

Genomic Mapping of Speciation-Related Traits in Hybridizing Sculpins (*Cottus*)



Dissertation

In fulfillment of the requirements for the degree

“Doctor rerum naturalium”

of the Faculty of Mathematics and Natural Sciences

at Kiel University

Submitted by

Jie Cheng

Research Group Evolutionary Genetics of Fishes

Max Planck Institute for Evolutionary Biology

Plön, February, 2013

First referee: Dr. Arne W. Nolte

Second referee: Prof. Dr. Thomas C.G. Bosch

Date of the oral examination: 7th, May, 2013

Approved for publication: 09th, July, 2013

Signed: Prof. Dr. Wolfgang J. Duschl (Dean)

Content

Zusammenfassung	I
Summary	III
Declaration	V
General Introduction	1
1. The build-up of reproductive isolation	1
2. Genetic and genomic signatures of speciation	5
3. The importance of hybridization.....	7
4. A model for initial hybrid speciation: <i>Cottus</i> system and current progress	10
5. Fishing for the secret of <i>Cottus</i> evolution: scope and study system of this dissertation.....	13
Chapter 1. The genomics of incompatibility factors and sex determination in hybridizing species of <i>Cottus</i> (Pisces)	17
Introduction	18
Materials and Methods	20
Results.....	25
Discussion	33
Supplementary Information.....	38

Chapter 2. The genetic architecture of skin prickling and body shape in hybridizing sculpins (<i>Cottus</i>)	45
Introduction	46
Materials and Methods	49
Results	55
Discussion	64
Supplementary Information	70
Conclusions and perspectives	79
References	83
Affidavit	99
Curriculum Vitae	101
Acknowledgements	103

Zusammenfassung

Aus den europäischen Süßwassergroppen, *Cottus rhenanus* und *Cottus perifretum*, sind im unteren Rhein Hybridlinien (invasive *Cottus*) entstanden. Obwohl ihr Genom vollständig durchmischt ist, ähneln die invasiven Groppen phänotypisch stärker ihrer einen Elternart, *C. perifretum*. Daraus ergibt sich ein Interesse, die genetischen Hintergründe der phänotypischen Merkmale, welche die Unterschiede zwischen den *Cottus* Arten bedingen, zu erforschen, sowie die Evolution dieser Merkmale in der Hybridlinie zu verstehen. In der vorliegenden Dissertation wird eine Genkarte erstellt, die wiederum zur Bestimmung von QTL (engl. *quantitative trait loci*) genutzt wird. Diese Ressourcen werden mit NGS Methoden (engl. *next-generation sequencing*) und Kandidatengenanalysen kombiniert, um die genomische Grundlage der in der Natur beobachteten phänotypischen Merkmale der *Cottus* Hybridlinie näher zu untersuchen.

Es wurden zwei unabhängige Gruppen zwischenartlicher F_2 Kreuzungen zwischen *C. perifretum* und *C. rhenanus* erzeugt. Mithilfe der Roche 454 GS-FLX Plattform wurden ESTs (Expressed Sequence Tags), die die Elternarten und die invasiven Groppen repräsentierten, sequenziert. Die entstehenden Sequenzabschnitte wurden überlappend zusammengefügt und mit einer Gen- und Proteindatenbank von fünf Fischarten (Teleostei), abgeglichen. Es wurden 252 SNPs ausgewählt und für alle F_2 Kreuzungen mittels des Illumina GoldenGate Verfahrens typisiert. Basierend auf einem vorliegendem Karyotyp von *Cottus* wurde eine genetische Karte mit 24 Kopplungsgruppen (engl. *linkage groups*, LG) und einer Gesamtlänge von 1575,4 cM erstellt. Die größte Ähnlichkeit wurde zu dem Stichlingsgenom gefunden, in welchem die Gruppierung von 97,3 % der homologen Marker mit den Kopplungsgruppen der Groppe überein stimmt. Das Stichlingsgenom ist damit eine hervorragende Referenz für die Erforschung des Groppengenoms.

Um sowohl Regionen mit Transmissionsdistorsion (engl. *transmission distortion*), als Zeichen für genetische Inkompatibilität in den Elternarten der Hybridarten, als auch geschlechtsbestimmende Regionen (engl. *sex determination*, SD) zu ermitteln, wurden Genkarten erstellt und QTL für SD identifiziert. Dadurch konnten mögliche Korrelationen zwischen SD und genetischer Inkompatibilität festgestellt werden. Auf diese Weise wurde getestet, ob auf Haldanes Regel basierende Effekte mitverantwortlich für beobachtete Unterschiede in natürlich hybridisierenden Groppen sind. In verschiedenen Kopplungsgruppen der Elternarten der Hybridlinie wurde ein geschlechtsbestimmender Abschnitt entdeckt, der heterogametischen sowie männlichen Ursprungs ist. Konservierte Gemeinsamkeiten mit bereits sequenzierten Fischgenomen zeigen, dass in *Cottus* geschlechtsbestimmende Regionen nicht an denselben Regionen wie in

den diesen Modellorganismen zu finden sind. Anzeichen für genetische Inkompatibilitäten waren nicht auf Männchen beschränkt, nicht mit SD Regionen gekoppelt und meistens spezifisch für einzelne Familien. Das weist auf einen schnellen, evolutiven Wechsel der SD in *Cottus* hin und gibt keinen Anhaltspunkt für auf Haldanes Regel basierende Fitnesseffekte. Insgesamt, ist es unwahrscheinlich, dass hier entdeckte Inkompatibilitätsfaktoren und SD Genorte intrinsische Selektion bedingen, welche den Genpool natürlicher Hybriden einfach und vorhersagbar beeinflusst.

Zwei morphologische Merkmale, Stachelbeschuppung und Körperform, variieren unter den *Cottus* Arten und erreichen eine extreme Ausprägung in der Hybridart. Der Ectodysplasin (EDA) Signalweg, welcher bekanntlich die Entwicklung von Schuppen beeinflusst, beinhaltet vielversprechende Kandidatengene, anhand derer sich der genetische Hintergrund der Stachelbeschuppung erforschen lässt. Um zu testen, inwiefern der EDA-Signalweg auch natürliche Variationen in *Cottus* kontrolliert, wurden EDA-Signalweg Bestandteile indentifiziert und kartiert sowie QTL Analysen durchgeführt. Ein hochsignifikanter QTL, der die Stachelbeschuppung in allen F₂ Kreuzungen beeinflusst, wurde in einem Abschnitt gefunden, welcher das EDA-Rezeptorgen (Edar) innerhalb der Kopplungsgruppe 3 beinhaltet. Darüber hinaus wurden keine Effekte für Gene des EDA-Signalwegs gefunden. Dieselbe QTL Region hat zusätzlich einen starken Effekt auf die Ausbildung der Körperform. Eine Untersuchung der Genstrukturen zeigt, dass die genetische Struktur des Edar-Gens es, eher als andere Komponenten des Signalwegs, zu einem adäquaten Ziel für evolutionäre Veränderungen macht. Angestammte Allelfrequenzen innerhalb der Genregion des EDA-Signalwegs im Genpool der Hybridline zeigen, dass Edar vornehmlich in *C. perifretum* vorkommt (85-96%). Dies korreliert mit der phänotypischen Ähnlichkeit zwischen invasiven Groppen und dieser Elternart. Der Edar-tragende QTL ist derzeit der beste Kandidat, die sichtbaren phänotypischen Varianzen in der Hybridline des Rheinsystems zu verursachen. EDA Signaltransduktion stellt ein adaptives Schlüsselmerkmal in Stichlingen dar und unsere Ergebnisse weisen darauf hin, dass derselbe Signalweg auch die auffällige Stachelbeschuppung in *Cottus* bedingen könnte.

Summary

European freshwater sculpins, *Cottus rhenanus* and *Cottus perifretum*, have formed hybrid lineages (invasive *Cottus*) in the Lower River Rhine. The hybrid *Cottus* particularly resembles one of its ancestral species, *C. perifretum*, in phenotypic features. This contrasts with the fact that the invasive genome is thoroughly admixed. It is hence of interest to investigate the genetic basis of phenotypic traits that represent differentiation among *Cottus* species, and then to learn about the evolution of these traits in the hybrid lineage. This dissertation combined genetic mapping and quantitative trait loci (QTL) mapping methods with next generation sequencing and candidate gene analysis to explore the genomic basis of *Cottus* hybrid traits that are observed in nature.

Two independent groups of interspecific F₂ crosses were generated between *C. perifretum* and *C. rhenanus*. EST libraries representing the parental species and invasive *Cottus* were sequenced using the Roche 454 GS-FLX platform. These reads were assembled into contigs that were mapped against 5 teleost model fish genomes and protein databases. 252 diagnostic SNPs were developed and typed for all F₂ crosses using Illumina GoldenGate genotyping Assay. A consensus map was constructed that contains 24 linkage groups (LG) with a total length of 1575.4 cM which is in agreement with available karyotype information for *Cottus*. The highest conserved synteny of the *Cottus* map was found with the stickleback genome in which 97.3% of the homologous markers agree among linkage groups. This underlines that the stickleback genome provides an excellent reference to explore the *Cottus* genome.

We have employed genetic and QTL mapping to detect transmission distortion regions as signs for hybrid genetic incompatibilities in the parental species as well as sex determination (SD) regions. This allows the exploration of correlations between SD and genetic incompatibilities to test whether effects related to Haldane's rule contribute to the observed differentiation in natural hybridizing *Cottus*. A male heterogametic (XY) SD region was found in different LGs in the two parental species of invasive *Cottus*. Conserved synteny with sequenced fish genomes reveals that SD in *Cottus* is not based on the same genomic regions found in model organisms. Signs for genetic incompatibilities were not restricted to males, not linked to SD regions and mostly specific to single mapping families. This indicates a fast evolutionary turnover of SD in *Cottus*, and does not provide evidence for fitness effects related to Haldane's rule. As a whole, the incompatibility factors and SD loci identified here are not likely to cause intrinsic selection that could affect the gene pool of natural hybrids in a simple and predictable way.

Two morphological traits, scale-like prickling and body shape, vary among *Cottus* species and are extreme in the hybrid lineage. The Ectodysplasin (EDA) signaling pathway provides promising candidates to investigate the genetic basis of prickling because it is well known to affect the development of scales in fishes. To test whether the EDA signaling pathway also controls natural variation of *Cottus*, we identified and mapped *Cottus* EDA signaling pathway components and performed QTL mapping analysis for prickling and body shape. A highly significant QTL that affects prickling in all F₂ crosses was detected in an interval that contains the EDA receptor (Edar) gene in *Cottus* LG 3 but none of the other EDA pathway genes. The same QTL region also has a strong effect on body shape. An investigation of gene structures suggests that the genomic architecture of the Edar gene makes it a more suitable target for evolutionary changes than other components in this pathway. Ancestral allele frequencies within EDA pathway gene regions in the invasive gene pool show that Edar has the genomic ancestry preference for *C. perifretum* (85%-96%). This correlates with the phenotypic similarity of invasive *Cottus* with that parental species. The Edar carrying QTL is currently the best candidate that strongly determines the visible phenotypic variation of hybrid *Cottus* in the River Rhine system. EDA signaling constitutes a key adaptive trait in sticklebacks and our results suggest that the same pathway may also cause conspicuous variation of prickling in *Cottus*.

Declaration of Author's Contribution

I present my doctoral research in this dissertation. The design of the whole project was done together with my direct supervisor Dr. Arne W. Nolte. The interpretations of all the different results were achieved during numerous discussions. Practical laboratory works as well as the major parts of the data analysis were conducted by me, unless stated below.

Chapter 1

The content of this chapter is accepted for publication as a research article in *Heredity*, authors contributed as follows. I performed the 454 transcriptome sequencing with Heike Buhtz. Raw sequences were analyzed together with Till Czypionka. F₂ Mapping families were bred in the lab together with Arne Nolte and Till Czypionka. Fish care was helped by Heik Harre, Sonja Dembeck and Elke Bustorf. I developed SNP markers and performed the GoldenGate genotyping together with Elke Bustorf. I performed the conserved synteny analysis, map construction and QTL analysis. K. Konrad, F. Chan, B. Haubold and D. Tautz helped with comments and discussions. Arne Nolte was the supervisor of this project and we wrote the final manuscript together. All authors have read and approved the final manuscript.

Chapter 2

This chapter is prepared as a research article, authors contributed as follows. F₂ Mapping families were bred in the lab together with Arne Nolte and Till Czypionka as described in Chapter 1. I collected morphology and genotype data for QTL mapping experiment together with Elke Bustorf. Arne Nolte performed the body shape analysis, Illumina sequences assembly and SNP calling. I performed the indel markers development, QTL mapping, candidate genes analysis and ancestral allele frequencies scanning. Arne Nolte was the supervisor of this project and we wrote the final manuscript together. All authors have read and approved the final manuscript.

General Introduction

Many evolutionary biologists attempt to explore two major features of the living world: the biological diversity between species and their adaptation to specific environments. Biological diversity presents mainly at species level and can be explained by the balance between extinction and speciation, and consequently, speciation is the central topic since the conception of evolutionary biology (The Marie Curie SPECIATION Network 2012; Santini *et al.* 2012). Given the recent advances of technology and theory, an enormous amount of research has been performed on speciation during the past years. In this general introduction, I will discuss several key speciation questions which are also the main focus of my doctoral dissertation: 1) the mechanisms that drive the build-up of reproductive isolation; 2) the genetics and genomics of speciation; 3) the importance of hybridization; 4) European freshwater sculpins as a model of hybrid speciation; 5) study system and main scope of this dissertation.

1. The build-up of reproductive isolation

Since Ernst Mayr (Mayr 1942) developed the biological species concept, to better understand the process of speciation, finding mechanisms that contribute to the build-up of reproductive isolating barriers between diverging populations is clearly the central target for speciation research. The mechanisms of reproductive isolation or hybridization barriers are collections of physical, behavioral and physiological processes that prevent the members of two different species that mate from producing offspring, or which ensure that any offspring that may be produced is less fit or even not fertile (Hvala and Wood 2012). The classic scheme for describing the origin of reproductive isolation is whether gene flow is prevented before (“prezygotic”) or after (“postzygotic”) hybrid zygote formation (The Marie Curie SPECIATION Network 2012). All these barriers to gene exchange may be extrinsic or intrinsic (Shaw 2012). Prezygotic barriers can be subdivided into ecological and non-ecological barriers, whereby the latter are generally assumed to be caused by sexual selection (Coyne and Orr 2004; Santini *et al.* 2012). Postzygotic barriers can be subdivided into intrinsic barriers in which developmental and genetic incompatibilities cause hybrid sterility and inviability, and extrinsic barriers, in which sterility and inviability will depend on the particular environments inhabited (Shaw 2012; Santini *et al.* 2012). Many studies suggest prezygotic isolation (often behavioural mate discrimination in animals or pollinator differences in plants) is a more important or earlier-evolving barrier to gene flow than postzygotic isolation and therefore is potentially more important to the process of speciation (Wolf *et al.* 2010; The Marie Curie SPECIATION Network 2012). However, this view does not properly consider extrinsic

postzygotic isolation owing to differential adaptation, which might be the first step in speciation (Schluter 2009). However, understanding the contributions of different components of reproductive isolation still is an open question. Speciation and hybrid zone theories have highlighted that intrinsic and extrinsic barriers can easily become coupled in the absence of natural barriers (such as rivers or mountains) in spatially subdivided populations, even if the loci involved are unlinked (Bierne *et al.* 2011). This makes it difficult to disentangle the effects between the intrinsic and extrinsic barriers, and the test of alternative hypotheses is warranted in order to assess which mechanisms contribute to emergence or maintenance of species.

Intrinsic barriers to reproduction

The genetics of intrinsic barriers to reproduction is a primary interest in speciation research. Several kinds of genetic changes have been identified as the likely causes of intrinsic genetic incompatibilities. Karyotypic changes are of major importance in plant speciation (Wolf *et al.* 2010). If two hybridizing species have different karyotypes, then hybrids may suffer from meiotic defects or produce gametes that are aneuploid, and sterile (Maheshwari and Barbash 2011). Chromosomal rearrangements, commonly inversions, have also been widely implicated in causing hybrid sterility in chromosomal heterozygotes through the production of unbalanced gametes due to crossing over in the rearranged regions during meiosis (Maheshwari and Barbash 2011). A key evolutionary effect of inversions is that they suppress recombination between alternative chromosomal arrangements in hybridizing populations, and they will spread if they capture alleles at multiple loci involved in divergent adaptation to contrasting environments (Kirkpatrick 2010; Lowry and Willis 2010). However, alternative views suggest that sterility is not directly caused by the inversions interfering with meiosis, but rather because inversions are more likely to harbor diverged Dobzhansky-Müller Incompatibility (DMI) genes compared with non-inverted regions (Rieseberg 2001; Hoffman and Rieseberg 2008). Nevertheless, genetic incompatibilities may be the most important cause of intrinsic isolation which play a common role in both hybrid sterility and inviability, and affect both animals and plants (Wu and Ting 2004; Wolf *et al.* 2010). Two geneticists, Theodosius Dobzhansky and Herman Müller dominated the early period of genetics of hybrid incompatibility, which now is termed as Dobzhansky-Müller incompatibility (DMI). They both realized that hybrid inviability or sterility was unlikely to arise from a single change at one locus but from more than one change and involved multiple genetic loci (Johnson 2008a). This means that an allele at one locus from one of the parental species does not interact well with an allele at another locus from the other parental species (Wu and Ting 2004; Figure 1). The emergence of reproductive isolation through DMI is expected to be a slow process that gets more efficient as time progresses (Wolf *et al.* 2010).

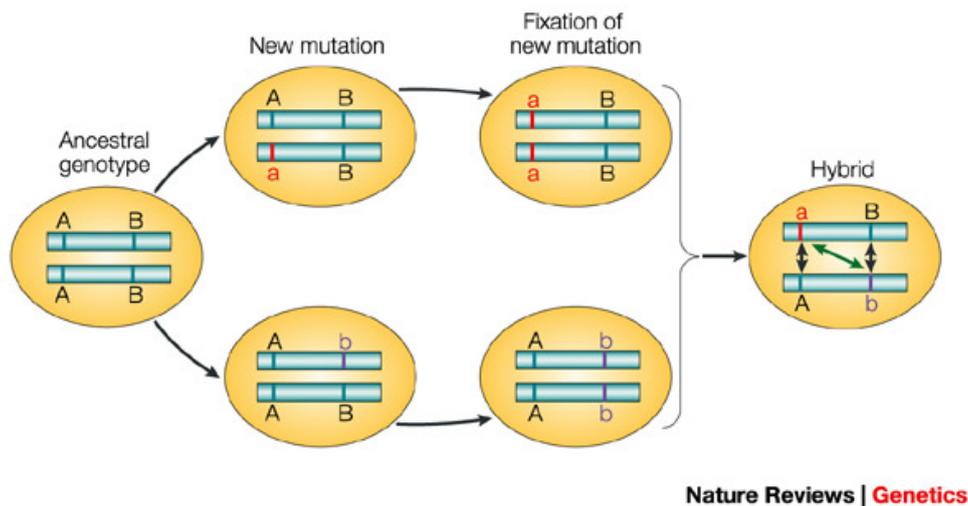


Figure 1. Dobzhansky-Müller incompatibility (DMI) models. In the ancestral population, the genotype is AA BB. When the population is split into two, A evolves into a in one population and B evolves into b in the other. a and b are mutually incompatible. As the a–b interaction is not present in the pure species, the evolution of incompatibility is possible. On the far right (hybrid), the divergence process is indicated by the black double-headed arrows and the incompatibility is indicated by the green double-headed arrow. The relationship between the black and green arrows should be the essence of the evolution of reproductive isolation. A deficiency of the DMI model is that it does not consider the divergence process that is indicated by the black arrows but focuses instead on the incompatibility, which is a byproduct of that divergence. (Figure was taken from Wu and Ting 2004).

Many of the researches on intrinsic barriers and genetic incompatibilities have been focused on Haldane's rule (Wolf *et al.* 2010), which is based on the observation that “when in the offspring of two different animal races one sex is absent, rare, or sterile, that sex is the heterozygous (heterogametic) sex” (Haldane 1922). Haldane's rule is one of the most general patterns in speciation biology and has been confirmed in many groups of animals (Schilthuizen *et al.* 2011). Of the various models that have been suggested for the action of Haldane's rule, there are strong supports for the dominance theory (heterogametic hybrids suffer fitness loss only for X-linked DMI genes that are recessive) and the faster male theory (male-expressed genes are subject to increased selection pressure which results in faster evolution for these loci), but weaker supports for some other models like the faster-X theory and meiotic drive (Wolf *et al.* 2010; Schilthuizen *et al.* 2011). Haldane's original observation concerned just the fertility and viability of hybrids, but Haldane's rule can also be extended to other traits. Indeed, in hybrids, the members of the heterogametic sex tend to suffer more adverse morphological effects than the homogametic counterparts (Johnson 2008b). While Haldane's rule is commonly viewed as important in the initial stage of speciation (Kulathinal and Singh 2008; Wolf *et al.*

2010), it may be argued that sex chromosomes are also comparatively important in the later stage of speciation, completing the process after the initial differentiation (Qvarnström and Bailey 2008; Wolf *et al.* 2010).

Natural selection and reproductive isolation

Besides the intrinsic barriers, the extrinsic reproductive isolation depends on the ecological settings, in which the hybrid phenotypes fare poorly in their interaction with the environment that each parental species have adapted to (Wolf *et al.* 2010). In some cases, these barriers can disappear if environmental conditions in nature have been changed (Nosil and Schluter 2011). There has been a great emphasis on the role of environment in speciation processes, especially whether resource-based divergent natural selection is responsible for much of the divergence between closely related species, now called “ecological speciation” (Schluter 2000; Santini *et al.* 2012). The concept of “ecological speciation” is the speciation process that reproductive isolation can evolve as a by-product of adaptation to different environments (Schluter 2001, 2009). An important idea on this is the significance of natural selection that can play on a set of few key traits associated with resource use or mate choice, which might be more effective at causing adaptive divergence in the face of homogenizing gene flow (Wolf *et al.* 2010). Therefore, in the early stage of the adaptive divergence process, populations or species are reproductively isolated only at a small number of specific adapted loci throughout the genome (“genomic islands of speciation”), while remaining gene flow is unaffected across the other parts of the genome (Harr 2006; Wolf *et al.* 2010). During the process of genomic divergence, gene flow is reduced in the regions surrounding the selected sites and divergence hitchhiking may increase differentiation of regions that physically linked to these under selection (Nosil and Feder 2012). As this proceeds slowly, effective gene flow gets further reduced across the genome with genome hitchhiking occurred even for regions unlinked to these under selection, and then it results in strong reproductive isolation and widespread genetic divergence (Wolf *et al.* 2010; Nosil and Feder 2012). Most of the hybrid incompatibility genes identified so far showed signatures of adaptive evolution (Orr *et al.* 2007; Wolf *et al.* 2010), and some recent work suggests that extrinsic isolation might be more common than intrinsic isolation, especially in the early stages of divergence (Schluter 2009). Cases of parallel evolution may also be used to learn about the relationship between adaptation to divergent ecological environments and the buildup of reproductive isolation. Populations in the similar ecological environments may have responded to similar selection pressures by parallel phenotypic changes underlying same or different genetic changes (Elmer and Meyer 2011). A prominent case is given by the divergence of different stickleback populations as limnetic vs. benthic and marine vs. fresh water (Schluter 2010; Hohenlohe *et al.* 2010). The relationships between adaptation to the environment and the build-up of reproductive isolation

may also vary depending on the genetic pathway taken to achieve the adaptation (Qvarnstrom *et al.* 2010).

2. Genetic and genomic signatures of speciation

To understand how divergent phenotypes of organisms can contribute to barriers to gene flow that leads to the origin of a new species requires a detailed understanding of the genetic basis of these phenotypes. While the genetics of reproductive isolation is one of the major areas in speciation research, it has been only since the 1980s that researchers have started answering the questions on speciation in terms of how many and what kind of genes controlling the traits that are responsible for reproductive isolation (Coyne and Orr 2004; Wu and Ting 2004; Santini *et al.* 2012).

Getting the genes under speciation

Speciation genes are commonly defined as genes involved in reproductive isolation. Conceptually, the genetics of speciation have a much broader definition with speciation genes being functional genomic elements that affect some degree of extrinsic, ecological based reproductive isolation, sexual isolation, pre- or post-mating, pre- or post-zygotic isolation (Wu and Ting 2004; Wolf *et al.* 2010; Nosil and Schluter 2011). Speciation genes can be associated with any form of reproductive isolating barriers, and the researches of finding speciation genes vary greatly and have different strength of evidence as described by Nosil and Schluter (2011). For example, the strongest evidence are the studies that have mapped reproductive isolation to a candidate gene and then used experimental methods, such as positional cloning, gene replacement or knockout, gene expression assay and transgenic manipulation, to confirm the function of the gene involved (Nosil and Schluter 2011). With few exceptions, most examples of these genes are yet known to be related to hybrid dysfunction like intrinsic hybrid sterility or inviability which could be readily examined in the laboratory and to be found in model species like *Drosophila* (Orr *et al.* 2007; Presgraves 2010). In a second case, reproductive isolation has been mapped to a genomic region like quantitative trait locus (QTL) and a candidate gene within the QTL interval was identified. The functional effects of the candidate gene still need to be proved to exclude genes in close linkage. In a third case, a few examples of specific gene have been found to affect a phenotypic trait known to be under divergent natural selection between contrasting environments which contribute to extrinsic reproductive isolation (Nosil and Schluter 2011). One well studied example of such genes is Ectodysplasin (Eda), which is the major locus controlling differences in the number of lateral armor plates between marine and freshwater stickleback populations (Colosimo *et al.* 2004, 2005). Lateral plate number is

probably under divergent selection, with a high plate number being advantageous in marine environments where toothed predators are abundance, and disadvantageous in freshwater habitats where low-plated fish have faster growth rates (Barrett *et al.* 2008, 2010; Schluter *et al.* 2010). Finally, the least evident cases are that reproductive isolation has been mapped to a QTL, but a candidate gene has not been identified (Nosil and Schluter 2011). However, the QTL analysis provide the experimental approach to determine both the number and location of loci, and their relative effects to phenotypic differences that contribute to reproductive isolation in natural populations (Peichel 2005). Further works to identify specific genes within the QTL and to test the functional role are always required to address questions involved in speciation.

The role of changes in gene expression and other genomic processes

Current evidence suggests that functional polymorphism in genes relevant to evolutionary divergence and speciation is not restricted to coding changes but also includes regulatory variations modulating the expression of a gene (Tautz 2000; Wittkopp *et al.* 2008). However, it remains a challenge to get the answer whether or not the differentiation in regulatory genomic factors are more common than that in coding sequences especially for causing reproductive isolation (The Marie Curie SPECIATION Network 2012). The study of gene expression divergence has also proven to yield valuable insights on the roles of *cis* and *trans* factors in the evolution of novel phenotypes (Wolf *et al.* 2010). Changes in *cis*-regulatory factors (e.g. enhancer) occur on the regulatory sequences of the gene itself and their effects are restricted to the sequences of their own DNA or RNA molecular, and *Trans* factors as transcription factors are separate molecules (proteins or RNAs) that can influence the activity of a broad variety of targets (Wray *et al.* 2003; Wolf *et al.* 2010). Studies in several model species suggested that expression divergence due to changes in *cis* factors were probably more common (Wittkopp *et al.* 2008; Wittkopp and Kalay 2012). Adaptation often occurs by changes in gene regulation and that *cis*-regulatory factors appear to play a special role on adaptive divergence (Fay and Wittkopp 2008). For example, in sticklebacks, *cis*-regulatory changes in the *Pitx1* gene seem to be associated with differences in pelvic skeleton structure between marine and freshwater forms (Shapiro *et al.* 2004; Chan *et al.* 2010). Moreover, the importance of other genomic-level phenomena such as gene duplication and loss, epigenetic effects, small RNAs, transposable element activity, creation of new exons or introns and perhaps many other unsuspected genomic processes might also contribute to different degrees of reproductive isolation (Maheshwari and Barbash 2011; Santini *et al.* 2012). Last but not least, great emphasis has been placed on the evo-devo aspect of speciation, like development mediated phenomena such as developmental phenotypic plasticity can also act as a facilitator for speciation (Pfennig *et al.* 2010; Santini *et al.* 2012).

3. The importance of hybridization

The phenomenon of hybridization has fascinated scientists for many decades (e.g. Darwin 1862; Haldane 1922; Dobzhansky 1937; Mallet 2005) and the effects of hybridization on the process of speciation may vary and has long been debated (Hochkirch 2013). Hybridization may slow or reverse speciation by allowing gene flow and recombination but also may accelerate speciation by adaptive introgression or instantaneous speciation by allopolyploidization (Abbott *et al.* 2013). All of these processes may generate novel phenotypes through interactions in hybrids as advantages of transgressive segregation and disadvantages of intrinsic or extrinsic incompatibilities (Abbott *et al.* 2013). Therefore the consequences of hybridization may vary widely depending on the genetic constitution, demography, ecology and spatial distribution of the species involved (Hochkirch 2013). As discussed by Hvala and Wood (2012, Figure 2), the admixture of genomes can result in the following four outcomes: 1) the populations may fuse and evolve as a single species; 2) reinforcement may complete the process of speciation initiated in allopatry; 3) the two populations may persist their genetic identity with hybrids restricted to zones of contact; and 4) new species can arise from the recombination of the differentiated genomes (Hvala and Wood 2012). Here, I will discuss the forth outcome in more detail because it could provide more perspectives on the question of how hybridization can contribute to the process of speciation.

Abbott *et al.* (2013) have proposed several points which suggest that hybridization and introgression could have important impacts on the origin of species. For example, hybridization is a common phenomenon in many taxa, although more common in plants (around 25%), but also in animals (around 10%), including fishes, amphibians and insects (Abbott and Rieseberg 2012; Mallet 2005, 2007); Mutations for novel adaptation to evolve are rare and hybridization can add more source of adaptive genetic variance than mutation (Whitney *et al.* 2010; Kunte *et al.* 2011); Adaptation is thought to be the most important process driving divergence and contributing to hybrid speciation (Coyne and Orr 2004; Sobel *et al.* 2010). Hybrid speciation may occur through the duplication of a hybrid's chromosomes (allopolyploid speciation) or by the stabilization of a fertile hybrid (homoploid hybrid speciation) without change of chromosome number (Abbott and Rieseberg 2012). Most cases of hybrid speciation studied so far concerned polyploid hybridization which seems to be more common in plants than in animals and usually leads to a direct genetic isolation of an allopolyploid hybrid population (Mallet 2007). Unlike polyploids, homoploid hybrids are not instantaneously reproductively isolated from their parents, and the hybrid lineages are likely to be homogenized by gene flow with their parents before they become established

(Abbott and Rieseberg 2012). Next I will discuss more about the homoploid hybrid speciation events.

