

Evolutionary analysis of gene expression in hybrid sculpins

Dissertation

**zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität
zu Kiel**

vorgelegt von

Till Czypionka

Kiel 2013

Erster Gutachter: Prof. Dr. Diethard Tautz

Zweiter Gutachter: Prof. Dr. Hinrich Schulenburg

Tag der mündlichen Prüfung: 8.4.2013

Zum Druck genehmigt: 7.6.2013

gez. Prof. Dr. Wolfgang Tuschel (Dekan)

Table of Contents

Summary.....	5
Zusammenfassung.....	7
Introduction.....	10
Evolutionary innovation and the speed of adaptation.....	10
Evolution through genetic exchange.....	11
Phenotypic plasticity and adaptation.....	14
Hybridization in Cottus.....	17
Scope of this thesis.....	19
Chapter 1	21
Transcriptome changes after genome wide admixture in invasive sculpins (Cottus).....	21
Introduction.....	22
Methods.....	24
Study populations and samples.....	24
Development of a microarray.....	26
Raw signal analysis and quality filtering.....	26
Data normalization.....	27
Patterns of gene expression.....	28
Annotation of genes and GO term enrichment analysis.....	30
Results.....	30
Microarray calibration and normalization.....	30
Differentiation in gene expression.....	31
Differential expression between species and populations.....	32
Variance in gene expression.....	33
Transgressivity.....	34
Functional annotation and GO term enrichment of candidate gene sets	35
Discussion.....	37
Microarray calibration.....	37
Hybrid origin and adaptive differentiation of invasive sculpins.....	38
From initial hybridization to the invasive phenotype	40
Acknowledgements.....	42
Chapter 2.....	43
Temperature dependent gene expression and the invasion of new thermal environments in Cottus.....	43
Introduction.....	44
Methods.....	47
Study populations in analysis and fish care.....	47
Temperature treatment	48
Analysis of gene expression.....	49
Signatures of RN-evolution.....	51
Differentiation of gene expression at individual temperatures.....	52
Functional characterization by GO term enrichment.....	52
Results.....	53
Temperatures in natural habitat.....	53
Plastic patterns of gene expression.....	53
Evidence for proposed modes of RN evolution.....	55

Differentiation of gene expression at different temperatures.....	57
Discussion.....	58
Characterization of thermal plasticity in <i>Cottus</i>	59
Evidence for proposed modes of RN evolution.....	61
Expression profile differentiation over temperature.....	62
Higher order phenotypes and future directions.....	64
Acknowledgements.....	65
Conclusion.....	67
Relating the results of the two chapters.....	67
Further characterization of relevant environmental conditions.....	68
Integration of results in the <i>Cottus</i> research program.....	68
Bibliography.....	70
Appendix A – supplementary material from Chapter 1.....	81
Appendix B – supplementary material from Chapter 2.....	82
Declaration of Author’s Contribution.....	86
Affidavit.....	87
Curriculum vitae	88
Acknowledgements.....	89

Figures and Tables

Introduction

Fig 1.....	12
Fig 2.....	16
Fig 3.....	17
Fig 4.....	18

Chapter 1

Table 1.....	25
Fig 1.....	31
Fig 2.....	32
Fig 3.....	33
Table 2.....	35

Chapter 2

Fig 1.....	54
Table 1.....	56
Fig 2.....	57

Summary

European freshwater sculpins of the genus *Cottus* receive attention as a model for studying the initial steps of a homoploid hybrid speciation. An invasive lineage of *Cottus*, which originated from the hybridization between *Cottus rhenanus* and *Cottus perifretum*, was able to invade a new habitat and occupy a new ecological niche not available to its parent species.

In this thesis several aspects of this process are studied at the transcriptome level. To this means, a *Cottus* specific oligonucleotide microarray was designed based on transcriptome sequencing. The binding behavior of the probes on the microarray was experimentally determined, revealing large variation in binding behavior between probes. A newly developed normalization approach was applied, to remove artifacts resulting from this variation.

As a first step, the phenotypic differentiation between invasive *Cottus* and its parent species was assessed. Expression level differentiation indicates that the invasive *Cottus* phenotype is overall intermediate between its parent species in agreement with its hybrid origin. However, invasive *Cottus* are characterized through a set genes with unique expression distinguishing it from its parent species. These are mainly genes with transgressive expression patterns, i.e. expression patterns outside the expression range observed in the parent species. Genes which are transgressively overexpressed in invasive *Cottus* are functionally correlated as evidenced through Gene Ontology term enrichment. These genes represent candidates for the adaptive phenotypic change which enabled the colonization of the new habitat by invasive *Cottus*. Some of the transgressive expression patterns were already observed in F_2 crosses between the parent species, which were analyzed as a proxy of the initial stages of hybrid speciation. This observation suggests that initial hybridization played a role in the formation of the unique invasive phenotype. Importantly, a large fraction of transgressive patterns of over expression were not observed in the in F_2 crosses and thus must have been acquired when invasive *Cottus* evolved after the initial hybridization. Overall these results are in line with an evolutionary process where natural selection acts on hybridization-induced increased phenotypic diversity to shape the new invasive phenotype, thus leading to ecological and spatial segregation from the parent species.

As a second aspect, the role of phenotypic plasticity in promoting adaptation and diversification was assessed. Evolutionary mechanisms involving phenotypic plasticity do

apply particularly in situations involving the adaptation to a new environment. This is the case for invasive *Cottus*, which colonized a new habitat with a different thermal regime. Temperature is a key factor structuring ecological communities along the fluvial ecosystems inhabited by *Cottus*. Thus, thermal plasticity was assessed by measuring temperature specific gene expression profiles over a temperature range from 14°C to 25°C. This temperature range reflects conditions in the originally occupied and the newly invaded habitat during summer when temperature differences between the two habitats are most pronounced. Expression profile differentiation between the invasive *Cottus* and the parent species was most pronounced at the lower half of the analyzed temperature range (i.e. < 21°C). This is surprising, given that these temperatures are not an exclusive property of the newly invaded habitat but are commonly observed both in original and newly invaded habitats. Comparative analysis of gene expression changes over temperatures revealed plastic responses shared by all the species of *Cottus* in analysis as well as changes in plasticity between parent species and invasive *Cottus*. In particular the augmentation of pre-existing plastic responses was shown to contribute to the phenotypic differentiation of invasive *Cottus*, which is in line with the prediction that plastic traits contribute to early steps of adaptive evolution. For the temperatures < 21°C, the augmentation of a plastic response was found in a mitochondrion related phenotype of potential adaptive value.

In summary, this thesis provides evidence in support of proposed modes of hybrid speciation as well as for proposed modes of adaptive evolution through the modification of plastic phenotypes. Furthermore, this thesis constitutes the first systematic phenotypic analysis of the *Cottus* study system and therefore is an important step towards the identification of the adaptive changes underlying the diversification process.

Zusammenfassung

Ein geeignetes Studiensystem zur Erforschung homoploider Hybridspeziation findet sich bei Süßwassergroppen der Gattung *Cottus*. Eine aus der Hybridisierung zwischen *Cottus rhenanus* und *Cottus perifretum* hervorgegangene invasive Hybridlinie konnte erfolgreich ein neues Habitat kolonisieren, in dem ihre Elternarten nicht vorkommen. Dieses Szenario entspricht der Eroberung eines neuen adaptiven Gipfels durch die invasiven Groppen.

In dieser Arbeit werden verschiedene Aspekte dieses Prozesses auf Transkriptomebene untersucht. Zu diesem Zweck wurde ein *Cottus* spezifisches Oligonukleotid-Microarray entworfen. Das Bindungsverhalten der Sonden auf dem Microarray wurde experimentell eruiert. Dabei wurde eine große Variation im Bindungsverhalten der verschiedenen Sonden festgestellt. Eine neu entwickelte Normalisierungsmethode zur Entfernung von aus dieser Variation resultierenden Artefakten wurde hier erstmalig angewandt.

Als erster Schritt wurde die phänotypische Differenzierung zwischen invasiven Groppen und ihren Elternarten charakterisiert. Die Differenzierung der Expressionslevel zeigt das invasive Groppen im Allgemeinen phänotypisch intermediär zu ihren Elternarten sind. Dies ist in Einklang mit der Differenzierung neutraler genetischer Marker und spiegelt die Hybridherkunft der invasiven Groppen wieder. Außerdem zeichnen sich invasive Groppen jedoch durch ein Gruppe von Genen mit spezifischen Expressionsmustern aus, die sie von den Elternarten unterscheiden. Dies sind vor allem Gene mit transgressiven Expressionslevel, also solche Gene, deren Expressionniveau außerhalb des Bereichs der in den Elternarten gefundenen Expressionslevel liegt. In invasiven Groppen transgressiv überexprimierte Gene sind für mehrere Gen Ontologie Termini angereichert, was auf eine funktionellen Zusammenhang hinweist. Diese Gene sind Kandidaten für die Charakterisierung der phänotypischen Veränderung zwischen invasive Groppen und ihren Elternarten, durch welche die Invasion des neuen Habitats ermöglicht wurde. Einige dieser transgressiven Expressionslevel konnten in F_2 Kreuzungen zwischen den Elternarten festgestellt werden, die als ein stellvertretend für die initiale Stufen eines Hybridspeziationsprozesses untersucht wurden. Dieses Ergebnis deutet darauf hin, dass der initiale Hybridisierungsprozess eine Rolle in der Entstehung des einzigartigen invasiven Phänotypen gespielt hat. Allerdings wurde ein Großteil der charakteristischen transgressiven Expressionsmuster der invasiven Groppen nicht in den in F_2 Kreuzungen festgestellt. Diese nicht unmittelbar durch Hybridisierung zu erklärenden Expressionsmuster müssen somit infolge sekundärer evolutionärer Prozesse nach der initialen Hybridisierung entstanden sein. Insgesamt sind diese Resultat im Einklang mit

einem evolutionären Szenario, in dem natürliche Selektion auf durch Hybridisierung erhöhte phänotypische Diversität bewirkt. Der daraus resultierende neue invasive Phänotyp bedingt dann die ökologische und räumliche Segregation der invasiven Groppen von ihren Elternarten.

