704	TGATGCACAGTGTGTAGCAGC	AATGCAGTGATTGCCGTTGGCT	1222
co705	CAGGACTCACGAAACAAATGCTGC	AAGCATCTGGAGATCCCTTGTGTT	1012
co708	GTCACACTGAAAACATGCTTGACC	AAATGTGTGGACTGTCTTGGT	1204

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Prod. Length bp
co716	GGTAGCCCTTCGTATGACTCAGCA	CAACGTCTTTCAAACTGTGGGA	532
co728	AGAACAATTTGGGCTGCCCCAGCT	GGTGAGATTCACATCCATTCACAC	1051
co739a	CCTACAAACACTTTGTAGGACTGG	TAAGAACATGACCTGGTCTCAC	397
co744a	ACCACTAATTGCTACCTGGAGTCG	CTGTACTGGTATGGGGAGTTCAG	321
co745	CAGCGTCATCACGGATGCGAG	CATATCTGTTGAGATGCACGTTGG	990
co749	TCCTCAGGATTACAGACTCAGTGG	CCGTGATATACCGTGAAACCGCCA	1121
co763	AGTCTGACAACTGTTATTCGCAGG	AAACAATCTTCCATGCCAGGTG	762
co779	GCCTCTGCAAAAATTGCCTGAGTG	CATGTTGTCAAACTCAAGGCCAGC	889
co796	AAGGTTGGCTGTTCAGTCCTCAGC	ACACAGCTCGAATGTTTCAGGCTC	1075
co797a	AGCGGTTGGTTGTCCGTAGCT	ATATTCCTGCAACTTGACCCCCAC	486
co830	ACAACACACGCGCTTAGTTGC	CTCTCAATTGACCAGCGCATGGAA	728
co587a	GCCGCTGATGTTCAACCGGCACGT	TCACGTGTGGGACATTGTGGA	577
co594	AAGACAGGTCTGTAGCCTATCTGC	GATCATCTTCAATCCTTAGCCACG	831
co643a	ATCGGCACCGCCTGGGTCAGAT	CTTGGTGACAGAACACATTGAGCT	597
co740a	AGCAACGACTGGCAACGCTTCGCT	TGACTTACTGACATACTGGCTG	584
co794	TATGCTGTAAACAAGCCCGGTGG	ATCACATCTATGTGGCTGCAT	843
co841a	TGTGTGCACTCAGCCCTGGCA	CTCGCACATTTCCTGTCTGGA	515
co935	TATGCTGTAAACAAGCCCGGTGG	GGCTGTCAGTATTGGCAACAC	805
co949r	TGGTGTGTTCACACCGAACGCGAC	GAGTGAGTTTACTCGCTCGAC	555
co580	AATACTTGCACACTGCTGGCGACG	GATCACCCAGTTTTCCAAGTTGG	1344
co593a	CTGATCAGCTGTTTCAAGGTCAGG	GACACAGATACACCTACCTCCTGG	423
co615	GAGAGCTGTCTTAGAGTGCTTCTC	GACGAGAAACCCGGTGCTCGA	904
co617	GTTAGCCTGGCATCATTCAGGCA	ACATAGCACTAACCAATGCCT	1222
co731	GGCTTTACTCGCGCCTGGAAGG	CACAACAGTCATGCTGACGTGTGC	1184
co758	CGATTGGAAAAACGTTGCCTGCTC	TACTTTATGCTGCACGGTCAA	708
co791	GCGATAAGGCACTACGCTGGTGTT	TAATGTGAGGCACGTTGTCTG	881
co804a	TCCTGCACAAATTAGTTGGCAGTG	CCCTTTCACAAGTACATGTGCAT	524
co808	GTTGATGATTGTGTGGTGGGCCTT	AGACATTGATCTCATTCGCCA	936
co815	TGTGTCGGGTACTTTGCATCAAGG	TGGGAGCCAGCTGATCAAGACAGG	1039
co824	GACTGACCAGAGCTGCTCACATCC	AGTGAGCGGTCTGTCTGCAGCACC	782
co837	AAGTCATGTCCAGTGTCATGTGCA	TCAAACACCTCCATGAGCGTC	1094

Supplement 2 Counts of parental allele states present in the three hybrid populations (P = *C. perifretum* derived, P.A. = ancestral state potentially from *C. perifretum*, R = *C. rhenanus* derived, R. A. = ancestral state potentially from *C. rhenanus*).