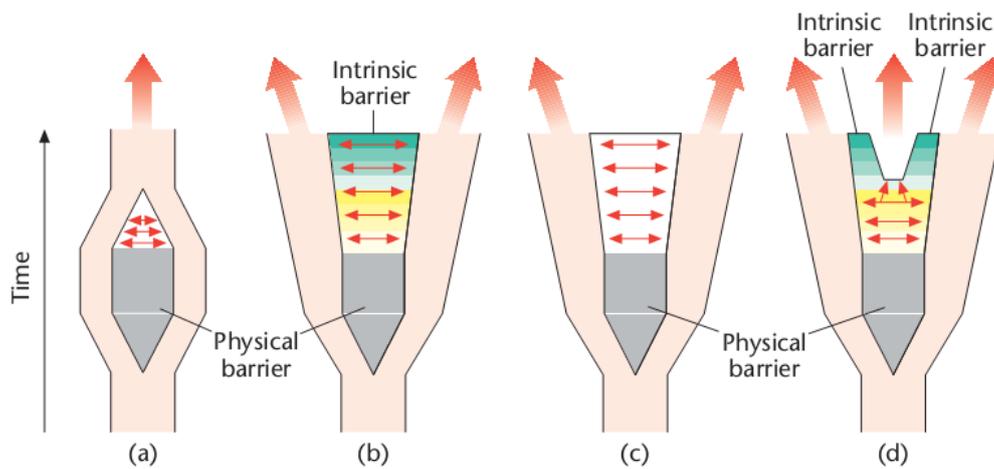


Figure 2. Possible outcomes of hybridization between populations that were previously isolated by a physical barrier. Broad red arrows represent evolutionary lineages (species). Narrow horizontal arrows indicate gene flow between independently evolving lineages. The coloration of the region between diverging lineages depicts the evolution of intrinsic isolating barriers (white, no intrinsic barrier; yellow, weak intrinsic barrier; dark green, strong intrinsic barrier). a) Populations merge and evolve as a single species; b) Reinforcement completes the formation of reproductive isolation; c) Populations retain genetic identity, but form hybrid swarms in contact zones; d) The recombination of differentiated genomes results in a new “hybrid” species. (Figure was taken from Hvala and Wood 2012)

Homoploid hybrid speciation

The evolutionary conditions required for homoploid hybrid speciation are much more stringent than that for polyploidy, and it is more difficult to explain that the hybrid lineage has to maintain itself in the face of gene flow with their parental species. Thus, models for homoploid hybrid speciation must explain how a new hybrid genotype can become reproductively isolated from its parental species which is also the line to distinguish hybrid speciation from adaptive introgression (Abbott and Rieseberg 2012; Abbott *et al.* 2013). To determine the difference between homoploid hybridization (general mixing) and limited introgression is difficult since introgression of genes from one species into another is common in nature and may also increase the genetic variation of the recipient (Baack and Rieseberg 2007; Nolte and Tautz 2010), *Heliconius* butterflies might provide an example (Salazar *et al.* 2010; Heliconius Genome Consortium 2012). Hence, important evidences in demonstrating hybrid speciation are identifying both extensive genomic admixture and unique hybrid traits that cause isolation from the parental species (Abbott *et al.* 2013). The most well studied examples of

homoploid hybrid speciation are the three sunflowers species *Helianthus anomalus*, *Helianthus deserticola* and *Helianthus paradoxus*, which are hybrids between *Helianthus annuus* and *Helianthus petiolaris*. All of these hybrid species exhibit complementary and favorably interacting (additive and epistatic) gene combinations that make them superior to their parents in extreme habitats (Rieseberg *et al.* 1996, 1999, 2003). The availability of an unoccupied habitat and the expanded ecological tolerances seem to be prerequisites for the establishment of a hybrid lineage (James and Abbott 2005; Abbott and Rieseberg 2012). Under a hybrid speciation scenario, an emerging hybrid lineage gives rise to novel traits and is protected from further backcrossing and direct competition with pure parental species by colonizing a new ecological habitat that is absent in the parental lineages (Barton 2001; Burke and Arnold 2001; Abbott and Rieseberg 2012). Evolutionary novelty can arise in hybrids through the segregation and recombination of additive genetic factors or epistasis alleles that originate from the parental lineages, and the new genetic combinations have not been tested by selection in the parental species (Abbott *et al.* 2013). This process is called transgressive segregation and is a very common outcome of hybridization in animals and plants (Rieseberg *et al.* 1999, 2003). However, hybridization could contribute to the origin of reproductive isolation but processes other than hybridization are still necessary to complete speciation in this scenario (The Marie Curie Speciation Network 2012).

The process of incipient hybrid speciation

To date, the genetic analyses of speciation have focused mostly on complete intrinsic reproductive isolation between species and on hybrid sterility or inviability rather than on ecological based barriers to gene flow (Via 2009). This only represents a retrospective look at the speciation process and will hinder a deeper understanding of the molecular basis of all aspects and stages of the speciation process (Wolf *et al.* 2010; Via 2009, 2012). Here we shall also focus on how quickly the process of hybrid speciation takes to complete (Abbott *et al.* 2010). Nolte and Tautz (2010) have recently emphasized the need to focus on the genetic changes occurring in the initial stages of hybrid speciation rather than on complete ones. Analysis of hybrid swarms or young hybrid lineages can play an important role in elucidating the first steps towards hybrid speciation because they focus on the most dynamic phase in the rise of a new species (Nolte and Tautz 2010, Figure 3). They can facilitate to test key predictions from models of hybridization and hybrid speciation (Barton 2001; Buerkle *et al.* 2000). For example, the first step after the secondary contact of two evolutionary lineages is the formation of a hybrid swarm (Figure 3). New genetic combinations occur in this hybrid swarm and hybrid lineages most likely emerge when hybrids can colonize a new ecological habitate that not utilized by their parental lineages (Nolte and Tautz 2010; Abbott and Rieseberg 2012). Extrinsic selection can then

maintain the distinct hybrid lineage even though the initial barriers to gene exchange with their parents are not complete (Abbott *et al.* 2013). Examples of very recently emerged homoploid hybrid lineages within the last 300 years, such as *Cottus* fishes (Nolte *et al.* 2005a; Stemshorn *et al.* 2011), Italian sparrows (Hermansen *et al.* 2011) and Appalachian swallowtail butterflies (Kunte *et al.* 2011), have shown that significant evolutionary changes towards hybrid speciation can emerge in relatively short historical timescales and the hybrid lineages can remain distinct even facing only partial reproductive barriers with their parents. While such hybrid lineages may not give rise to well defined species at the end, by studying the genetic basis of partial reproductive isolation between these lineages and their parents, early barriers to gene flow can be identified before they become confounded with other species differences.

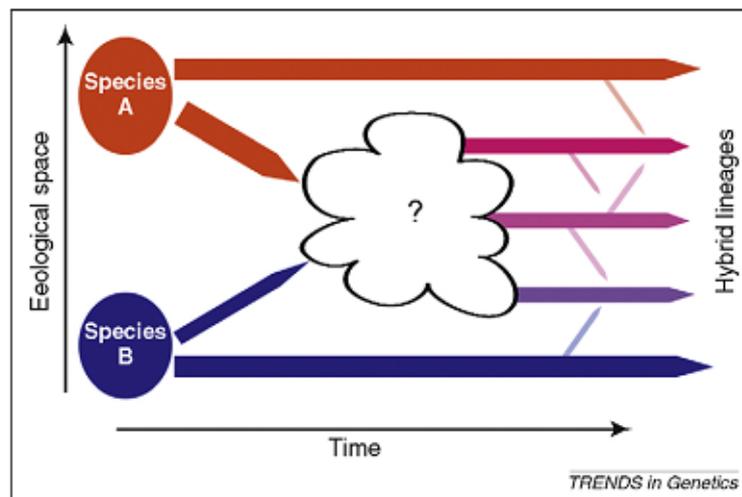


Figure 3. The process of homoploid hybrid speciation. Individuals from two established species meet in an ecological space that is not part of the normal ecological range of their parental species. They can form a hybrid swarm, whereby the contributions from the parental species will normally be uneven. Hybrid lineages can emerge from these hybrid swarms and are expected to occupy different ecological niches than their parental species. Gene flow can still continue to some extent among them and their parental species. The processes occurring within the hybrid swarm until the formation of a hybrid lineage are largely unknown, but are expected to differ genetically from other processes of speciation. (Figure was taken from Nolte and Tautz. 2010).

4. A model for initial hybrid speciation: *Cottus* system and current progress

In Central Europe, freshwater sculpins (genus *Cottus*) are inhabitants of well oxygenated cold streams and lakes, and they are generally absent from

downstream habitats such as large rivers or artificial canals (Nolte *et al.* 2005a). Most *Cottus* species tend to be isolated in their particular habitats as ecological factors do not allow them to migrate across main rivers (Kottelat and Freyhof 2007). Studies on the phylogeography of sculpins across most of their European range found that deeply split lineages occur in non-overlapping ranges (Englbrecht *et al.* 2000) and these were later described as separate species (Freyhof *et al.* 2005). The molecular divergence among these species suggests that they originated before the Pleistocene's glaciation cycles and persisted for a long time, which illustrates that long term isolation mechanisms generally prevent admixture (Englbrecht *et al.* 2000). The only exception to this general pattern was observed in the lower Rhine River where the presence of mitochondrial haplotypes was detected that could only be explained by secondary contact of different species (Nolte *et al.* 2005a). An important taxonomic character for *Cottus* is the presence, absence and extent of scale-like skin prickles (prickling) cover the body of different species (Koli 1969; Freyhof *et al.* 2005). On the other hand, patterns of prickling do not reflect the phylogeny of freshwater *Cottus* through enormous geographic distance (Kinziger *et al.* 2005), which implies that the phenotype has evolved independently on several occasions and suggests that the trait is subject to parallel, recurrent or long term stabilizing evolutionary processes.

Here our initial interest was based on the simple but well founded observation that in Central Europe, sculpins began to spread within the lower reaches of the River Rhine and particularly within the Dutch canals and artificial lake systems (De Nie 1997; Nolte *et al.* 2005a). Sculpins were not known from these bodies of water before, which is in line with the general ecological requirements of them. However, records of sculpins within the last ~30 years indicate a population increase within manmade or strongly perturbed habitats in the Netherlands (Nolte *et al.* 2005a). They are referred to as “invasive”, because they invaded perturbed habitats that are not occupied by other *Cottus* species (Nolte *et al.* 2005a). The emergence of the invasive lineage was inferred briefly as that man-made environmental changes within the past 200 years have increased the connectivity of the rivers Rhine and Scheldt and massively altered the ecological conditions in parts of these drainages (Figure 4). Integrated molecular, morphological and ecological data demonstrated that the invasive lineage of *Cottus* originated from hybridization between two *Cottus* species (*Cottus rhenanus* and *Cottus perifretum*) from smaller tributaries of different river systems (Nolte *et al.* 2005a, Figure 4). While a phylogeographic analysis indicates that the parental species have been separated for up to 2 million years (Englbrecht *et al.* 2000, but see Volckaert *et al.* 2002 for an alternative view), the hybrid lineage are possibly less than 200 generations old and were first detected at a massive scale in the Netherlands after 1980 (De Nie 1997; Nolte *et al.* 2005a). Intriguingly, the particular fish that colonized the new habitats are hybrids whereas ancestral populations are still confined to small streams in the study area. This leads to the hypothesis that

hybridization may have contributed to generating the new ecological potential that apparently distinguishes invasive *Cottus* from their ancestors.

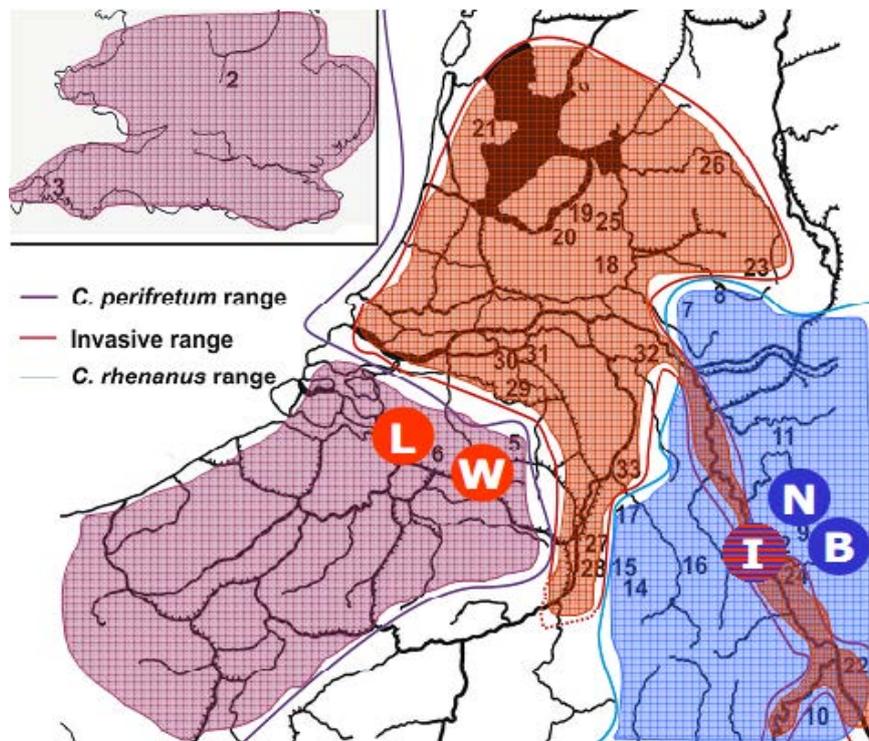


Figure 4. Distribution of *Cottus* species studied in this dissertation. The area shows centers around the Rhine delta and includes Belgium, Netherlands, Luxembourg and parts of France and Germany. The inset represents part of Great Britain at a different scale. The approximate extent of the three *Cottus* groups that are considered here (the parental species *C. perifretum* and *C. rhenanus*, as well as the Invasive hybrid lineages) are indicated by coloured areas. The capital letters represent the populations of each parental species and the invasive lineage that I mainly studied in this dissertation (*C. perifretum*: L: Laarse beek; W: Witte Nete; *C. rhenanus*: N: Naaf; B: Broel; Invasive: I: Sieg). (Figure was taken and amended from Stemshorn *et al.* 2011)

The ranges of different lineages of *Cottus* abut where small streams disemboque into larger rivers. The analysis of these contact zones provides evidence that natural selection prevents free admixture of *Cottus* at these habitat transitions (Nolte *et al.* 2006, 2009). The view that invasive *Cottus* indeed differ ecologically also receives support from studies at hybrid zones. Invasive *Cottus* have expanded their range by colonizing large rivers in an upstream direction (Nolte *et al.* 2005a, 2006). This range expansion has led to secondary contact and hybridization between the invasive sculpins and one of their ancestors, the Rhine stream sculpins (*C. rhenanus*), since the middle 1990s (Nolte *et al.* 2006). Hybrid zones now found at these regions are narrow and coupled with steep ecological gradients, which strongly suggests ecological differentiation and selection prevent

admixing of invasive *Cottus* and their ancestors (Nolte *et al.* 2006). So far, the above results were discussed only with a strong focus on ecological settings but it is still difficult to disentangle the effects of intrinsic and extrinsic barriers. In regard to the evolution of the invasive lineage itself, a small fraction of the invasive genome appears to be subject to genotypic selection and required to maintain the invasive lineage as a whole (Stemshorn *et al.* 2011). An analysis of changes in gene expression provides evidence that novel phenotypes have evolved in the invasive gene pool (Czypionka *et al.* 2012).

One of the most conspicuous features of invasive *Cottus* is the fact that their phenotypes particularly resemble one of their ancestors, *C. perifretum*, more than the other one, *C. rhenanus* (Nolte *et al.* 2005a). This includes life history related traits (allocation of energy into growth versus early maturation) but also morphological features like the degree to which spine-like scales cover the body, the number of vertebrae and also body shape (Nolte 2005; Nolte *et al.* 2005a). Clearly, the similarity of invasive *Cottus* with *C. perifretum* contrasts with the fact that the invasive genome is thoroughly admixed. A comprehensive study of invasive ancestry based on SNP markers has revealed an ancestry coefficient of 0.6 in favor of *C. perifretum* with few markers that are differentially fixed (Stemshorn *et al.* 2011). It is hence of interest to identify the genetic basis of phenotypic traits in *Cottus* and to study the evolution of these traits in the hybrid lineage. Moreover, skin prickling constitutes a major source of variation among several species of freshwater sculpins (Freyhof *et al.* 2005) which suggests that it represents a fast evolving character in these fishes and makes it a character of interest to understand evolution of sculpins in general.

5. Fishing for the secret of *Cottus* evolution: scope and study system of this dissertation

Over the last 10 years, the technology behind genomic studies has undergone rapid development. This includes the introduction of microarrays for high-throughput gene expression analyses, and most importantly, the development of next-generation sequencing (NGS) allowing large-scale genome or transcriptome sequencing even for non-model organisms. The improvement of these technologies means that the genomic methods are becoming more and more useful for addressing questions about speciation in wild populations than ever before. As illustrated in figure 5 by Rice *et al.* (2011), different methods can be combined to identify genomic regions that are important in speciation. Both the genome scan and microarrays can test a large number of loci throughout the genome and find many potential genes that show greater differentiation than expected under neutral model. Once the candidate loci are identified, target

sequencing may be performed on these loci to allow additional tests for signatures of selection. With the availability of a pedigree, linkage mapping becomes possible, and conserved synteny can be compared between different populations or species. Regions with an inversion detected are especially likely to harbor speciation genes. With the further information of phenotypic data, quantitative trait locus (QTL) mapping can be performed to link genomic regions with specific phenotypes and even candidate gene can be identified (Rice *et al.* 2011).



Figure 5. An illustration of combining a variety of genomic methods to identify genomic regions that are important in speciation. Moving from the left to the right, an increasing amount of information is required for each method, which results in an increasing ability to tie specific genomic regions to specific traits. (Figure was taken from Rice *et al.* 2011)

Ecological or speciation-related hybrid traits often contribute strongly to the maintenance or ecological expansion of the new lineage, and genomic mapping of the hybrid traits is not yet available for any animal case. The hybridizing *Cottus* now is rapidly emerging as a new model system to study the questions about the initial stages of hybrid speciation. In my dissertation, I combined several genomic methods as described above to infer answers to these questions under the hybrid speciation scenario of *Cottus*. The goal of this dissertation is to characterize the genetic basis of evolutionary differentiation between *Cottus* species and to learn about how genetic variations that originated from the ancestral lineages can affect the fitness of a newly emerging hybrid lineage. This dissertation consists two chapters.

In chapter 1, I used interspecific F_2 genetic maps to identify genomic regions in the parental species that might cause hybrid incompatibilities and test whether these regions manifest in a sex specific manner to learn about processes that affect natural hybridization in *Cottus*. Meanwhile conserved synteny between *Cottus* and sequenced fish genomes permits to genetically study the *Cottus* genome through the transfer of genomic information from model organisms. The ability to anchor the *Cottus* map with sequenced fish genomes was quantified.

In chapter 2, I combined QTL mapping with a candidate gene approach to learn about traits that are divergent among *Cottus* species and are found to be conspicuous between invasive *Cottus* and its parental species. Candidate genes from Ectodysplasin (EDA) signaling pathway were isolated and mapped using the same F₂ crosses from chapter 1. QTL mapping was performed to learn the genetic basis of skin prickling and body shape of *Cottus*. Ancestral allele frequencies from the invasive gene pool were analyzed along the candidate regions to test whether or not these regions are in line with the neutral model and might be subject to selection.

Chapter 1

The genomics of incompatibility factors and sex determination in hybridizing species of *Cottus* (Pisces)

Jie Cheng, Till Czypionka and Arne W. Nolte

Abstract

Cottus rhenanus and *Cottus perifretum* have formed hybrid lineages and narrow hybrid zones that can be best explained through the action of natural selection. However, the underlying selective forces as well as their genomic targets are not well understood. This study identifies genomic regions in the parental species that cause hybrid incompatibilities and tests whether these manifest in a sex specific manner to learn about processes that affect natural hybridization in *Cottus*. Meanwhile we also explore the possibilities to genetically study the *Cottus* genome through the transfer of genomic information from model organisms. Interspecific F₂ crosses were typed for 255 markers for genetic mapping and to detect transmission distortion as signs for genetic incompatibilities. The *Cottus* map consists of 24 linkage groups with a total length of 1575.4 cM. A male heterogametic (XY) sex determination region was found on different linkage groups in the two parental species. Conserved synteny with sequenced fish genomes reveals homology relationships which indicate that sex determination in *Cottus* is not based on the same genomic regions found in model organisms. Genetic incompatibilities were incomplete, varied among populations and individuals and were not associated with the heterogametic sex. The results suggest a fast evolutionary turnover of sex determination and do not provide evidence for fitness effects related to Haldane's rule. As a whole, the candidate loci identified here can slow down gene flow. However, the great variance between populations and individuals makes it unlikely that incompatibility factors and loci affecting sex differentiation cause intrinsic selection that could affect the gene pool of natural hybrids in a simple and predictable way.

Key words: scorpaeniformes; postzygotic isolation; reference genome; hybrid speciation; hybrid zone.

Introduction

Evolutionarily young and ecologically distinct lineages that diverge although introgressive hybridization, albeit limited, is still possible (Schluter 2000; Via 2009; Presgraves 2010) are of high interest to understand how new species form. In such systems, the nature of barriers to reproduction and genetic factors that affect the fitness of hybrid offspring represent a central aspect to explain evolutionary processes. For example, an important debate relates to the relative importance of exogenous and endogenous barriers as species diverge (Coyne and Orr 2004; Wolf *et al.* 2010) or whenever secondary hybridization occurs (Barton and Gale 1993; Bierne *et al.* 2011). Besides contributing to barriers of gene flow, genetic factors that affect the fitness of hybrids and their offspring are crucial to understand the evolutionary dynamics that take place when a lineage of hybrid origin emerges. An admixed gene pool initially comprises a diverse mixture of alleles with varying fitness effects in different genotype combinations. Accordingly, selection can set evolutionary trajectories that determine which combination of parental alleles is purged from an admixed gene pool. To date, this process has received little attention in empirical studies on the first steps of hybrid speciation (Nolte and Tautz 2010). Like the discussion that revolves around barriers to reproduction it is furthermore possible that loci are affected by purely intrinsic or extrinsic selective pressures.

We are studying European freshwater sculpins (*Cottus* spp.) because they can yield insights into the early evolutionary processes that lead to hybrid speciation (Nolte *et al.* 2005a; Nolte and Tautz 2010; Stemshorn *et al.* 2011) and to learn about the genetic factors that mold narrow hybrid zones (Nolte *et al.* 2006, 2009). Briefly, man-made environmental changes within the past 200 years have increased the connectivity of the rivers Rhine and Scheldt and massively altered the ecological conditions in parts of these drainages. These perturbations have fostered hybridization of two fish species (*Cottus rhenanus* and *Cottus perifretum*) from smaller tributaries (Nolte *et al.* 2005a). While a phylogeographic analysis indicates that the parental taxa have been separated for up to 2 million years (Englbrecht *et al.* 2000, but see Volckaert *et al.* 2002 for an alternative view), the hybrid sculpins are possibly less than 200 generations old and were first detected at a massive scale in the Netherlands after 1980 (De Nie 1997; Nolte *et al.* 2005a). They are referred to as “invasive”, because they invaded perturbed habitats that are not occupied by their parental species. The ranges of different lineages of *Cottus* about where small streams disembody into larger rivers. Our analysis of these contact zones (Nolte *et al.* 2006, 2009) provides evidence that natural selection prevents free admixture at these habitat transitions. With respect to the evolution of the invasive lineage itself, a small fraction of the invasive genome appears to be subject to genotypic selection (Stemshorn *et al.* 2011). Further, an analysis of changes in gene expression provided evidence that novel phenotypes have evolved (Czypionka *et al.* 2012), which highlights the need to identify the evolutionary processes at work. To date, the above results were discussed with a strong focus on ecological settings. However, it can be difficult to disentangle the effects of endogenous and exogenous barriers as correlations are expected to arise between unlinked loci in structured populations (Bierne *et al.* 2011) and

because hybrid zone clines maintained by endogenous or exogenous barriers are difficult to distinguish (Barton and Gale 1993; Kruuk *et al.* 1999; Taylor *et al.* 2012). Hence, testing of alternative hypotheses is warranted in order to assess which mechanisms contribute to emergence or maintenance of species.

The classical mechanism that is thought to critically affect hybrid fitness is given by Dobzhansky-Müller incompatibilities that cause intrinsic postzygotic isolations (Coyne and Orr 2004). These result from genetic changes that, while functional in their normal genetic backgrounds, reduce the viability or fertility when recombined in F_1 hybrids. In F_2 crosses or later generation hybrids, intrinsic traits could cause selective mortality if essential regulatory interactions are disrupted when ancestral genotypes segregate (Maheshwari and Barbash 2011). The evolution of endogenous barriers to reproduction is believed to be a slow process that is outpaced by ecologically driven adaptive evolution (Coyne and Orr 2004; Wolf *et al.* 2010). For fishes, F_1 hybrid inviability is thought to evolve more slowly than in other taxa and estimates range between 11.6 - 24.81 myr (Russel 2003; Bolnick and Near 2005). However, partial effects, especially those that manifest only in the F_2 generation, are expected to evolve more quickly (Edmands 1999; White *et al.* 2011). The evolution of such incompatibilities can be explained through different hypothesis related to Haldane's rule (Kulathinal and Singh 2008; Wolf *et al.* 2010). Accordingly, the first incompatibilities between young species would manifest in the heterogametic sex. Although it is not clear whether this applies to fishes in general (Russel 2003), the genetic regions that determine the sex and the identification of the heterogametic sex are of particular interest to study the evolution of intrinsic barriers to reproduction.

Genetic mapping experiments can complement or challenge the view that differentiation in *Cottus* is determined by ecological factors. Progress in mapping experiments relies on linking markers with underlying genes which remains a challenge in non-model organisms. A possible path to accelerate mapping in unexplored species is given when genomic information from fully sequenced reference organisms can be transferred. Provided that there is sufficient conservation in the sequence and in the gene order (collinearity), evolutionary research can turn towards study systems that are of high interest for evolutionary genetic questions, but which are essentially unexplored to date. We have previously generated low resolution genetic maps by typing 154 random microsatellite loci in F_1 crosses among invasive (hybrid lineage) *Cottus* and *C. rhenanus* populations from Germany. Although the map order could not be well resolved within linkage groups, the maps demonstrated a general utility of distantly related model organisms as reference genomes for *Cottus* and found no evidence for major chromosomal rearrangements (Stemshorn *et al.* 2005, 2011). In this study, we generated an evenly covered genetic map from F_2 interspecific crosses among pure *C. rhenanus* and pure *C. perifretum* to scan the whole genome for the presence of genetic factors that could affect the fitness of hybrids and to study sex determination in *Cottus*. We evaluated whether incompatibility factors manifest in the heterogametic sex and discussed the relevance of the results in the light of evolutionary processes that determine the outcome of hybridization and admixture of *Cottus* within the Lower River Rhine basin.

Moreover, the refined map served to quantify synteny conservation with model fish genomes which permits transferring positional information and to tentatively explore the gene content of candidate regions affecting sex differentiation in the *Cottus* genome.

Materials and Methods

Mapping strategy and taxa used

We first identified SNPs that have a high allele frequency differences between the parental species and used these to analyze interspecific F_2 crosses in a mapping experiment. This ensures a high proportion of fully informative markers in F_2 families, i.e. where linkage phase can be traced from the parental generation through to the F_2 crosses. Interspecific F_2 crosses can be used to construct linkage maps and to map regions that determine sex differentiation with the limitation that species specific chromosomal rearrangements are difficult to resolve, if occurring. Although combining mapping data is challenging, the use of several families and populations provides us with independent replicates to test the generality of patterns. The focus of this study is on four separate populations representing the ancestral species (Freyhof *et al.* 2005) that gave rise to invasive sculpins. *C. rhenanus* were sampled from the small streams Broel (GIS: 50°50'N 7°22'E; Stream Broel between Bröl and Winterscheidt, Germany) and Naaf (GIS: 50°53'N 7°15'E; east of Wahlscheidt, Germany); *C. perifretum* were sampled from the small streams Witte Nete (WN) (GIS: 53°14'N 5°04'E; Flanders, Belgium) and Laarse Beek (LB) (GIS: 51°17'N 5°04'E; Flanders, Belgium). *C. rhenanus* from the Broel and Naaf populations display significant differentiation with an F_{ST} of 0.3 (Nolte *et al.* 2006) and the same holds for *C. perifretum* populations from the Witte Nete and Laarse Beek ($F_{ST} = 0.368$, Knapen *et al.* 2003). The study of gene expression patterns also underlined that the same populations we are studying here have diverged within and between species (Czypiona *et al.* 2012).

We obtained two independent groups of interspecific F_2 crosses from pure *C. perifretum* and *C. rhenanus* and also performed alternative cross directions with respect to the sex of the parents (Table 1). The F_2 crosses were generated from different outbred grandparents. All fishes were kept in laboratory aquaria and fed with invertebrates. The temperature and light regime mimicked the conditions in Central Europe with a winter temperature of 4°C for at least one month. Spawning occurred in artificial shelters when water temperatures were raised to 8-10°C (March 2010). A first group of 625 individuals were sacrificed when they had reached a total length of approximately 3 cm (September 2010). The genotypes of these fish were used to estimate recombination rates for map construction but sex was not determined. Another 273 individuals were raised until the gonads were mature for visual examination in March 2011 and used to map the genetic fragments underlying sexual differentiation. Fishes were sacrificed after anesthesia with CO₂ and preserved in RNAlater or 70% ethanol at -20°C. All F_2 mapping families used for map construction and QTL analysis were shown in Table 1.

Table 1. Interspecific F_2 crosses between *C. rhenanus* and *C. perifretum* were used for map construction and to identify the genetic basis of sex differentiation. The cross type indicates the origin of the parental populations (Broel, Naaf, LB, WN) and the gender of the parental generation. Individual mapping families are listed together with the total number of offspring that were raised and used for map construction or QTL mapping of sex. The number of males and females is given for each mapping family. Sex ratios were found to be significantly skewed towards a male excess in family ZG46 but this pattern was not generally observed.

Cross type	Mapping family	Total	Map construction	Mapping sex	Male	Female	Sex ratio	balanced sex ratio(p-value)
Broel(♂) × WN(♀)	ZG11	104	104	NA	NA	NA	NA	NA
	ZG18	76	76	NA	NA	NA	NA	NA
	ZG33	52	NA	52	24	28	0.462	0.579
	ZG42	97	NA	97	51	46	0.526	0.612
WN(♂) × Broel(♀)	ZG40	86	58	28	18	10	0.643	0.131
	ZG46	75	55	20	15	5	0.750	0.025
LB(♂) × Naaf(♀)	ZG26	117	96	21	10	11	0.476	0.827
	ZG32	115	98	17	10	7	0.588	0.467
	ZG35	38	NA	38	21	17	0.553	0.516
Naaf(♂) × LB(♀)	ZG38	70	70	NA	NA	NA	NA	NA
	ZG39	68	68	NA	NA	NA	NA	NA

Cottus transcriptome sequencing

Individual transcriptome libraries were generated from four individuals of each parental population as well as six invasive *Cottus* from the river Sieg (GIS: 50°48'N 7°9'E; Sieg west of Siegburg, Germany). The latter samples were solely included in the assembly of EST contigs but not used to ascertain SNPs. Total RNA from whole fishes was extracted using the TRIzol Reagent protocol (Invitrogen). Enrichment for poly(A) mRNA was conducted using uMACSTM mRNA isolation kits (Miltenyi Biotec) and full-length complementary DNA was synthesized from each poly(A) mRNA sample following the SMARTTM PCR cDNA synthesis protocol (Clontech). All cDNA samples were PCR amplified using Advantage[®] 2 PCR kit (Clontech) with PCR conditions as follows: 1min @ 95°C, followed by 16 - 18 cycles of (30s @ 95°C, 30s @ 65°C, 6min @ 68°C). Amplification products were quantified using Agilent RNA chips (Agilent Technologies). We followed the 454 GS FLX Titanium General Library Preparation Method Manual (April 2009) from Roche, but used adapters with individual tags to assign reads back to individuals after sequencing. Equimolar amounts of all libraries were pooled and sequenced on a 454 GS-FLX DNA Sequencer following the protocol from Roche.

EST contig assembly and sequence comparison with model fish genomes

Sample specific tags were removed from the sequences using Roche 454 analysis software SFF Tool and reads were assigned to the individuals of origin. Base calling was performed using PyroBayes (Quinlan *et al.* 2008). Primer

trimming and *de novo* assembly were performed with 454 analysis software Newbler with default criteria. The resulting EST contig sequences (>500bp) were blast mapped against publicly available genomes of 5 fully sequenced teleost model fish species: zebrafish (*Danio rerio*), fugu (*Takifugu rubripes*), tetraodon (*Tetraodon nigroviridis*), medaka (*Oryzias latipes*), and three-spined stickleback (*Gasterosteus aculeatus*). The nucleotide, EST and protein sequences of the reference genomes were downloaded from the Ensembl Genome Browser (<http://www.ensembl.org/index.html>) and the International Fugu Genome Consortium (<http://www.fugu-sg.org/project/info.html>). Blast searching was tuned for distant homologies with parameters (-r 1 -q -1 -G 1 -E 2 -W 9 -F /"m D/" -U) described by Korf *et al.* (2003) using an E value < 1e-5 as a threshold to accept a hit. The genomes were searched using blastn and blastx and the number of hits was noted at the nucleotide as well as at the protein sequence level. The similarity of the *Cottus* transcriptome to the EST data sets from model fishes was further explored using the program *kr* (<http://guanine.evolbio.mpg.de/kr2/>) by Domazet-Lošo and Haubold (2009). This implements a method for efficiently estimating the Jukes-Cantor distances between sets of unaligned sequences based on the distribution of exact match lengths. The EST data sets investigated here differed in size by up to 5.2 – fold. To exclude the majority of random matches between non-homologous regions arising from the comparison of such unbalanced data sets we ran *kr* with the -P parameter set to 0.7, as discussed by Haubold *et al.* (2008).

SNP selection

To identify polymorphic SNP markers, individual reads were mapped against the EST contigs obtained from the *de novo* assembly using the Mosaik suite of programs (http://bioinformatics.bc.edu/marthlab/Main_Page). Assemblies were screened for SNPs using Gigabayes. The output from Gigabayes was filtered using a custom Perl (<http://www.perl.org/>) script to identify candidate informative SNPs if 1) the SNP distinguished all samples at the species level 2) the SNP was covered at least once for each of the 4 ancestral populations and 3) at least four reads covered each of the 2 species. The value of the resulting SNP loci to distinguish *C. rhenanus* and *C. perifretum* was validated for 220 candidate SNP loci by PCR amplification and Sanger sequencing of independent pooled DNA samples representing populations of *C. perifretum* (Witte Nete, Laarse Beek, Zwane Beek) and *C. rhenanus* (Broel, Naaf, Flaumbach) (comp. Stemshorn *et al.* 2011). Sequences from each parental pool were visually inspected for alternatively fixed SNPs (data not shown).