Als ein weiterer Aspekt wurde der Beitrag plastischer Phänotypen zur Adaptation und Diversifizierung in *Cottus* untersucht. Auf phänotypische Plastizität beruhende evolutionäre Mechanismen sind besonders in Situationen, die Adaptation an neue Umweltfaktoren beinhalten, relevant. Dies ist für die invasive Groppen, die ein Habitat mit einer distinkten Temperatur invadiert haben, der Fall. Temperatur ist ein wichtiger Umweltfaktor, der eine Schlüsselrolle bei der Strukturierung der Biozönosen entlang fluvialer Ökosysteme einnimmt. Deshalb wurde phänotypische Plastizität in *Cottus* in Relation zum Faktor Temperatur untersucht. Dafür wurden temperaturspezifische Genexpressionsprofile für einen Bereich von 14°C bis 25°C analysiert. Unterschiede der Wassertemperatur zwischen dem neu invadierten und dem von den Elternarten bewohnten Habitat sind vor allem im Sommer ausgeprägt. Der analysierte Temperaturbereich umfasst typische Sommertemperaturen beider Habitate. Die ausgeprägteste Expressionsprofil-Differenzierung wurde in der unteren Hälfte (<21°C) des analysierten Temperaturbereichs festgestellt. Dies ist unerwartet, weil diese Temperaturen nicht für das neu invadierte Habitat charakteristisch sind, sondern während des Sommers in beiden Habitaten vorkommen. Durch Analyse der Änderung des Genexpressionsniveaus in Bezug auf Temperatur und durch funktionelle Analyse der relevanter Kandidatengene konnten plastische Phänotypen vorgeschlagen werden. Dies beinhaltete sowohl solche plastischen Phänotypen, die allen analysierten Arten gemein waren, als auch plastische Phänotypen, in denen sich die invasiven Groppen von den Elternarten unterscheiden. Insbesondere die Übersteigerung bereits etablierter plastischer Reaktionen in den invasiven Groppen trug zur phänotypischen Differenzierung derselben von den Elternarten bei. Dies ist im Einklang mit der Vorhersage, dass plastische Phänotypen eine wichtige Rolle in den frühen Phasen von Adaptationsprozessen spielen. In der unteren Hälfte des untersuchten Temperaturbereichs (<21°C), wurde eine Übersteigerung einer bereits vorhandenen plastischen Reaktion für einen mitochondrialen Phänotypen von potentieller adaptiver Relevanz festgestellt.

Zusammenfassend stützen die in dieser Arbeit dargestellten Daten sowohl derzeitige Modelvorstellungen über den Verlauf von Hybridspeziationen, als auch Annahmen über die Rolle von plastischen Phänotypen in Adaptationsprozessen. Diese Arbeit stellt die erste systematische phänotypische Analyse des *Cottus* Studiensystems dar. Sie ist

deshalb ein wichtiger Schritt hin zur Identifikation der adaptiven Veränderungen, die den Diversifikationsprozess in *Cottus* ermöglicht haben.

Introduction

In this work, I am exploring the contribution of two processes, evolution through genetic exchange (Arnold 2007) and evolution through modification of plastic phenotypic responses, to the formation of a new invasive hybrid lineage in the genus *Cottus*. As a general introduction, a background on these processes is provided. Especially their potential role in facilitating (and hindering) adaptation is discussed in comparison to traditional models of evolution. This general introduction is followed by a paragraph that introduces the specific case of hybrid speciation in *Cottus*.

Evolutionary innovation and the speed of adaptation

Adaptation is one of the key processes in evolution and its importance in diversification and speciation is well understood (Schluter 2000, Dieckmann *et al.* 2004). It is now accepted by evolutionary biologists that adaptation happens due to natural selection (Barton *et al.* 2007). Adaptations have been characterized down to the molecular level (e.g. Chan *et al.* 2010, Hoekstra *et al.* 2006, Lang *et al.* 2012) but the actual speed with which adaptations can be realized has received comparably little attention (but see Carrol *et al.* 2007 and cites therein). This is currently changing due to a debate about the adaptability of organisms in the face of climate change (e.g. Hoffmann & Sgrò 2011). One of the main concerns is that even organisms, which have the potential to adapt, might go extinct because the rate of environmental change could exceed the rate of adaptation (Visser 2008). In addition, the speed of adaptation is not only relevant for avoiding extinction via evolutionary rescue (Bell & Gonzalez 2011) but also critical for the successful invasion of new habitats (Prentis *et al.* 2008). In the ecological literature priority effects, i.e. the monopolization of resources by early colonists to a new habitat and the subsequent exclusion of later colonists are well documented (e.g. Shulman *et al.* 1983, Miller-Pierce & Preisser 2012). Urban & de Meester (2009) showed that evolutionary priority effects can also be expected when early colonists quickly adapt to the new habitat before the arrival of potential competitors. In this case, the capability of an early colonist to occupy an adaptive peak and to thereby exclude competitors with a later arriving species with comparable potential for adaptation (de Meester *et al.* 2002) crucially depends on how rapid the early colonist can adapt. In conclusion, not only the potential of a population to evolve adaptive phenotypes, but also the speed by which such evolutionary innovations can be archived are relevant. Adaptation via selection on standing variation can occur fast compared to adaptation due to new mutations (Barrett & Schluter 2008). The latter has

been characterized as a “waiting for a mutation process” (Phillips 1996), emphasizing that beneficial mutations can take prohibitively long to occur and that the probability of subsequent loss through drift can be high (Pfennig *et al.* 2010).

Thus, the ability of a population or species to react to new selection pressures by *fast* adaptation largely depends on pre-existing phenotypic and underlying genetic variation. However, evolutionary “mechanisms” exist, which extend this simplified view and allow for fast adaptation and evolutionary innovation. In the next paragraph I will therefore discuss “evolution through genetic exchange” (Arnold 2007) with a focus on homoploid hybrid speciation. Thereafter, I will review the special role phenotypic plasticity can play in the fast evolution of complex phenotypes, but also in facilitating or hindering adaptation.

Evolution through genetic exchange

Genetic exchange can enable fast adaptation and allow for the expression of phenotypic novelty (Arnold 2007). Genetic exchange, e.g. through hybridization (Abbott *et al.* 2013), leads to an increase in genetic variation and thereby phenotypic variation on which selection may act. It therefore provides an additional source of new and potentially beneficial alleles. Even though these new alleles also must have resulted from mutation processes in their respective source populations at one point, this process is very different from a “waiting for a mutation process” (Phillips 1996). Hybridization can provide many new alleles simultaneously (Arnold 2007). The risk of losing new beneficial alleles due to drift can also be overcome, when these alleles are present at high frequency in the source population and hybridization takes place repeatedly. Moreover, the increase of genetic variation through hybridization is not only due to the recombination of ancestral alleles. Hybridization leads to a “genomic shock” (Comai *et al.* 2003, Hegarty *et al.* 2006) which can cause changes in epigenetic markers (Comai *et al.* 2003). This entails the activation and proliferation of transposable elements (Fontdevila 2005, Ungerer *et al.* 2006, Kawakami *et al.* 2010) which displays a hybridization related mutation process and independent source of genetic variation. This increase in genetic variation (together with potential epigenetic changes) translates into an increase in phenotypic variation (Nolte & Tautz 2010). The potential of hybridization to allow for adaptation then depends on ecological context (Mallet 2007). This is illustrated in Figure 1 which shows a simplified adaptive landscape containing three adaptive peaks. The two peaks at the bottom of the graph are occupied by a red and a blue species respectively. Colored dots represent individuals and the spread of the clouds of dots indicates the phenotypic variation within the population. While the red and the blue species are trapped at their adaptive peaks, the

hybrids between them (represented as two-colored dots) occupy a different space on the adaptive landscape. Most of the hybrid individuals in Fig 1 are found far away from an adaptive peak. This indicates a common outcome of hybridization: the production of unfit offspring (Barton 2001, Coyne and Orr 2004).

In contrast, some of the hybrids are positioned near the still unoccupied adaptive peak at the top of Figure 1. Selection may act on these individuals so that a subpopulation of hybrids are eventually dragged to the new adaptive peak (yellow arrow in Fig 1) (Barton 2001). As can be seen in Figure 1, hybrid phenotypes near the new adaptive peak are extreme compared to the parental phenotypes. The generation of such extreme phenotypes is referred to as transgressive segregation (Rieseberg 1999). Transgressive phenotypes can result from “complementary gene action” (Rieseberg *et al.* 2003a), i.e. the recombination of genes with additive or epistatic effects from the parental species in the hybrid. While transgressive segregation might produce new adaptive phenotypes (Seehausen 2004), allelic combinations of adaptive value might be lost in the next generation due to sexual reproduction with unfit hybrids or by backcrossing to parental lineages (Barton 2001, Buerkle *et al.* 2000). A key question is therefore how an emerging hybrid lineage can archive at least partial reproductive isolation from their parent species (Abbott *et al.* 2013). Different not mutually exclusive solutions to this problem have been described.