	Sieg				Ijsselmeer				Mosel			
	P	P.A.	R	R.A.	P	P.A.	R	R.A.	P	P.A.	R	R.A.
cand3e	1			1	1			1				2
co522		1	1			1	1			1	1	
co82f		2				1	1			2		
co444	1			1	2				2			
co470		2				2				2		
co325		2				2				2		
co46r		1	1			1	1			1	1	
co542	1			1	1			1	1			1
co411		1	1			1	1			1	1	
co445		2				2				2		
cand34		2				2				2		
co306	1			1	1			1				2
co421		1	1			1	1			1	1	
co527		2				2				2		
co311				2				2	1			1
co413		2				2				2		
co552		1	1			1	1			2		
co572-m13	1			1	1			1				2
co407-sp6				2		2						2
co476-sp6	1			1	2				1			1
co355		2				2				1	1	
co391-m13		2				2				1	1	
co373-sp6	2				2				2			
Cott68	1			1				2				2
co569				2	1			1	2			
Cand6				2				2				2
co302-sp6				2				2				2
Cott313		1	1			1	1			1	1	
LCE68	1			1	1			1				2
CottE9-1	2				2				2			
co264		1	1			1	1			1	1	
cand64	1			1	2				1			1
co531	1			1	1			1	2			
co534	1			1	2				2			
co376		1	1			1	1			1	1	
co78r		1	1			1	1			2		
co484				2				2				2
LCE78				2				2	1			1
cand24	1			1	1			1	1			1
LCE27	1			1	1			1				2
Cott153		2				2					2	
co577		1	1			1	1			2		
co830		1	1			2				2		

	Sieg				Ijsselmeer				Mosel			
	P	**P.A.**	**R**	**R.A.**	**P**	**P.A.**	**R**	**R.A.**	**P**	**P.A.**	**R**	**R.A.**
co540	1			1	2							2
co492	1			1	1			1	1			1
cand38		1	1			1	1			2		
Cott580	1			1	1			1	2			
co555		1	1			1	1			1	1	
LCE87	1			1	2				2			
Cott205	1			1	1			1	1			1
co316	1			1	1			1	2			
co539-m13		2				2				1	1	
co539-sp6	1			1	1			1	1			1
co349-m13	1			1	2				2			
co624a	1			1	1			1				2
co705		1	1			2				2		
co346-sp6	1			1	2				2			
co547-m13		2				2				2		
cand54	1			1	1			1	1			1
cand13	1			1	1			1				2
co293-sp6	1			1	1			1	1			1
CottE2	1			1	1			1	2			
co521	1			1	1			1	2			
co40f	1			1	1			1	2			
Cott43		1	1			1	1			2		
co804		1	1			1	1		2			
co824		1	1			1	1			2		
co545		1	1			1	1			1	1	
CottE7	1			1	1			1				2
co414		1	1			2				1	1	
co481	1			1	1			1	1			1
co352-m13	1			1	1			1				2
cand27		2				2				2		
co528		2				2				1	1	
co379		1	1			1	1			2		
co426		2				2					2	
cand26	1			1	1			1				2
co395	1			1				2				2
co449			2				2				2	
co485	2				2				1			1
co564-m13	2				1			1	2			
co541		1	1			1	1			2		
co425		1	1			1	1			1	1	
co468		1	1			1	1			1	1	
Cott173	1			1	1			1				2
cand29	1			1				2				2
co491		2				2				1	1	
co331-m13				2				2				2
co403				2		1	1			1	1	
co417				2				2				2

	Sieg				Ijsselmeer				Mosel			
	P	**P.A.**	**R**	**R.A.**	**P**	**P.A.**	**R**	**R.A.**	**P**	**P.A.**	**R**	**R.A.**
LCE25	2				2				2			
LCE21-1		2				2				2		
co340		2				2				2		
co317-m13	1			1	1			1	1			1
co93f	1			1	1			1				2
co303-sp6		1	1			2				2		
CottE1	2				2				2			
co320	1			1	1			1				2
co525		1	1			2					2	
co312	1			1	1			1				2
cand39		1	1			1	1				2	
cand21		1	1			1	1			1	1	
Cott197	2				2				2			
Cott108	1			1	1			1				2
Cott228	2				2				1			1
co434-m13	2				1			1	2			
co434-sp6	2				2				2			
CottE31	2				2				2			
	63	63	29	61	67	70	26	53	59	63	29	65

Supplement 3 Gene content of ancestry informative marker loci (P = *C. perifretum* derived, P.A. = ancestral state potentially from *C. perifretum*, R = *C. rhenanus* derived, R. A. = ancestral state potentially from *C. rhenanus*). Loci names typed in bold indicate, that a polymorphic SNP with a potential private allele is found in the hybrid lineage.