Linkage map construction

SNP carrying loci were blast mapped against the stickleback genome to identify a set of 384 SNPs (334 EST-based loci and 50 loci from Stemshorn *et al.* 2011) with inter locus intervals of about 1.5 Mb according to the stickleback genome. A custom GoldenGate genotyping assay (Illumina) was genotyped for all F₂ mapping families using the Illumina BeadXpress system following the recommendations of the manufacturer. Linkage analysis was carried out using

R/qtl (Broman *et al.* 2003; <http://www.rqtl.org>) following the latest version of the manual (<http://www.rqtl.org/tutorials/geneticmaps.pdf>). Mapping families of different sizes were combined according to their population of origin and the direction of sex of the parents as shown in table 1. This assumes that there is no between family variation in the genetic maps and that recombination does not differ among families, which is an assumption that can hardly be verified. However, combining families is necessary in order to obtain better overall estimates of recombination fractions among loci. For map comparisons with different teleost fish genomes a consensus map was generated based on the combined genotype data for all mapping families.

Individuals and markers with more than 50% missing data or with obvious genotyping problems were removed from the raw genotype data. Because the SNP markers we used might not be entirely diagnostic for the ancestral species, some SNPs were not fully informative in all of the mapping families (60 loci were not fully informative in Broel × WN crosses and 58 loci were not fully informative in Naaf × LB crosses). Genotypes for loci in non-fully informative families were marked as missing data in order to reconstruct genetic maps and to map QTL. We retained only markers that showed no strong indication for segregation distortion in the F₂ offspring (p -value < 1e-10) for the map construction. Independent linkage maps were constructed for the groups of crosses as defined by their population of origin. Linkage was estimated using the function `est.rf` to obtain pairwise recombination fractions and to calculate a LOD score for a test of $rf = 1/2$. Two markers were placed in the same linkage groups when they have an estimated recombination fraction ≤ 0.35 and LOD score ≥ 4 . We ordered the markers within each linkage group, then compared all possible orders and choose the best order using the `ripple` function. A consensus *Cottus* map was constructed by combining all mapping families for comparison with reference genomes.

Synteny and collinearity analysis

Synteny is defined as consistent linkage between certain genes across species genomes and this definition does not require conservation of gene order or orientation within chromosomes (Gaut 2001). This condition was visualized using oxford grids in this study. The comparison was based on the map positions of SNPs in the consensus *Cottus* map versus the blast hit locations for the respective loci in reference genomes. Collinearity refers to the conservation of the order of orthologous loci within chromosomes between two species (Gaut 2001; Lukens *et al.* 2003). Permutation of map data permits testing whether a block of n collinear markers indicates identical genome organization between two genomes. The method adopted here to assess the significance of collinear regions is based on the approaches outlined by Gaut (2001) and Lukens *et al.* (2003) but has some modifications. We considered the strict definition, in that loci in the collinear region must be in the same order between species and not be interrupted by other markers. Accordingly, it does not matter which genome is taken as a “reference” or as a “tester” to detect collinear blocks. If a collinear region with n shared markers was detected, it was scored. The score is defined as the average distance (centimorgan, cM) between the loci within the collinear region. The

physical distance (Megabase, Mb) of model fish genomes was transferred into genetic distance (cM) according to the mean ratio of genetic distance to physical distance for each *Cottus* linkage group and corresponding model fish chromosome. To test whether the score of a collinear region could be expected by chance or not, markers were randomly and uniformly assigned new map positions within the respective linkage group for 10,000 times. In the permuted data, the largest collinear block with the reference genome was identified and the score was calculated as above for each permutation. The probability that the observed collinear region could be observed due to random association between loci was evaluated against the distribution of p – values ($p < 0.05$) from the permutations. The test for collinear regions was implemented using a custom Perl script.

Mapping of regions affecting sex and survival

Sex was examined for 273 F_2 offspring after gonad maturation by raising the water temperature after the winter (Table 1). Male and female gonads morphology was inspected in ethanol preserved specimens. Male fishes have a pair of flat and prolonged white gonads while the female gonads are rounded and have an orange or brown color, sometimes with discernible eggs inside. Available sex phenotypes were used for map based QTL analysis in R/qtl whereby we followed Broman and Sen (2009). Sex was mapped independently for crosses combined according to population of origin (Broel \times WN and Naaf \times LB) and considering the paternal sex direction of the cross (*C. perifretum* ♂ vs. *C. rhenanus* ♂). We have carefully checked whether all loci were informative for individual mapping families (i.e. if grandparents were differentially fixed and where both parents from the F_1 generation were heterozygous) before combining families into mapping populations. The trait considered was sex (coded as a binary character: male - 0; female - 1), and no other covariate was included in the model. Standard interval mapping and a binary model were used for the genome scan. Likelihood of odds (LOD) significance thresholds for each cross was determined by permutation testing (1000 permutations). Significant QTL was above the genome-wide LOD threshold ($\alpha=0.05$). The confidence interval was examined by 95% Bayes interval flanking the QTL peak.

Evidence that genomic regions affect the survival of F_2 crosses was sought by testing all loci included in the consensus map for segregation distortion. This test can reveal deviations from mendelian inheritance induced by differential survival of certain genotypic classes. A chi-square test was used to detect deviations from the expected genotypic proportions of 1:2:1 (homozygotes AA, heterozygotes AB and homozygotes BB, respectively) for each single mapping family. The significance threshold to identify candidate loci for segregation distortion was set to $p < 0.05$ and not corrected for multiple testing. We expected a corresponding number of false positive results due to chance and we have reported patterns of segregation distortion only when minimally two neighboring loci show a significant and congruent signal which we take as evidence that the underlying genetic loci have a sufficiently strong effect. Further, it was assessed whether segregation distortion was observed only in single crosses or whether the pattern appeared in parallel crosses.

Both sex and differential survival have in common that the underlying loci cannot be easily included in the generation of a genetic map. Sex linked loci may occur in a hemizygous state and loci that affect survival will show deviations from mendelian inheritance. Moreover, deviations from mendelian inheritance can also result from genotyping artifacts. These phenomena bias the estimation of recombination fractions. Excluding the affected loci is necessary to apply mapping functions used here (comp. linkage map construction) but implicates a danger to lose particularly interesting candidates related to sex differentiation or genetic incompatibilities. Our strategy to explore loci excluded from the map construction was to transfer positional genomic information by tentatively assigning a position on the *Cottus* map based on blast hits in the stickleback genome (comp. collinearity analysis). While the observed patterns can contain artifacts due to genotyping error and differences in marker order between species, they can also add evidence for segregation distortion of candidate regions. To complement QTL mapping and to confirm hemizygous inheritance of sex determining regions, raw genotypic data for each locus, including loci where the F₁ generation was not fully heterozygous in some families, was screened for associations of parental genotypes with the offspring's sex using the marker regression method from R/qtl. Moreover, additional markers were typed that could be placed based on blast searches against the stickleback genome. This included three length polymorphism markers (CottE1-indel, Ctg06004i-indel, Cott108-indel, Table S1) and two microsatellite markers (CottE1, Cott108) from Nolte *et al.* (2005b). These markers were typed by separating PCR products from fluorescently labeled primers on a capillary sequencer. The actual map position of the three length polymorphism markers was confirmed through patterns of segregation in our F₂ mapping families. The genomic fragments carrying the microsatellites CottE1 and Cott108 both contain SNP or indel markers that were already included in our SNP based map. Hence, patterns of inheritance of parental alleles could be analyzed for the actual microsatellites CottE1 and Cott108 in sexed offspring to test for the presence of a heterogametic sex. To test whether recombination was suppressed at sex-linked markers, we calculated male and female meiotic recombination frequencies for loci spanning the QTL interval for groups of F₂ crosses.

Results

Sequence similarity between *Cottus* and model species

A total of 1,124,134 reads with an average length of 299 nucleotides were obtained from 454 sequencing of cDNA libraries. The assembly yielded 11,548 contigs larger than 500bp, which were mapped against 5 teleost model fish genomes and protein databases. The number of blastx hits that detect similarity at the amino acid level was relatively similar for all reference genomes (46.2%-51.2%). However, there was a conspicuous variance in the number of significant blast hits at the level of the nucleotide sequence. The highest number of hits was found when comparing *Cottus* ESTs to the stickleback genome with significant hits for 94.9% of the contigs. The medaka (71.6%), tetraodon (69.3%), fugu (75.3%) and zebrafish (52.2%) genomes also permit to locate homologues of the *Cottus* contigs but their genomes are less conserved. The genetic distance at the

nucleotide sequence level was also evaluated for the combined set of *Cottus* EST contigs compared to model fish EST datasets based on the number of substitutions per site between pairs of sequences. It showed again that the transcriptome sequences of *Cottus* are more similar to the stickleback than to the other model fish species with the lowest genetic distance (estimated per nucleotide substitution rate = 0.08) to the stickleback as opposed to values greater than 0.10 in all other comparisons (Figure S1).

SNP discovery and map construction

The assembly of individual reads against the EST contigs yielded 1,302 SNPs that were ancestry-informative according to our criteria. These, together with SNPs from a previous study (Stemshorn *et al.* 2011), were used to identify a set of 384 SNPs that were evenly distributed across the 21 chromosomes contained in the most recent version (2006) of the stickleback genome assembly. A custom 384-plex GoldenGate SNP genotyping assay was typed for 898 F₂ individuals and their parents. After excluding SNPs found to be monomorphic (9) or in which genotyping had a low success rate for most of the mapping families (77), data for 298 SNPs were available for further analysis. Genotypes for loci in non-fully informative families were marked as missing data in order to reconstruct genetic maps and to map QTL. Independent linkage maps were constructed for the groups of crosses as defined by their population of origin (Figure S2). For the Broel × WN map, we had 234 SNPs that mapped to 24 linkage groups totaling a length of 1437.2 cM with an average interval of 6.8 cM. For the Naaf × LB map, 238 SNPs mapped to 24 linkage groups of a length of 1319.0 cM with an average mapping interval of 6.2 cM. The two maps share a total of 201 markers.

The conservation of synteny and collinearity between the two *Cottus* maps was analyzed. All markers show a 1:1 representation on single linkage groups and the maps are highly collinear (Figure S2). Only 4 single markers do not show the same order between the two maps. Notably, all of them mapped to the end of one linkage group in one of the *Cottus* maps. The test for significance of collinearity shows that the two *Cottus* maps share 27 significantly collinear linkage blocks containing 2-14 markers, which covers 98.75% of the genetic map. Given the fully conserved synteny between the two *Cottus* maps and under the assumption that the observed differences between maps are minor, we combined data for all mapping families to generate a consensus map for comparisons with reference genomes. A total of 252 markers (Table S1) mapped to 24 linkage groups in the *Cottus* consensus map at a LOD threshold of 4 with a total map length of 1575.4 cM and an average distance between markers of 6.8 cM (Figure 1). In comparison with the preliminary maps of *Cottus* (Stemshorn *et al.* 2005, 2011) the maps generated here are shorter and do not have an exceedingly large linkage group.

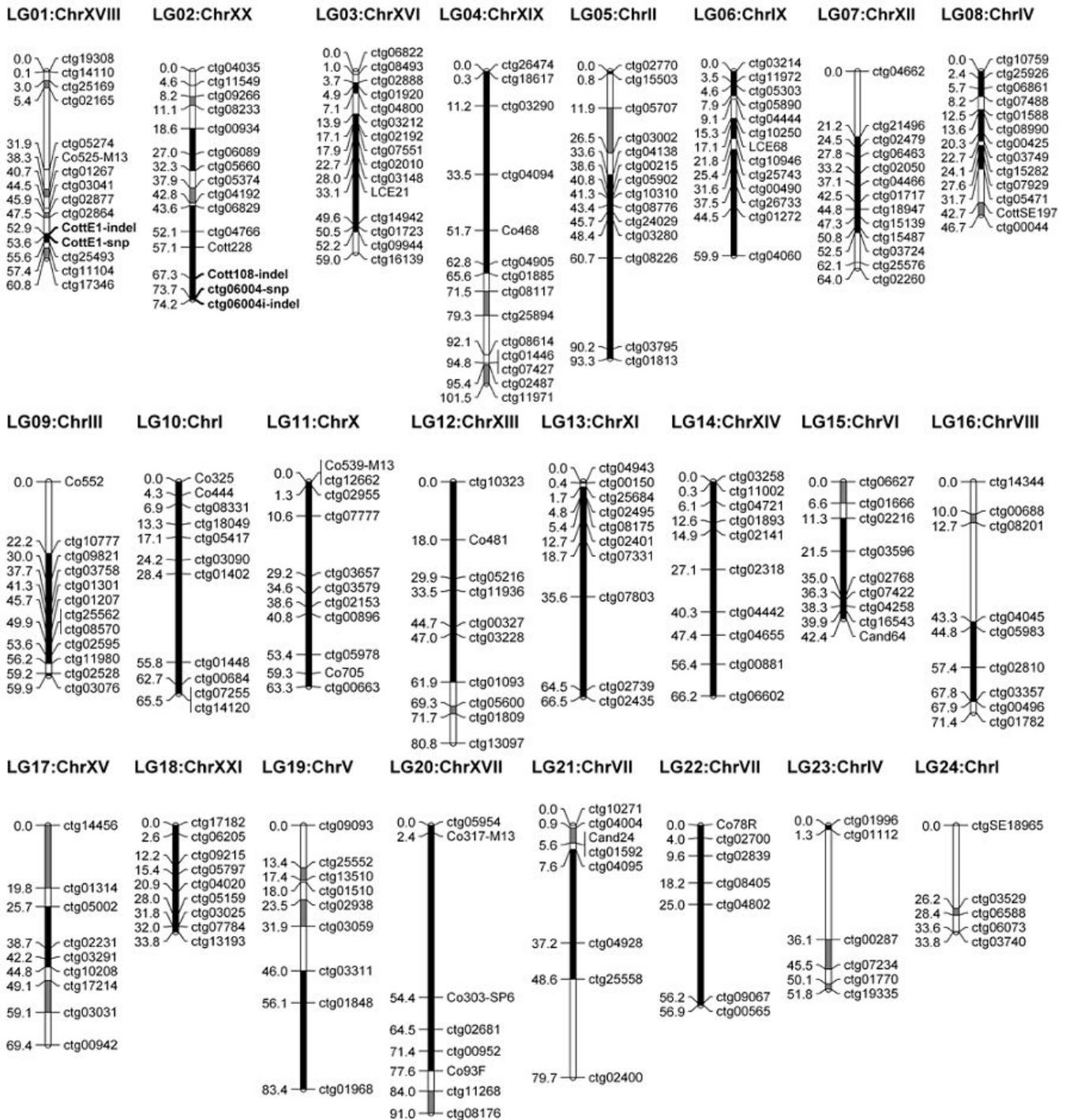


Figure 1. A consensus genetic map derived from interspecific F_2 crosses between *C. renaeus* and *C. perifretum* can be best anchored in the stickleback (*Gasterosteus*) genome which permits the transfer of positional information to explore the *Cottus* genome. Filled black blocks within linkage groups (LG) represent significantly collinear regions ($p < 0.05$) with a perfectly conserved marker order to relative stickleback chromosomes (Chr). Grey blocks are collinear but this was not found to be significant. Each of the parental species has a single independent heterogametic region that determines sex. These regions were identified on linkage groups 1 and 2 (see text) and are highlighted through bold marker names.

Conservation of synteny and colinearity between *Cottus* and model species.

Blast searches permit establishing syntenic relationships of *Cottus* with the reference species which were summarized in the form of Oxford grids (Figure 2 and Figure S3). A one-to-one relationship among many known chromosomes and linkage groups can be observed. The majority of *Cottus* linkage groups are associated with single stickleback chromosomes with only 7 markers that do not follow this rule (Figure 2). Some chromosomal rearrangements suggestive of fusion/fission events distinguish *Cottus* from the stickleback which appears to have fused chromosomes 1, 4 and 7 that fall into separate linkage groups in *Cottus*. The correlation of genetic maps gets weaker when comparing *Cottus* with the medaka, tetraodon, fugu and zebrafish genomes which correspond with the decreasing pattern of nucleotide sequence conservation (Figure S3). Testing for significance revealed that a number of collinear blocks could be expected by chance given the number of sampled markers per chromosome (Table 2). *Cottus* shares 34 significantly collinear blocks with stickleback, which account for 60.62% of the *Cottus* genetic map, followed by medaka ($n = 26$, 39.15%), fugu ($n = 21$, 32.44%), tetraodon ($n = 21$, 31.59%) and the zebrafish genome ($n = 7$, 7.36%). Finally, the correlation of the physical distance in stickleback and genetic distance in *Cottus* within collinear blocks is shown in figure 3. The average physical distance in stickleback of a recombination unit in *Cottus* is 4.25 cM/Mb (range: 0.000006 cM/Mb to 11.67 cM/Mb).

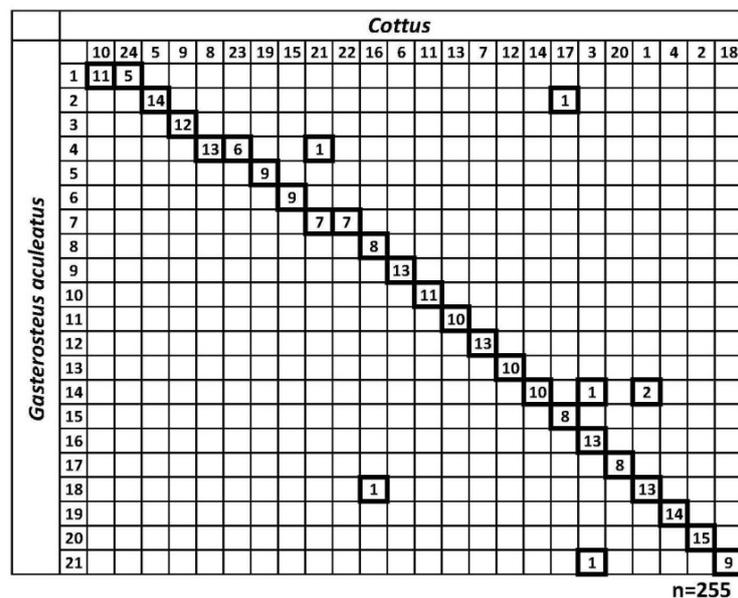


Figure 2. Oxford grid of synteny relationships between *Cottus* linkage groups (x – axis) and stickleback chromosomes (y – axis) based on BLAST mapping of *Cottus* EST contigs from which markers were derived. Numbers in fat squares indicate shared markers among genetic regions. Six instances of possible inter - chromosomal rearrangements were noted and chromosomal fusion events are suggested for stickleback chromosomes 1, 4 and 7.

Table 2. Conservation of marker order between genomes was assessed by randomizing markers within linkage groups. The number of significantly collinear regions, the number of markers contained in them and the proportion of the genome that is covered were determined between *Cottus* and sequenced fish genomes. Besides the almost perfectly conserved marker order between the two *Cottus* maps studied here, the stickleback shows the highest degree of genome collinearity with *Cottus*.

	Collinear Blocks (n)	Markers per Block	percentage of <i>Cottus</i> map
<i>Cottus</i> (within)	27	2-14	98.75%
Stickleback	34	2-11	60.62%
Medaka	26	2-8	39.15%
Fugu	21	3-9	32.44%
Tetraodon	21	2-9	31.59%
Zebrafish	7	2-3	7.36%

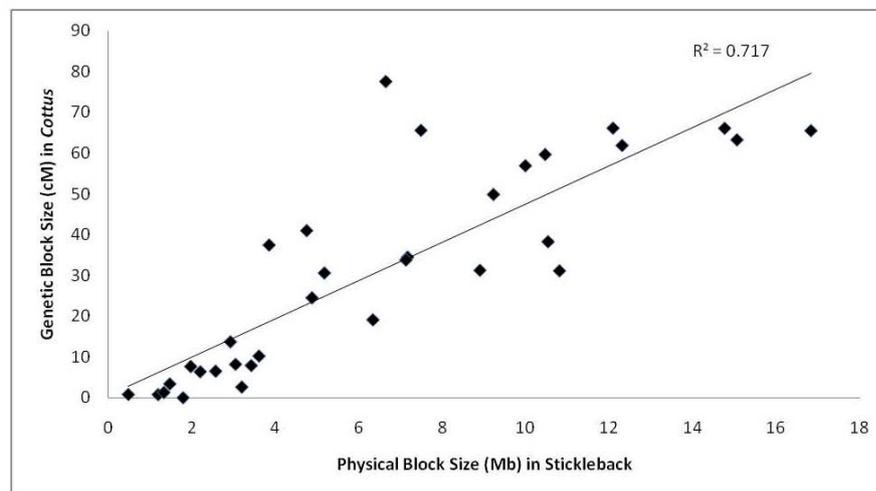


Figure 3. Correlation of physical map distance with genetic map distance within significantly collinear fragments in *Cottus* and the stickleback. Pairwise physical distances (base pairs) in the stickleback genome and genetic distance (centimorgan) in *Cottus* are correlated ($r^2 = 0.717$) and permit a rough prediction of *Cottus* map distances from the stickleback genome. This relationship is useful for the analysis of small genomic regions as covered in this graph, while it gets distorted across whole chromosomes due to the presence of intra chromosomal rearrangements between stickleback and *Cottus*.

Genomic regions affecting sex and genetic incompatibilities

Sex could be determined for 273 F₂ individuals. The QTL analysis, which serves to detect genomic regions determining sex differentiation, revealed one significant region on LG01 (LOD=6.91) for Broel(♂) × WN (♀) crosses. In contrast, the two other types of crosses both revealed a single significant QTL on LG02 ((WN(♂) × Broel (♀) crosses (LOD=4.8) and LB(♂) × Naaf(♀) crosses (LOD=9.67)) (Table 3 and Figure S4). Note that the first type of cross involves male *C. rhenanus*, while crosses that suggest an alternative QTL location involve male *C. perifretum*. The significance of these candidate regions was further supported when additional

fragment length polymorphism markers were typed and analyzed. CottE1-indel is linked to sex QTL locus CottE1 on LG01. Likewise, ctg06004i-indel and Cott108-indel are linked to sex QTL locus ctg06004 on LG02. The LOD scores of the initial QTL candidates combined with the added markers increased for both candidate regions (Table 3 and Figure S4). To complement this inference, raw genotype data including those for loci that were excluded from the mapping were tested for single locus association with sex. One additional marker (ctg05522, Table S1) was found only for WN(♂) × Broel (♀) crosses. The flanking sequence of this locus had a blast hit on stickleback genome indicating that the locus is linked to the candidate region detected on LG02. Hence, it does not suggest the presence of additional genomic regions that affects sex differentiation. Sex specific recombination rates were estimated for both candidate sex determining regions in the respective crosses. Recombination was observed in both sexes and was not consistently smaller in either sex (Table 3). Parental microsatellite alleles permitted to trace the inheritance of alleles from the F₁ into the sexed F₂ offspring in six mapping families (Table 4). The segregation patterns revealed that males inherit an allele that is strongly associated with sex which indicates a heterogametic (XY) sex determination system. Seven families from three different types of crosses were available to compare the sex ratios in the F₂ generation (Table 1). The sex ratios did not deviate significantly from balanced proportions except for one family (ZG46) that had a significant male excess which corresponds with an excess of the heterogametic sex.

Table 3. QTL mapping results for sex differentiation in interspecific *Cottus* F₂ crosses according to the origin and gender of the parental populations. QTLs that exceed the significance threshold (LOD) with imputation method are reported with their linkage group (LG), interval size, maximum LOD score and the most closely linked marker. Recombination fractions for male and female F₂ offspring were determined for the markers spanning the significant QTL intervals.

Crosses	LG	Interval Size (cM)	LOD	Linked Marker	Recombination fraction (♂:♀) [cM]
Broel(♂) × WN(♀)	1	6.1	6.91	CottE1-indel	7 : 8.8
WN(♂) × Broel(♀)	2	25.16	4.8	ctg06004i-indel	24.4 : 23.6
LB(♂) × Naaf(♀)	2	6.9	9.67	ctg06004i-indel	15.1 : 12.5

To determine if the candidate regions in *Cottus* have similarities with sex determination regions in other fish species, we used Oxford grids (Figure 2 and Figure S3) to determine homology of the *Cottus* sex QTL with stickleback, medaka and fugu chromosomes. This analysis was complemented with published records (Kikuchi *et al.* 2007, Takehana *et al.* 2008, Kondo *et al.* 2009, Ross *et al.* 2009, Myosho *et al.* 2012, Kamiya *et al.* 2012) of sex determination loci accessible through the reference species (Table S2). The region represented by CottE1 on *Cottus* - LG01 is homologous with parts of stickleback chromosome 18 and medaka chromosome 24, while the region represented by ctg06004 on *Cottus* - LG02 is homologous with parts of stickleback chromosome 20 and medaka chromosome 16. None of these regions are associated with the sex determination regions in stickleback (chromosome 19) and medaka (chromosome 1). The same holds for additional closely related species that are accessible through the reference genomes of this study and for which the sex determination system has

been mapped (Table S2). As an exception, *Cottus* - LG02 is homologous with the sex chromosome in *Oryzias javanicus* (medaka chromosome 16) (Takehana *et al.* 2008), however with the difference that the latter species has a female heterogametic sex.

Table 4. The segregation of dam and sire microsatellite alleles could be traced for six mapping families. Each sub table summarizes results for cross type (origin and gender of the parental populations), the concerned linkage groups with the marker name and genotypic data for one of six mapping families. The segregation of alleles (labeled A, B, C, D) from the dam and sire of the F_1 into the F_2 generation reveals a strong correlation of the offspring sex with the presence of a single paternal allele (numbers indicate instances of observed allele per genotype; note that the inheritance of identical alleles cannot be resolved). This pattern is observed in a single family (ZG42) for the paternal allele on linkage group 1 from *C. rhenanus* and five times for the paternal allele on linkage group 2 which originates from *C. perifretum*.

Cross type	Broel(♂) × WN(♀)				WN(♂) × Broel(♀)				LB(♂) × Naaf(♀)						
Microsatellite	LG01 - CottE1				LG02 - Cott108				LG02 - Cott108						
Parents	ZG42	Dam (♀)		Sire (♂)		ZG40	Dam (♀)		Sire (♂)		ZG26	Dam (♀)		Sire (♂)	
		A	B	A	C		B	A	C	A		A	A	A	B
Offspring	♂	25	26	3	48	♂	7	10	13	4	♂	7	0	0	7
	♀	20	24	43	1	♀	4	6	1	9	♀	10	0	10	0
						ZG46	Dam (♀)		Sire (♂)		ZG32	Dam (♀)		Sire (♂)	
							B	A	C	D		A	A	A	B
						♂	10	0	10	0	♂	10	0	0	10
						♀	0	3	0	3	♀	7	0	6	1
											ZG35	Dam (♀)		Sire (♂)	
												A	A	A	B
											♂	21	0	0	21
											♀	17	0	16	1

A general pattern of viability and fertility of F_1 crosses between *C. rhenanus* and *C. perifretum* was observed for replicated pairs of populations and directions of crosses. Likewise, our experiments did not suggest that F_2 crosses were strongly impaired, but showed allele frequency distortions that suggest partial incompatibilities being associated with varying genomic regions. Chi-square tests for deviation from the expected 1:2:1 mendelian segregation indicated that groups of linked markers deviated significantly in the same manner within mapping families (Table 5). However, none of these effects was common or even fixed for the parental species. Of the 27 candidate regions detected among all families, 19 regions were not shared whereas four were shared by maximally two out of seven mapping families. Importantly, none of the regions, for which evidence for transmission distortion was detected, corresponded with the candidate genomic regions that affect differentiation of sex (Table 3). Accordingly, there are no shifts in genotype frequencies in F_2 offspring that would suggest the presence of the hemizygous allele determining sex affecting survival. Markers that were assigned map positions based on synteny relations with the stickleback added only limited evidence confirming patterns observed for markers that were placed directly on the *Cottus* map. We detected four instances where strong segregation distortion

(Table S3) coincides with patterns at regularly mapped loci in *Cottus* (Table 5). The remaining patterns did not correspond to signals in directly mapped markers. The marker ctg05522 is located in the region on LG2 that was found to be involved in sex determination in the cross WN(♂) × Broel(♀) (ZG40,ZG46). This locus shows signs of distortion in a single family ZG46 but the sex ratio in this family is skewed in favor of males (Table 1).

Table 5. Genotypic segregation distortion in individual mapping families. Genomic regions (linkage groups and marker names were given) spanning two or more linked markers with a congruent and significant ($p < 0.05$) deviation from mendelian inheritance are reported. Distorted regions were shared only for pairs of families (marker names bold). Most regions were not shared and no region had a general effect in all crosses.

Cross type	Mapping families	LG	Linked distorted markers
Broel(♂) × WN(♀)	ZG18	10	ctg00684, ctg14120, ctg07255
		12	ctg05216, ctg11936, ctg00327, ctg03228
		13	ctg04943, ctg00150, ctg25684 , ctg02495
		20	ctg00952, ctg02681
		24	ctg03529, ctg06588, ctg03740
	ZG33	11	ctg12662, Co539-M13, ctg02955
		12	ctg01809, ctg13097
		18	ctg07784, ctg03025
		20	ctg02681 , Co303-SP6
		WN(♂) × Broel(♀)	ZG40
3	ctg07551, ctg02192, ctg03212		
5	ctg05707, ctg03002		
ZG46	22		ctg08405, ctg04802
	9		Co552, ctg10777
Naaf(♂) × LB(♀)	ZG39	9	ctg03758, ctg01301
		9	ctg25562, ctg02595
		19	ctg01510, ctg02938, ctg03059
		15	ctg02768, ctg07422, ctg04258, ctg16543
	ZG32	6	ctg03214, ctg11972
LB(♂) × Naaf(♀)	ZG38	6	ctg05890, ctg04444
		10	ctg08331, ctg18049, ctg05417, ctg03090, ctg01402
	ZG38	13	ctg00150, ctg25684
		22	ctg02839, ctg08405, ctg04802 , ctg09067, ctg00565

Discussion

A genetic map can play a central role to integrate results on the genetics of speciation (Bernatchez *et al.* 2010) and linkage maps have become established as important genetic tools to gain knowledge about evolutionary processes such as speciation, hybridization and adaptation (Schluter 2009; Stern and Orgogozo 2009). Interspecific comparisons of genomes aim at identifying the gene content of candidate genetic regions (Lehmann *et al.* 2008) but also contribute to understand broad-scale patterns of genome and chromosomal evolution (Backstroem *et al.* 2008; Sarropoulou *et al.* 2008; McGraw *et al.* 2010). This study has followed these paths for the exploration of a non-model species. We have generated a *Cottus* linkage map that agrees with the haploid chromosome number of 24 as described for the closely related *Cottus gobio* (Vitturi and Rasotto 1990). The map contains an evenly dispersed set of markers and the conserved synteny and collinearity with sequenced reference genomes can serve to bridge considerable phylogenetic distances. It was a central goal of this study to test whether genetic factors exist that may contribute to intrinsic barriers to reproduction between hybridizing species of *Cottus*. Diverse genetic factors that cause incompatibilities were identified in F₂ crosses. Moreover, we have identified the genomic regions that determine sex differentiation in the ancestral species and did not find evidence that incompatibility loci reduce the viability of the heterogametic sex in hybrids. Here we discuss how genomic information from reference genomes contributes to understand the results and discuss the role that the candidate regions identified here may play for the evolutionary processes following hybridization of *Cottus* in the River Rhine system.

Conserved synteny between *Cottus* and stickleback

We found that the zebrafish genome was the least conserved with *Cottus* which agrees with the fact that it is the phylogenetically most distant reference species in our comparisons (Steinke *et al.* 2006). A low degree of conserved synteny has also been reported in other comparisons among Ostariophysi with Acanthopterygii (Kucutas *et al.* 2009). The genome structure of *Cottus* is more conserved between morphologically and ecologically deeply divergent acanthopterygian fishes which was also reported for other Acanthopterygii (Kasahara *et al.* 2007; Sarropoulou *et al.* 2008). *Cottus* shares the highest collinearity and nucleotide sequence similarity with the stickleback. This result agrees with recent phylogenetic studies that have placed the Gasterosteidae into a clade that contains *Cottus* among other Scorpaeniformes fishes (Smith and Wheeler 2004; Kawahara *et al.* 2008; Li *et al.* 2009). Therefore, the stickleback genome could serve as an excellent reference for the entire Scorpaeniformes clade. It appears that the physical distance in the stickleback genome is a reasonable estimator of map distances in *Cottus* with the limitation that this relationship is expected to be fuzzy. This is because the conservation between the two taxa is not perfect and because variation in recombination rates is common among chromosomes in eukaryotic organisms (Petes 2001). On the other hand, finding approximate genome positions of random genomic reads from *Cottus* was highly successful in this study and will

enhance QTL studies through positional and functional candidate gene approaches (Erickson *et al.* 2004; Lehmann *et al.* 2008).