Reproductive isolation by polyploidization can occur in animals, but is a speciation mode especially well established in plants (Abbott 2013, Mallet

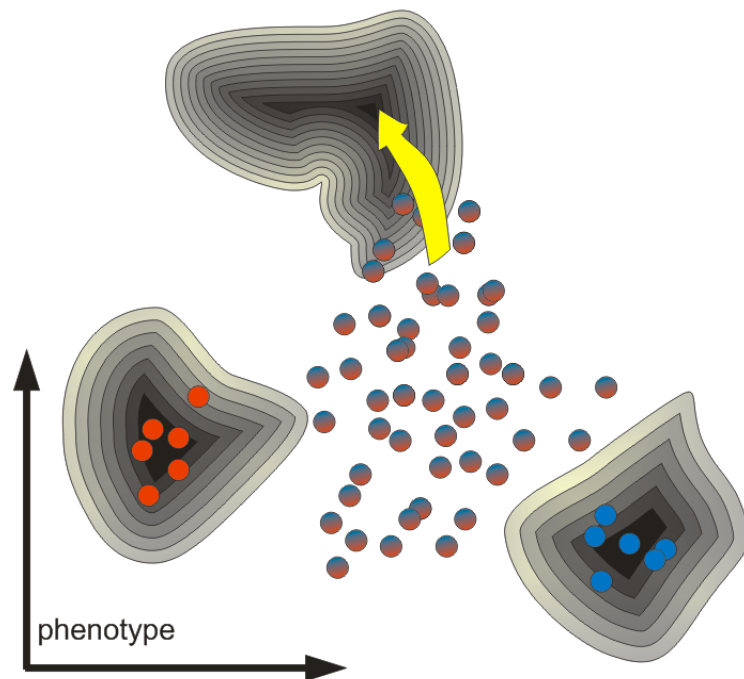


Fig 1: An adaptive landscape with 3 adaptive peaks is shown. The two peaks at the bottom are occupied by a blue and a red species respectively. Colored dots indicate phenotypes of individuals representative for the species. Note that phenotypic variation exists, but is small. Thus, the species are trapped on their adaptive peaks. Phenotypes of hybrids between the two species are represented as two-colored dots. The cloud of two-colored dots thus represents a hybrid swarm. Note the increased phenotypic variation compared to the parent species. Many hybrid phenotypes are located in an adaptive valley, i.e. are unfit. However, a fraction of the hybrid phenotypes is located near to the unoccupied adaptive at the top of the figure. Directional selection (yellow arrow) might act on such hybrid subpopulations. Thus hybridization can release adaptive phenotypic novelty and help to invade adaptive peaks inaccessible to the parent species (adapted from Mallet 2007).

2007). Polyploidization can lead to reproductive isolation, because polyploid species produce aneuploid and infertile offspring when crossing back to the parent species (Ainouche & Jenczewski 2010). Allopolyploid hybrids thus still would suffer from 'minority cytotype disadvantage', because they would produce mostly unfit offspring. Most plants have the capacity for self-fertilization, often are perennial and sometimes can reproduce vegetatively and therefore can overcome minority cytotype disadvantage by establishing local polyploid populations (Mallet 2007). Reproductive isolation due to sexual selection with hybrids showing a mate preference based on certain hybrid specific traits has been suggested (Mavárez *et al.* 2006, Meyer *et al.* 2006). The breakup of adaptive allelic combinations is also avoided, when the sexual reproduction is replaced by an asexual reproduction mode (Kearney 2005), including sperm-dependent parthenogenesis (Lamatsch and Stöck 2009). Finally, extensive chromosomal repatterning in the hybrid species can lead to genetic incompatibilities with its parent species causing its reproductive isolation (Lai *et al.* 2005). All these modes of reproductive isolation can protect beneficial allelic combinations underlying adaptations of an emerging hybrid lineage from genetic swamping through backcrossing to the parent species (Buerkle *et al.* 2000). Still, they are themselves largely independent of any adaptive value of the phenotypic novelty released through the hybridization process. Another situation is given, when the adaptation to a new adaptive peak (Fig 1) itself causes reproductive isolation. Seehausen (2004) characterizes such a scenario as an 'ecological hybrid speciation'. This can for example happen when hybrid species occupy a different spatial or temporal niche (Seehausen 2004). This mode of reproductive isolation is especially relevant for homoploid hybrid speciation (HHS). HHS does not involve a change in ploidy and therefore relies on an alternate source of reproductive isolation. HHS is observed both in plants and animals (Mavárez & Linares 2008, Abbott *et al.* 2010). The best known example for HHS is found in *Helianthus* desert sunflowers. I will use this as a case study to exemplify how the above mentioned processes apply and interact in nature.

The three species *Helianthus anomalus*, *H. deserticola* and *H. paradoxus* resulted from hybridization between *H. annuus* and *H. petiolaris* between 60,000 to 20,000 years ago (Rieseberg 1991, Rieseberg *et al.* 2003a). While the parent species *H. annuus* and *H. petiolaris* occupy mesic habitats (Rieseberg *et al.* 2003a), the hybrid species occupy different desert habitats, which display more extreme environments. Rosenthal *et al.* (2002) identified several transgressive phenotypes in the different hybrid species, which distinguish them from their parent species. Rieseberg *et al.* (2003a) could experimentally verify that artificial hybridization between the two parent species can recreate the

transgressive expression of some of these traits. They also presented evidence that transgressive segregation in the hybrid species was largely due to complementary gene action between parental alleles with opposing effects. Rieseberg *et al.* (2003a) also suggest that many of these transgressive phenotypes (e.g. succulent leaves in *H. anomalus*) are adaptive for the invasion of the different desert habitats. Hybridization induced adaptation to the desert habitats can happen really fast (~ 3 generations; Rosenthal *et al.* 2005). This adaptation also constitutes an ecological and spatial separation, which provides a mean of reproductive isolation. Lai *et al.* (2005) found genomic repatterning to contribute to reproductive isolation between present-day *Helianthus* hybrid species and their parent species. For the incipient hybridization process, the necessary reproductive isolation between the parent and hybrid species can additionally be provided via spatial and ecological separation through hybridization induced adaptive change (Rosenthal *et al.* 2005). The case of the *Helianthus* desert sunflowers illustrates how hybridization can provide the phenotypic novelty and variation necessary to adapt to environmental challenges and to invade new habitats. However, especially for the invasion of a new ecological niche with new environmental conditions, the role of phenotypic plasticity in adaptation becomes relevant. This is discussed in the next paragraph.

Phenotypic plasticity and adaptation

Another factor especially relevant for fast adaptation is the evolution of plastic phenotypes. Phenotypic plasticity is the expression of different phenotypes by a single genotype depending on environmental conditions (DeWitt and Scheiner 2004). For the purpose of this introduction, I do not distinguish developmental plasticity and phenotypic plasticity in labile traits (Crispo 2007). Adaptations to a new adaptive peak often coincide with changes in environmental parameters, which in turn can alter the expression of plastic phenotypes. Therefore phenotypic plasticity can have profound effects on the adaptive process (Moczek *et al.* 2011). Above, I identified two key challenges, which populations can encounter when facing new selective pressures. Firstly, they may be trapped on an adaptive peak by too low standing phenotypic variation (Fig 1). Secondly, even if the phenotypic and underlying genetic variation in principle allows for an adaptive evolutionary response to selection, this response might be too slow. This can result in the extinction of the population before evolutionary rescue might take place (Bell 2012). Phenotypic plasticity might help to overcome both constraints by what has become known as the “Baldwin effect” (Crispo 2007). The Baldwin effect describes a situation where a

(partially) adaptive response is archived through environmentally induced plastic phenotypic change, which places the population near enough to the unoccupied adaptive peak for selection to act on it (similar to but not identical with the situation in Fig 1 and independent of hybridization). Since plastic phenotypic change can happen fast (< 1 generation), the Baldwin effect is thought to facilitate and speed up adaptation. However, phenotypic plasticity can under other conditions also slow down and prevent adaptation (Ghalambor *et al.* 2007). The relevance of phenotypic plasticity for adaptive evolution extends beyond the Baldwin effect. Plastic phenotypes are expressed as a function of environment and genotype (Schlichting & Pigliucci 1998). This implies that genetic variation which is neutral (i.e. not causing a phenotypic difference) in one environment can have a phenotypic effect under different environmental conditions. Such environmental induced variation is referred to “cryptic variation” (Rutherford 2000). Its expression constitutes an increase in heritable phenotypic variation, even though genetic variation remains initially unchanged. Similar to the increase of heritable phenotypic variation due to genetic exchange, the release of cryptic variation therefore can facilitate fast adaptation to a new environment (Fig 1; Rutherford 2000). While in this introduction I focus on the release of cryptic variation due to environmental stimuli, it should be mentioned that genetic change, like e.g. the dysfunction of chaperones can also cause the release of cryptic variance (Rutherford & Lindquist 1998). Related to the concept of cryptic variation is the concept of canalization. Canalization describes the capacity of biological system to express predictable adaptive phenotypes despite of environmental and genetic perturbations (Masel & Siegal 2009). Accordingly, the release of cryptic variance is also referred to as decanalization. Plastic phenotypes are only conditionally expressed and exposed to natural selection depending on environmental context. Especially these genes controlling plastic responses under rarely encountered environmental conditions experience a reduced selection pressure and therefore can accumulate more variability (van Dyken and Wade 2009). This is illustrated in Figure 2, which displays the reaction norms for three different genotypes from one population. A reaction norm (RN) is the representation of a plastic phenotype (y-axis in Fig 2) as a function the environment (x-axis in Fig 2) (Schlichting & Pigliucci 1998). Under other environmental conditions rarely or never experienced by the population (novel high & novel low), natural selection did not have the chance to form the RN, allowing phenotypic variation to accumulate. As a consequence, the release of cryptic variation is especially observed under novel environmental conditions, increasing phenotypic variability under conditions associated with new selection pressures (Ghalambor *et al.* 2007).