Locus	*Hybrid Ancestry*	*SNP/Indel*	*Gene content*
co445	fixed P.A	SNP	Brain-derived neurotrophic factor precursor
co547-m13	fixed P.A	SNP	CAMP-dependent protein kinase type II regulatory chain
Cand6	fixed R.A.	SNP	COILED COIL DOMAIN CONTAINING 53
co434-sp6	fixed P	SNP	Cullin-2
co470	fixed P.A	Indel	Echinoderm microtubule associated protein like 4
co407-sp6	fixed P	SNP	Ephrin-B3 precursor.
co340	fixed P.A	SNP	GDP-mannose pyrophosphorylase A
co484	fixed R.A.	SNP	Pim-1 oncogene
CottE9	fixed P	SNP/Indel	Potassium voltage-gated channel subfamily H member 3
co417	fixed R.A.	SNP	pyruvate dehydrogenase kinase, isozyme 1
co527	fixed P.A	SNP	undescribed gene
co302-sp6	fixed R.A.	SNP	upstream of genescan transcript
cand27	fixed P.A	SNP	upstream of Dystrophin
co413	fixed P.A	SNP	upstream of genescan transcript
co331-m13	fixed R.A.	SNP	upstream of Insulin-like growth factor-binding protein-4.
LCE21	fixed P.A	SNP/Indel	upstream of short-chain dehydrogenase
CottE1	fixed P	SNP	upstream of undescribed gene
LCE25	fixed P	indel	upstream of undescribed gene
co325	fixed P.A	SNP	downstream of hypothetical protein
cand34	fixed P.A	SNP/Indel	downstream of undescribed gene-scan transcript
co373-sp6	fixed P	SNP	no gene
Cott197	fixed P	indel	no gene
CottE31	fixed P	SNP	no gene
co449	fixed R	SNP	no gene
co564-m13	mixed	SNP	Aggrecan core protein precursor
co705	mixed	SNP	apoptotic chromatin condensation inducer 1
co316	mixed	SNP	between alpha-catenin-like protein and Glycine max protein
Cott108	mixed	SNP	CDK5 regulatory subunit associated protein 1-like 1(putative ortholog)
co434-m13	mixed	SNP	Cullin 2 (Intron)
co577	mixed	SNP	cytoplasmic tyrosine kinase
co804	mixed	SNP	Diacylglycerol kinase alpha
co421	mixed	SNP	glycolipid synthetase
co93f	mixed	SNP	guanine nucleotide exchange factor (GEF) 10-like
cand38	mixed	SNP	high density lipoprotein binding protein
Cott580	mixed	SNP	Homolog of Homo sapiens "Jumonji domain containing protein 2B
co521	mixed	Indel	Laminin subunit gamma-3 precursor
co46r	mixed	SNP	LIM/homeobox protein Lhx5
co411	mixed	SNP	LIN-7 homolog 2 (MALS-2)
co352-m13	mixed	Indel	Major facilitator superfamily domain-containing protein 3
Cott153	mixed	SNP	Mast/stem cell growth factor receptor precursor
Cott205	mixed	SNP	Membrane-associated DHHC26 zinc finger protein
Cott173	mixed	SNP	MyosinX-Intron
co552	mixed	SNP	Nicastrin
LCE27	mixed	SNP	PDZ and LIM domain 4
co395	mixed	SNP	Peroxisomal Ca-dependent solute carrier-like protein.
co303-sp6	mixed	SNP	peroxisome proliferator-activated receptor gamma binding protein
co824	mixed	Indel	plectin 1, intermediate filament binding protein 500kDa
cand24	mixed	SNP	postsynaptic density protein
co492	mixed	SNP	Receptor-type tyrosine-protein phosphatase S precursor(R-PTP-sigma).
cand13	mixed	Indel	Relaxin 3a
cand39	mixed	SNP	Retinitis pigmentosa 1-like 1 protein (RP1L1)