Sex determination differs between species of *Cottus*

A single genetic region was significantly associated with the differentiation of sexes in different crosses. Further, the sex ratios were generally not imbalanced which suggests that sex is heritable and has a simple genetic basis (Devlin and Nagahama 2002). When a male from *C. perifretum* was mated with a *C. rhenanus* female to generate F₂ crosses, a region on LG02 was recovered whereas the opposite cross direction (male *C. rhenanus* mated with *C. perifretum* female) revealed a region on LG01. This finding, together with patterns of segregation of microsatellite alleles, suggests that sex is determined by different male heterogametic regions in the two ancestral species. No consistent sex specific reduction of recombination rates at QTL regions was detected which suggests that the heterogametic region should be rather small as also observed in *O. latipes* (Devlin and Nagahama 2002; Kondo *et al.* 2006). A lack of large heterogametic regions or distinct Y chromosomes corresponds to Vitturi and Rassotto's (1990) observation that the closely related *C. gobio* has no discernible sex chromosome in its karyotype, although heterogametic XY systems have been reported for related species (Devlin and Nagahama 2002).

Conserved synteny among species represents a bridge to complement initial QTL experiments with results from homologous chromosomal locations. A lack of homology of *Cottus* LG01 and LG02 with the sex chromosomes of medaka (Chr01) and stickleback (Chr19) suggests that the genetic basis of sex determination in *Cottus* differs from the reference genomes and also from patterns observed in closely related species (Kikuchi *et al.* 2007; Takehana *et al.* 2008; Kondo *et al.* 2009; Ross *et al.* 2009; Myosho *et al.* 2012; Kamiya *et al.* 2012). Interestingly, LG02 corresponds to medaka chromosome 16 which is the sex chromosome in *Oryzias javanicus* (Takehana *et al.* 2008) albeit with the limitation that, unlike *Cottus*, *O. javanicus* has a ZW sex determination system. Moreover, the XY sex determination of African fish *Nothobranchius furzeri* is also located in a region that is homologous to medaka chromosome 16 (Valenzano *et al.* 2009), suggesting the possibility that sex in *C. perifretum* and *N. furzeri* could be determined by a common system. However, this association has to be reassessed once the candidate regions have been studied in detail. It is not likely that this system is shared due to common ancestry as the species are phylogenetically distant (Table S2). Sex is an important evolutionary component and some key elements of sex determining pathways are conserved across the animal kingdom (Herpin and Schartl 2011). However, the mechanism by which the sex differentiation pathway is triggered varies widely (Marin and Baker 1998; Herpin and Schartl 2011). The fact that two different regions determine sex in closely related species of *Cottus* indicates that factors controlling a conserved sex determination pathway evolve rapidly and have a fast turnover, which is not astonishing given the results from other fish species (Mank and Avise 2009). Detailed genetic studies will have to reveal whether the same genes are repeatedly recruited in different lineages. If this occurs, then similarity of sex determining regions across large phylogenetic

distances as found for *Cottus* LG02 in this study could be explained through convergent evolution. The sex ratios observed here for different crosses varied and there was one significant deviation from a balanced ratio which involved an increase of the heterogametic males. Imbalanced sex ratios have been explained through the disruption of regulatory networks that diverged between species (Devlin and Nagahama 2002; Shinomya *et al.* 2006). On the other hand, it is possible, that environmental factors may influence sex differentiation (Baroiller *et al.* 2009) in different aquaria and these possibilities will have to be assessed in dedicated experiments.

Implications for the genetics of speciation and hybridization in *Cottus*

The presence of genetic incompatibilities may be relevant to understand the persistence of the ancestral *C. perifretum* and *C. rhenanus* throughout glacial cycles (Englbrecht *et al.* 2000) and could affect more recent processes related to hybrid speciation (Nolte and Tautz 2010) or gene flow between the parental species and invasive *Cottus* (Stemshorn *et al.* 2011). No effects on F₁ hybrids viability were observed which is in line with the inference that inviability in fish hybrids was estimated to take 2.2 to 10 times longer to evolve than the assumed divergence between *C. rhenanus* and *C. perifretum* (Englbrecht *et al.* 2000; Russel 2003; Bolnick and Near 2005; Stelkens *et al.* 2009). On the other hand, the time to evolve effects that manifest in the F₂ generation is expected to be shorter (Edmands 1999; White *et al.* 2011) and studies in F₂ crosses represent a powerful approach to detect emerging barriers to reproduction (Maheshwari and Barbash 2011; White *et al.* 2011). We described evidence that the males are the heterogametic sex in *Cottus* and the screen for patterns of segregation distortion in F₂ crosses of *Cottus* revealed a number of candidate regions that are associated with offspring survival in our mapping families (Table 5). However, the absence of signs for segregation distortion at the sex determining regions and the absence of signs for a reduced viability of the males suggest that none of the incompatibilities appeared to manifest specifically in the heterogametic sex. Hence, the incompatibility factors we have identified between *C. perifretum* and *C. rhenanus* cannot be explained through effects related to Haldane's rule. Genetically differentiated sex chromosomes are expected to contribute to intrinsic barriers to reproduction (Wolf *et al.* 2010) but the evolution of incompatibilities associated with heteromorphic sex chromosomes requires some evolutionary stability (Devlin and Nagahama 2002; Bachtrog *et al.* 2008). Hence, it is likely that a small size of the heterogametic regions and a rapid evolutionary turnover prevented that they contribute to barriers to reproduction between *C. rhenanus* and *C. perifretum*.

Incompatibilities in our experiments were hardly consistent between crosses with a strong representation of effects that were encountered in single families. This strongly suggests that incompatibilities vary at the level of differentiated populations and individuals. It is possible, that unknown environmental factors differed among our laboratory tanks or that intrinsically caused transmission distortion (Dermitzakis *et al.* 2000; Koide 2008) played a role which warrants further studies. However, with respect to the evolutionary processes triggered by

hybridization in *Cottus*, our results suggest that these are neither affected by species specific nor by very strong genetic factors. Invasive *Cottus* have emerged from a complex pattern of admixture from diverse sources (Nolte *et al.* 2005a; Stemshorn *et al.* 2011). Hence, a diverse set of genetic incompatibility factors will have entered the admixed gene pool, but most likely none at very high frequencies. According to this study, such factors could cause fluctuating and genotype specific patterns of selection in natural hybrids. The incompatibility loci identified here may thus constitute a fitness burden for admixed individuals, affect hybrid zone structure and slow down secondary gene flow into the invasive gene pool (Nolte *et al.* 2006, 2009; Stemshorn *et al.* 2011). On the other hand they are unlikely to be purged or fixed rapidly in the admixed gene pool due to the family specific nature of the effects. Likewise, natural hybrids of *Cottus* may carry alternative suites of genetic factors that interact to determine sex depending on their paternal origin. The genetic basis for sex differentiation can be subject to strong balancing selection (Conover and Voorhees 1990) when a dominant factor biases sex ratios. However, 6 out of 7 of the crosses analyzed here suggest that both sex determining regions cause a balanced sex ratio in a hybrid genetic background. Moreover, effects on the viability of the heterogametic sex appear to be absent. Accordingly, it would be possible that a mosaic genetic architecture of sex determination prevails in the admixed lineage or in hybrid zones for some time without inducing strong selection.

The results from this screen have not found evidence for prevalent genetic factors that trace back to the ancestral species and determine general evolutionary trends in invasive *Cottus*. This result confirms our previous inference that patterns of introgression vary among individual loci for different hybrid zones (Nolte *et al.* 2009), which is in line with that the genetic incompatibilities are unique to local demes. On the other hand general evolutionary trends exist that distinguish invasive *Cottus* from all ancestral populations studied to date. We have previously described strong genotype-environment associations for *Cottus* in the River Rhine basin. Particularly, that general patterns of adaptive divergence are replicated in nature indicates that they are independent from local genotypic variance. This is true for the parallel molding of hybrid zone clines across habitat borders (Nolte *et al.* 2006) and the evolution of multiple lineages of invasive sculpins in ecologically perturbed habitats (Stemshorn *et al.* 2011). This implies that local genetic variance must interact with more general ecological factors to determine evolutionary processes after hybridization in *Cottus* within the Lower River Rhine basin.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank K. de Gelas, A. Kobler and L. Bervoets and F. Volckaert for their benevolent help in obtaining breeding stocks of *Cottus perifretum*. K. Lessenich and A. Mellin (Berirksregierung Köln) and T. Heilbronner from the Siegfischereigenossenschaft have given permits to conduct the research. *Cottus* were kept and bred in the laboratory of the MPI in Plön with permission from S. Hauschildt (Veterinaeramt in Plön). E. Bustorf, H. Buhtz, S. Dembeck, H. Harre, G. Augustin, D. Mertens and D. Lemke have contributed to the laboratory work and fish care. We thank K. Konrad, F. Chan, B. Haubold and D. Tautz for comments and discussions and D.T. and the Max-Planck Society for generous support. The project was funded by a DFG grant to A.N.

Supplementary Information

Table S1. Loci included in the consensus map for *Cottus* with locus name, segregating alleles that distinguish *C. perifretum* and *C. rhenanus* populations, the sequence from which the marker was developed and that was used for BLAST searches (variable position marked in square brackets).

Table S1 is available as an Excel file and deposited in the Dryad repository (<http://datadryad.org>): doi:10.5061/dryad.7s393

Genotype and phenotype data for QTL mapping is deposited in the Dryad repository (<http://datadryad.org>): doi:10.5061/dryad.7s393

Table S2. Sex determination systems in fully sequenced model fishes or closely related species and the synteny relationships with *Cottus*. Known linkage groups (LG) that are underlying sex differentiation are given and homology with *Cottus* is established based on synteny relationships between *Cottus* and model organisms. For species that were not directly compared to *Cottus*, published homology relationships with the latter species were used to identify corresponding *Cottus* linkage groups (literature cited in text).

	Cross/Species	System	LG	Corresponding LG in <i>Cottus</i>
<i>C. rhenanus</i> x <i>C. perifretum</i>	Broel(♂) x WN(♀)	XY	1	1
<i>C. perifretum</i> x <i>C. rhenanus</i>	WN(♂) x Broel(♀)	XY	2	2
<i>C. perifretum</i> x <i>C. rhenanus</i>	LB(♂) x Naaf(♀)	XY	2	2
Stickleback	<i>G. aculeatus</i>	XY	19	4
	<i>G. aculeatus</i> -Japan	XY	9	6
	<i>G. wheatlandi</i>	X ₁ X ₂ Y	12,19	7,4
	<i>P. pungitius</i>	XY	12	7
	<i>A. quadracus</i>	ZW	#	#
Medaka	<i>O. latipes</i>	XY	1	6
	<i>O. curvinotus</i>	XY	1	6
	<i>O. luzonensis</i>	XY	12	14
	<i>O. mekongensis</i>	XY	2	24
	<i>O. minutillus</i>	XY	8	13
	<i>O. dancena</i>	XY	10	8
	<i>O. hubbsi</i>	ZW	5	20
	<i>O. javanicus</i>	ZW	16	2
Fugu	<i>T. rubripes</i>	XY	19	20

Table S3. A screen for possible candidate loci among markers excluded from map construction because of non-mendelian inheritance. While these loci are possibly affected by genotyping error, they can be assigned a tentative map position based on transferring positional information from the stickleback genome. The table lists markers, genome positions for *Cottus* and Stickleback and indicates for which mapping family significant segregation distortion was detected (\checkmark). Inferred map positions agree for 4 markers that are neighboring with loci in mapped distorted regions for specific mapping families (black squares) (comp. Table 3). These loci add evidence that the transmission distortion at the candidate regions is stronger than revealed by the mapped loci. The remaining patterns do not correspond with signals in directly mapped markers (Table 3) and are likely to contain strong artifacts. Further, the marker ctg05522 is located in the region on LG2 that was found to be involved in sex determination in cross WN(σ) \times Broel (ρ) (ZG40,ZG46) (see results). Although this locus shows signs of distortion in ZG46, the sex ratio in this family is skewed in favor of males (Table 1) which are the heterogametic sex.

Distorted Loci	<i>Cottus</i> LGs	Stickleback Chr.	ZG18	ZG33	ZG40	ZG46	ZG32	ZG38	ZG39
ctg02154	1	XVIII				\checkmark			
ctg05522	2	XX				\checkmark			
ctg01878	3	XVI			\checkmark				
ctg06214	3	XVI	\checkmark		\checkmark				
ctg25574	9	III	\checkmark		\checkmark		\checkmark	\checkmark	
Co522-SP6	10	I			\checkmark				
ctg03568	12	XIII				\checkmark			
ctg12684	12	XIII				\checkmark			
ctg02559	13	XI	\checkmark						
ctg01973	15	VI			\checkmark				
ctg02006	15	VI	\checkmark		\checkmark		\checkmark	\checkmark	\checkmark
ctg03756	15	VI							\checkmark
ctg06311	16	VIII	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark	
ctg04208	17	XV		\checkmark					
ctg01408	21	VII	\checkmark		\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
ctg26734	21	VII		\checkmark					
ctg04762	22	VII							\checkmark
Co476-SP6	23	IV	\checkmark				\checkmark	\checkmark	
ctg00943	23	IV				\checkmark			
ctg03678	23	IV					\checkmark	\checkmark	
ctg01009	24	I	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
ctg03146	24	I	\checkmark		\checkmark	\checkmark			
ctg18413	24	I				\checkmark			

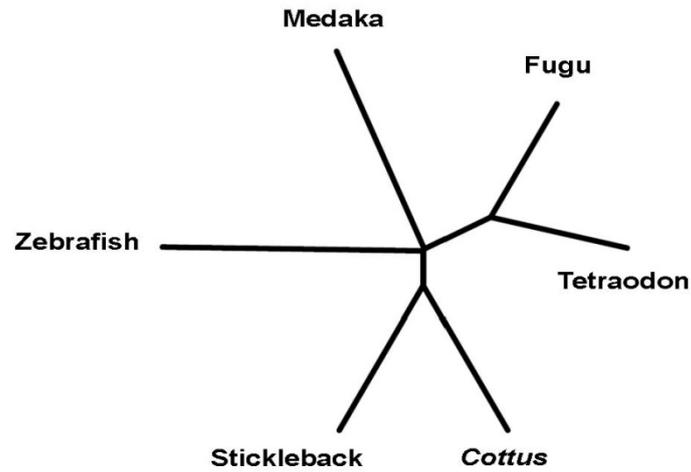


Figure S1. Genetic distance tree based on transcriptome sequences of *Cottus* EST contigs and five model fish EST datasets. Note that the estimated per nucleotide substitution rate used here is not useful to reflect distant phylogenies and yet illustrates that the stickleback transcriptome nucleotide sequence is more similar to *Cottus* than the sequences of other model fish species.

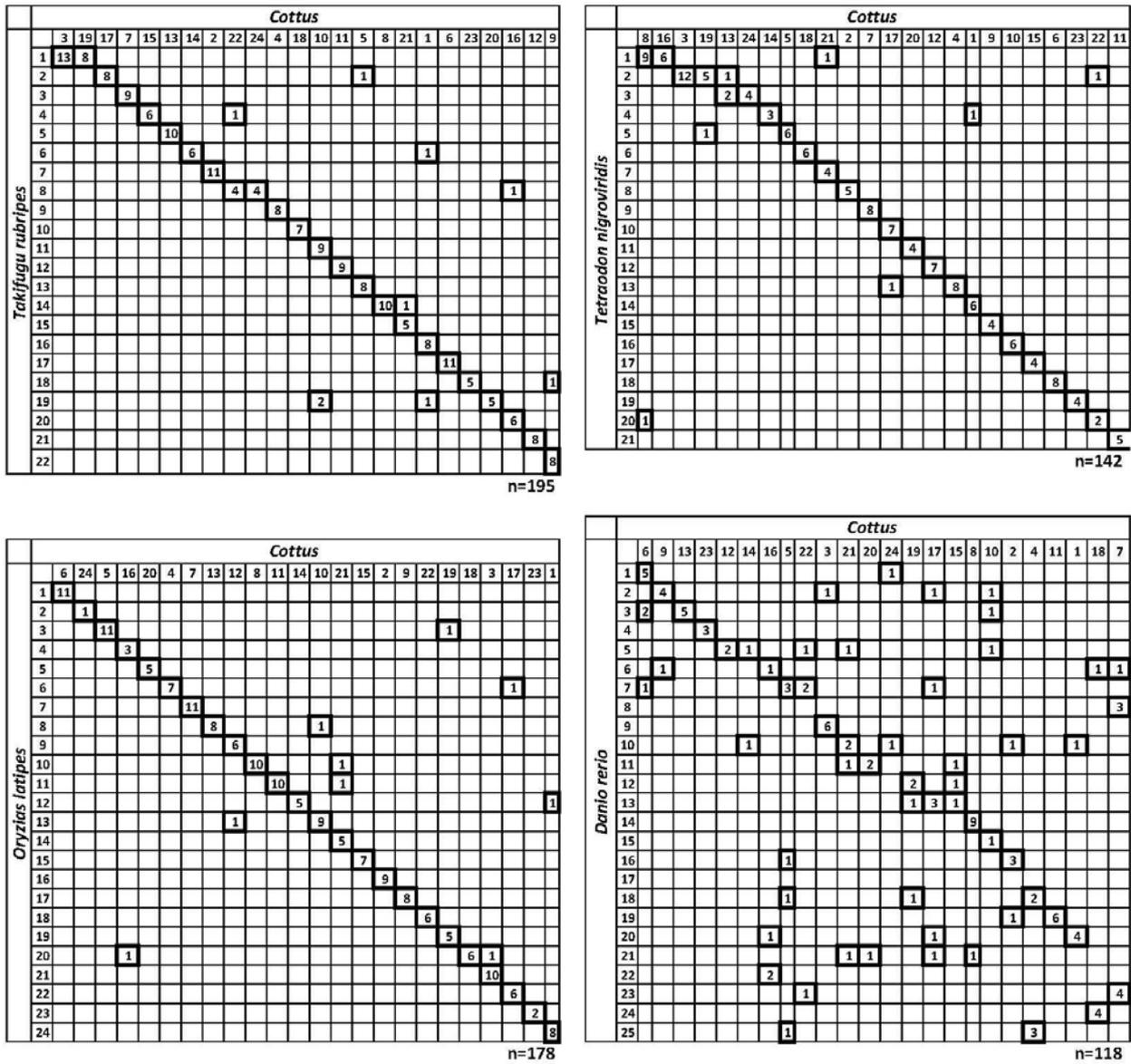


Figure S3. Oxford grids of synteny relationships between *Cottus* linkage groups (x – axis) and chromosomes of the medaka, fugu, tetraodon and zebrafish genomes (y – axis) based on BLAST mapping of *Cottus* EST contigs from which markers were derived. Numbers in fat squares indicate shared markers among genetic regions. None of the genomes is as well conserved with *Cottus* as the stickleback and the zebrafish genome is clearly the most divergent both in terms of chromosome organization and in the low number of markers (n) that could be mapped using BLAST.

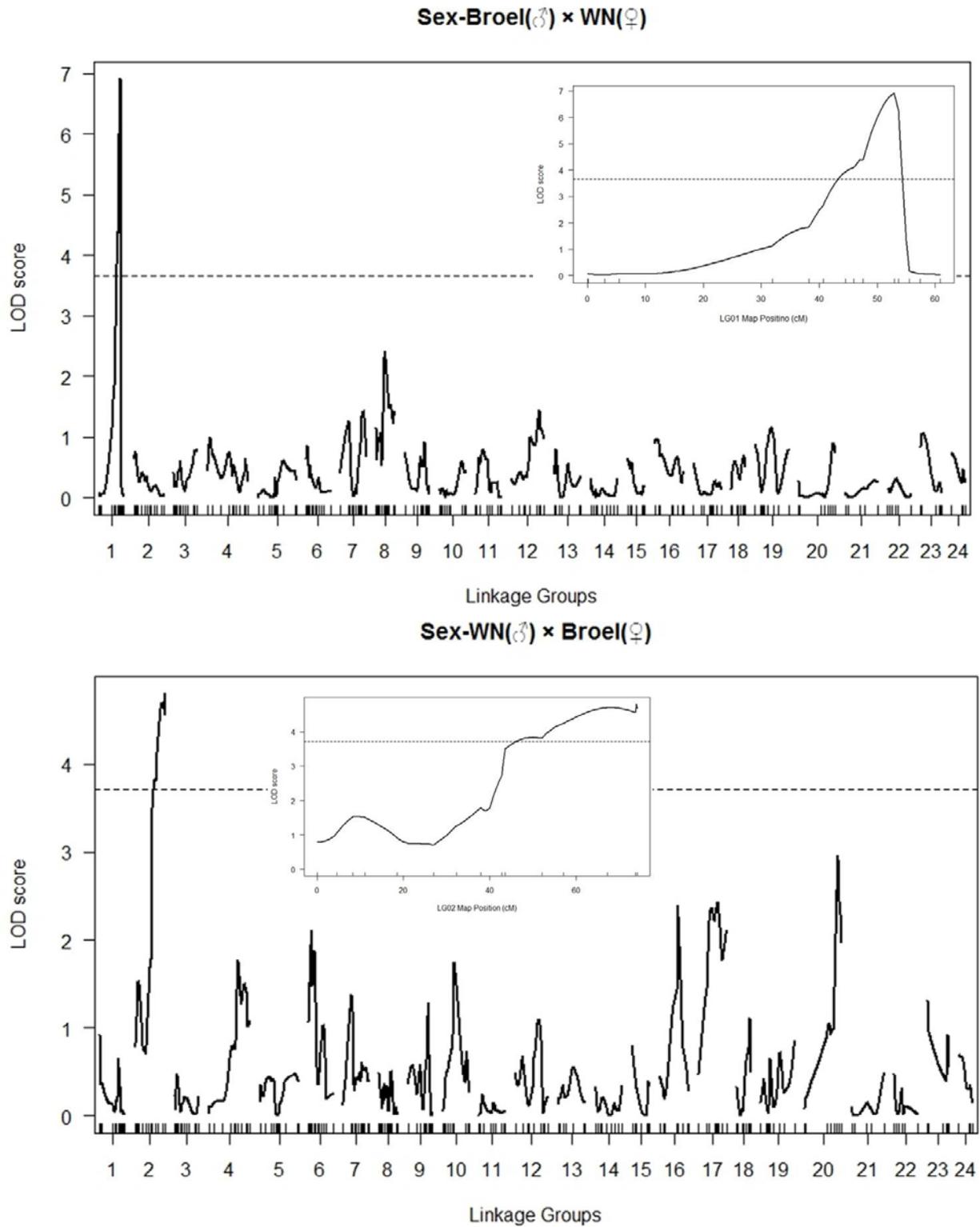


Figure S4. QTL associated with sex determination for each cross type (origin and gender of the parental populations). Genome-wide significant LOD score at $\alpha < 0.05$ with permutation test is indicated as a dashed line for all linkage groups and enlarged for each linkage group that contains the significant QTL of sex determination.

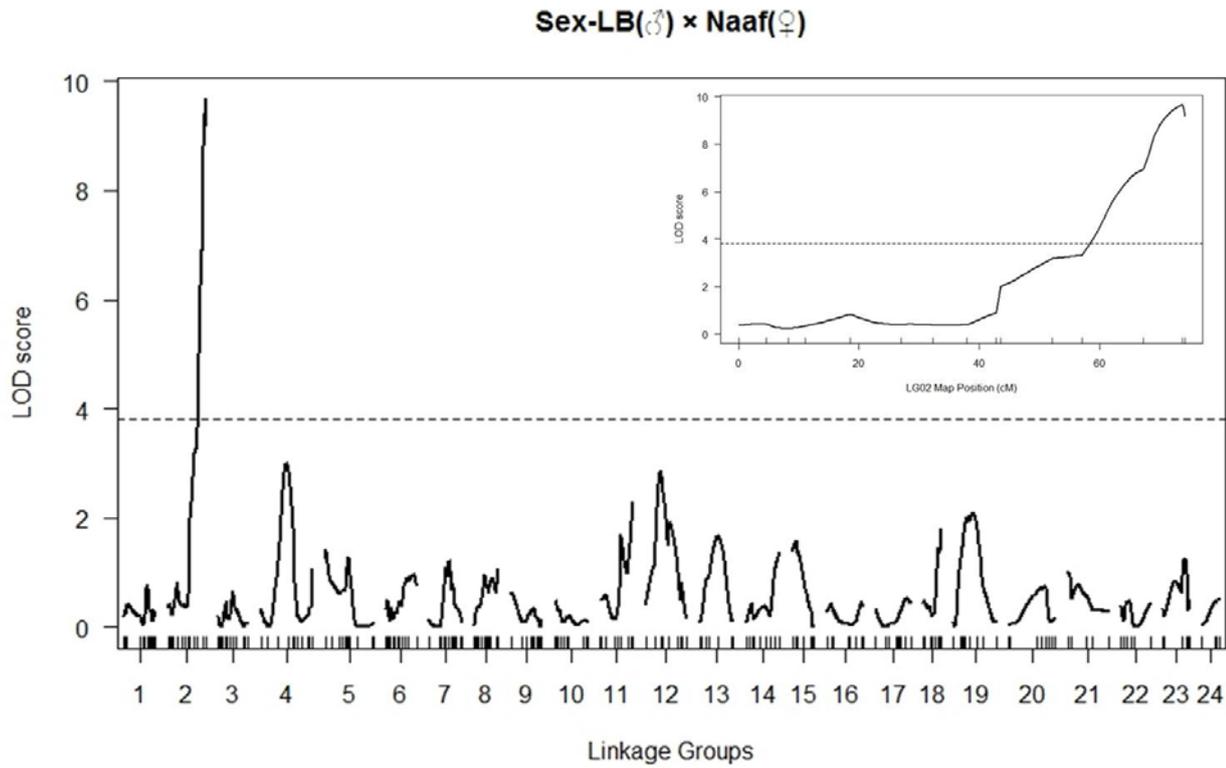


Figure S4. (Continued) QTL associated with sex determination for each cross type (origin and gender of the parental populations). Genome-wide significant LOD score at $\alpha < 0.05$ is indicated as a dashed line for all linkage groups and enlarged for each linkage group that contains the significant QTL of sex determination.

Chapter 2

The genetic architecture of skin prickling and body shape in hybridizing sculpins (*Cottus*)

Jie Cheng and Arne W. Nolte

Abstract

A hybrid lineage (invasive *Cottus*) of sculpins particularly resembles one of its ancestral species in phenotypic features. This contrasts with the expectation of hybrid intermediacy, which is suggested by the fact that the invasive genome is thoroughly admixed. Among these traits, the scale-like skin prickling and body shape represent a common source of variation among a number of *Cottus* species. The Ectodysplasin (EDA) signaling pathway provides promising candidates to investigate the genetic basis of prickling and body shape because it is known to affect the development of dermal bones and scales in fishes. We identified and mapped *Cottus* EDA signaling pathway components and performed quantitative trait loci (QTL) mapping for prickling and body shape in *Cottus*. A single highly significant QTL that affects prickling in all F₂ mapping families was detected in an interval that contains the EDA receptor (Edar) on *Cottus* linkage group 3 but none of the other EDA pathway genes. The same QTL region also shows effects on body shape. An investigation of gene structure from 6 individuals suggests that the genomic architecture of the Edar gene makes it a more likely contributor to evolutionary changes than other components in this pathway. An analysis of ancestral allele frequencies within EDA pathway gene regions in the invasive gene pool shows that Edar genomic ancestry is biased towards *C. perifretum* (85%-96%), which is correlated with the phenotypic similarity of invasive *Cottus* with that species. The Edar carrying QTL is currently the best candidate that strongly determines phenotypic variation of hybrid *Cottus* in the River Rhine. EDA signaling constitutes a key adaptive trait in stickleback and our results suggest that the same pathway may contribute to conspicuous phenotypic variation in *Cottus*.

Key words: prickling; body shape; hybrid speciation; QTL mapping; parallel evolution.

Introduction

Identifying the genetic and molecular basis of phenotypic variation in natural populations is a prerequisite to analyze underlying evolutionary processes in detail. According to Schluter (2009) “the most obvious shortcoming of our current understanding of speciation is that the threads connecting genes and selection are still few” which is also a call to link phenotypes with their underlying genes. One reason for this continuous challenge is that all species are somewhat unique with respect to their environmental and evolutionary history. The best way to increase power in evolutionary biology research is to study evolution in replicate and in a well-defined phylogenetic context by the comparative method (Elmer and Meyer 2011). Therefore, focusing on cases of parallel phenotypic evolution in closely related populations or species is a rigorous framework to infer the genetic basis of adaptive traits (Wolf *et al.* 2010; Elmer and Meyer 2011).

Famous examples of parallel phenotypic evolution used for examining the genetics of adaptation in vertebrates include sticklebacks (Colosimo *et al.* 2005; Hohenlohe *et al.* 2010), whitefishes (Bernatchez *et al.* 2010) and cichlid fishes in Africa (Salzburger 2009). Another good candidate species for parallel phenotypic study is the small freshwater sculpins (genus *Cottus*) which are widely distributed in inland and coastal aquatic habitats across North America and Eurasia. Studies on the phylogeography of *Cottus* across most of their European range found that deeply split lineages occur in non-overlapping ranges (Englbrecht *et al.* 2000) and these were later described as separate species (Freyhof *et al.* 2005). An important diagnostic character for *Cottus* species is the extent to which prickles cover the body (Koli 1969; Freyhof *et al.* 2005). Prickles are remnants of reduced scales, more or less deeply embedded in the skin. In the brackish coastal waters of the Baltic Sea of Northern Europe, non-prickled and prickled forms of *Cottus* have been distinguished in the species *Cottus gobio* and *Cottus koshewnikowi* (Kontula and Vainola 2001). In Central Europe, a strongly prickled lineage arose through hybridization of two *Cottus* species, namely the mostly non-prickled *Cottus rhenanus* from smaller tributaries of River Rhine and the prickled *Cottus perifretum* from smaller tributaries of River Scheldt (Nolte *et al.* 2005a). Additional species with pronounced prickling (e.g. *Cottus petiti*, *Cottus metae*) or lack thereof (e.g. *Cottus rondeletii*, *Cottus scaturigo*) have been described from the European *C. gobio* species complex (Freyhof *et al.* 2005). Likewise, prickled and mostly naked examples also occur in distantly related species from Northern America, like prickled *Cottus asper* and non-prickled *Cottus aleuticus* in Alaska (Scott and Crossman 1973; Baumsteiger *et al.* 2012). Taken together, the phylogenetic relationships among taxa of the freshwater Cottidae do not correspond with the differentiation into prickled and non-prickled *Cottus* species (Kinzinger *et al.* 2005). Therefore, skin prickling constitutes a major source of variation among several

species of freshwater sculpins. It represents a fast evolving character that is relevant to understand evolution of sculpins in general.

Skin prickling of *Cottus* is homologous to fish scales because it represents dermal bones like scales and also shares the same mode of development and location on the fish body. The ectodysplasin (EDA) signaling pathway is a promising candidate to study prickling in *Cottus* since this pathway is involved in early organogenesis of vertebrate skin appendages, such as teleost fish scales or teeth, bird feathers or mammal teeth, hair and glands (Pantalacci *et al.* 2008). At least five different genes in this pathway have previously been shown to produce ectodermal dysplasia phenotypes in mice and human (Thesleff and Mikkola 2002; Morlon *et al.* 2005; Knecht *et al.* 2007). In fish species, ectodysplasin (Eda) and its receptor Edar are specific for the development of adult scales, skeletal and dental structures in zebrafish (Harris *et al.* 2008). Edar is also required for scale development in medaka (Kondo *et al.* 2001). Moreover, Eda is the major locus controlling variation of bony lateral plates between wild marine and freshwater stickleback populations (Colosimo *et al.* 2004, 2005). As mutations in multiple components of EDA signaling pathway produce similar phenotypes in both human, mice and fish species, it suggests that the EDA signaling pathway is highly conserved in vertebrates (Thesleff and Mikkola 2002; Morlon *et al.* 2005; Pantalacci *et al.* 2008). The repeated recruitment of Eda alleles in cases of parallel adaptive evolution of stickleback species (Colosimo *et al.* 2005; Hohenlohe *et al.* 2010) suggests that the same pathway might be recruited repeatedly in *Cottus* species as well. Moreover, the similarity of high and low plate morphs in sticklebacks with strongly and lowly prickled phenotypes in *Cottus* possibly extends the idea of parallel evolution as described for stickleback to a much higher level of divergence.