Fig 2 also shows, that for the “normal range of environments” usually experienced by the population, the RNs converge due to similar selection pressures (Ghalambor *et al.* 2007). Indeed, Schlichting & Pigluicci (1998; p. 79) propose that the reaction norm itself is a trait under selection. RNs often vary genetically within populations and can

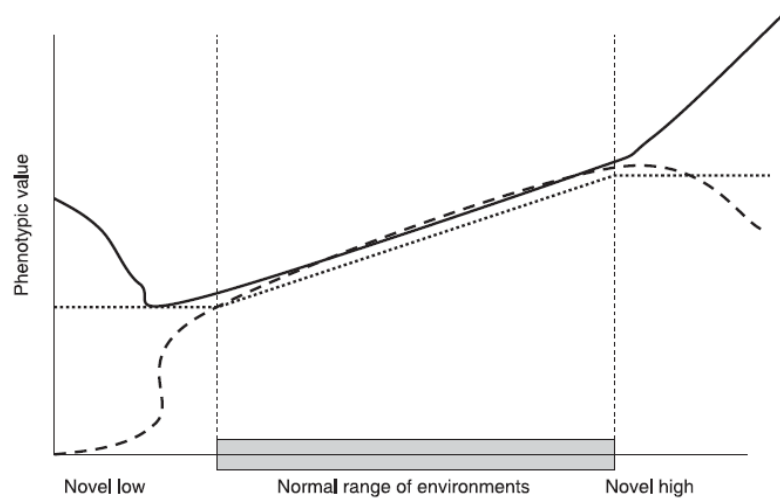


Fig 2: Reaction norms for 3 genotypes from one population. The x-axis depicts environment. The y-axis depicts phenotype. For the ‘normal range of environments’, i.e. the range of environments to which the population is usually exposed, reaction norms converge. This canalization of the plastic phenotype is the result of an adaptive response to natural selection in these environments. Outside the normal range of environments (novel high & novel low on the x-axis), individuals are not exposed to natural selection. This allows for the accumulation of cryptic phenotypic variation which might be expressed once the population becomes exposed to these environmental conditions (taken from Ghalambor *et al.* 2007).

themselves evolve (David *et al.* 2004). The maybe most renowned process of RN evolution is genetic assimilation. Genetic assimilation describes a process during which a plastic phenotype becomes expressed constitutively, i.e. independent of the environment (Figure 3 a; Crispo 2007). Genetic assimilation is often referred to as the Waddington effect, due to a series of pioneering experiments by Conrad Hal Waddington (e.g. Waddington 1953, 1956). In these experiments, Waddington showed that a “new” trait can reach fixation in a population via genetic assimilation in only ~ 23 generations. Thus, adaptation via genetic assimilation can be especially fast. Using a quantitative genetic model, Lande (2009) proposed fast adaptation to an abrupt change in environment via RN evolution should consist of two phases. In the first phase, fast adaptation to the new environment is accomplished via an increase in plasticity (Fig 3 b), i.e. a widening of the RN. Genetic assimilation then occurs in a second phase as a further adaptive modification of the new widened RN. In summary, RN evolution enables fast adaptive change. This is especially relevant for complex plastic phenotypes. Unlike the slow evolution of complex phenotypes through stepwise accumulation of *de novo* mutations, RN evolution can provide complex evolutionary innovations almost instantaneously through the expression of already established plastic phenotypes under new environmental conditions (Moczek *et al.* 2011). Therefore, like evolution through genetic exchange, RN evolution extends the range of realizable phenotypic novelty and the speed by which adaptive change can be

archived.
 Moreover, both processes are interdependent. For certain *Drosophila* species Kellermann *et al.* (2009) suggest that the degree of resistance to desiccation or

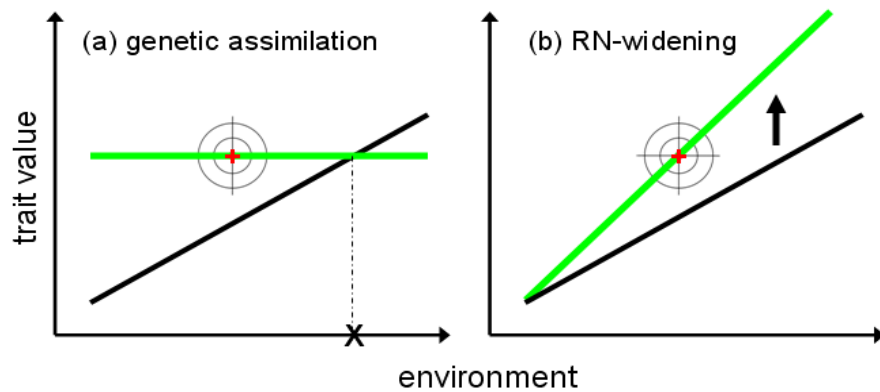


Fig 3: Illustration of two modes of RN evolution: genetic assimilation and RN widening in sensu Lande (2009). Graphs show ancestral (black) and evolved (green) RNs. The x-axis depicts environment. The y-axis depicts phenotype. (a) Genetic assimilation: the ancestral plastic phenotype becomes expressed constitutively, i.e. irrespective of environments. In the case shown, the phenotypic response for an environment marked with an x on the x-axis becomes assimilated. (b) RN widening: the ancestral plastic phenotype becomes even more responsive to the environmental cue, i.e. the RN is widened (see arrow). While both processes are different, they can result in the same phenotypic change for a certain environmental condition. This is indicated by the cross-hairs in both panels.

cold (i.e. survival RNs for temperature or moisture) fails to evolve because a lack of underlying genetic variation. Knowing that hybridization increases genetic variation, hybridization might be thought of as a trigger of RN evolution or, depending on one own's perspective, RN evolution might be seen as an important aspect in hybrid speciation.

Hybridization in *Cottus*

An invasive lineage of European freshwater sculpins (*Cottus*) has been found to colonize the River Rhine and its tributary rivers, a habitat formerly devoid of sculpins, since approximately three decades (Nolte *et al.* 2005). Nolte *et al.* 2005 could show that this invasive lineage originated from hybridization between *Cottus rhenanus* and *Cottus perifretum*. The invasive lineage is thought to have originated only ~ 200 years ago following anthropogenic habitat connection of the parent species habitats, i.e. the watersheds of the rivers Rhine and Scheldt (Fig 4). *C. rhenanus* and *C. perifretum* are restricted in their distribution to small streams. *C. rhenanus* only occurs in small oxygen-rich and summer cold headwater and tributary streams of the rhithral region. In turn, the invasive lineage occupies turbid and summer warm rivers belonging to the potamal region (Nolte *et al.* 2005; Nolte *et al.* 2006). The invasive hybrid lineage is currently invading the River Rhine and its tributaries at a speed of approx. 4-8 km /year (Nolte *et al.* 2005). Where it establishes secondary contact with populations of its parent species *C. rhenanus*, it forms narrow hybrid zones. These hybrid zones localize at the transition from stream to river habitat (Nolte *et al.* 2006). Intrinsic factors do not seem to play major role for reproductive isolation as suggested by the fact that parent species and invasive

lineage readily hybridize in the lab. Also, Stemshorn *et al.* (2011) did not detect any chromosomal rearrangements which could cause reproductive isolation. In summary the situation in *Cottus* is in agreement with an ‘ecological hybrid speciation’ (Seehausen 2004). It therefore constitutes an animal study system analogous to *Helianthus* sunflowers discussed above, in which current open questions regarding hybrid speciation (Abbott *et al.* 2013) can be addressed. Additionally, the habitats occupied by the parent species and the invasive lineage are differentiated in various environmental factors like e.g. temperature (Nolte *et al.* 2006) which are known to control the expression of various plastic phenotypes (Angiletta 2009).

The process of hybrid speciation in *Cottus* therefore lends itself to study the role of hybridization as well as the role of RN evolution in creating adaptive phenotypic novelty. Since the system fits an ecological speciation scenario with ecological and spatial separation of the ancestral and new species, relevant adaptations to the new habitat should have occurred already at the beginning of the speciation process (Nolte & Tautz 2010). These adaptations can be detected by assessing phenotypic differentiation between the new invasive lineage and its extant parent species. Given the recent origin of the invasive lineage, this initial adaptive changes should still be detectable and not be concealed due to the accumulation of secondary evolutionary change (Nolte & Tautz 2010).