Locus	*Hybrid Ancestry*	*SNP/Indel*	*Gene content*
co539-m13	mixed	SNP	Ribosomal protein S18
co539-sp6	mixed	SNP	Ribosomal protein S18
co555	mixed	Indel	Sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)
co531	mixed	SNP	serine/threonine kinase (gamma-PAK)
co78r	mixed	SNP	Serine/Threonine Kinase EC_2.7.11.1
cand26	mixed	SNP	Small nuclear ribonucleoprotein Sm D3
Cott43	mixed	SNP	Sperm plasma membrane calcium transporting ATPase.
cand64	mixed	SNP	thyroid adenoma associated
co376	mixed	SNP	toxin-1
co82f	mixed	SNP	undescribed gene
co317-m13	mixed	SNP	undescribed gene
co391-m13	mixed	SNP	vascular cadherin-2
co481	mixed	SNP	WD repeat domain 31 (WD repeat domain 31, isoform CRA_b)
co572-m13	mixed	SNP	genescan transcript
CottE2	mixed	SNP	genescan-transcript
co425	mixed	SNP	genescan-transcript
co468	mixed	SNP	genescan-transcript
cand21	mixed	Indel	genescan-transcript
co320	mixed	SNP	upstream of (cytosine-5)-methyltransferase 3A (EC 2.1.1.37)
co346-sp6	mixed	SNP	upstream of (positive cofactor 2, multiprotein complex) glutamine/Q-rich-associated protein
co830	mixed	SNP	upstream of Acetyl-CoA Acetyltransferase, Mitochondrial precursor
co624a	mixed	SNP	upstream of collagen, type VI, alpha 3
Cott68	mixed	SNP	upstream of Ficolin-2 precursor
co542	mixed	SNP	upstream of genescan transcript
co534	mixed	SNP	upstream of genescan transcript
co540	mixed	SNP	upstream of genescan transcript
co528	mixed	SNP	upstream of genescan transcript
co312	mixed	SNP	upstream of myc target 1 (predicted)
co414	mixed	SNP	upstream of NADH dehydrogenase
co541	mixed	Indel	upstream of transmembrane 4 superfamily member 8
LCE87	mixed	SNP	upstream of undescribed gene
cand54	mixed	SNP	upstream of undescribed gene
co426	mixed	Indel	upstream of Zinc finger and BTB domain-containing protein 24
cand29	mixed	SNP	downstream of Arylacetamide deacetylase-like 1
CottE7	mixed	SNP	downstream of ATP-binding cassette sub-family A member 2
co349-m13	mixed	SNP	downstream of gene-scan transcript
co403	mixed	SNP	downstream of gene-scan transcript
Cott313	mixed	SNP	downstream of glucose transporter 3
Cott228	mixed	SNP	downstream of protein tyrosine phosphatase, receptor type, (putative ortholog)
co444	mixed	SNP	downstream of RNA-binding protein Nova-1 (Neuro-oncological ventral antigen 1)
cand3e	mixed	SNP	no gene
co522	mixed	SNP	no gene
co306	mixed	SNP	no gene
co311	mixed	SNP	no gene
co476-sp6	mixed	SNP	no gene
co355	mixed	SNP	no gene
co569	mixed	SNP	no gene
LCE68	mixed	SNP	no gene
co264	mixed	SNP	no gene
LCE78	mixed	SNP	no gene
co293-sp6	mixed	SNP	no gene
co40f	mixed	SNP	no gene
co545	mixed	Indel	no gene
co379	mixed	SNP	no gene
co485	mixed	SNP	no gene
co491	mixed	SNP	no gene
co525	mixed	SNP	no gene

8 Digital Supplement

- Supplement 1 Genotyping data of mapping families
- Supplement 2 Sequences of the *Cottus* genomic library
- Supplement 3 PDF files and sequences of ancestry-informative marker loci
- Supplement 4 PDF files and sequences of polymorphic SNP loci in the hybrid lineage

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Prof. Dr. Diethard Tautz betreut worden.

Köln, den 28.04.2007

Teilpublikationen:

Die folgende Publikation basiert auf Teilen dieser Arbeit, die die Konstruktion der vorläufigen genetischen Karte und die Untersuchungen zur konservierten Synteny zwischen Groppe und Pufferfisch betreffen:

Stemshorn, K. C., Nolte, A. W. and Tautz, D. (2005) A genetic map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the physical map of *Tetraodon nigroviridis*. *Journal of Evolutionary Biology* 18, 1619-1624.

Lebenslauf

Name:	Kathryn Stemshorn
Anschrift:	Urftweg 3 50354 Hürth
Geburtsdaten:	21.08.1978 in Köln
Staatsangehörigkeit:	deutsch

Schulausbildung:

1985 – 1989	Deutschherrenschule in Hürth (Grundschule)
1989 – 1998	Gymnasium Bonnstraße in Hürth
1995 – 1996	Austauschschülerin an der Littleton High School, Littleton, Massachusetts, USA
1998	Abitur

Hochschulausbildung:

WS 1998/99 – WS 2003/2004	Studium der Biologie Universität zu Köln
Januar – April 2001	Auslandssemester am University College of the Cariboo, Kamploops, British Columbia, Kanada
12. Januar 2004	Abschluss des Diploms in Biologie Institut für Genetik Prof. Dr. Diethard Tautz Universität zu Köln
April 2004 – Juni 2007	Promotion am Institut für Genetik Lehrstuhl für Evolutionsgenetik Prof. Dr. Diethard Tautz, Universität zu Köln
11. Juni 2007	Abschluss der Promotion