We are focusing on the Central European freshwater sculpins *C. rhenanus* and *C. perifretum* because they can yield insights into the early evolutionary processes of hybrid speciation (Nolte *et al.* 2005a; Nolte and Tautz 2010; Stemshorn *et al.* 2011). A hybrid lineage of *Cottus* between these two species has been found in nature and referred to as “invasive”, because it invaded new habitats that are not occupied by its parental species (Nolte *et al.* 2005a). This is due to a raised fitness of invasive *Cottus* in large river habitats and can be best explained through the action of natural selection (Nolte *et al.* 2006; Stemshorn *et al.* 2011). However, the underlying selective forces as well as their genomic targets are not well understood. One of the most conspicuous features of invasive *Cottus* is the fact that their phenotypes resemble one ancestor (*C. perifretum*) more than the other one (*C. rhenanus*) (Nolte *et al.* 2005a). This includes the intensity of skin prickling, body shape, number of vertebrae and life history traits like allocation of energy into growth versus early maturation (Nolte *et al.* 2005a; Nolte and sheet 2005). The similarity to one parental species seems paradoxical given that the invasive

Cottus comprise a thoroughly admixed gene pool (Stemshorn *et al.* 2011) and the commonly believed notion that hybrids are typically intermediate between their parental taxa (Wilson *et al.* 1992; Estabrook 1996). Therefore, it is interesting to investigate the genomic basis of the invasive *Cottus* phenotypes which probably originated from only one of the parental species. With the development of the genomic methods, it becomes more and more feasible to identify genotype-phenotype associations particularly in a hybrid lineage. Ancestral genetic variation in an admixed population can be used to identify candidate loci associated with specific phenotypes. For example, male nuptial color and body shape in a hybrid population of stickleback were genetically studied and shown to be associated with the same region that affects plate number (Malek *et al.* 2012). Therefore, identifying the genetic basis of invasive *Cottus* phenotypes by investigating ancestral allelic frequency distributions permits the important inference about the evolutionary processes that have shaped the hybrid lineage.

In this study, we employed the possibility to transfer genomic information between *Cottus* and stickleback (Cheng *et al.* 2013) to investigate the genetic basis of phenotype variation in *Cottus*. QTL analysis from independent interspecific F₂ crosses between pure *C. perflretum* and *C. rhenanus* was performed to identify genomic regions that affect skin prickling and body shape. This was complemented with analyses of targeted candidate genes known from other organisms. Genes of the EDA signaling pathway interact in skin appendage development in vertebrates (Knecht *et al.* 2007; Pantalacci *et al.* 2008). *Cottus* homologues of these genes, including Ectodysplasin (Eda), EDA Receptor (Edar), the EDAR-Associated Death Domain adaptor (Edaradd), Tumor Necrosis Factor Receptor (TNFR) superfamily member19 (Troy), its adaptor TNFR-Associated Factor 6 (Traf6) and the downstream regulator nuclear factor kappa B Essential modulator (Nemo) were identified from next generation sequencing data. The map position of each candidate gene was determined in *Cottus* for comparison with the QTL mapping results. Further, detailed sequence analysis of the regions of each gene can suggest whether these genes have a genomic architecture that may contribute to different evolutionary utilization and be particularly amenable to producing modular regulatory changes in natural populations. Furthermore, we scanned ancestral allele frequencies in the candidate regions of the invasive gene pool to test whether or not these regions show extreme ancestral preference and contain outlier loci that might be subject to selection. The results suggest that a *Cottus* homologue of an adaptive trait in stickleback (Schluter *et al.* 2010) may also be involved to generate conspicuous evolutionary divergence in *Cottus*.

Materials and Methods

Cottus crosses and phenotype analysis

The focus of this study is on two Central European *Cottus* species, namely *C. perifretum* and *C. rhenanus* and the natural hybrid lineage (invasive *Cottus*) that has emerged from them. Breeding stocks of different populations of *C. rhenanus* were sampled from small streams of Broel (GIS: 50°50'N 7°22'E; between Bröl and Winterscheidt, Germany) and Naaf (GIS: 50°53'N 7°15'E; east of Wahlscheidt, Germany) which are part of the German River Rhine basin; while different populations of *C. perifretum* were sampled from small tributaries of Witte Nete (WN) (GIS: 53°14'N 5°04'E; Flanders, Belgium) and Laarse Beek (LB) (GIS: 51°17'N 5°04'E; Brasschaat, Antwerp, Flanders, Belgium) which are part of the River Scheldt drainage in Belgium. Populations from each species display significant differentiation with an F_{st} of 0.3 (Nolte *et al.* 2006) between population Broel and Naaf and 0.368 (Knapen *et al.* 2003) between population Witte Nete and Laarse Beek. Previous gene expression analysis in *Cottus*, using the same populations described here, also underlined divergence within and between species (Czypionka *et al.* 2012). Invasive *Cottus* were sampled from the lower River Sieg (GIS: 50°48'N 7°9'E; west of Siegburg, Germany). Two independent groups of F_2 crosses were obtained from each population of pure *C. perifretum* and *C. rhenanus* and we performed alternative cross directions with respect to sex of the parents for each pair of populations (Table 1). All fishes from the F_2 crosses

Table 1. Interspecific F_2 crosses between different populations of *C. rhenanus* and *C. perifretum* that were used for QTL mapping. The cross type indicates the origin of the parental populations (Broel, Naaf, LB, WN) and the gender of the parental generation. Individual mapping families are listed together with the total number of offspring that were raised and used for QTL mapping of skin prickling and body shape.

Cross Type	Mapping family	Total No.	Prickling	Body Shape
Broel(♂) × WN(♀)	ZG11	103	97	97
	ZG18	74	74	74
	ZG33	52	52	0
	ZG42	97	97	0
WN(♂) × Broel(♀)	ZG40	69	69	56
	ZG46	70	64	47
LB(♂) × Naaf(♀)	ZG26	108	12	95
	ZG32	96	96	85
	ZG38	51	46	46
Naaf(♂) × LB(♀)	ZG39	55	41	41

(775 specimens) as well as additional 132 individuals representing the pure ancestral species (*C. rhenanus*: Broel 30, Naaf 30; *C. perifretum*: LB 40, WN 32) were used to score phenotypic traits. 96 invasive *Cottus* specimens from lower River Sieg were included only to analyze prickling but not body shape.

(a) Prickling. Sculpins vary in the degree to which spine like scales (prickling) cover the lateral sides of their body. The skin prickling phenotype was quantified as described by Nolte *et al.* (2005a). Prickling categories were chosen to best capture patterns of variation that we have encountered in different species of *Cottus* in extensive surveys of this phenotype among populations in Europe (Freyhof *et al.* 2005). The categories used included 0 = absence of any prickling; 1 = less than 10 prickles present beneath pectoral fin; 2 = more than 10 prickles but all covered by the pectoral fin; 3 = prickling extends beyond pectoral fin but ends anterior to the middle of the second dorsal fin; and 4 = prickling extends back beyond the middle of the second dorsal fin (Figure 1).

(b) Body shape. We have previously documented significant differentiation in body shape among invasive *Cottus* and their parental species. From a lateral view, most of the variation in body shape was found to occur in the length and height of the abdomen and the head but also the length of the caudal peduncle (Nolte *et al.* 2005a; Nolte and Sheets 2005). In this study, we chose to analyze a two-dimensional body shape from a dorsal view as the known length differences in shape also manifest from a dorsal perspective. This approach permits efficient and rapid image acquisition from live fishes as these are dorsal-ventrally compressed and tend to sit still on the bottom of a container. The fish were mildly anesthetized with CO₂ and placed into shallow dishes filled with water with a millimeter graduated paper included as a scale (Figure 2a). A digital picture was taken from each fish of the F₂ mapping families and pure ancestral species (*C. rhenanus*: Broel 30, Naaf 30; *C. perifretum*: LB 40, WN 32) before the individuals were sacrificed and preserved. A set of 9 anatomical landmarks was chosen to capture the shape from a dorsal view. As seen in figure 2a, the 9 landmarks highlight important differences in shape with a particular emphasis on the proportions of the head (tip of mouth - #1, posterior ends of mouth - #4 and #7, eyes - #5 and #8) and three landmarks that cover the abdomen (origins of the first dorsal fins - #6 and #9, the pectoral fin - #2 and the caudal fin - #3). The data was analyzed as described in Nolte and Sheets (2005) using the software packages TPS (Rohlf 2003) and Imp (Sheets 2002). Canonical variate analysis (CVA) was used in order to identify those shape vectors that best discriminate the ancestral species *C. perifretum* and *C. rhenanus* without the data from the F₂ mapping families. The corresponding function was then used in a second step to calculate the axis scores of the F₂ individuals of the mapping families. F₂ individual CV axis scores were then used as variables in the QTL mapping analysis. Although the CV axis scores can be defined as an independent trait for

mapping, they do not correspond to independent dimensions of genetic variation and might cause problems when two or more shape variations that together loaded a given CV were found to map to non-overlapping QTLs when tested separately (Albert *et al.* 2007). In order to disentangle individual shape coordinates variation between specific landmarks, traditional morphometric variables (length measurements) from different geometric landmarks were also collected to represent head length (HL, average of the two measurements between tip of mouth and dorsal fin, $[(4-6) + (7-9)]/2$), eye width (EW, distance between the eyes, landmark 5-8) and body length without head (BL, measurement between the pectoral fin and caudal fin, landmark 2-3) as these show significant differentiation between different species (Figure 2c). The scores of HL/EW, HL/BL, BL/(HL \times EW) were calculated and used for QTL analysis separately.

Cottus EDA signaling pathway genes isolation

Genomic sequences of *Cottus* EDA signaling pathway genes (Eda, Edar, Edaradd, Troy, Traf6 and Nemo) were obtained from Illumina sequencing of genomic DNA from 6 *Cottus* individuals that represent separate natural populations of the two parental species. 3 individuals of *C. rhenanus* from population Broel, Naaf (see above) and Fokkenbach (GIS: 50°32'N 7°25'E; east of Niederbreitbach, Germany) and 3 individuals of *C. perifretum* from population Witte Nete, Laarse Beek (see above) and Zwanebeek (GIS: 51°14'N 4°34'E; Antwerp, Flanders, Belgium) were sequenced to a read depth of 10 - 11 \times using an Illumina GAII sequencer. Raw paired end reads (maximal 150 bp) of each individual were quality trimmed using the popoolation data analysis pipeline (Kofler *et al.* 2011) and assembled independently by the program SOAPdenovo (31mer version: 1.05) (Li *et al.* 2010; soap.genomics.org.cn/soapdenovo.html), with a *kmer* size set to 29, into scaffolds (Nolte *et al.* unpublished data). As the stickleback genome serves as an excellent reference for *Cottus* (Cheng *et al.* 2013), genomic DNA and cDNA sequences of the stickleback EDA pathway genes were downloaded from Ensemble Genome Browser (<http://www.ensembl.org/index.html>) and blasted against each of the 6 *Cottus* genomic DNA assemblies to identify scaffolds that contain *Cottus* EDA pathway gene homologs. Blast searching was used for distant homologies with parameters of bit score > 80 , E value $< 1e-5$ and an alignment length > 100 bp as thresholds to accept a hit. Scaffolds assembled from paired end reads contained stretches of "N" where paired reads were not completely sequenced. This can cause assembly problems. As a result, scaffolds from single individual often did not cover the full length of each gene, whereby successfully assembled fragments differed between individuals. Then such regions were replaced by only limited number of "N" and scaffolds covered the EDA pathway genes were combined from all individuals and co-assembled using the CodonCodeAligner software

(<http://www.codoncode.com/aligner/>) to get the full length of each EDA pathway gene. cDNA sequences for *Cottus* EDA pathway genes were identified by blast searching each gene of stickleback cDNA sequence against EST sequences assembly of one *C. rhenanus* individual (Population Fokkenbach, see above) (data unpublished).

Mapping of EDA pathway genes and QTL analysis

We previously generated a *Cottus* genetic map using independent F₂ mapping families between different populations of the two *Cottus* species which showed highly conserved synteny with stickleback genome (Cheng *et al.* 2013). Here, we used the same F₂ mapping families as described by Cheng *et al.* (2013, Table 1). We analyzed the available mapping families for the presence of QTL that affect prickling and body shape with R/qtl. This analysis is complemented with an analysis of candidate genes for which we developed markers that permitted genetic mapping to confirm the map position as inferred from the stickleback genome.

Diagnostic Length polymorphic indel markers that can distinguish the two parental species were developed from introns or UTRs of *Eda*, *Edar*, *Edarrad*, *Troy*, and *Nemo* by looking for variable length insertion from the gene alignments of the 6 individual assemblies described above (Table S1). Diagnostic indel markers for the 5 genes (without *Traf6*) were genotyped for all F₂ mapping families and genotyping was performed and analyzed by separating PCR products generated from fluorescently labeled primers on a capillary sequencer to determine the fragment length. Peak sizes for each marker were analyzed using GeneMapper software (Applied Biosystems). The actual map positions of these length polymorphism markers were confirmed through patterns of segregation in our F₂ mapping families (see below). We did not find length polymorphism in gene *Traf6* assembly hence we cannot localize the position of this gene through genetic mapping. Therefore, the position of *Traf6* was inferred by blast searching the gene sequence against the stickleback genome to infer a likely map position, thus exploiting the highly conserved synteny between these species (Cheng *et al.* 2013).

Genotype data of 298 SNPs for all F₂ mapping families was obtained after filtering and prepared as described by Cheng *et al.* (2013). We combined genotype data of all mapping families from crosses between Naaf × LB and Broel × WN, and have carefully checked whether a given locus was informative for individual mapping families before combining the data. Genotypes for loci which were not fully informative (i.e. where grandparents were not fully differentially fixed and either parent from the F₁ generation was not heterozygous) were marked as missing data for QTL mapping. The previously generated genetic map (Cheng *et*

al. 2013) including mapped indel markers representing genes of the EDA pathway was used for QTL mapping. In order to guard against artifacts caused by combining mapping families and to obtain independent estimates for QTL intervals and LOD scores, we reconstructed LG 3 which contains the major QTL for prickling (see results) for crosses Broel × WN and Naaf × LB separately. This was complemented with 3 additional diagnostic indel markers (Table S1) located in the QTL interval and developed as the same way as described above for indel markers of EDA pathway genes. The phenotypes of prickling and body shape were mapped independently for crosses combined according to population of origin (Broel × WN and Naaf × LB). The QTL analysis was performed with package R/qtl (Broman *et al.* 2003, 2009). The multiple imputation method was used for the genome scan. Likelihood of odds (LOD) significance thresholds for each trait was determined by permutation testing (1000 permutations). Significant QTL was above the genome-wide LOD threshold ($\alpha=0.05$). The confidence interval was examined by 95% Bayes interval flanking the QTL peak.

Cottus EDA signaling pathway gene structures analysis

Multiple factors could contribute to whether a particular locus is selected in natural populations, as for example of copy number variation, mutational target size and types of mutations (coding or regulatory) (Knecht *et al.* 2007). Here we ask whether different genes in the EDA signaling pathway from *Cottus* show significant differences in genomic structure that might contribute to differential evolutionary utilization. Precise locations of exons of the EDA pathway genes were identified by blasting *Cottus* cDNA sequence against genomic sequences of each gene. The gene size (exons + introns + UTRs), transcript size (exons + UTRs), protein size (exons) were determined for each of the EDA pathway components. In order to estimate the number of evolutionary conserved regions (ECR) which might correspond to regulatory sequences in these genes, homologous regions between stickleback and *Cottus* EDA pathway genes were analyzed in pairwise comparisons of the genomic sequences of each gene from the two species by mVISTA (Frazer *et al.* 2004; <http://genome.lbl.gov/vista/mvista/submit.shtml>) using the LAGAN alignment program (Brudno *et al.* 2003) with translated anchoring and a probability of 0.5. Sequence homology displays used the following parameters: 50bp calculation window, 50bp minimum conserved width, and 70% conserved identity. ECRs were defined as regions of at least 50 bp having at least 70% identity. Multiple alignments of Edar protein sequence with other model organisms were produced by using Seqman (<http://www.dnastar.com/t-sub-products-lasergene-seqmanpro.aspx>) combined with Boxshade (http://www.ch.embnet.org/software/BOX_form.html) to compare the amino acid identity and conservation between two *Cottus* species and also with other model organisms.

Genomic ancestry in the hybrid lineage

Since the phenotype of invasive *Cottus* resembles that of *C. perifretum*, we test whether EDA signaling pathway genes in the invasive gene pool show an excess of alleles that trace back to *C. perifretum*. For this purpose, a DNA pool containing equal concentration of 48 invasive individuals from River Sieg was sequenced to a read depth of 22 × with an Illumina GAII sequencer to infer allele frequencies of ancestry informative loci in the invasive *Cottus* lineage (Nolte *et al.* unpublished data). The genetic map we have generated to date (Cheng *et al.* 2013) is not detailed enough to place the SNPs as obtained from massively parallel sequencing experiments. However, we have shown that the *Cottus* genetic map and the stickleback genome have a well conserved synteny and the marker sequence along chromosomes is collinear in a large proportion of the two genomes (Cheng *et al.* 2013). Accordingly, blast searches permit a tentative assignment of map positions for *Cottus* genomic DNA based on the available stickleback genome. Hence, inferred linkage between directly neighboring loci based on the stickleback genome is more likely to be true for *Cottus* than linkage across longer map distances because the two genomes differ in some intra chromosomal rearrangements. All genomic DNA scaffolds assembled from the Naaf individual (*C. rhenanus*) describe above were blast mapped against the stickleback genome (version BROADS1.56) with blast parameters set for distant homologies (parameters: -e 0.0000000001 -m 9 -F T -G 2 -E 2 -r 2 -q -3 -W 9). This excluded sequences shorter than 500 bp. Since many genomic scaffolds exceeded the length that is useful for blast searches, we have cut them into 500mers of which a maximum of 25 were randomly chosen and blast searched against the stickleback genome. Significant blast hits obtained for all 500mers from a given scaffold were evaluated for the presence of unambiguous hits that suggest homology with a genomic region in the reference genome and the orientation of the scaffold with respect to the reference genome was determined. In a second step, all raw *Cottus* reads of 6 sequenced individuals (see above) were mapped against the *Cottus* scaffolds with bwa (Version: 0.6.1-r104) (Li and Durbin 2009) using default settings to map single reads. SNP calling was done using the mpileup function of SAMtools (version 0.1.18) (Li *et al.* 2009) which outputs nucleotide states for all bases and groups defined in the analysis. A custom perl script was used to filter this data for candidate ancestry informative loci. When all six ancestral individuals (3 × *C. rhenanus* and 3 × *C. perifretum*) were represented for a given base in the assembly and a fully diagnostic SNP distinguishes the six samples at the species level, this locus was assumed to be ancestry informative with the limitation that the sample size of the parental species is low but comprises deeply divergent subpopulations. The list of diagnostic SNPs obtained in the allele calling step was reduced to those for which a map position was identified in the stickleback genome. The position of the SNP was approximated relative to the center of the blast hit in the stickleback genome. The

direction and absolute value of the offset were based on the orientation of the scaffold relative to the reference genome. This procedure only approximates the map positions for individual SNP loci but orders of SNP loci correctly relative to one another. SNP allele frequencies for ancestry informative loci were determined for assembled reads of the invasive *Cottus* DNA pool and the ratio of *C. perifretum* over *C. rhenanus* alleles was used as an estimate of the ancestry proportions at a given locus. When SNP loci covered less than 15 reads from the invasive gene pool they were removed from the analysis. Here we used a threshold to detect extreme ancestries for each chromosome based on an earlier estimation that 2.5% of the admixed invasive genome might be under selection (Stemshorn *et al.* 2011). The top 1.25% and bottom 1.25% of the loci with ancestry excess in favor of *C. perifretum* and *C. rhenanus* were considered as outliers. The regions used for outlier loci detection include the EDA pathway genes and 1Mb of the up- and down- stream region unless a candidate was too close to the end of a chromosome.

Results

Variation in *Cottus* morphology patterns

132 individuals representing the pure ancestral species (*C. rhenanus*: Broel 30, Naaf 30; *C. perifretum*: LB 40, WN 32) together with all F₂ mapping families were studied to determine the extent to which skin prickling covers the body and for variations in body shape. As shown in figure 1, a particularly prominent diagnostic difference between the two parental species is the occurrence of skin prickling, which is strong in *C. perifretum* and virtually absent in *C. rhenanus*. The invasive *Cottus* show skin prickling comparable to or even stronger than what is seen in pure *C. perifretum* (Figure 1). Another significant differentiation among the two species can be observed in a geometric morphometric analysis. Both populations of each species form distinct clusters in the canonical variate analysis (CVA), which display the greatest separation of the species (Figure 2b). The differentiation in shape as captured by CV axes can be visualized as displacement vector for different measurements between landmarks. As seen in figure 2c, *C. rhenanus* have a smaller value of Head Length/Body Length than *C. perifretum* which suggests *C. perifretum* have a longer head and a shorter body compared to *C. rhenanus*; *C. rhenanus* have a higher value of Head length/Eye width than *C. perifretum* which suggests that *C. perifretum* have a wider head compared to *C. rhenanus*; Finally, *C. rhenanus* have a higher value of Body Length/(Head Length × Eye Width) than *C. perifretum* which suggests *C. perifretum* have a bigger head compared to *C. rhenanus*. Generally, *C. perifretum* differs from *C. rhenanus* by having a shorter body and a larger head (Figure 2c).

The above results agree with what was previously described for skin prickling and body shape in the species studied here (Nolte *et al.* 2005a; Nolte and Sheet 2005).

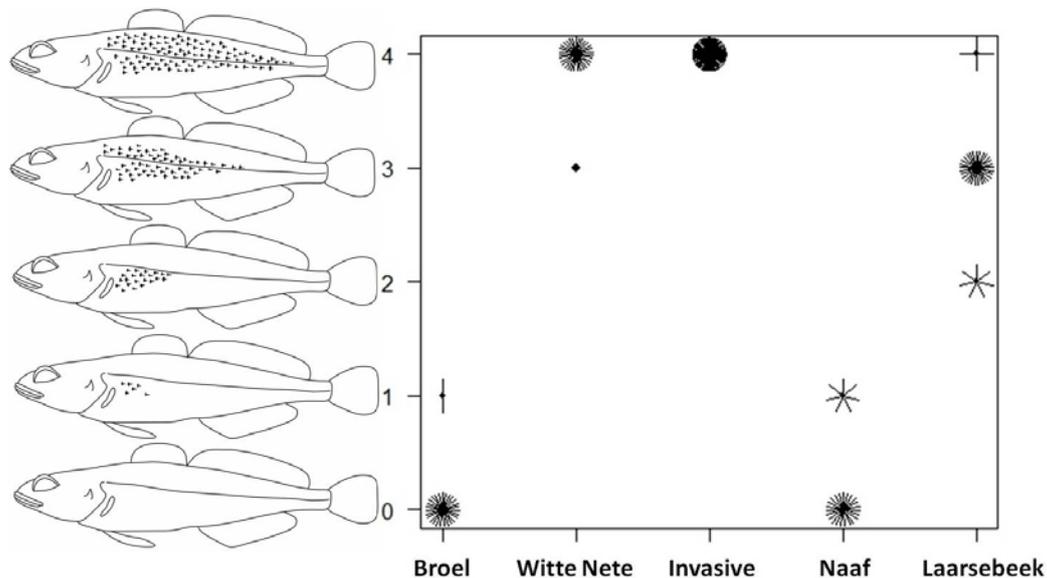


Figure 1. Prickling analysis of sculpins. The degree to which prickling covers the lateral sides of the body in *Cottus* differs between species. The number of leaves from the sunflower graph indicates the number of individuals falls into each category from different parental populations and invasive *Cottus*. The skin prickling categories were quantified for all F_2 mapping families and used for QTL mapping.

Identification and mapping of *Cottus* EDA pathway genes

Genomic sequences of *Cottus* EDA pathway genes were identified from assemblies of 6 sequenced *Cottus* individuals. cDNA sequences were identified from assembly of sequenced ESTs from one individual of *C. rhenanus* (Fokkenbach). Blast searches indicate that each of the isolated genes correspond to a single *Cottus* homolog of Eda, Edar, Edaradd, Nemo, Troy and Traf6. The genetic locations of EDA pathway genes were mapped with all F_2 mapping families by identifying informative length polymorphisms indel markers in introns or UTRs of each EDA pathway gene (Table S1). All *Cottus* EDA pathway components except for Traf6 were genetically mapped to locations of *Cottus* linkage groups that correspond to the locations of the stickleback homologs thus confirming a highly conserved synteny between these two species (Figure 3). While we did not find useful genetic markers do directly map *Cottus* Traf6, this gene is presumably localized on *Cottus* linkage group 5 based on transferring positional information from the stickleback genome. Note that the synteny with stickleback is significantly conserved (perfectly collinear) in the genomic region that carries the Traf6 gene (Cheng *et al.* 2013; Figure 3).

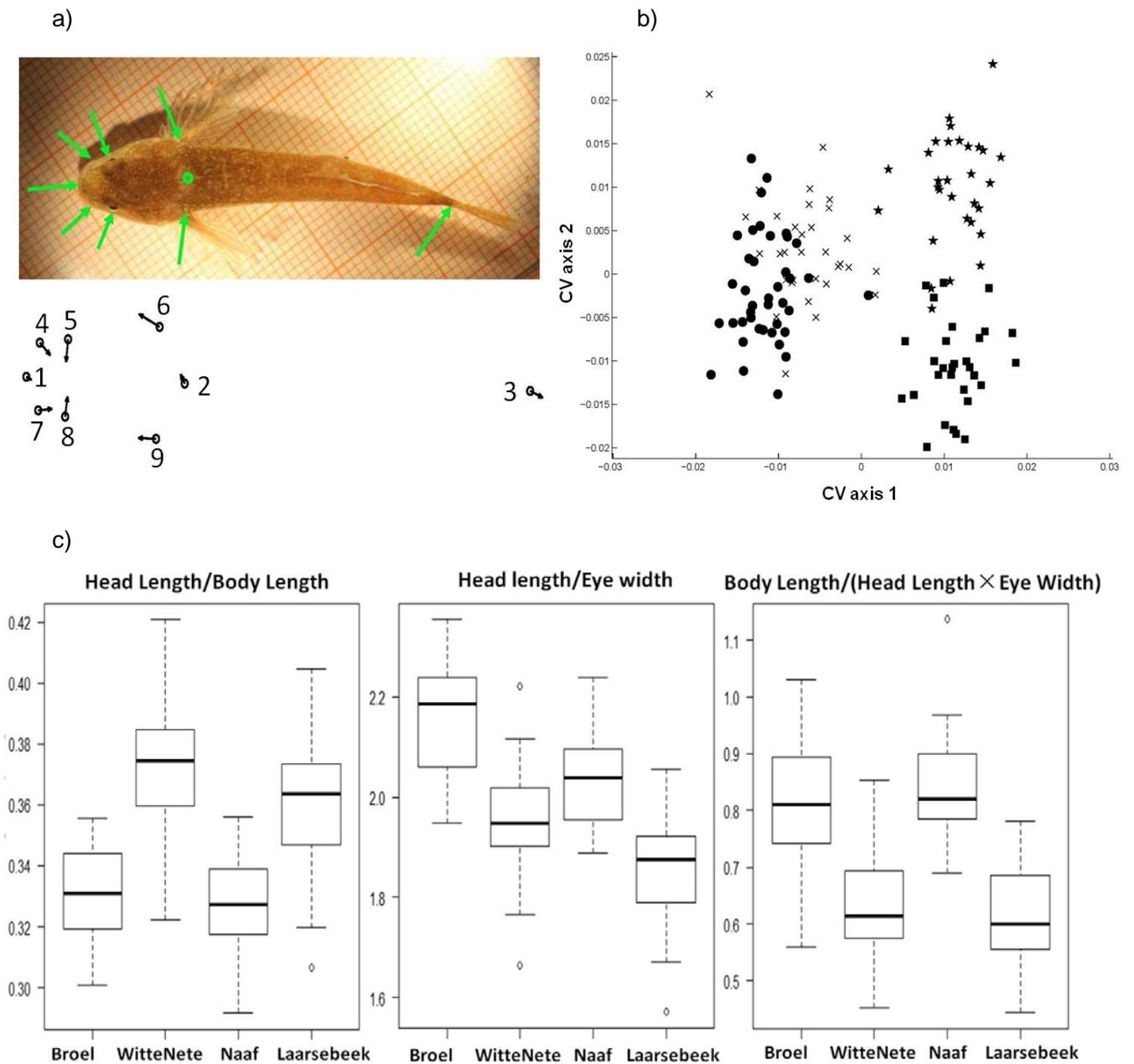


Figure 2. Differentiation of sculpins in body shape. (a) A set of 9 anatomical landmarks was chosen to capture the important shape differences between *Cottus* species from a dorsal view. (b) Each population of the two *Cottus* parental species comprises a distinct cluster that separates along two CVA axes (*C. rhenanus*: Broel – crosses, Naaf - black circles; *C. perifretum*: LB - stars, WN – black squares). The two parental species separate along the first CV axis. (c) Differentiation in body shape with different measurements of traditional length morphometric variables from different landmarks. All of these phenotypes segregate in F_2 hybrids and can be used for QTL mapping.

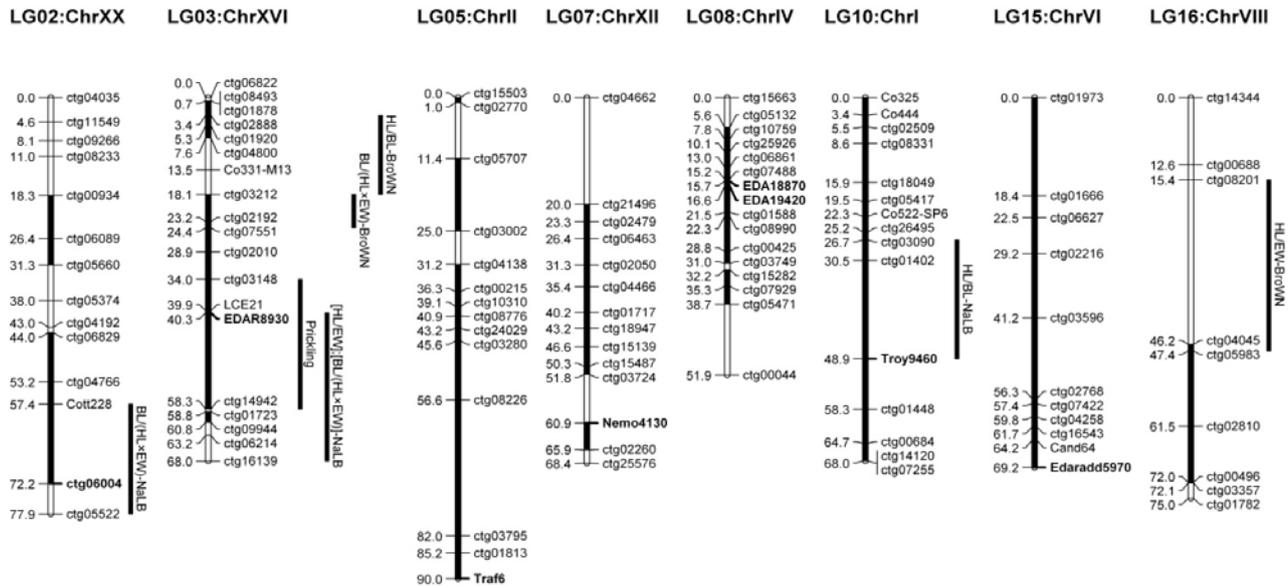


Figure 3. Linkage map positions of *Cottus* ectodysplasin (EDA) signaling pathway genes which are highlighted through bold marker names and significant QTL intervals for prickling and body shape which are shown with black bars beside *Cottus* linkage groups (LG) represent collinear regions with a perfectly conserved marker order with stickleback chromosomes (Chr) as described by Cheng et al. (2013).

QTL analysis

We used the *Cottus* genetic map as a reference and the multiple imputation method from R/qtl to scan the genome for the presence of loci that affect prickling and body shape. A highly significant QTL for prickling was detected on linkage group (LG) 3 (LOD=9.49/21.36 for Broel x WN and Naaf x LB crosses separately, Table 2, Figure S1). Both crosses have the highest QTL peak around the locus EDAR8930 which is located at the second intron of the Edar gene. For the genotypes at EDAR8930, the regression of prickling phenotype on genotype shows strong association in Naaf x LB crosses but less strong in Broel x WN crosses (Figure S2). Unlike for the Broel x WN crosses, a second significant QTL was detected at the end of LG 3 for Naaf x LB crosses (ctg1639, LOD=22.43) which was linked to the first QTL detected. The interval flanking the QTL has a length of 13.8 cM and 48.6 cM for Broel x WN and Naaf x LB respectively. This conspicuous difference is associated with signals for different marker orders that manifest only when the Edar containing LG3 is reconstructed independently for each family (see below). We observed congruent but non-significant trends for separate families that supported the presence of a QTL in that region. Therefore we used the combined dataset and a consensus genetic map for trait mapping because separate mapping families were too small to detect significant QTL

signals. According to our analysis, none of the other EDA pathway genes had a significant effect on prickling.