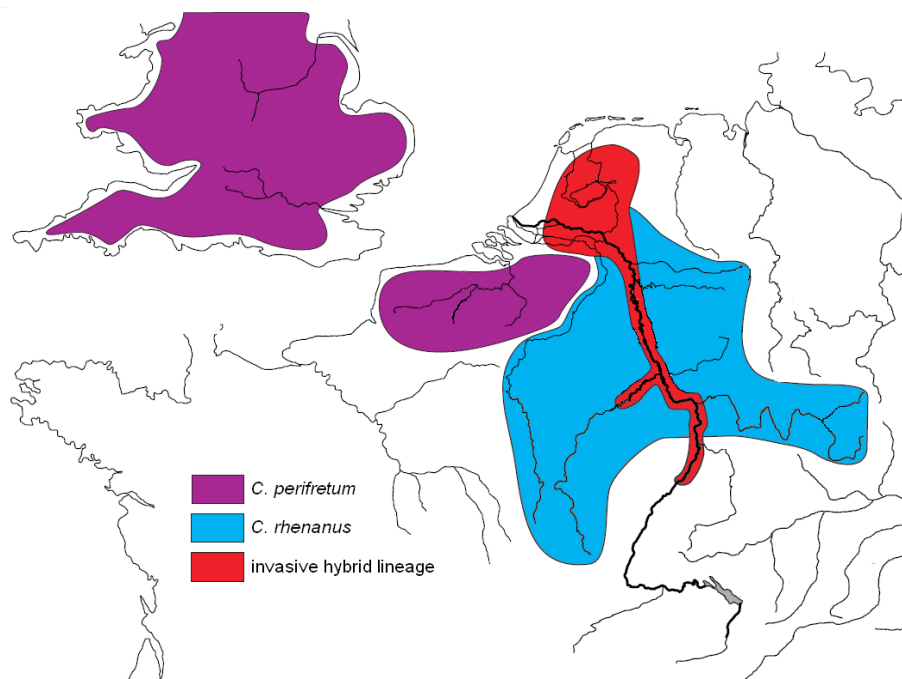


Fig 4: Distribution of *C. rhenanus*, *C. perifretum* and the invasive hybrid lineage in Europe. *C. perifretum* is found in parts of Western Europe and in Great Britain. *C. rhenanus* occupies small streams of the River Rhine watershed. The invasive hybrid lineage is colonizing the River Rhine and its larger potamal tributaries (amended from Nolte *et al.* 2005).

Scope of this thesis

This thesis has two main objectives.

The first main objective is to assess the phenotypic differentiation between the invasive hybrid lineage and its parent species. While not all the phenotypic differentiation is expected to be of adaptive relevance, this analysis sets stage for the prospective identification of the relevant adaptations of the parent species and the invasive hybrid lineages to their respective habitats. The second main objective is to assess if and how the two mechanisms described above, i.e. evolution through genetic exchange and RN evolution, did contribute to the formation of the new invasive phenotype. Up to date, research on the *Cottus* hybrid speciation has been mainly carried out with neutral markers (Nolte *et al.* 2005, 2006, 2009a, Stemshorn *et al.* 2011) except for the assessment of some morphological traits (Nolte *et al.* 2005). In this thesis, this gap is filled and *Cottus* are phenotypically characterized through gene expression profiling. Gene expression profiling is ideally suited for revealing phenotypic differentiation during the course of ecological speciation (Pavey *et al.* 2010). Testing a large number of genes simultaneously also reduces the bias caused by focusing on only few candidate traits. It has therefore the potential to uncover 'hidden phenotypes', including changes in physiological pathways which often constitute relevant adaptations to new environments (Larsen *et al.* 2011).

This thesis is structured in two chapters.

In the first chapter ("Transcriptome changes after genomewide admixture in invasive sculpins (*Cottus*)"), the phenotypic differentiation is assessed by comparing gene expression profiles between populations of the parent species *C. rhenanus* and *C. perifretum* and the invasive hybrid lineage. Furthermore, the role of hybridization in the formation of the invasive phenotype is assessed by analyzing the expression profiles of artificial created F₂ crosses between the parent species. This chapter was published in *Molecular Ecology* (Czypionka *et al.* 2012) and contains relevant technical details about the development of a *Cottus* specific microarray including a newly developed normalization procedure which forms the basis of the work in the second chapter.

The second chapter ("Temperature dependent gene expression and the invasion of new thermal environments in *Cottus*") deals with the phenotypic effects of environmental influences and the role of RN evolution in the formation of the invasive phenotype. Specifically we account for the effect of temperature, a key environmental factor of ecological relevance, on gene expression. Expression profiles are assessed over four

temperatures chosen to reflect critical differences between the habitats occupied by the invasive lineage and its parent species. Thereby the temperature range in which the phenotypic differentiation of the invasive hybrid lineage is most pronounced is determined. Phenotypes displaying plasticity towards temperature are identified. Also, phenotypes for which a change in thermal plasticity between parent species and the invasive lineage is detected are characterized. Based on this, the contribution of proposed modes of RN evolution, i.e. genetic assimilation or RN widening, to the phenotypic differentiation between invasive lineage and ancestral parent species is evaluated.

Chapter 1

Transcriptome changes after genome wide admixture in invasive sculpins (*Cottus*)

Till Czypionka¹, Jie Cheng¹, Alexander Pozhitkov^{1,2}, Arne W. Nolte¹

1) Max Planck Institute for Evolutionary Biology, August-Thienemannstrasse 2, 24306 Plön, Germany.

2) University of Washington, Department of Periodontology, Box 357444, Seattle, WA 98195, USA.

Abstract: Models on hybrid speciation assume that hybridization generates increased phenotypic variance that is utilized to invade new adaptive peaks. We test to what extent this prediction can be traced using gene expression data in the fish species *Cottus perifretum* and *Cottus rhenanus* as well as a natural hybrid lineage referred to as invasive sculpins. In addition, interspecies crosses were used to explore evolutionary trajectories from initial stages to the hybrid lineage. EST libraries were sequenced to design an oligonucleotide microarray that was calibrated for probe specific differences in binding behavior. Levels of gene expression divergence between species correlate with genetic divergence at neutral markers and, accordingly, invasive sculpins were intermediate between the parental species overall. However, the hybrid lineage is distinguished through unique patterns of gene expression that are enriched for biological functions which represent candidates for the fitness properties of invasive sculpins. We compare F₂ crosses with natural invasive sculpins to show that the variance in gene expression decreases in invasives. Moreover, few of the transgressive patterns of gene expression that distinguish invasives can be directly observed in F₂ crosses. This suggests that the invasive transcriptome was subject to secondary changes after admixture. The result is in line with an evolutionary process that reduces maladaptive variance and optimizes the phenotype of an emerging hybrid lineage.

Introduction

There is a recent interest in the origin of new species through the process of hybridization (Mallet 2007, Abbott *et al.* 2010, Nolte & Tautz 2010) and, from a different perspective, the merging of gene pools induced by ecological change (Seehausen *et al.* 1997, Taylor *et al.* 2006) or by the artificial translocation of alien taxa (Laikre *et al.* 2010). However, to date most studies on natural hybridization are relatively uninformed about the functional genetic and ecological consequences of admixture and mostly draw conclusions based on anonymous genetic markers. One alternative to obtain broad insight into functionally relevant traits is given by gene expression profiling. The possibility to simultaneously test many genes reduces the unavoidable bias imposed by selecting phenotypic traits. Therefore, it uncovers “hidden phenotypes” that were not given attention before (Larsen *et al.* 2011). Given a list of differentially expressed and annotated genes, gene ontology (GO) terms permit inference of functions that have evolved (Conesa *et al.* 2005). Although GO term analysis cannot account for evolutionarily young but important orphan genes (Domazet-Lošo & Tautz, 2010), there is a growing number of studies that employ gene expression analysis to study adaptive evolutionary change. For example, a series of studies on whitefish (*Coregonus clupeaformis*) have shown that candidate adaptive traits inferred from gene expression data can be confirmed using independent physiological experiments and convey fitness advantages in a known ecological context (Bernatchez *et al.* 2010). Although it has been documented that gene expression levels differ in a very characteristic fashion between tissues and under different ecological conditions (Xia 2007, Cheviron *et al.* 2008), it is reassuring, that a considerable fraction of candidate adaptive transcriptomic patterns can be found in different tissues independently of the environment (Nolte *et al.* 2009b).

Gene expression data have been recognized as a rich source of information to study fish hybrids (Mavárez *et al.* 2009, Normandeau *et al.* 2009, Renaut *et al.* 2009) and permit testing of predictions related to the process of hybrid speciation. The evolutionary outcome of hybridization is critically determined by the genetic interactions which result from the fusion of different gene pools (Landry *et al.* 2007). Notably this includes both effects that can reduce or raise fitness. For example, Michalak and Noor (2003) have successfully employed a microarray screen to identify genes that show disrupted patterns of expression in *Drosophila* hybrids and many of these candidate genes indeed contribute to postzygotic isolation (Michalak and Noor 2004). A fitness enhancing effect of initial hybridization is given by heterosis (Charlesworth and Willis 2009) or when transgressive

segregation generates phenotypes that lie outside the range observed in the parental species (Rieseberg et al 1999, 2003a). This increased phenotypic variance might facilitate the invasion of a new adaptive peak not available to the parents (Mallet 2007, Barton 2001). However, only a limited number of the admixed genotypes will have a high fitness and an admixed population goes through a lag phase in which selection optimizes the composition and overall fitness of the admixed gene pool (Barton 2001). This evolutionary change during the initial phase is what distinguishes this mode of speciation most from other modes of speciation (Nolte & Tautz 2010). We are studying a hybrid lineage of invasive sculpins (*Cottus*, Cottidae, Teleostei) that lends itself to identify the gene expression changes associated with the ecological transition away from the habitats occupied by the parental species. Invasive sculpins emerged in the wake of man-made environmental changes within the past 200 years that have increased the connectivity of the central European rivers Rhine and Scheldt and massively altered the ecological conditions in the lower parts of these drainages. These perturbations have fostered hybridization of two previously isolated fish species (*Cottus rhenanus* and *Cottus perifretum*) from smaller tributaries to the River Rhine and River Scheldt systems (Nolte et al. 2005, Stemshorn et al. 2011). While the parental taxa have been estimated to be separated for 2 million years (Englbrecht et al. 2000, but see Volckaert et al. 2002), the hybrid sculpins are possibly less than 200 generations old and were first detected at a massive scale in the Netherlands after 1980. They are referred to as “invasive”, because they invaded new habitats that are not occupied by the parental species (Nolte et al. 2005). Population genetic studies indicate that this is due to a raised fitness of invasive sculpins in large river habitats (Nolte et al. 2006, Stemshorn et al. 2011). A recent systematic screen for genetic incompatibilities between the ancestral species has revealed no evidence that effective intrinsic barriers to reproduction exist that explain the population structure of *Cottus* in the River Rhine basin (Cheng et al. in prep). Taken together, these findings suggest that ecological selection pressures constitute an essential component to explain the evolution of invasive sculpins and stress the need to assess the adaptive relevance of genetic changes for the adaptation of hybrids (Abbott et al. 2010).