Table 2. QTL mapping results for prickling and body shape in interspecific *Cottus* F₂ crosses according to the origin and gender of the parental populations. All QTLs that exceed the significance threshold (LOD) with permutation test ($\alpha < 0.05$) are shown with their respective Linkage Group (LG), map position (cM), Maximum LOD score, and the most closely linked markers.

Trait	Broel × WN				Naaf × LB			
	LG	cM	LOD	Marker	LG	cM	LOD	Marker
Prickling	3	70	9.49	EDAR	3	83	22.43	ctg16139
					3	36	21.36	EDAR
Body CVA	NA	NA	NA	NA	3	84.5	17.11	ctg16139
					3	36	8.07	EDAR
HL/EW	16	46.2	3.769	ctg04045	3	84.5	9.77	ctg16139
					3	36	4.38	EDAR
HL/BL	3	15.6	4.856	ctg01920	10	33	3.605	ctg01402
	3	36.2	4.71	ctg02192	3	84.5	10.745	ctg16139
BL/(HL×EW)					3	36	5.76	EDAR
					2	5	3.777	ctg06004

We identified significant QTLs for different measurements of body shape differences (Table 2, Figure S3). For CVA scores, no significant QTL was detected for Broel × WN crosses, but two major QTLs were mapped on LG 3 and linked to EDAR8930 (LOD=8.07) and ctg16139 (LOD=17.11) in Naaf × LB crosses which cover the QTL regions of skin prickling. For the ratio between head length and eye width, one QTL was detected on LG 16 (ctg04045, LOD=3.769) in Broel × WN crosses and the same two QTL were also mapped on LG 3 and linked to EDAR8930 (LOD=4.38) and ctg16139 (LOD=9.77) in Naaf × LB crosses. For the ratio between head length and body length, one QTL was detected on LG 3 (ctg01920, LOD=4.856) in Broel × WN crosses and one QTL was mapped on LG 10 (ctg01402, LOD=3.605) in Naaf × LB crosses. For the ratio between body length and head size (head size = head length × eye width), one QTL was mapped on LG 3 (ctg02192, LOD=4.71) in Broel × WN crosses. In Naaf × LB crosses, on LG 3, two QTLs were detected at EDAR8930 (LOD=5.76) and ctg16139 (LOD=10.745). Interestingly, another minor QTL was detected in Naaf × LB cross on LG 2 and located at one sex determination region (ctg06004, LOD=3.777) previously reported (Cheng *et al.* 2013). The QTL on LG 3 affected prickling as well as body shape indicating pleiotropy or tight linkage. None of the other EDA signaling pathway genes shows significant association with body shape.

Fine mapping of the *Cottus* Edar region

The significant QTL interval containing the Edar gene on LG 3 turned out to be the most interesting candidate in this study as it had the most consistent effect on prickling and body shape. Our previous analysis suggested for this region that the synteny is significantly conserved with stickleback chromosome 16 (Cheng *et al.* 2013; Figure 3), but there is a large gap between the Edar locus EDAR8930 and its neighboring locus ctg14942 (18 cM of *Cottus* map to 2 Mb of stickleback genome) comparing to that between EDAR8930 and ctg03148 on the other side (6.3 cM of *Cottus* map to 1.5 Mb of stickleback genome) (Figure 3). This gap might be caused by conflicting marker orders among different mapping families when combined them together. In order to investigate this region in more detail, we developed three diagnostic indel markers from random genomic DNA fragments (Illumina sequenced genomic DNA) as described for EDA pathway genes (Table S1). According to the stickleback genome, ctg00996i is located between EDAR8930 and ctg03148; ctg04162i and ctg02117i are located between EDAR8930 and ctg14942. The whole region between ctg03148 and ctg14942 covers about 3.5 Mb with average intervals of 0.7 Mb between each neighboring markers according to the stickleback genome. The new genotypes were combined with our existing data for all F₂ mapping families and the mapping procedure was repeated as described by Cheng *et al.* (2013). The three indel markers were located on LG3 as expected (Table S2).

The marker order for Edar region was reconstructed for each of the 8 mapping families as defined by their population of origin and grand paternal sex. As shown in table S2, the 7 markers in the Edar region were assigned to three blocks. LCE21 and EDAR8930 were tightly linked and assigned to block A (blue); ctg03148, ctg04162i and ctg02010 were assigned to block B (green) and shown suppressed recombination in LB (♂) × Naaf (♀) crosses (ZG26, ZG32, ZG38) but not in other crosses; ctg00996i and ctg02117i were assigned to block C (yellow) except for the crosses in WN (♂) × Broel (♀) (ZG40, ZG46) and one cross in LB (♂) × Naaf (♀) (ZG38) where the two markers were separated on each side of the Edar block A. The three blocks showed specific order patterns according to different origins of population and sex of their parents. In Broel (♂) × WN (♀) crosses, the block pattern was B-C-A; in Naaf (♂) × LB (♀) cross, the block pattern was C-B-A; in WN (♂) × Broel (♀) crosses, it showed pattern B-C-A-C with the two markers from block C separated by the Edar block A; and in LB (♂) × Naaf (♀) crosses, it had pattern B-A-C with one exception from ZG38 which had the same pattern with WN (♂) × Broel (♀) of B-C-A-C. In summary, fine mapping of markers covering the Edar locus suggests marker orders that are highly specific to population and sex origin of the parental generation.

Genomic architecture of EDA pathway genes

Comparison of numbers of introns and exons, gene size (exons + introns + UTRs), transcript sizes (exons + UTRs) and protein sizes (exons) of different EDA signaling pathway components (Table 3) showed that Nemo and Edar have a higher number of exons (14 and 11 separately) comparing to other 4 genes which have less than 10 exons (least for Troy with only 5 exons). Traf6 is the smallest gene spanning about 3 kb and Edar is the largest gene spanning almost 20 kb, the other 4 genes are intermediate and spanning about 10-15 kb. The Edar locus spans the largest genomic region and also encodes a relatively large transcript (3815 bp, only smaller than Nemo, 4055 bp) and large protein (1545 bp for coding exons, only smaller than Nemo and Traf6) (Table 3). The total number of noncoding evolutionarily conserved regions (ECRs) in the intronic regions (Table 3) is considerably higher in Edar (33) than for any of the other genes. Eda, Edaradd and Troy have the next highest numbers of ECRs (20, 20 and 22), with Traf6 and Nemo having relatively few ECRs (6 and 8). Generally, the Edar locus has a larger target size and more complex regulatory apparatus than other components of the EDA signaling pathway. Comparison of the protein sequences of Edar gene showed high level of *Cottus* sequence identity with human, mouse and fish homologs and three mutations lead to amino acid changes between *C. rhenanus* (non-prickled) and *C. perifretum* (prickled) (Figure 4).

Table 3. Summary of genomic structure of *Cottus* EDA signaling pathway genes. Transcript size refers to the sum of exons and UTRs; Total Coding refers to the sum of exons; Gene Size refers to the sum of exons, introns and UTRs. UTR, untranslated region; ECR, evolutionary conserved region.

Eda Pathway Gene	Transcript Size (bp)	Total Coding (bp)	No. Exons	Gene Size (bp)	Noncoding ECRs (vs. Stickleback)
Eda	2062	1107	8	10412	20
Edar	3815	1545	11	19564	33
Edaradd	2428	693	6	15038	20
Troy	609	609	5	11814	22
Traf6	3664	1731	8	3360	6
Nemo	4055	1695	14	10195	8

Genomic ancestry of the admixed invasive gene pool

The invasive ancestral allele frequency analysis revealed different patterns of ancestry for different components of the EDA pathway genes (Figure 5; Figure S4). On average, the genomic regions within the candidate genes Eda, Edar, Edaradd, and Nemo showed excess of *C. perifretum* ancestry (84-92.6%, 85-96%, 75-85.3% and 91.2-97.9% respectively). In contrast, the regions within Troy and Traf6 appeared to be well mixed on average (55-64.3%, 34.6-39.3%). The Edar

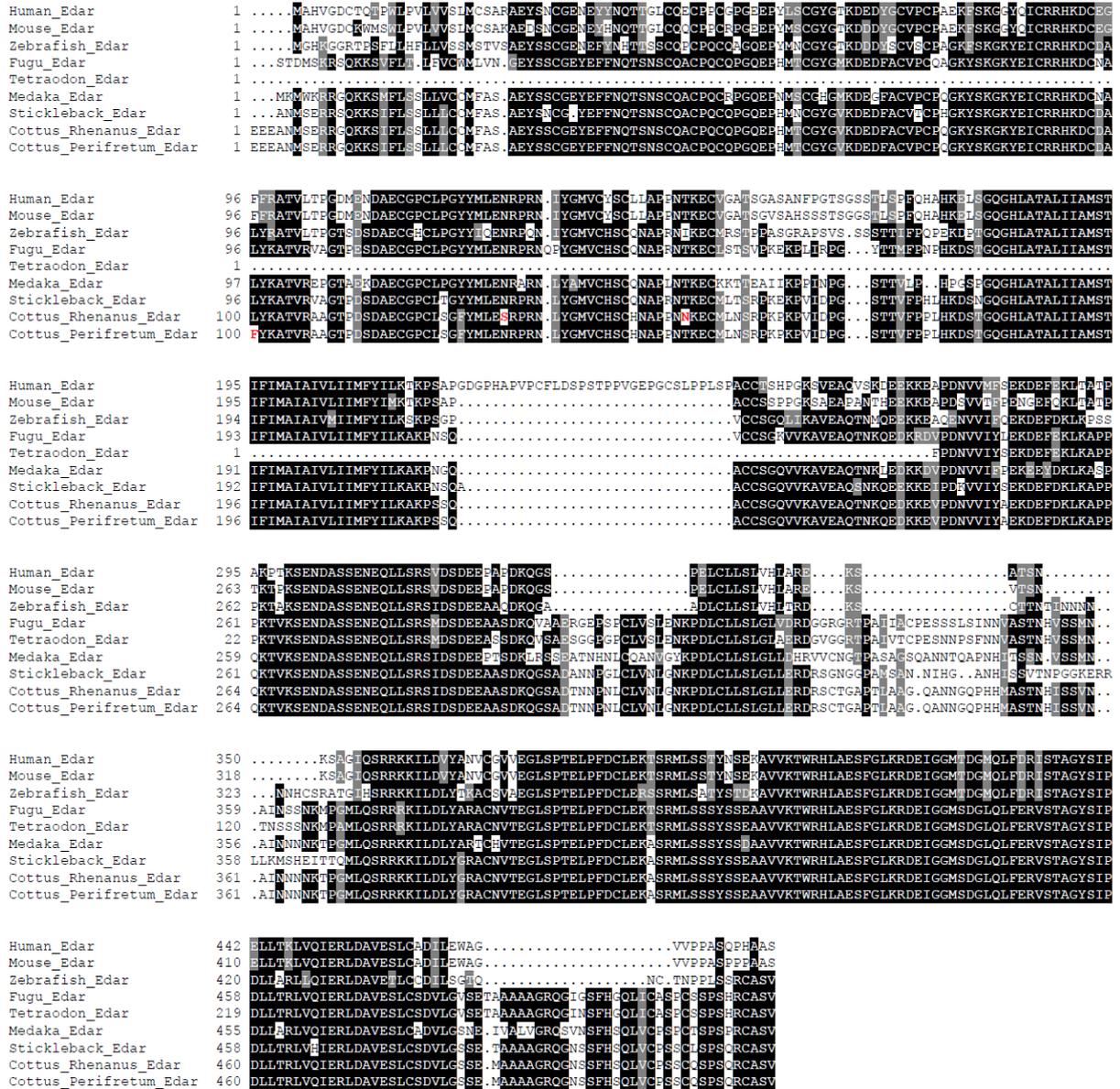


Figure 4. Alignment of Ectodysplasin receptor (Edar) protein sequences from human, mouse, Zebrafish, Fugu, Tetraodon, Medaka, stickleback, *C. rhenanus* (non-prickled) and *C. perifretum* (prickled). Identical amino acids are shaded black, and conservative substitutions are shaded grey. Three amino acids different between *C. rhenanus* and *C. perifretum* are highlighted in red.

gene was the only one showing ancestry excess for *C. perifretum* when compared to its immediate neighboring region (Figure 5). Allele frequencies of the other 5 genes were not distinguishable from the background when compared to their neighboring regions (Figure S4). According to the estimation that 2.5% of the admixed invasive *Cottus* genome could be outliers that are required for the maintenance of the hybrid lineage, thresholds of 97-100% and 0 - 4.5% ancestral of *C. perifretum* alleles were inferred for different chromosomes. When applying these thresholds, individual outlier loci can be found in all genomic regions of EDA

Table 4. Outlier loci that have extreme ancestries in the invasive gene pool are detected in *Cottus* EDA signaling pathway genes and their neighboring regions. According to the estimation that 2.5% of the invasive *Cottus* genome could have extreme ancestries that are required for the maintenance of the hybrid lineage, thresholds of 97-100% and 0-4.5% ancestry of *C. perifretum* alleles were inferred for different chromosomes (genes).

EDA Pathway Genes	No. outlier (<i>C.perifretum</i>)	No. outlier (<i>C.rhenanus</i>)
Eda	1	4
Edar	4	1
Edaradd	1	2
Troy	2	2
Traf6	1	4
Nemo	15	0

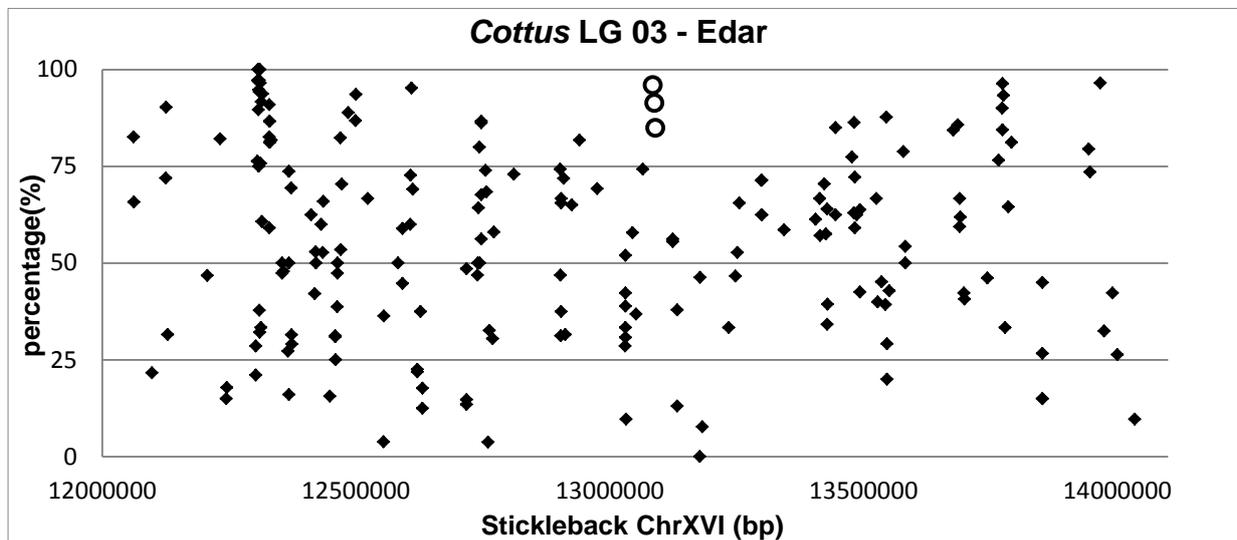


Figure 5. Genomic ancestries in hybrid *Cottus* mapped against the stickleback reference genome for Edar gene (circles) and 1 Mb up-stream and down-stream of the gene (diamonds). The x axis corresponds to the genomic position [bp] of stickleback chromosome 16 and the genomic ancestry (percentage of the genome derived from *C. perifretum*) is plotted on the y axis. Individual points represent genomic ancestries for SNP loci after data filtering. It is possible to detect genomic regions in which groups of loci deviate strongly from the genome wide average ancestry of 60%. The regions with the most extreme ancestries represent candidates that may be subject to genotypic selection and play a role in adaptive evolution and hybrid speciation.

signaling pathway genes (Table 4). In Edar region, 4 outlier loci were detected that are conspicuous 100% of *C. perifretum* ancestry and clustered together in a 3 kb region which is located 774 kb upstream of the Edar gene (Figure 5). The neighboring region of Nemo gene had the highest number of outlier loci detected (15 loci). 7 outliers were clustered in a 248 kb region which contains the Nemo gene, another 8 outliers were clustered in a 126 kb region which is about 0.5 Mb

up-stream of Nemo gene (Figure S4). Other than that, there were only 3 to 5 outlier loci detected for each other 4 genes and all these outlier loci were spread individually in the gene neighboring region but not in the gene itself (Table 4; Figure S4). Here, among the 6 EDA pathway components, Edar and Nemo showed strong ancestral excess for *C. perifretum* within the genes as well as having outlier loci clustered together, but except for Edar, we did not detect any QTL signal on Nemo gene that can explain the phenotypic differences on prickling or body shape.

Discussion

This study exemplifies to what extent candidate genes can be studied for a hitherto unexplored fish species through the transfer of genomic information from model organisms. Gibson used the term “super-model” to describe the three-spined stickleback as an organism conducive to the total integration of the disparate field of ecology, molecular and developmental biology and the evolution of biodiversity (Gibson 2005). Our previous study has revealed that the stickleback genome has a highly conserved synteny and serves as an excellent reference for *Cottus* (Cheng et al. 2013). In this study we followed this path in establishing how genomic resources created for stickleback and other model organisms can be transferred and used to target candidate genes and functional aspects in evolutionary and ecological studies. Here, we discuss the evidence that, despite considerable evolutionary differentiation between *Cottus* and stickleback, phenotypic variations between the two fishes might be affected by the same developmental pathway. Interestingly, the *Cottus* homologue of an adaptive trait in stickleback (Schluter et al. 2010) shows intriguing evolutionary patterns among distantly related species of *Cottus* but also within the invasive lineage of *Cottus* that has recently colonized disturbed habitats in the River Rhine.

A possible role of EDA signaling in sculpins prickling phenotype

The EDA signaling pathway was shown to critically affect the development of scales in *Oryzias* and *Danio* (Harris et al. 2008; Kondo et al. 2001). Further, Colosimo et al. (2005) have shown that the Ectodysplasin molecule (Eda) from this pathway is the key locus that affects the lateral plate reduction in sticklebacks. Moreover, in a later study Knecht et al. (2007) have shown that the Eda receptor (Edar) also has a minor effect on the lateral plate phenotype. Skin prickling in sculpins develops much like the dermal armor plates in sticklebacks. This includes their general position on the lateral region of the body, the pattern of reduction observed in some populations which occurs in a posterior to anterior fashion whereby reduced patterns are limited to the anterior part of the body, and finally the fact that both are composed of calcified dermal bone material. This homology

of prickling and lateral plates as scale derivatives already suggests that *Edar* is a key gene in the observed QTL interval for *Cottus* prickling. By screening the assembled genomic scaffolds and an EST collection, we have been able to identify and map *Cottus* homologs of the EDA signaling pathway genes. QTL analysis of skin prickling showed that the region carrying *Edar* has a significant effect on the phenotype between different crosses while there were no such signals associated with other genes from this pathway. While it is still possible that genes are not represented by a linked marker to *Edar* play a role, or that other EDA pathway genes may have small effects on the *Cottus* prickling phenotype, this study provided strong evidence that the QTL carrying a very likely candidate gene indeed plays a similar role in *Cottus* as in other fish species.

Linkage of body shape with prickling phenotype

Our results provide evidence that two seemingly independent phenotypes, like the degree to which modified scales cover the body and body shape, are linked in *Cottus* genome. We cannot determine whether these are caused by a single gene with pleiotropic effects or multiple genes in close linkage from QTL mapping results alone. However, the body shape phenotypes we observed in *Cottus* correspond to what is known from the stickleback that QTL regions affects both head and pelvic shape traits were strongly linked to *Eda* gene, which is the major locus responsible for reduction in the number of lateral plates in sticklebacks (Albert *et al.* 2007; Malek *et al.* 2012). Hence the shape and derived scale phenotypes that mapped to the same genomic region in *Cottus* are also correlated in stickleback. A key difference between *Cottus* and stickleback exists in that the phenotypes in *Cottus* are associated only with *Edar* whereas *Eda* plays a dominant role in stickleback with only a minor role of *Edar* in the latter (Knecht *et al.* 2007). The fact that the effects associated with EDA signaling in stickleback has a different but functionally related genetic basis in *Cottus* suggests that this pathway could be the central one to understand the phenotypic variation in a number of species. Interestingly, one shape QTL was mapped to one sex determination region of one *Cottus* species reflecting sexual dimorphism for shape in Na×LB crosses which agrees with the observation that male and female *Cottus* do show divergence in head shape. A similar divergence in body shape between male and female *Cottus* has also been described in stickleback populations (Jones *et al.* 2006; Albert *et al.* 2007; Reid and Peichel 2010).

Evolutionary patterns at the *Edar* locus

Different loci may represent very different target sizes for mutation, and genes that encode very large proteins or have large numbers of exons could be expected to show higher rate to mutations that alter coding regions or splicing (Knecht *et al.* 2007). Comparison of the numbers of introns and exons, gene sizes,

transcript sizes and protein sizes of different *Cottus* EDA pathway components showed similar results from that of stickleback (Knecht *et al.* 2007) in which Edar homologs from both fishes have larger target sizes than the other genes in this pathway. The key difference between *Cottus* and stickleback is from the number of regulatory elements (ECRs) detected from non-coding regions. Among the 6 EDA pathway genes, Edar has the highest ECRs (33) in *Cottus* whereas Eda has the highest ECRs (57) in stickleback which correspond to the phenotypic effects of different genes for each species. Most mouse and human ectodermal dysplasia mutations are coding region changes that disrupt the function of the gene and produce severe phenotypic defects in multiple tissues (Thesleff and Mikkola 2002; Morlon *et al.* 2005; Knecht *et al.* 2007). In contrast, the alterations of Eda in sticklebacks are regulatory potential as there are no conserved coding changes in any population studied (Colosimo *et al.* 2005). Many conserved noncoding regions correspond to regulatory sequences that control expression at specific times and places during development (Boffelli *et al.* 2004; Woolfe *et al.* 2005). Selection in wild populations may favor regulatory mutations that make it possible to alter function in particular regions of the body, while preserving other functions of a gene that are required for viability and fitness (Carroll 2000; Tautz 2000). A final proof for the function of the Edar locus and whether coding or regulatory change contribute to the variation still requires a detailed genetic analysis in *Cottus*. Still the complexity of this gene relative to other members of the pathway suggests that the Edar locus may have a more complex regulatory apparatus and may be preferentially utilized over the other EDA pathway components to evolve phenotypic changes in natural populations.

We observed different QTL interval sizes for prickling between the two crosses and have found evidence that marker orders may differ at a finer scale among different mapping families. This suggests that chromosomal rearrangements may have occurred in the Edar region among different populations of *Cottus*. Such a situation can also explain the comparably large gap in the consensus genetic map relative to the physical map of the stickleback as an artifact, namely the inflation of the map due to conflicting signals in a heterogeneous mapping population. The possible rearrangement is smaller than the marker distance we have obtained in previous studies (Stemshorn *et al.* 2011; Cheng *et al.* 2013) and was detected only when including additional markers developed for this study (Table S2). Closely related species frequently differ by chromosomal rearrangements such as inversions, which have been thought to play a critical role as genomic prerequisite in adaptation and speciation (McCarthy *et al.* 1995; Rieseberg 2001; Livingstone and Rieseberg 2004; Kirkpatrick 2010) and also hybridization (McCarthy *et al.* 1995; Buerkle *et al.* 2000). The precise order of loci around the Edar locus still requires detailed genetic analysis with more markers but the Edar locus containing QTL strongly becomes the candidate for phenotypic changes in *Cottus* species.

Parallel evolution for prickling in freshwater sculpins

Particular phenotypic traits can evolve repeatedly when independent populations are exposed to similar ecological conditions (Bernatchez *et al.* 2010). Prickling has an intriguing prevalence in some species of sculpins that makes it an important taxonomic character (Koli 1969; Freyhof *et al.* 2005). On the other hand patterns of prickling do not reflect the phylogeny of freshwater *Cottus* across enormous geographic distance. Kinzinger *et al.* (2005) have assigned *Cottus* species across Europe, Asia and North America to 5 clades according to mitochondrial sequences. Prickled and non-prickled *Cottus* can be found in different clades, like *C. asper* (prickled) and *C. aleuticus* (non-prickled) from the Cottopsis Clade (Western North America); *C. rhotheus* (prickled) and *C. bairdii* (non-prickled) from the Uranidea Clade (North America); and in addition to the species studied here, *C. sibiricus* (prickled) and *C. gobio* (non-prickled) from the Cottus Clade (Europe). The phylogeny implies that the phenotype has evolved independently on several occasions and suggests that the trait is subject to parallel, recurrent or long term stabilizing evolutionary processes much like it was suggested for sticklebacks (Colosimo *et al.* 2005; Hohenlohe *et al.* 2010). It is because of the tight coupling of *Eda* alleles with ecological conditions that the EDA signaling pathway is thought to play a key role in the adaption of sticklebacks (Barret *et al.* 2008; Schluter *et al.* 2010). The *Eda* controlled lateral plate number is probably under divergent selection, with a high plate number being advantageous in marine environments where toothed predators are in abundance, and disadvantageous in freshwater habitats where low-plated fish have faster growth rates and possibly better survival in the presence of invertebrate predators (Barrett *et al.* 2008; Schluter *et al.* 2010). Likewise, a coupling of prickling phenotypes with different ecological conditions can be observed as prickling varies conspicuously across ecotones of *Cottus* hybrid zones between invasive *Cottus* and *C. rhenanus* in Germany (Nolte *et al.* 2006), the hybrid zone between *C. gobio* and *C. koshevnikovi* in western Finland (Kontula and Vainola 2001; Freyhof *et al.* 2005) or between species that utilize different ecological niches like *C. asper* and *C. aleuticus* in Alaska (Scott and Crossman 1973). In marine sculpin fishes from North American pacific coast, phylogenetically independent changes in body size and number of scales also suggest selection operated on these traits in novel intertidal and original subtidal habitats for oxygen usage (Knape and Scales 2013). Unfortunately, a precise adaptive role or function of prickling or body shape are not evident from what is known about the freshwater *Cottus* and we are not aware of a general pattern of association of the phenotype with a particular ecological factor. Since similar prickling phenotypes segregate between several species of *Cottus*, the same genetic region is probably responsible for conspicuous patterns of species level divergence in other species of sculpins. Future studies will have to address, whether EDA signaling causes the phenotypic

variance in other species of *Cottus* and if an evolutionary ecological context of this phenotype can be identified.

Ancestry of Edar locus in a hybrid lineage

We have previously studied the genetic makeup and spatial distribution of hybrid “invasive” *Cottus* within the River Rhine system. The general pattern fits that of an ongoing process of hybrid speciation (Nolte *et al.* 2005a; Nolte and Tautz 2010; Stemshorn *et al.* 2011). However, the forces that affect the evolving genome have not been studied beyond the interplay of migration and genetic drift (Stemshorn *et al.* 2011). This is the first time that we can discuss how genomic architecture could contribute to evolutionary processes that resulted in a fixed phenotype in invasive *Cottus*. Out of the 6 *Cottus* EDA pathway genes, only Edar and Nemo showed high ancestry preference for *C. perifretum* within the gene itself as well as outlier loci clustered in neighboring regions. Of these two, only the Edar gene was covered by the most conspicuous QTL interval and associated with prickling and body shape phenotypes. Within the Edar region, 4 outlier loci have a conspicuous 100% ancestry excess of *C. perifretum* alleles which are all clustered in a single 3 kb region that is located 774 kb up stream of the Edar locus. It is possible that this region directly affects the Edar locus as outlier loci were found to be separated from their closest QTL as far as 10 cM from other studies (Hohenlohe *et al.* 2010; Via *et al.* 2012; Renaut *et al.* 2012) but functional studies are essential to test the roles of other genes in this region.

The high similarity of invasive *Cottus* to one of their parental species (*C. perifretum*) is intriguing given the thoroughly admixed genome of the hybrids (Nolte *et al.* 2005a; Stemshorn *et al.* 2011). Our results imply a simple evolutionary scenario that only the fixation of a single genetic region could explain the invasive phenotypic similarity to only one parental species. The alternative scenario would require that favorable alleles at multiple loci acting in concert to determine invasive sculpin phenotypes would be raised in frequency from an admixed gene pool. However, although it is tempting to suspect that the genetic factor has been subject to environmentally induced selection in invasive *Cottus*, there is a chance that this may have evolved by drift alone as we have demonstrated that much of the variance in ancestry among different loci in the invasive gene pool can be explained by drift alone, especially an excess of *C. perifretum* ancestry (Stemshorn *et al.* 2011). Clearly, further studies are needed to address the potential adaptive value and precise mode of regulation of EDA signaling in *Cottus* and to identify the genetic architecture of other traits in invasive *Cottus*. On the other hand, this study reveals details about why invasive *Cottus* are more similar to one of their ancestors, *C. perifretum*, and suggests that the notion that they are identical with *C. perifretum* applies only for a single genetic region. If a genetic architecture as found here for *Cottus* is common in

lineages of hybrid origin, then many hybrid lineages may go undetected after some evolutionary change that blurs hybrid intermediacy has occurred.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank K. de Gelas, A. Kobler, L. Bervoets and F. Volckaert for their benevolent help in obtaining breeding stocks of *Cottus perifretum*. K. Lessenich, A. Mellin (Berirksregierung Köln) and T. Heilbronner from the Siegfischereigenossenschaft have given permits to conduct the research. *Cottus* were kept and bred in the laboratory of MPI in Plön with permission from S. Hauschildt (Veterinaeramt in Plön). E. Bustorf, S. Dembeck, H. Harre, G. Augustin, D. Mertens and D. Lemke have contributed to the laboratory work and fish care. We thank T. Czypionka, S. Ellendt and D. Tautz for comments and discussions and D.T. and the Max-Planck Society for generous support. The project was funded by a DFG grant to A.N.

Supplementary Information

Table S1. Diagnostic indel markers developed from *Cottus* EDA signaling pathway genes and 3 additional markers from the Edar QTL interval, with locus name, segregating PCR product size that distinguishes *C. perifretum* and *C. rhenanus* populations and the primers used for PCR amplification.

Locus	Product Size	Primers (5' - 3')
EDA 18870	294-310 bp	L- ACACGCTGTCACATCTCTCG R- ATCTCTTCCCATTGCTGCAC
EDA19420	188-400 bp	L- TTGAAGGGTACGAAGGCTGT R- GTCCAGATGTTCCCATGAGC
EDAR 8930	332-345 bp	L- TCTGAGAGTCCGCCTGGTAT R- GCAAAGACGCATATGCACAC
Edaradd 5970	303-305 bp	L- TTGACTTCCTCAGCCAGGTT R- CCAGGACACTGCTGTTTGTG
Troy 9460	231-252 bp	L- GGAGACTTTTTGGCAGAGCA R- TGCACCGATAGACCAATGAG
Nemo 4130	191-202 bp	L- CTGTCTCCGTCCCGTTCTAA R- AACGCCTGTGTTGTGTTTCA
ctg04162i	355-359 bp	L- CAGAACTGCGTGCGCATA R- CCCGGGAAAACCTACTTCAG
ctg00996i	282-287 bp	L- GCTTTTGGTGGCAGTAGAGC R- GGCATGCTTGGAACCTCATCT
ctg02117i	200-203 bp	L- GTTGAAGAGCGACTGTGCTG R- TGAGAGGCCAGCATTGTATG

Table S2. Fine mapping of *Cottus* Edar region. The 7 markers in the *Cottus* Edar region were mapped to three blocks (blue, green and yellow) with their map distances (cM) among 8 F₂ mapping families. Marker that is not fully informative in single family was marked as missing date. The markers covering the Edar locus show order patterns highly specific to the origin and gender of their parental populations. This suggests there might be a chromosomal rearrangement occurred in the Edar region among different populations of *Cottus* species.