Transcriptome analyses are subject to a number of biases that, while not erasing biologically relevant signal, can distort inferred patterns of gene expression massively (Jeukens et al. 2010). For oligonucleotide microarrays this concerns non responsive or sticky probes (Pozhitkov et al. 2006) that do not have a linear relationship between signal intensity and target transcript concentration (Pozhitkov et al. in prep). We analyzed gene expression data using custom oligonucleotide microarrays that were calibrated for the

samples analyzed in this study. Gene expression divergence was compared among populations representing the parental species *C. perifretum* and *C. rhenanus* as well as their natural hybrids, the invasive sculpins. These comparisons identified gene functions that are correlated with the invasion of new habitats but also to assess functional differentiation at the within and between species level. Since the rise of invasive sculpins is very recent, the *Cottus* system is suitable to recreate initial steps that have led to the formation of invasive sculpins and to test hypotheses about the origin of novelty through hybridization. We compare F₂ crosses with natural invasive sculpins to test to what extent the features that distinguish invasives can be explained as a direct consequence of hybridization or, alternatively, have emerged independently in a secondary evolutionary process.

Methods

Study populations and samples

The populations used in this study were also described and used by Cheng *et al.* (in prep). Here, we compared gene expression among parental species (*Cottus rhenanus* and *Cottus perifretum*), F₂ crosses between these and invasive sculpins to infer patterns of transcriptome evolution. We analyzed two independent biological replicate crosses to account for within and between species variance. Previous studies confirmed that the populations used here (Table 1) are representative of the ancestral species that gave rise to invasive sculpins (Freyhof *et al.* 2005; Nolte *et al.* 2005; Stemshorn *et al.* 2011, Knapen *et al.* 2003). *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 Witte Nete (WN) and Laarse Beek (LB) (F_{ST} = 0.368, Knapen *et al.* 2003).

Adult fishes of the study populations were transferred to laboratory aquaria in climate chambers and fed with frozen and live insect larvae and brine shrimp. The room temperature and light regime mimicked the conditions in central Europe. During the winter the temperature was lowered to 4 °C for at least one month. Spawning occurred readily in artificial shelters partially buried in sand when water temperatures were raised to 8-10°C. We have generated two independent groups of F₂ crosses from interspecies crosses between pure *C. perifretum* and *C. rhenanus*: Broel(♂) × WN(♀) and WN(♂) × Broel(♀) as well as Naaf(♂) × LB(♀) and LB(♂) × Naaf(♀). Larvae were fed initially using live *Artemia* nauplii, and later with frozen insect larvae until a majority had reached a length of

approximately 3 cm (September to October 2010). Water temperatures were raised to a maximum of 17°C during the summer and fish were kept at this temperature for several weeks before they were sacrificed. Only individuals that weighed 0.4 to 0.5 g, appeared to be normally developed and healthy were selected for analysis of gene expression. Fishes were anaesthetized with CO₂ and killed by immersing the whole fish in RNA_{later} until RNA extraction. We were concerned that analyses based on single families from aquarium stocks yield biased patterns of gene expression. Thus, we combined unrelated fishes from our aquaria and ones that have grown in natural habitats where possible (Table 1). For this purpose wild fish representing invasive *Cottus* as well as the two populations of *C. rhenanus* were collected in the field and adapted to our aquaria for one week before they were sacrificed. Our study is imbalanced with respect to controlling for gene expression changes due to long term aquarium care effects for ancestral *C. perifretum* as these fish are protected in their native range and could not be resampled. Likewise, synthetic F₂ crosses cannot be obtained from natural habitats.

Table 1: Origin and affinity of *Cottus* samples used for transcriptome analysis. Two populations were chosen to represent the parental species *C. rhenanus* (Broel and Naaf) and *C. perifretum* (Laarse Beek and Witte Nete). A hybrid lineage that resulted from natural admixture between the two species is represented by the invasive *Cottus* from the Sieg population. Initial stages of hybridization were recreated in the lab as F₂ crosses between the parental species. The samples column shows that each group of F₂ crosses comprises individuals (number given) from two independent families (alternative cross directions) and that all experimental groups contained individuals from unrelated families reared in the lab (LR = lab reared) or outbred fishes collected in nature (WC = wild caught).

species	population	samples	origin of populations
<i>C. rhenanus</i>	Broel	2 x WC, 6 x LR	Broel (GIS: 50°50'N 7°22'E; River Sieg system, North Rhine-Westphalia, Germany)
	Naaf	2 x LR, 6 x WC	Naaf (GIS: 50°52'N 7°16'E River Agger system, southeast of Hausdorp, North Rhine-Westphalia, Germany)
<i>C. perifretum</i>	Laarse Beek	4 x LR 4 x LR	Laarse Beek (GIS: 51°17'N 5°04'E; River Scheldt system, Flanders, Belgium)
	Witte Nete	4 x LR 4 x LR	Witte Nete (GIS: 53°14'N 5°04'E; River Scheldt system, Flanders, Belgium)
invasive <i>Cottus</i>	Sieg	2 x LR, 6 x WC	Sieg (GIS: 50°48'N 7°9'E; River Rhine system, North Rhine Westphalia, Germany)
F ₂ crosses	Broel X Witte Nete	4 x LR 4 x LR	Synthetic lab crosses between Broel and WN
F ₂ crosses	Naaf X Laarse Beek	4 x LR 4 x LR	Synthetic lab crosses between Naaf and LB

Development of a microarray

In order to develop a custom oligonucleotide microarray, we sequenced transcriptome libraries of 22 individuals representing all populations mentioned above (Table 1) using a Roche GS-FLX DNA Sequencer. 27243 EST contigs were reconstructed from the transcriptome reads using the Roche analysis software Newbler with default criteria and were available together with 87320 unassembled reads to design an Agilent oligonucleotide microarray which requires consideration of the read direction. We performed a BlastX (Altschul *et al.* 1990) homology search of *Cottus* sequences against the stickleback protein sequences (Version 5.17; www.ensembl.org/) and were able to annotate 11430 sequences (e-value < 10^{-3}). In a second step, a custom database was built for protein coding nucleotide sequences annotated to the Actinopterygii from the NCBI Taxonomy Browser (<http://www.ncbi.nlm.nih.gov/Taxonomy/>) (as of 7/6/2010). 713 additional sequences could be annotated by blasting against this database (e-value < 10^{-3}). An alternative criterion to identify the directionality of target sequences is given when the ends of the sequences contain a poly-A tail, which agreed in 86% of the cases with the inference from blast results. All reads for which the directionality could be inferred from one of the two methods were submitted to the Agilent e-array software (<https://earray.chem.agilent.com/earray/>) to design 60-mer oligonucleotide probes that are specific to the target (mRNA-) sequences. A total of 13329 different probes and 1879 replicates of these were synthesized to fill the 15208 features available on an Agilent 15k Expression microarray. We refer to probes and the levels of transcript abundance detected for these as “gene” and “gene expression” throughout this study, but acknowledge that some probes may target alternative splice variants, copies or fragments of the same gene.

Raw signal analysis and quality filtering

Total RNA was extracted from whole fishes using a Trizol protocol and labeled using the Agilent Low Input Quick Amp Labeling kit. The microarray was loaded with labeled cRNA as described in the Agilent protocol for One-Color Microarray-Based Gene Expression Analysis (version 6.5, May 2010). Care was taken to randomize the distribution of the 8 individuals representing different populations (see table 1) between different slides. Microarrays were scanned with an Agilent DNA Microarray Scanner type C. Feature extraction was carried out using the Agilent Feature Extraction software (version 10.7.3.1) and the “gprocessed signal” was used as raw signal for gene expression analysis. Signals flagged for bad quality or saturated intensity were treated as missing data. Probes with

missing data for more than two individuals per population were removed from the dataset.

An implicit assumption in normalization procedures and microarray analysis is that there exists a one-to-one relationship between signal intensity and target transcript abundance. However, the binding behavior of surface-bound oligonucleotides varies greatly (Pozhitkov *et al.* 2006) and a realistic assessment of gene expression changes ideally requires consideration of the binding behavior of probes (Pozhitkov *et al.* in preparation). Secondly, our inference of the antisense probe sequence (required to produce the microarray) may have been wrong. For these two reasons, we have validated and calibrated our custom microarray for a range of concentrations that is appropriate for our experiment. The binding behavior between a probe and target transcript depends on the template concentration, but also on the affinity of the probe and can be described using a dose response curve that describes the increase of signal intensity as a function of template concentration (Pozhitkov *et al.* in preparation). A calibration pool containing equimolar amounts of total RNA from representatives of both parental species and the invasive hybrid lineage was prepared and a calibration series of seven microarrays were loaded with 0.125, 0.25, 0.5, 1, 2, 4, 8 times the amount of recommended cRNA (600 ng) to detect the hybridization signal. The binding behavior of the individual probes was assessed using custom R code (R Development Core Team 2010). For the purpose of this study we were interested in identifying probes with a positive and approximately linear dose response and to exclude probes that did not detect a change in concentration or for which the signal approaches saturation. A linear function

$$(1) SI_{\text{calibration}} = m_{\text{probe}} * c_{\text{calibration}} + b_{\text{probe}}$$

was fitted for each probe, where $SI_{\text{calibration}}$ is the raw signal and $c_{\text{calibration}}$ is the amount of calibration pool [ng] for which the $SI_{\text{calibration}}$ was obtained. The slope m_{probe} and the intercept b_{probe} derived from this calibration step were used in the normalization of raw-signal measurements (see below). Probes with a non linear binding behavior (R^2 of the linear fit < 0.95) and the most responsive probes with the 2.5% highest dose response slopes were excluded. Further, probes with slopes smaller than 0.3 were not used due to their unfavorable noise-to-signal ratios (see Figure 1). The noise-to-signal ratio was defined as twice the average of the residual absolute values divided by the overall change in signal.

Data normalization

In order to make changes of signal intensity comparable among target transcripts, raw signal intensities were normalized between probes to correct for different binding

behaviors. For this purpose raw signal intensities were transformed such that they have a common signal to concentration change relationship in two steps.

First, C_{detected} (in ng cRNA) corresponding to the raw signal intensities obtained in an experiment (SI_{detected}) were determined based on the coefficients from the dose response functions derived in the calibration step (Equation 1).

$$(2) C_{\text{detected}} = (SI_{\text{detected}} - b_{\text{probe}}) / m_{\text{probe}}$$

This step already enables unbiased calculation of the gene expression fold changes from the same array. Secondly, based on C_{detected} we calculated signal intensities expected ($SI_{\text{corrected}}$) under a uniform linear dose response behavior with slope one and intercept 0 for all probes.

$$(3) SI_{\text{corrected}} = C_{\text{detected}} * 1/(600 \text{ ng}) * SI_{600\text{ng}}$$

where $SI_{600\text{ng}}$ is the signal intensity expected at 600 ng calibration pool ($SI_{600\text{ng}} = m_{\text{probe}} * 600 \text{ ng} + b_{\text{probe}}$). Thus e.g. for $C_{\text{detected}} = 300 \text{ ng}$, $SI_{\text{corrected}}$ becomes exactly half $SI_{600\text{ng}}$. This last step reintroduces signal intensity as a proxy for transcript abundance to the normalized expression values, which allowed us to perform the between array normalization by dividing all normalized expression signal intensities of an array by their 75 percentile as recommended in the *One-Color Microarray-Based Gene Expression Analysis. Low Input Quick Amp Labelling. Protocol* (Version 6.0. December 2009 Agilent Technologies). Negative normalized signal intensity values were set to zero but not removed from the dataset. The normalization procedure was performed with custom R code.

Patterns of gene expression

Our experimental design introduces genotype and environmentally induced effects which are controlled for by 1) using a covariate in the analysis and 2) by visually checking resulting candidate genes of special interest for remnant effects associated with the origin of the samples. Variance in gene expression was explored using a principal component analysis (PCA) in R in order to identify major axis of variation in gene expression and to remove sources of variation that are not of interest in this study (Pickrell *et al.* 2010). The R-package R/maanova (Wu 2010) was used to perform a two stage ANOVA to test for differentiation in gene expression between “species” (*C. rhenanus*, *C. perfretum* and the invasive hybrid lineage) and “populations” (Broel, Naaf, Laarse Beek, Witte Nete, Sieg). Differential gene expression was assessed using fixed effect models with the factor species or population respectively. The scores of PC 1 were used as a cofactor in the

models to remove effects obscuring the signal suggestive of fixed between species differences. Significance of the pairwise comparisons was determined using a contrasting function implemented in R/maanova ("t-test"-option with 2000 permutations, Fs test option) with FDR (False Discovery Rate) estimation as described by Storey *et al.* (2002).

The invasive sculpins in our analysis comprised six individuals sampled from nature and two individuals raised in our experimental aquaria. Genes that were transgressively over expressed in the invasive turned out to be of major interest in the functional analysis (see below). Therefore, we examined for these genes if transgressive expression patterns were due to a plastic response to environmental conditions or are a ubiquitous feature of the invasive lineage. According to scenario 1 (plastic response) individuals derived from nature should show transgressive levels, while lab raised fish should not. Scenario 2 (static expression level) posits that no general difference should be observed in expression between individuals derived from nature or individuals raised in the lab. In order to remove the effects of PC 1 from gene expression values, a linear model was fitted with PC 1 being the only explanatory variate and the residuals from each sample were used as a surrogate of its PC 1 independent expression value. These surrogates were plotted and the genes transgressively overexpressed in the invasive were examined by eye into the categories "genes showing a plastic response" (scenario 1) and "genes with static expression level" (scenario 2) and reanalyzed separately.

To test whether hybridization affects the variance of gene expression we calculated the variance in gene expression on a per gene basis for all experimental groups. The variance describes how far data points are spread around their mean on an absolute basis and thus increases with expression value. To disentangle the differences in variance between the tested groups from the effect of the mean expression value on variance we included the average expression value for each gene and group tested as a cofactor in the analysis. First, we fitted the same model which already was used for species level analysis with the R/maanova package in R. The variance in gene expression was calculated individually per gene for each group from the residuals of the fitted model. The mean expression values per group were calculated for each gene excluding the influence of the first principal component from the same model. To prevent that outliers blur the general patterns in the downstream analysis, we excluded the genes responsible for the 5% highest variances and/or mean expression values for every group. A linear model was fitted in R for the remaining genes with the explanatory variable being "group" as a categorical variable with the levels (*C.rhenanus* and *C.perifretum*, invasives and F₂crosses) and "mean expression value excluding PC 1" of the respective group as a

covariate. The response variable was the variance in gene expression of the genes for the respective groups.

In this study genes were considered to show a transgressive pattern of gene expression when gene expression levels were significantly different from both parental species (*C. rhenanus*, *C. perifretum*) and when the levels of gene expression were not intermediate between the parental species. We distinguished between genes that were transgressively up and down regulated relative to the parental populations. Transgressivity was tested for F_2 crosses as well as for invasive sculpins in order to analyze to what extent the patterns in the artificial hybrids and natural hybrids agree. To test whether the sets of genes transgressively expressed in the invasives and in the F_2 crosses overlapped more than expected by chance, we performed a one-tailed Fisher's exact test using the `phyper` function in R.

Annotation of genes and GO term enrichment analysis

For the purpose of the GO term enrichment analysis, we have lowered the stringency of the statistical significance thresholds ($p < 0.05$ and $q < 0.05$) to use larger sets of genes to test for the overrepresentation of functional groups based on biological effects. EST contigs that were targeted by probes on our microarray were annotated using the software Blast2GO (Conesa *et al.* 2005) using the default parameters. A homology search was performed with BlastX against the NCBI nr database (as of 2/17/2010) and hits with an e-value $< 10^{-6}$ were used to extract GO annotations from several repositories (Conesa *et al.* 2005). We used the one-tailed Fisher's exact test with FDR correction implemented in Blast2GO to test if any GO terms were significantly overrepresented in a candidate set of genes compared to all genes for which we had gene expression data after removing bad quality probes.

Results

Microarray calibration and normalization

Fitting of a linear model to the raw signal intensities obtained in the calibration experiment revealed that 11630 out of the 13329 probes on the custom microarray showed an approximately linear dose response behavior. The slopes of the linear dose response relationship with an $R^2 > 0.95$ ranged from 0.001 to 48.5. We excluded those with a slope smaller than 0.3 below which unfavorable noise-to signal ratios increased (Figure 1). Probes with a dose response slope of above 20 were also removed thus excluding the

2.5% most responsive probes from the analysis. As a last step probes with more than two (out of eight) individuals per population having missing data were removed. 7479 out of the 13329 probes could be analyzed for differentiation in gene expression (Supplementary Table 1).

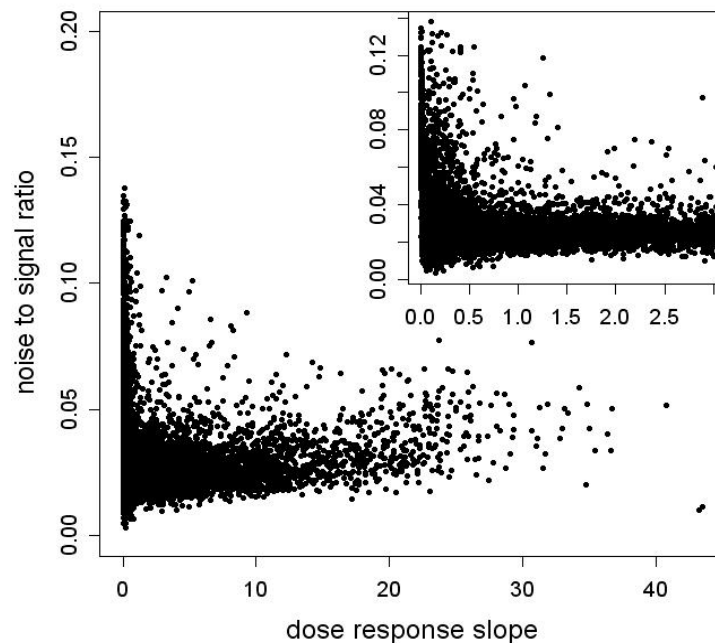


Fig 1: Relation between slopes of the linear dose response functions (x-axis) to the noise to signal ratio (y-axis) for all probes with a linear fit ($R^2 > 0.95$). Dots represent individual probes. Dose response slopes vary greatly and slopes that exceed a value of 1 are common. The highest noise to signal ratios are found mostly at probes with a slope of less than ~ 0.3 (enlarged in small graph).