<i>C. rhenanus</i> (♂) × <i>C. perifretum</i> (♀)					
Broel(♂) × WN(♀)		Broel(♂) × WN(♀)		Naaf(♂) × LB(♀)	
ZG11	cM	ZG18	cM	ZG39	cM
ctg04162i	57.69	ctg04162i	0.00	ctg02117i	0.00
ctg02010	39.42	ctg02010	9.05	ctg00996i	7.03
ctg03148	33.70	ctg03148	12.81	ctg02010	33.17
ctg00996i	26.94	ctg00996i	18.92	ctg03148	36.19
ctg02117i	19.89	LCE21	24.74	ctg04162i	40.47
LCE21	11.58	EDAR8930	25.66	EDAR8930	47.55
EDAR8930	11.58	ctg14942	52.78	LCE21	47.55
ctg14942	0.00			ctg14942	61.20

<i>C. perifretum</i> (♂) × <i>C. rhenanus</i> (♀)					
WN(♂) × Broel(♀)		WN(♂) × Broel(♀)			
ZG40	cM	ZG46	cM		
ctg02010	0.00	ctg02010	0.00		
ctg03148	4.73	ctg03148	3.53		
ctg00996i	7.62	ctg00996i	8.17		
EDAR8930	14.53	EDAR8930	10.31		
ctg02117i	19.35	ctg02117i	17.96		
ctg14942	37.15	ctg14942	27.66		

LB(♂) × Naaf(♀)		LB(♂) × Naaf(♀)		LB(♂) × Naaf(♀)	
ZG26	cM	ZG32	cM	ZG38	cM
ctg02010	0.00	ctg02010	0.00	ctg02010	0.00
ctg04162i	7.66	ctg04162i	5.16	ctg04162i	4.55
ctg03148	8.76	ctg03148	5.71	ctg03148	4.55
LCE21	12.78	LCE21	11.38	ctg00996i	7.63
EDAR8930	14.35	EDAR8930	12.46	EDAR8930	10.70
ctg00996i	28.78	ctg00996i	25.38	LCE21	10.70
ctg02117i	40.20	ctg02117i	35.71	ctg02117i	25.21
ctg14942	54.31	ctg14942	50.70	ctg14942	31.72

Table S3. Diagnostic SNP loci included in the ancestral allele frequency analysis in invasive *Cottus* gene pool with locus name, map position according to the stickleback genome, allele counts and the percentage of *C. perifretum* ancestry.

Table S3 is available as an Excel file and will be uploaded to the journal's homepage as soon as the paper is accepted for publication.

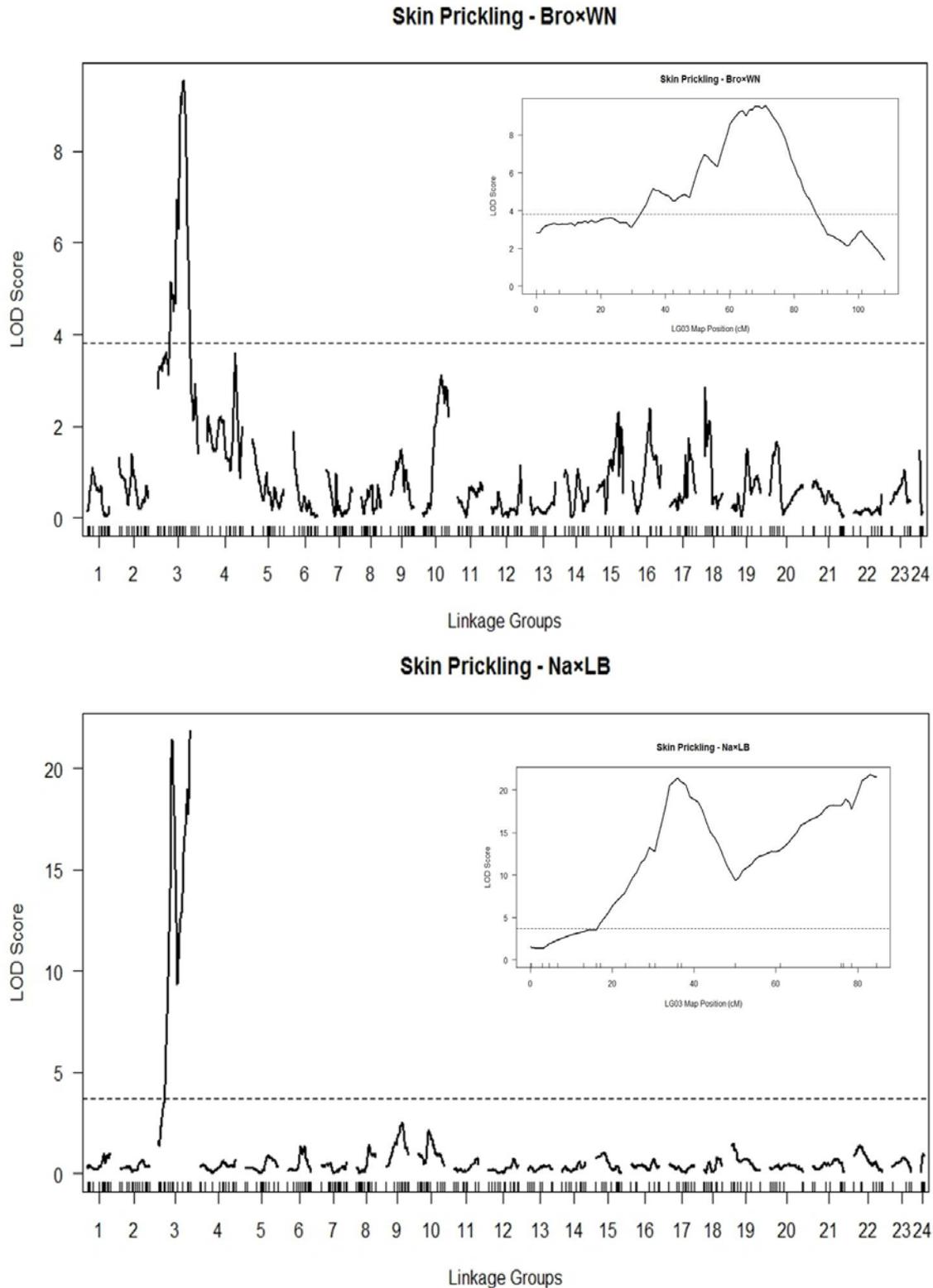


Figure S1. QTL associated with skin prickling for each cross type (origin of the parental populations). Genome-wide significant LOD score at $\alpha < 0.05$ with permutation test is indicated as a dashed line for all linkage groups and enlarged for linkage group 3 that contains the significant QTL.

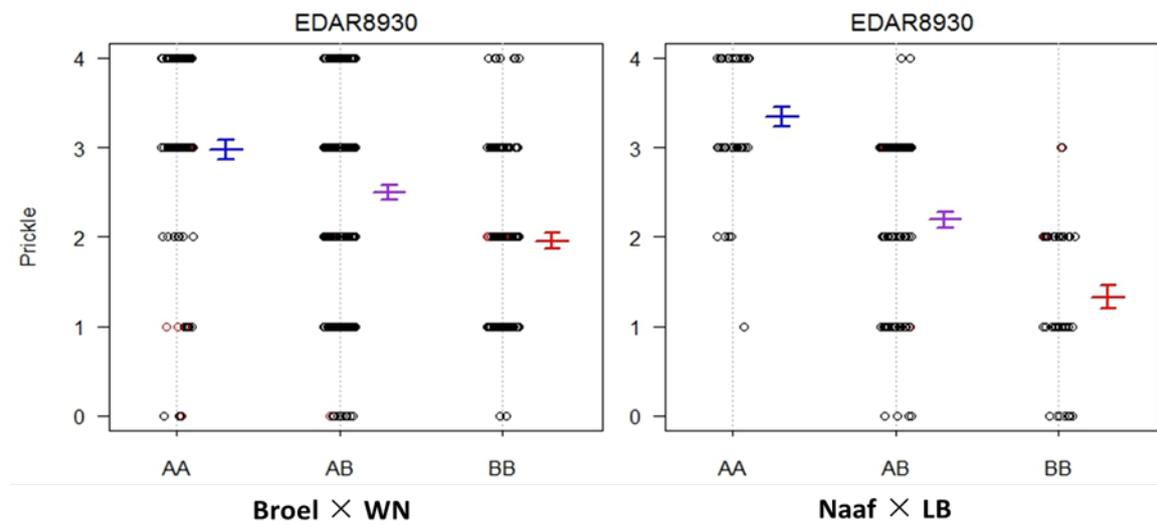


Figure S2. Marker regression from each cross (origin of the parental populations) between the skin prickling phenotype and the genotype at the indel marker EDAR8930 which is nearest to the QTL.

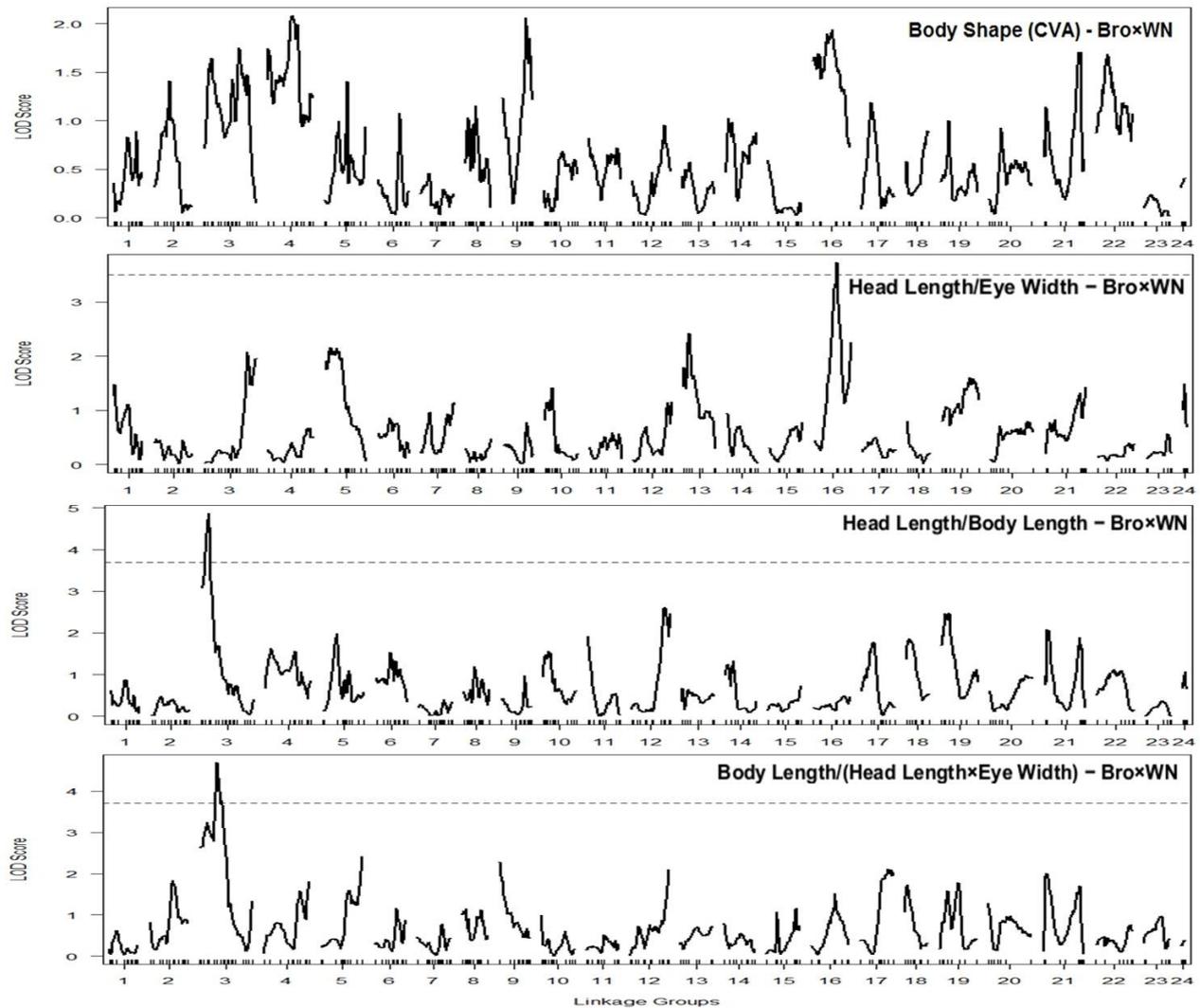


Figure S3. QTLs associated with Canonical variate analysis (CVA) and different morphometric variables (length measurements) of body shape for each cross type (origin of the parental populations) (BroxWN). Genome-wide significant LOD score at $\alpha < 0.05$ with permutation test is indicated as a dashed line for all linkage groups.

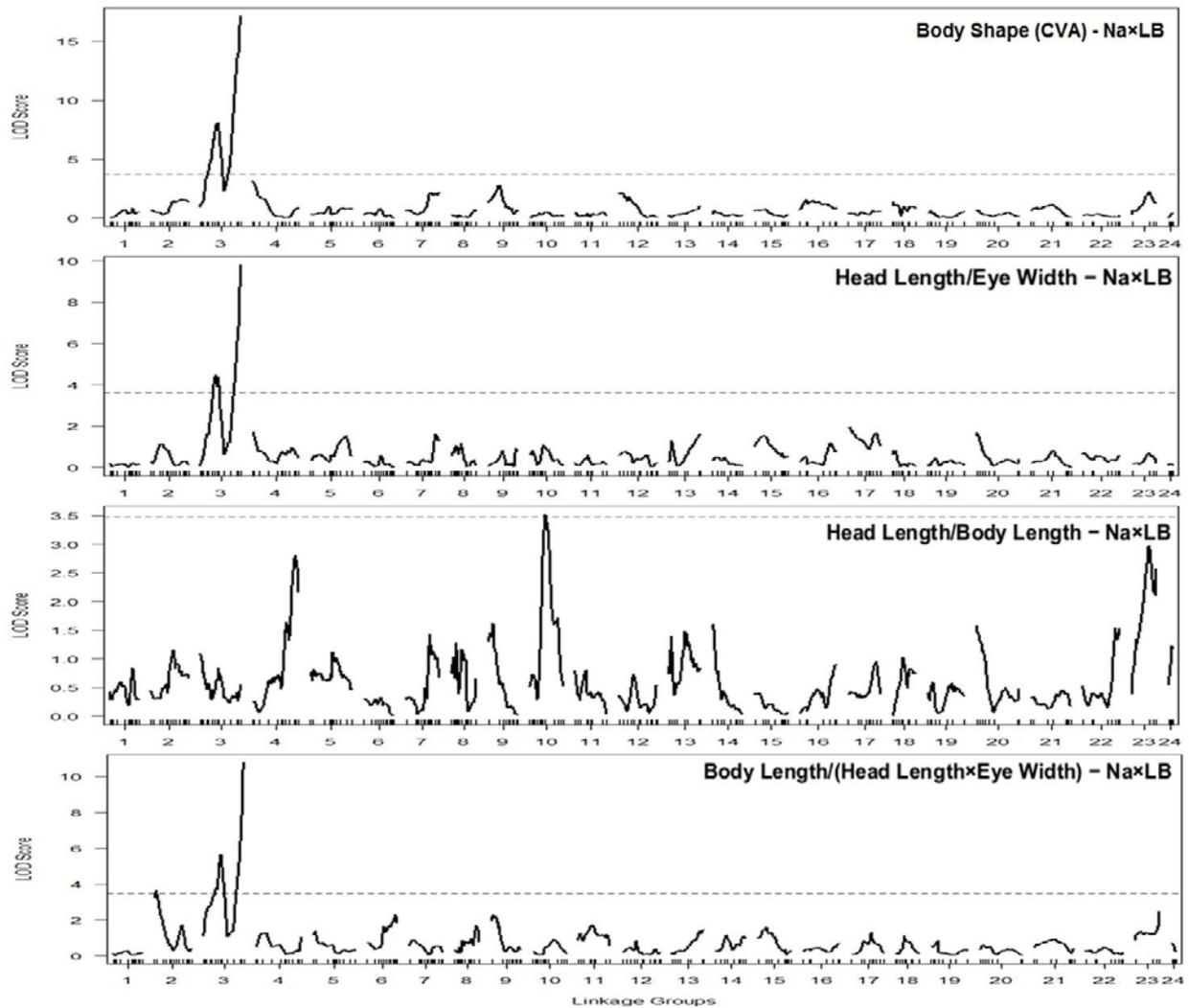
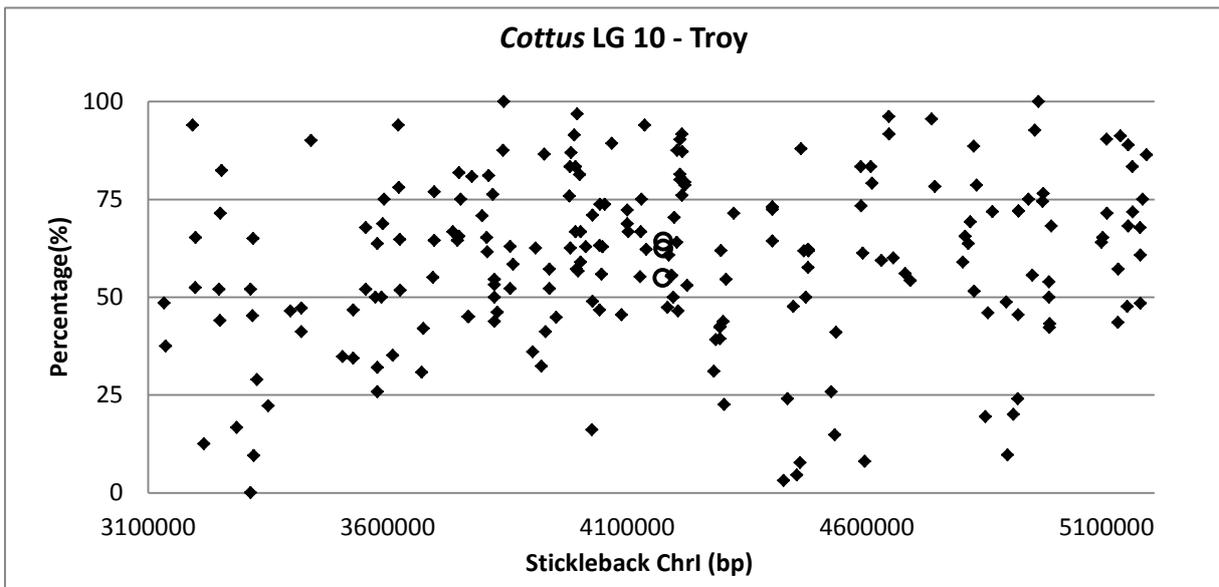
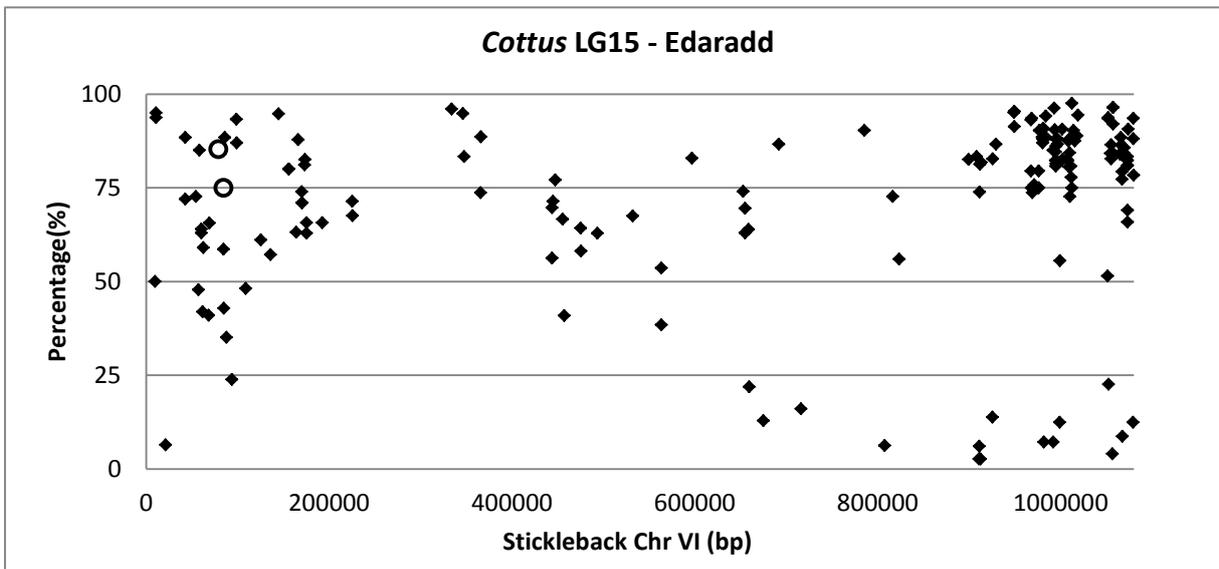
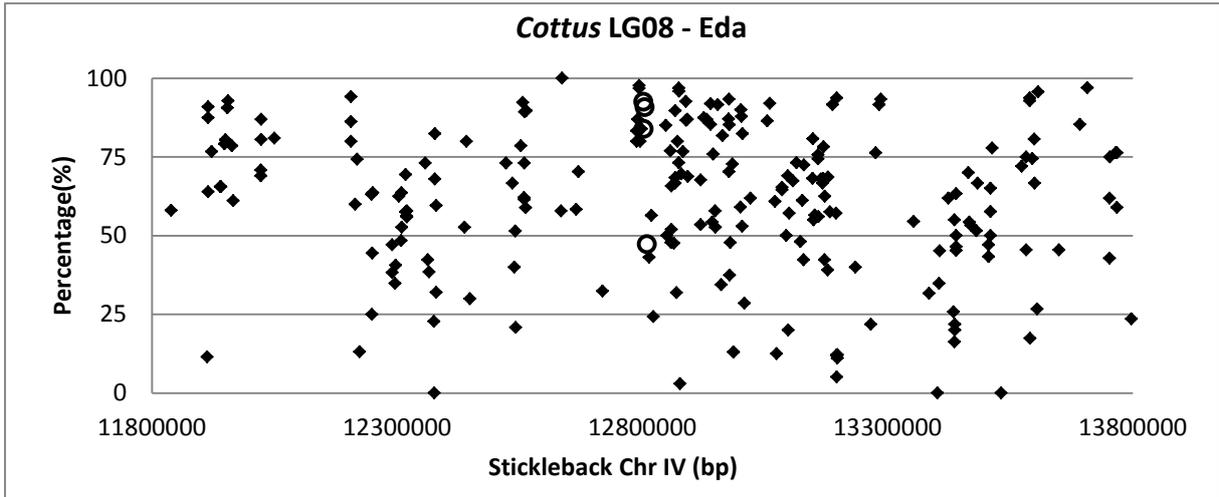


Figure S3. (Continued) QTLs associated with Canonical variate analysis (CVA) and different morphometric variables (length measurements) of body shape for each cross type (origin of the parental populations) (Na×LB). Genome-wide significant LOD score at $\alpha < 0.05$ with permutation test is indicated as a dashed line for all linkage groups.



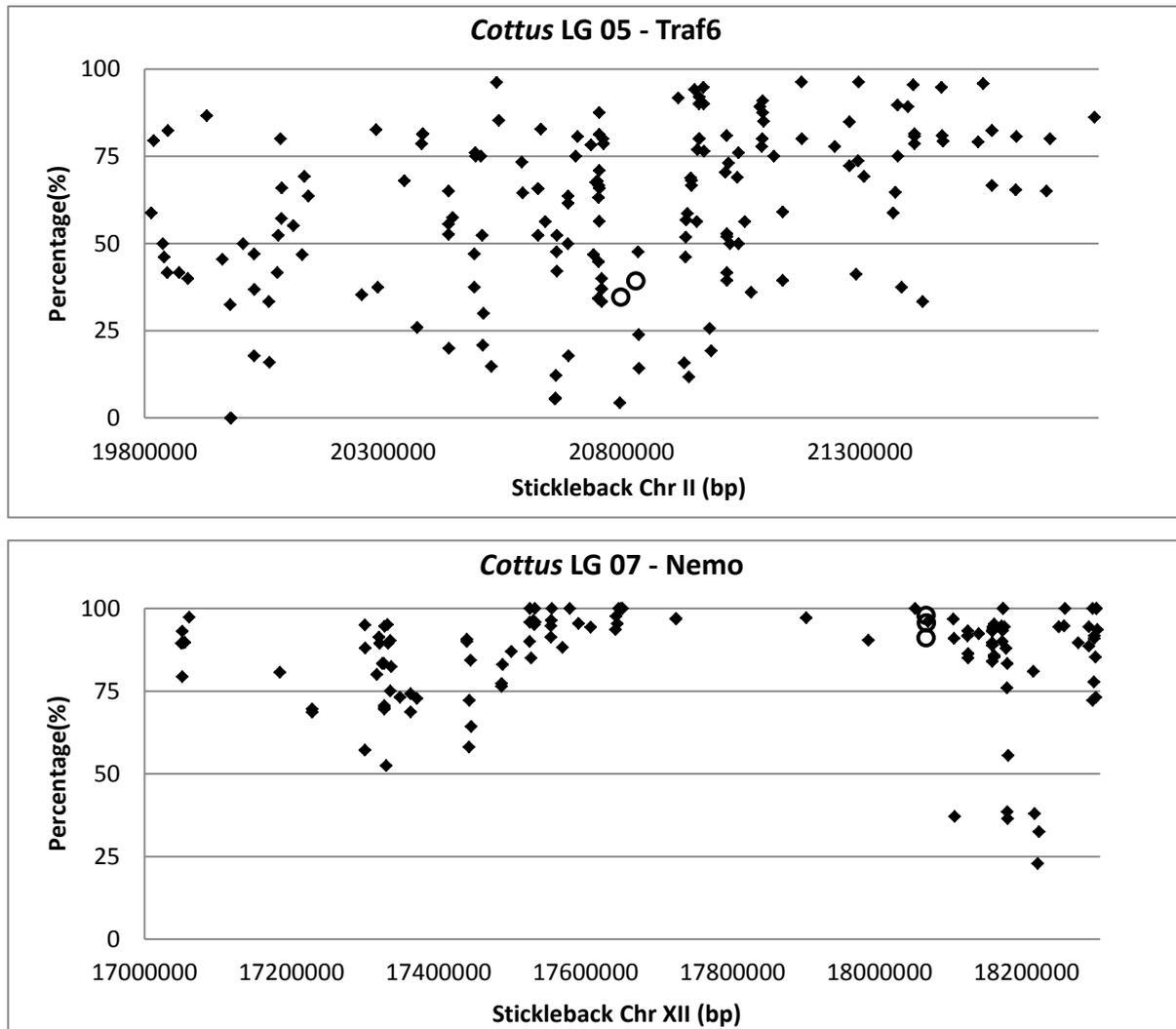


Figure S4. Genomic ancestries in hybrid *Cottus* mapped against the stickleback reference genome for EDA signaling pathway genes (circles) and 1 Mb up-stream and down-stream of the gene (diamonds) except for Edar (see Figure 5). Edaradd and Nemo do not have 1 Mb neighboring region because they are at one end of the chromosome according to stickleback genome. The x axis corresponds to the genomic position [bp] of one stickleback chromosome and the genomic ancestry (percent of the genome derived from *C. perifretum*) is plotted on the y axis. Individual points represent genomic ancestries for SNP loci after data filtering. It is possible to detect genomic regions in which groups of loci deviate strongly from the genome wide average ancestry of 60%. The regions with the most extreme ancestries represent candidates that may be subject to genotypic selection and play a role in adaptive evolution and hybrid speciation.

Conclusions and Perspectives

The European freshwater sculpins (*Cottus*) are rapidly emerging as a new model genetic system to study the questions about the initial stages of hybrid speciation. This dissertation relied heavily on the development of genomic tools for the largely unexplored *Cottus* and the transfer of resources developed for model organisms. We have characterized the genetic basis of morphological variation in natural populations of *Cottus* and traced how genetic variation that originated from the ancestral lineages can affect the fitness of a newly emerging hybrid lineage. It turned out that the stickleback serves as an excellent model for *Cottus*, not only for the transfer of genomic information but also for its well established model of integration of several evolutionary fields (Gibson 2005; Peichel 2005). Here, I will discuss several aspects that offer exciting perspectives for future studies in *Cottus*.

How many genes underline morphological variation in natural populations?

Does evolutionary differentiation accumulate from many small changes with minor phenotypic effects or mutations of large phenotypic effect that underlie evolutionary changes? This has been debated for a long time. Experimental evidence and theoretical prediction suggest that the distribution of loci fixed during adaptation will be exponential, with a single locus of large effect and many more loci of smaller effect, like the pelvic girdle and plate number variation in sticklebacks (Orr 1998; Shapiro *et al.* 2004; Colosimo *et al.* 2004). The results in this dissertation suggest that the Edar locus is the best candidate that strongly determines prickling phenotypic variation of *Cottus*, and also has effects on body shape. We have not detected minor effects of other unlinked loci. Further studies are needed to identify the genetic architecture of other divergent traits in invasive *Cottus* for comparison. For example, the number of vertebrae is another trait that varies between different *Cottus* species (Nolte 2005). Learning about the genetic basis of the number of vertebrae can help to complement whether different divergent traits in *Cottus* are controlled by a single genomic region or interactions of multiple genomic factors.

Do coding or regulatory mutations underlie morphological evolution?

Many cases of morphological differentiation are the result of regulatory evolution, but there are also examples of dramatic morphological changes in natural populations that result from coding changes. In this dissertation, we did find some clues for both coding and possible regulatory potentials from the Edar locus. Regulatory evolution of developmental genes can provide a simple mechanism for natural selection without affecting the overall fitness and viability of

the organisms (Carroll *et al.* 2000). In the future It will be exciting to get a final proof for the function of the Edar gene and find out whether coding or regulatory change contribute to the variation in *Cottus*. Evolutionary developmental experiments like whole mount *in situ* hybridization would help to address the temporal and spatial gene expression pattern of the Edar locus, and transgenic manipulations could finally confirm the function of Edar locus in *Cottus*.

What is the molecular and genetic basis of parallel morphological evolution?

The evolution of similar phenotypes in independent populations with similar environments implies the trait evolves in response to natural selection rather than by genetic drift (Peichel 2005; Elmer and Meyer 2011). Since similar prickling phenotypes segregate between several species of *Cottus*, the same genetic region identified in this dissertation is probably responsible for conspicuous patterns of species level divergence in other *Cottus* species. Future studies will have to address, whether EDA signaling causes the phenotypic variance in other species of *Cottus* and if an evolutionary ecological context and a precise adaptive role of this phenotype can be identified. The best candidates in distantly related *Cottus* species are the prickled *Cottus asper* and non-prickled *Cottus aleuticus* from Northern America. These species are well described *Cottus* lineages (Baumsteiger *et al.* 2012) that were assigned to a different clade from European *Cottus* and a lot of genetic resources in our *Cottus* system are going to be useful to study other species of *Cottus* as well.

How important is the hybridization?

Populations adapt to novel environments in two distinct ways, selection on standing genetic variation or on new mutations (Barrett and Schluter 2007; Wolf *et al.* 2010). This dissertation paves the way to discuss how genomic architecture could contribute to evolutionary processes that resulted in a fixed phenotype in hybrid *Cottus*. It is interesting to ask, whether pre-existing genetic variation or recombination from the ancestors contributed to the evolution of invasive hybrid novelty and their colonization of a new habitat. This question can be explained by exploring the genetic basis of prickling, but natural hybrids of *Cottus* may also carry alternative suites of genetic factors that interact to determine sex depending on their paternal origin. Hence characterizing the genetic basis of sex determination in invasive *Cottus* is highly interest in the future with the fact that the invasive gene pool comprises a complex mixture from different *Cottus* lineages.

What is the role of gene expression ?

Gene expression data can reveal the hidden phenotypes of potential ecological relevance and be used to identify regions of genomic divergence (Wolf *et al.* 2009;

Renaut *et al.* 2012). In this dissertation, the focus is only on morphological traits which may biased our view about the number and nature of phenotypic traits involved in the evolution of hybridizing *Cottus*. Our group has successfully established the high-throughput microarray gene expression approach and identified a set of genes that are responsible for hybridization induced phenotypic variation in invasive *Cottus* (Czypionka *et al.* 2012). Expression QTL (eQTL) could be applied in the future to localize genomic islands of expression divergence and provide more insights into the speciation process of hybridizing *Cottus*.

The conclusions on phenotypic differentiation between *Cottus* parental species and invasive hybrid lineage in this dissertation provide one important step towards all of these questions. The most exciting aspect of the research in *Cottus* perhaps is not what we have learned so far, but the things we have to learn in the future. More efforts on this subject should be promoted both to improve our theoretical understanding and expand the relevant empirical data available for future researchers.

References

- Abbott RJ, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J *et al.* (38 co-authors) 2013. Hybridization and speciation. *Journal of Evolutionary Biology*. 26: 229-246.
- Abbott RJ, Hegarty MJ, Hiscock SJ and Brennan AC. 2010. Homoploid hybrid speciation in action. *Taxon*. 59(5): 1375-1386.
- Abbott RJ and Rieseberg LH. 2012. Hybrid speciation. eLS. DOI: 10.1002/9780470015902.a0001753.pub2.
- Albert AYK, Sawaya S, Vines TH, Knecht AK, Miller CT, Summers BR, Balabhadra S, Kingsley DM and Schluter D. 2008. The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution*. 62(1): 76-85.
- Baack EJ and Rieseberg LH. 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*. 17: 513-518.
- Bachtrog D, Hom E, Wong KM, Maside X and de Jong P. 2008. Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biology*. 9(2): R30.
- Backstroem N, Karaiskou N, Leder EH, Gustafsson L, Primmer CR, Qvarnstrom A and Ellegren H. 2008. A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics*. 179: 1479-1495.
- Baroiller JF, D'Cotta H and Saillant E. 2009. Environmental Effects on Fish Sex Determination and Differentiation. *Sexual Development*. 3: 118–135.
- Barrett RDH, Rogers S and Schluter D. 2008. Natural selection on a major armor gene in threespine stickleback. *Science*. 322: 255-257.
- Barrett RDH and Schluter D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*. 23: 38-44.
- Barrett RDH. 2010. Adaptive evolution of lateral plates in three-spined stickleback *Gasterosteus aculeatus*: a case study in functional analysis of natural variation. *Journal of Fish Biology*. 77: 311-328.

- Barton NH and Gale KS. 1993. Genetic analysis of hybrid zones. Pp. 13-45 in R. G. Harrison, ed. *Hybrid zones and the evolutionary process*. Oxford University Press, Oxford.
- Barton NH. 2001. The role of hybridization in evolution. *Molecular Ecology*. 10: 551-568.
- Baumsteiger J, Kinziger AP and Aguilar A. 2012. Life history and biogeographic diversification of an endemic western north American freshwater fish clade using a comparative species tree approach. *Molecular Phylogenetics and Evolution*. 65: 940-952.
- Bernatchez L, Renaut S, Whiteley AR, Derome N, Jeukens J, Landry L, Lu G, Nolte AW *et al.* (11 co-authors). 2010. On the origin of species: insights from the ecological genomics of lake whitefish. *Phil. Trans. R. Soc. B*. 365: 1783-1800.
- Bierne N, Welch J, Loire E, Bonhomme F and David P. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*. 20: 2044-2072.
- Boffelli D, Nobrega MA and Rubin EM. 2004. Comparative genomics at the vertebrate extremes. *Nature Reviews Genetics*. 5: 456-465.
- Bolnick DI and Near TJ. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). *Evolution*. 59: 1754–1767.
- Broman KW and Sen Ś. 2009. *A guide to QTL mapping with R/qtl*. New York: Springer.
- Broman KW, Wu H, Sen Ś and Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. 19: 889-890.
- Brudno M, Chuong BD, Cooper GM, Kim MF, Davydov E *et al.* 2003. LAGAN and Multi-LAGAN: efficient tools for large scale multiple alignment of genomic DNA. *Genome Research*. 13: 721-731.
- Buerkle CA, Morris RJ, Asmussen MA and Rieseberg LH. 2000. The likelihood of homoploid hybrid speciation. *Heredity*. 84: 441-451.
- Burk JM and Arnold ML. 2001. Genetics and the fitness of hybrids. *Annual Reviews Genetics*. 35: 31-52.
- Carroll SB. 2000. Endless form: the evolution of gene regulation and morphological diversity. *Cell*. 101:577-580.

- Chan YF, Marks ME, Jones FC, Villarreal Jr. G, Shapiro MD, Brady SD, Southwick AM and Absher DM *et al.* 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science*. 327: 302-305.
- Cheng J, Czyplionka T and Nolte AW. 2013. The genomics of incompatibility factors and sex determination in hybridizing species of *Cottus* (Pisces). *Heredity*. *Accepted*.
- Colosimo PF, Hosemann KE, Balabhadra S, Villarreal Jr. G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D and Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*. 307: 1928-1933.
- Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D and Kingsley DM. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biology*. 2(5):635-641.
- Conover DO and Van Voorhees DA. 1990. Evolution of a balanced sex ratio by frequency-dependent selection in a fish. *Science*. 250: 1556-1558.
- Coyne JA and Orr HA. 2004. *Speciation*. Sunderland, MA, USA: Sinauer Associates.
- Czyplionka T, Cheng J, Pozhitko Av and Nolte AW. 2012. Transcriptome changes after genome wide admixture in invasive sculpins (*Cottus*). *Molecular Ecology*. 21: 4797-4810.
- Darwin CR (Ed.). 1862. *Notes on the Causes of Cross and Hybrid Sterility*. Cambridge University Press, Cambridge.
- Davies N, Aiello A, Mallet J, Pomankowski A and Silberglied RE. 1997. Speciation in two neotropical butterflies: extending Haldane's rule. *Proc. R. Soc. Lond. B*. 264: 845-851.
- De Nie HW. 1997. *Atlas van de Nederlandse Zoetwatervissen*. Doetinchen: Media Publishing pp. 151.
- Dermitzakis ET, Masly JP, Waldrip HM and Clark AG. 2000. Non-Mendelian Segregation of Sex Chromosomes in Heterospecific *Drosophila* Males. *Genetics*. 154: 687-694.
- Devlin RH and Nagahama Y. 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological and environmental influences. *Aquaculture*. 208: 191-364.
- Dobzhansky T. 1937. *Genetics and the Origin of Species*. Columbia university Press, New York.

- Domazet-Lošo M and Haubold B. 2009. Efficient estimation of pairwise distances between genomes. *Bioinformatics*. 25: 3221–3227.
- Edmands S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution*. 53: 1757–1768.
- Elmer KR and Meyer A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*. 26 (6): 298-306.
- Englbrecht CC, Freyhof J, Nolte AW, Rassmann K, Schliewen U and Tautz D. 2000. Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-pleistocene origin of the major central European populations. *Molecular Ecology*. 9: 709-722.
- Erickson DL, Fenster CB, Stenoiien HK and Price D. 2004. Quantitative trait locus analyses and the study of evolutionary process. *Molecular Ecology*. 13:2505-2522.
- Estabrook GF, Gil-ad NL and Reznicek AA. 1996. Hypothesizing Hybrids and Parents Using Character Intermediacy, Parental Distance and Equality. *Taxon*. 45: 647-662.
- Fay JC and Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*. 100: 191-199.
- Frazer KA, Pachter L, Poliakov A, Rubin EM and Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Research*. 32: W273-W279.
- Freyhof J, Kottelat M and Nolte AW. 2005. Taxonomic diversity of European *Cottus* with description of eight new species (Teleostei:Cottidae). *Ichthyological Exploration of Freshwaters*. 16: 107-172.
- Gaut BS. 2001. Patterns of chromosomal duplication in Maize and their implications for comparative maps of the grasses. *Genome Research*. 11: 55-66.
- Gibson G. 2005. Evolution: the synthesis and evolution of a supermodel. *Science*. 307: 1890-1891.
- Haldane JBS. Sex-ration and unisexual sterility in hybrid animals. *Journal of Genetics*. 12: 101-109.
- Harr B. 2006. Genomic islands of differentiation between house mouse subspecies. *Genome Research*. 16: 730-737.

- Harris MP, Rohner N, Schwarz H, Perathoner S, Konstantinidis P and Volhard CN. 2008. Zebrafish *eda* and *edar* mutants reveal conserved and ancestral roles of ectodysplasin signaling in vertebrates. *PLoS Genetics*. 4(10): e1000206.
- Haubold B, Domazet-Lošo M and Wiehe T. 2008. An alignment-free distance measure for closely related genomes. In C. R. Nelson and S. Vialette, editors, *RECOMB-CG*, volume 5267 of *Lecture Notes in Bioinformatics*, pages 87–99, Heidelberg, 2008. Springer Verlag.
- Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. 487: 94-98.
- Hermansen JS, Sæther SA, Elgvin TO, Borge T, Hjelle E and Sætre GP. 2011. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology*. 20: 3812-3822.
- Herpin A and Schartl M. 2011. *Dmrt1* genes at the crossroads: a widespread and central class of sexual development factors in fish. *FEBS Journal*. 278: 1010–1019.
- Hochkirch A. 2012. Hybridization and the origin of species. *Journal of Evolutionary Biology*. 26: 247-251.
- Hoffmann AA and Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation. *Annual Reviews Ecology & Evolution system*. 39: 21-42.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA and Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*. 6(2): e1000862.
- Hvala JA and Wood TE. 2012. Speciation: Introduction. eLS. DOI:10.1002/9780470015902.a0001709.pub3.
- James JK and Abbott RJ. 2005. Recent, allopatric, homoploid hybrid speciation: the origin of Oxford ragwort, *Senecio squalidus* (Asteraceae), in the British isles from a hybrid zone on Mount Etna, Sicily. *Evolution*. 59: 2533-2547.
- Johnson N. 2008a. Hybrid Incompatibility and Speciation. *Nature Education* 1(1).
- Johnson N. 2008b. Haldane's rule: the heterogametic sex. *Nature Education* 1(1).
- Jones FC, Brown C, Pemberton JM and Braithwaite VA. 2006. Preproductive isolation in a threespine stickleback hybrid zone. *Journal of Evolutionary Biology*. 19: 1531-1544.

- Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, Fujita M, Suetake H, Suzuki S, Hosoya S, Tohari S, Brenner S, Miyadai T, Venkatesh B, Suzuki Y and Kikuchi K. 2012. A Trans-species missense SNP in *Amhr2* is associated with sex determination in the Tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*. 8(7): e1002798.
- Kasahara M, Naruse K, Sasaki S *et al.* (33 co-authors). 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature*. 447: 714–719.
- Kawahara R, Miya M, Mabuchi K, Lavoue S, Inoue JG, Satoh TP, Kawaguchi A and Nishida M. 2008. Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes and their relatives): A new perspective based on whole mitogenome sequences from 75 higher teleosts. *Molecular Phylogenetics and Evolution*. 46: 224–236.
- Kikuchi K, Kai W, Hosokawa A, Mizuno N, Suetake H, Asahina K and Suzuki Y. 2007. The sex-determining locus in the tiger pufferfish, *Takifugu rubripes*. *Genetics*. 175: 2039-2042.
- Kinziger AP, Wood RM and Neely DA. 2005. Molecular systematic of the genus *Cottus* (Scorpaeniformes: Cottidae). *Copeia*. 2: 303-311.
- Kirkpatrick M. 2010. How and why chromosome inversions evolve. *PLoS Biology*. 8(9): e1000501.
- Knapen D, Knaepkens G, Bervoets L, Taylor MI, Eens M *et al.* 2003. Conservation units based on mitochondrial and nuclear DNA variation among European bullhead populations (*Cottus gobio* L., 1758) from Flanders, Belgium. *Conservation Genetics*. 4(2): 129-140.
- Knecht AK, Hosemann KE and Kingsley DM. 2007. Constraints on utilization of the EDA-signaling pathway in threespine stickleback evolution. *Evolution & Development*. 9(2): 141-154.
- Knope ML and Scales JA. 2013. Adaptive morphological shifts to novel habitats in marine sculpin fishes. *Journal of Evolutionary Biology*. doi:10.1111/jeb.12088.
- Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V *et al.* 2011. PoPoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* 6(1): e15925.
- Koide Y, Ikenaga M, Sawamura N, Nishimoto D, Matsubara K, Onishi K, Kanazawa A and Sano Y. 2008. The Evolution of Sex-Independent Transmission Ratio Distortion Involving Multiple Allelic Interactions at a Single Locus in Rice. *Genetics*. 180: 409-420.

- Koli L. 1969. Geographical variation of *Cottus gobio* L. (Pisces, Cottidae) in Northern Europe. *Annales Zoologici Fennici*, 6: 353-390.
- Kondo S, Kuwahara Y, Kondo M, Naruse K, Mitani H, Wakamatsu Y, Ozato K, Asakawa S, Shimizu N and Shima A. 2001. The medaka *rs-3* locus required for scale development encodes ectodysplasin-A receptor. *Current Biology*. 11: 1202-1206.
- Kondo M, Hornung U, Nanda I, Imai S, Sasaki T, Shimizu A, Asakawa S, Hori H, Schmid M, Shimizu N and Schartl M. 2006. Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Research*. 16: 815–826.
- Kondo M, Nanda I, Schmid M and Schartl M. 2009. Sex determination and sex chromosome evolution: insights from medaka. *Sexual Development*. 3: 88-98.
- Kontula T and Vainola R. 2001. Postglacial colonization of Northern Europe by distinct phylogeographic lineages of the bullhead, *Cottus gobio*. *Molecular Ecology*. 10: 1983-2002.
- Korf I, Yandell M and Joseph B. 2003. BLAST. L. LeJeune, O'Reilly.w
- Kottelat M and Freyhof J. 2007. Handbook of European freshwater fishes. Kottelat, Cornol, Switzerland and Freyhof, Berlin, Germany.
- Kruuk LE, Baird SJ, Gale KS and Barton NH. 1999. A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*. 153: 1959-1971.
- Kucuktas H, Wang S, Li P *et al.* (16 co-authors). 2009. Construction of Genetic Linkage Maps and Comparative Genome Analysis of Catfish Using Gene-Associated Markers. *Genetics*. 181: 1649–1660.
- Kulathinal RJ and Singh RS. 2008. The molecular basis of speciation: from patterns to processes, rules to mechanisms. *Journal of Genetics*. 87: 327–338.
- Kunte K, Shea C, Aardema ML, Scriber JM, Juenger TE, Gilbert LE *et al.* 2011. Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. *PLoS Genetics*. 7(9): e1002274.
- Lehmann J, Stadler PF and Prohaska SJ. 2008. SynBlast: Assisting the analysis of conserved syntenic information. *BMC Bioinformatics*. 9: 351.

- Li B, Dettai A, Cruaud C, Couloux A, Desoutter-Meniger M and Lecointre G. 2009. *RNF213*, a new nuclear marker for acanthomorph phylogeny. *Molecular Phylogenetics and Evolution*. 50: 345–363.
- Li H and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25: 1754-1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. 25: 2078-2079.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J and Wang J. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research*. 20: 265-272.
- Livingstone K and Rieseberg LH. 2004. Chromosomal evolution and speciation: A recombination-based approach. *New Phytologist*. 161: 107-112.
- Lowry DB and Willis JH. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation and reproductive isolation. *PLoS Biology*. 8(9): e1000500.
- Lukens L, Zou F, Lydiate D, Parkin I and Osborn T. 2003. Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics*. 164: 359-372.
- Maheshwari S and Barbash DA. 2011. The genetics of hybrid incompatibilities. *Annual Reviews Genetics*. 45: 331-355.
- Malek TB, Boughman JW, Dworkin I and Peichel CL. 2012. Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Molecular Ecology*. 21: 5265-5279.
- Mallet J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*. 20: 229-237.
- Mallet J. 2007. Hybrid speciation. *Nature*. 446: 279-283.
- Mank JE and Avise JC. 2009. Evolutionary Diversity and Turn-Over of Sex Determination in Teleost Fishes. *Sexual Development*. 3: 60–67.
- Marie Curie Speciation Network. 2012. What do we need to know about speciation? *Trends in Ecology & Evolution*. 27(1): 27-39.

- Marin I and Baker BS. 1998. The evolutionary dynamics of sex determination. *Science*. 281: 1990–1994.
- Mayr E. 1942. *Systematics and the origin of species*. (Columbia Univ Press. New York).
- McCarthy EM, Marjorie AA and Anderson WW. 1995. A theoretical assessment of recombinational speciation. *Heredity*. 74: 502–509.
- McGraw L A, Davis JK, Lowman JJ, ten Hallers BFH, Koriabine M, Young LJ, Jong PJ, Rudd MK and Thomas JW. 2010. Development of genomic resources for the prairie vole (*Microtus ochrogaster*): construction of a BAC library and vole-mouse comparative cytogenetic map. *BMC Genomics*. 11: 70.
- Morlon A, Munnich A and Smahi A. 2005. TAB2, TRAF6 and TAK1 are involved in NF- κ B activation induced by the TNF-receptor, Edar and its adaptator Edaradd. *Human Molecular Genetics*. 14(23): 3751-3757.
- Myosho T, Otake H, Masuyama H, Matsuda M, Kuroki Y, Fujiyama A, Naruse K, Hamaguchi S and Sakaizumi M. 2012. Tracing the emergence of a novel sex-determining gene in Medaka, *Oryzias luzonensis*. *Genetics*. 191: 163-170.
- Nolte AW. 2005. Evolutionary genetic analysis of an invasive population of sculpins in the Lower Rhine. PhD thesis. Köln University.
- Nolte AW, Freyhof J, Stemshorn KC and Tautz D. 2005a. An invasive lineage of sculpins, *Cottus* sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. *Proc. R. Soc. B*. 272: 2379-2387.
- Nolte AW, Stemshorn KC and Tautz D 2005b. Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. *Molecular Ecology Notes*. 5: 628-636.
- Nolte AW and Sheets HD. 2005. Shape based assignment tests suggest transgressive phenotypes in natural sculpin hybrids (Teleostei, Scorpaeniformes, Cottidae). *Frontiers in Zoology*. 2: 11.
- Nolte AW, 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.
- Nolte AW, Gompert Z and Buerkle CA. 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*. 18: 2615–2627.

- Nolte AW and Tautz D. 2010. Understanding the onset of hybrid speciation. *Trends in Genetics*. 26 (2): 54-58.
- Noor MAF and Feder JL. 2006. Speciation genetics: evolving approaches. *Nature Reviews Genetics*. 7: 851-861.
- Nosil P and Feder JL. 2012. Genomic divergence during speciation: causes and consequences. *Phil. Trans. R. Soc. B*. 367: 332-342.
- Nosil P and Schluter D. 2011. The genes underlying the process of speciation. *Trends in Ecology & Evolution*. 26(4): 160-167.
- Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*. 52: 935-949.
- Orr HA, Masly JP and Phadnis N. 2007. Speciation in *Drosophila*: from phenotypes to molecules. *Journal of Heredity*. 98: 103-110.
- Pantalacci S, Chaumot A, Benoit G, Sadier A, Delsuc F, Douzery E and Laudet V. 2008. Conserved features and evolutionary shifts of the EDA Signaling pathway involved in vertebrate skin appendage development. *Molecular Biology and Evolution*. 25(5): 912-928.
- Peichel CL. 2005. Fishing for the secrets of vertebrate evolution in threespine sticklebacks. *Developmental dynamics*. 234: 815-823.
- Petes TD. 2001. Meiotic recombination hot spots and cold spots. *Nature Reviews Genetics*. 2: 360-369.
- Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD and Moczek AP. 2010. Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*. 25(8): 459-467.
- Presgraves DC. 2010. The molecular evolutionary basis of species formation. *Nature Reviews Genetics*. 11: 175-180.
- Quinlan AR, Stewart DA, Strömberg MP and Marth GT. 2008. PYROBAYES: An improved base-caller for SNP discovery in pyrosequences. *Nature Method*. 5: 179-181.
- Qvarnström A and Bailey RI. 2008. Speciation through evolution of sex-linked genes. *Heredity*. 102: 4-15.

- Qvarnström A, Rice AM and Ellegren H. 2010. Speciation in *Ficedula* flycatchers. *Phil. Trans. R. Soc. B.* 365: 1841–1852.
- Ramsey J, Bradshaw HD and Jr. Schemske DW. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution.* 57: 1520-1534.
- Reid DT and Peichel CL. 2010. Perspectives on the genetics architecture of divergence in body shape in sticklebacks. *Integrative and Comparative biology.* pp.1-10.
- Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM and Bernatchez L. 2012. Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Phil. Trans. R. Soc. B.* 367: 354-363.
- Rice AM, Rudh A, Ellegren H and Qvarnstrom A. A guide to the genomics of ecological speciation in natural animal populations. *Ecology Letters.* 14: 9-18.
- Rieseberg LH, Sivervo B, Linder CR, Ungerer MC and Arias DM. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science.* 272: 741-745.
- Rieseberg LH, Archer MA and Wayne RK. 1999. Transgressive segregation, adaptation and speciation. *Heredity.* 83: 363-372.
- Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution.* 16: 351-358.
- Rieseberg LH, Widmer A, Arntz AM and Burke JM. 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Phil. Trans. R. Soc. B.* 358: 1141-1147.
- Roberts HF. 1929. *Plant Hybridization before Mendel.* Princeton: Princeton University Press.
- Rohlf FJ. TpsDig, digitize landmarks and outlines, version 1.39. Department of Ecology and Evolution, State University of New York at Stony Brook; 2003.
- Ross JA, Urton JR, Boland J, Shapiro MD and Peichel CL. 2009. Turnover of sex chromosomes in the stickleback fishes (*Gasterosteidae*). *PLoS Genetics.* 5(2): e1000391.
- Russell S. 2003. Evolution of intrinsic post-zygotic reproductive isolation in fish. *Ann. Zool. Fennici.* 40: 321-329.

- Salazar C, Baxter SW, Pardo-Diaz C, Wu G, SurrIDGE A, Linares M, Bermingham E and Jiggins CD. 2010. Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLoS Genetics*. 6(4): e1000930.
- Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*. 18: 169-185.
- Santini F, Miglietta MP and Faucci A. 2012. Speciation: Where Are We Now? An Introduction to a Special Issue on Speciation. *Evolutionary Biology*. 39: 141-147.
- Sarropoulou E, Nousdili D, Magoulas A and Kotoulas G. 2008. Linking the genomes of nonmodel teleosts through comparative genomics. *Marine Biotechnology*. 10: 227-233.
- Schilthuizen M, Giesbers MCWG and Beukeboom LW. 2011. Haldane's rule in the 21st century. *Heredity*. 107: 95-102.
- Schluter D. 2000. *The ecology of adaptive radiation*. Oxford: Oxford University Press.
- Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution*. 16(7): 372-380.
- Schluter D. 2009. Evidence for Ecological Speciation and Its Alternative. *Science*. 323: 737-740.
- Schluter D, Marchinko KB, Barrett RDH and Rogers SM. 2010. Natural selection and the genetics of adaptation in threespine stickleback. *Phil. Trans. R. Soc. B*. 365: 2479-2486.
- Scott WB and Crossman EJ. 1973. *Freshwater fishes of Canada*. Fisheries research board of Canada, Ottawa (1973) Bulletin 184.
- Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D and Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*. 428: 717-723.
- Shaw KL. 2012. Species and Speciation: Overview. eLS. DOI:10.1002/9780470015902.a0001742.pub2.
- Sheets HD. 2002. IMP-Integrated Morphometrics Package. Department of Physics, Canisius College, Buffalo, New York;
- Shinomiya A, Kato M, Yaezawa M, Sakaizumi M and Hamaguchi S. 2006. Interspecific Hybridization Between *Oryzias latipes* and *Oryzias curvinotus* Causes XY Sex

- Reversal. *Journal of Experimental Zoology Part A: Comparative Experimental Biology*. 305A(10): 890–896.
- Smith WL and Wheeler WC. 2004. Polyphyly of the mail-cheeked fishes (Teleostei: Scorpaeniformes): evidence from mitochondrial and nuclear sequence data. *Molecular Phylogenetics and Evolution*. 32: 627–646.
- Sobel JM, Chen GF, Watt LR and Schemske DW. 2010. The biology of speciation. *Evolution*. 64: 295-315.
- Steinke D, Salzburger W and Meyer A. 2006. Novel Relationships Among Ten Fish Model Species Revealed Based on a Phylogenomic Analysis Using ESTs. *Journal of Molecular Evolution*. 62: 772–784.
- Stelkens R, Young KA and Seehausen O. 2009. The accumulation of reproductive incompatibilities in African cichlid fish. *Evolution*. 64(3): 617–633.
- Stemshorn KC, Nolte AW 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.
- Stemshorn KC, Reed FA, Nolte AW and Tautz D. 2011. Rapid formation of distinct hybrid lineages after secondary contact of two fish species (*Cottus* sp.). *Molecular Ecology*. 20: 1475-1491.
- Stern DL and Orgogozo V. 2009. Is genetic evolution predictable? *Science*. 323: 746-51.
- Takehana Y, Hamaguchi S and Sakaizumi M. 2008. Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi*. *Chromosomes Research*. 16: 801-811.
- Taylor SJ, Rojas LD, Ho SW and Martin NH. 2013. Genomic collinearity and the genetic architecture of floral differences between the homoploid hybrid species *Iris nelsonii* and one of its progenitors, *Iris hexagona*. *Heredity*. 110(1): 63-70.
- Tautz D. 2000. Evolution of transcriptional regulation. *Current Opinion in Genetics & Development*. 10: 575-579.
- The Marie Curie SPECIATION Network. 2012. What do we need to know about speciation? *Trends in Ecology & Evolution*. 27 (1): 27-39.
- Thesleff I and Mikkola ML. 2002. Death receptor signaling giving life to ectodermal organs. *Sci.STKE*. 131:pe22.

- Valenzano DR, Kirschner J, Kamber RA, Zhang E, Weber D, Cellerino A, Englert C, Platzer M, Reichwald K and Brunet A. 2009. Mapping loci associated with tail color and sex determination in the short-lived fish *Nothobranchius furzeri*. *Genetics*. 183: 1385-1395.
- Via S. 2009. Natural selection in action during speciation. *Proc. Natl. Acad. Sci. USA*. 106: 9939-9946.
- Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Phil. Trans. R. Soc. B*. 367: 451-460.
- Via S, Conte G, Mason-Foley C and Mills K. 2012. Localizing F_{st} outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Molecular Ecology*. 21: 5546-5560.
- Vitturi R and Rasotto MB. 1990. Karyotype analysis of *Cottus gobio* L. Pisces Cottidae. *Cytobios*. 62: 81–86.
- Volckaert FAM, Haenfling B, Hellemans B and Carvalho GR. 2002. Timing of the population dynamics of bullhead *Cottus gobio* (Teleostei: Cottidae) during the Pleistocene. *Journal of Evolutionary Biology*. 15: 930-944.
- White MA, Steffy B, Wiltshire T and Payseur BA. 2011. Genetic dissection of a key reproductive barrier between nascent subspecies of house mice, *Mus musculus domesticus* and *Mus musculus musculus*. *Genetics*. 169: 289-304.
- Whitney KD, Randell RA and Rieseberg LH. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *helianthus annuus*. *New Phytologist*. 187: 230-239.
- Wiley C, Qvarnstrom A, Andersson G, Borge T and Saetre PG. 2009. Postzygotic isolation over multiple generations of hybrid descendants is a natural hybrid zone: how well do single-generation estimates reflect reproductive isolation? *Evolution*. 63: 1731–1739.
- Wilson P. 1992. On Inferring Hybridity from Morphological Intermediacy. *Taxon*. 41: 11-23.
- Wittkopp PJ, Haerum BK and Clark AG. 2008. Regulatory changes underlying expression difference within and between *Drosophila* species. *Nature Genetics*. 40: 346-350.
- Wittkopp PJ and Kalay G. 2012. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*. 13: 59-69.

Wolf JBW, Lindell J and Backstro N. 2010. Speciation genetics: Current status and evolving approaches. *Phil. Trans. R. Soc. B.* 365: 1717-1733.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T *et al.* 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biology.* 3: e7.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV and Romano LA. 2003. The Evolution of transcriptional regulation in Eukaryotes. *Molecular Biology and Evolution.* 20(9): 1377-1419.

Wu CI and Ting CT. 2004. Genes and speciation. *Nature Reviews Genetics.* 5: 114-122.

Affidavit

I hereby declare that the dissertation in form and content and except for advices given by my supervisor constitutes my own work. Contributions by collaborators have been accurately specified in the “Declaration of Author’s Contribution”.

The content of first chapter “The genomics of incompatibility factors and sex determination in hybridizing species of *Cottus* (Pisces)” was accepted for publication as a research article in *Heredity* on 1st of July, 2013. The second chapter “The genetic architecture of skin prickling and body shape in hybridizing sculpins (*Cottus*)” is prepared as a research article for publication as well. This work has been undertaken in compliance with the German Research Foundations (Deutsche Forschungsgemeinschaft, DFG) rules of good academic practice.

Plön, 09th of July 2013

Jie Cheng

Curriculum Vitae

Jie Cheng

Date of Birth: 19th April, 1983.

jiecheng@evolbio.mpg.de

chegjie98301@hotmail.com

Nationality: China

Home Town: Qingdao, Shandong

Education

9/2001---6/2005

Bachelor of Marine Biology

College of Marine Life Science, Ocean University of China (OUC), Qingdao, China.

Thesis "AFLP analysis of the genetic variations of four different geographic populations and cultured populations of Zhikong Scallop (*Chlamys farreri*)".

9/2005---6/2009

M.S./M.A. leading to Ph.D. Programme. Major: Molecular Genetics of Mollusk

College of Marine Life Science, Ocean University of China (OUC), Qingdao, China.

Project "Construction and Application of Large Insert Libraries for Zhikong Scallop (*Chlamys farreri*)".

10/2006---2/2007

Sino - German Master Programme on Marine Sciences.

Major: Biological Oceanography.

IFM-GEOMAR (Helmholtz Centre for Ocean Research since 2012), Kiel University, Kiel, Germany.

7/2009---2/2013

Doctoral Candidate. Major: Evolutionary Genetics of Fishes.

Research Group Evolutionary Genetics of Fishes, Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön, Germany.

Publications

1. Czypionka T, **Cheng J**, Pozhitko Av and Nolte AW. 2012. Transcriptome changes after genome wide admixture in invasive sculpins (*Cottus*). *Molecular Ecology* 21: 4797-4810.
2. **Cheng J**, Czypionka T and Nolte AW. 2013. The genomics of incompatibility factors and sex determination in hybridizing species of *Cottus* (Pisces). *Heredity*. *Accepted*.
3. **Cheng J** and Nolte AW. The genetic architecture of skin prickling and body shape in hybridizing sculpins (*Cottus*). *In preparation*.

Presentations, workshops and conferences

1. 1st Joint Congress on Evolutionary Biology. 6th-10th, July, 2012. Ottawa, Canada. **Oral Presentation Title:** Independence of sex determination and genetic incompatibilities in hybridizing species of *Cottus*.
2. Aquavit meeting 2012, Annual scientific meeting of MPI for Evolutionary Biology. 2nd -3rd, May, 2012. Plön, Germany. **Oral Presentation Title:** Fast turnover of sex determination and intrinsic barriers to reproduction in hybridizing species of *Cottus*.
3. Otto Warburg international summer school and research symposium on Evolutionary Genomics. 14th-22nd, September, 2011. Berlin, Germany. **Poster Title:** From stickles to prickles: studying genes in *Cottus* through the stickleback genome.
4. 13th Congress of the European Society for Evolutionary Biology (ESEB). 20th-25th, August, 2011. Tübingen, Germany. **Poster Title:** From stickles to prickles: genetic access to *Cottus* through the stickleback genome.
5. Aquavit meeting 2011, Annual scientific meeting of MPI for Evolutionary Biology. 30th, March-1st, April, 2011. Plön, Germany. **Oral Presentation Title:** Gene-based genetic mapping in two species of sculpins (*Cottus*) and the conservation of synteny with fish model species.
6. PhD-workshop: Evolutionary Genetics - from genes to ecosystems, organized by the Gen-AG, "Evolutionsgenetik" a work group of the Gesellschaft für Genetik, Germany. 28th September - 1st October, 2010. Kappelsberg, Plön, Germany. **Oral Presentation Title:** Genomic mapping of hybrid speciation-related traits in invasive sculpins (*Cottus*).
7. Aquavit meeting 2010, Annual scientific meeting of MPI for Evolutionary Biology. 6th-7th, May, 2010. Plön, Germany. **Oral Presentation Title:** Genomic mapping of hybrid speciation-related traits in invasive sculpins (*Cottus*).

Acknowledgements

I would like to express my gratitude to my supervisor Dr. Arne W. Nolte and Prof. Dr. Diethard Tautz for giving me the opportunity to develop the fascinating *Cottus* project, especially for Arne's guidance and support throughout my studies at Max Planck Institute for Evolutionary Biology. I am also grateful to my two referees Prof. Dr. Diethard Tautz and Prof. Dr. Thomas Bosch for giving their time to evaluate my dissertation and improve my research.

I would like to thank all the people who participated in the *Cottus* project: Dr. Kathryn Stemshorn, Till Czypionka, Sunna Ellendt, Xiangyi Li and all the people who contributed to the laboratory work and fish care: Elke Bustrof, Heik Harre, Sonja Dembeck, Heinke Buhtz, Gerhard Augustin and Daniel Mertens.

As a newcomer to Evolution Biology, I need to learn a lot about new theories and technologies in this field. I thank all current and former members of the Evolutionary Genetics Department for what I have gained in a pleasant, productive and helpful working environment. And give my sincere thanks to all my friends and colleagues at the institute for all their help and support throughout my long years in Plön.

Special thanks to all the proof readers of my thesis for their time, patience with my usage of English and all helpful comments: Till Czypionka, Sunna Ellendt and Henrik Krehenwinkel.

I would like to thank my parents far from home for their extraordinarily support throughout my study in Germany these years. It always feels good knowing to have a home to be welcomed unconditionally. Finally, my love to Wei. I know she is always there supporting me and making me feel nothing to be worried about. Thanks for all your help and support.

At the end I want to dedicate this thesis to my grandpa. Hope he would be proud of me and have a happy time in the paradise.