Differentiation in gene expression

A principal component analysis using the individuals as variables and expression data (4622 probes without missing data) as observations was used to explore major axes of variation in gene expression. PC 1 explained $\sim 28\%$ of the total variance while PC 2 explained $\sim 10\%$ of the total variance. Individuals belonging to all experimental groups overlap greatly along axis one (Figure 2). The ancestral species *C. perflretum* and *C. rhenanus* separate along axis two and F_2 crosses as well as invasive *Cottus* assume an intermediate position. Figure 2 suggests that the environment in which individuals have been raised contributes to PC 1, which is supported when comparing the PC 1 scores of the lab reared individuals with those from the wild (Welch's test p-value < 0.001). However, the natural environment did not affect artificial F_2 crosses and *C. perflretum* but

these groups still vary along PC axis 1. Thus, it is likely that PC 1 is affected by unknown factors. Taken together, these factors obscure the differences between species. We thus use PC 1 scores as covariates to test for differences between species.

Differential expression between species and populations

The number of genes differentially expressed between *C. rhenanus* and *C. perifretum* was assessed for the species level and compared with invasive sculpins (Figure 3a, Supplementary Table 1) but also at the population level (Figure 3b). Both analyses revealed that the number of genes differentially expressed between the parental species

were in the order of two fold higher than the differentiation with the invasive hybrid lineage. Moreover, the comparison at the population level demonstrated that the within species differentiation (448 and 664 differentially expressed genes) is lower than the between species comparison (2100 and 2820 differentially expressed genes) and also lower than any comparisons with the invasive *Cottus* (1114 to 1803 differentially expressed genes) (Figure 3b). The relative proportions of differentially expressed genes reflect the current classification of species and

population and suggest that the hybrid lineage shows intermediate levels of differentiation to both of its parental taxa. This pattern corresponds with the differentiation observed along PC axis two.

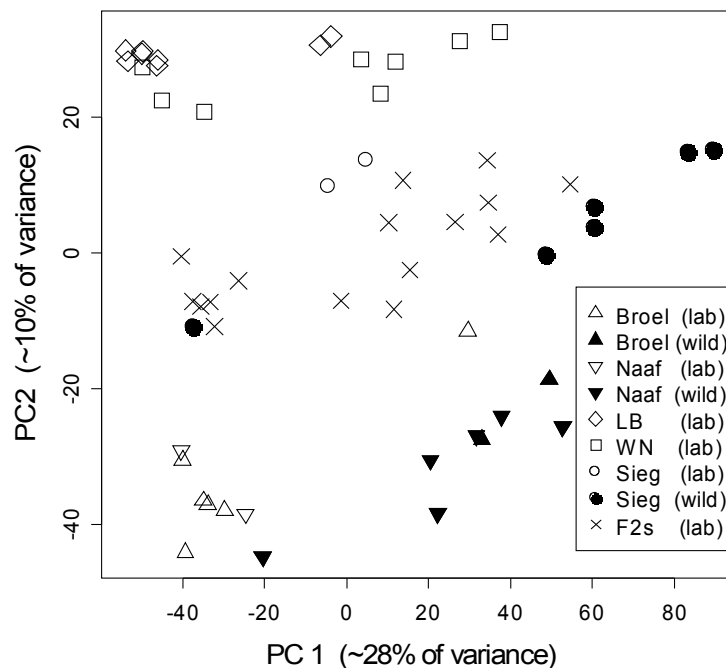


Fig 2: Plot of individual scores for principal components of gene expression. *C. perifretum* is represented by the populations WN and LB and *C. rhenanus* is represented by the populations Naaf and Broel. Symbols are filled or empty depending on whether individuals grew up in the laboratory or in natural environments (legend). The environment contributes to the variance within groups along PC axis 1 (x-axis) while PC axis 2 (y-axis) separates species. Artificial F₂ crosses as well as invasive sculpins take an intermediate position between the parental species along PC axis 2.

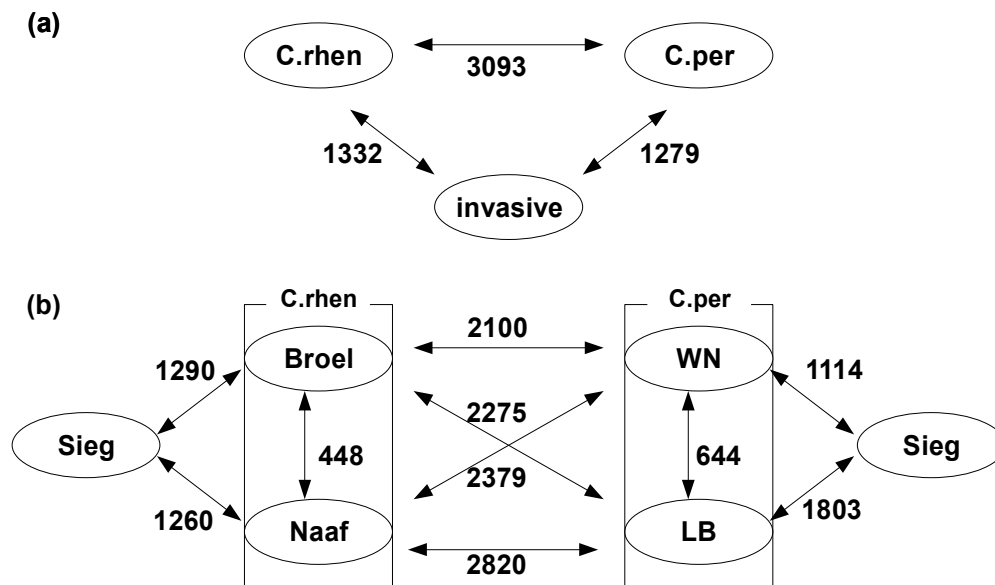


Fig 3: Number of differentially expressed genes ($p < 0.01$; $q < 0.01$) between species (a), and populations (b). Broel and Naaf = *C. rhenanus*; Witte Nete (WN) and Laarse Beek (LB) = *C. perifretum*; A single population from the River Sieg represents the invasive hybrid lineage and is used in all comparisons. Numbers near arrows indicate differentially expressed genes.

Variance in gene expression

A linear model was used to assess differences in the variance in gene expression. This model tested for the relationship between the categorical variable “group” (levels: *C. perifretum*, *C. rhenanus*, invasives, F_2 crosses) and the covariate “expressionvalue” (mean level of gene expression per gene and group after removal of PC 1) versus variance in gene expression. *C. rhenanus* was used as the base category from which the other terms deviated. The model, is highly significant ($p < 2 * 10^{-16}$) with an adjusted R^2 -value of 0.4055. The estimated parameters indicate a significant effect of the “expressionvalue” term ($p < 2 * 10^{-16}$; estimate = 0.1150252) and indicate significant differences between the different levels of the “group” term. The estimates for the group term reveal that the variance in gene expression is lowest in the parental species, being slightly lower for *C. perifretum* (intercept + estimate = -0.0303738) than for *C. rhenanus* (estimate = -0.0226103). The highest variance in gene expression was found for the F_2 crosses (intercept + estimate = 0.0158769), while for the invasive lineage variance in gene expression (intercept + estimate = 0.0024742) was intermediate between parental species and F_2 crosses. The p-values indicate that all estimates are significantly different from *C. rhenanus*. Using the invasive lineage as the base of the model, significant differences in gene expression variance are detected from both parental species ($p < 2 * 10^{-16}$) and from the F_2 crosses ($p = 1.12 * 10^{-10}$).

Transgressivity

The number of genes that were transgressively expressed in the invasive lineage and the F₂ crosses relative to the parental species was assessed separately for genes that have an up regulated or down regulated pattern (Supplementary Table 1). A balanced proportion of 237 and 264 genes were up and down regulated in F₂ crosses while an excess of 580 genes was transgressively up regulated in invasive sculpins as opposed to 120 genes that were transgressively down regulated in that group. When the intersecting sets of transgressively expressed genes among F₂ crosses and invasive sculpins were identified 18 and 15 genes were found to be shared between up and down regulated genes respectively. Interestingly, transgressively down regulated genes are shared more often between F₂ crosses and invasive sculpins than expected by chance ($p = 4.23 * 10^{-6}$) while sets of transgressively up regulated genes do not have more genes in common than expected by chance ($p = 0.48$).

Functional annotation and GO term enrichment of candidate gene sets

We were able to annotate 4942 of the 7479 features entering analysis (Supplementary Table 1) based on database searches (distant homologies) with model organisms and this set of features served as a reference data when testing for the overrepresentation of GO terms. The largest fraction of 3952 features was found to be differentially expressed between populations representing the ancestral species *C. rhenanus* and *C. perifretum*.

Table 2: GO term enrichment of sets of differentially expressed genes. The column “gene set” specifies pairwise comparisons and the nature of transgressive patterns while the numbers of genes with GO-annotation are given in column two. Significant enrichment (cutoff: FDR-corrected $p < 0.01$) of specific GO-terms of these sets compared to all genes entering analysis is reported. GO-terms are indexed as pertaining to the categories cellular component (C), molecular function (F) or biological process (P). There is no enrichment for genes differentially expressed between parental species. However, gene sets in which the invasive lineage has a higher expression than both parental species are enriched for several GO terms. This holds for genes with a with a static expression level and for genes showing a plastic response. In stark contrast, genes transgressively under expressed in invasives or genes in which the invasives differ from only one parental species do not show signs for enrichment. Note that differential expression was called at $p < 0.05$ and $q < 0.05$ for the purpose of this analysis.

gene set	features /genes (with GO annotation)	enriched GO-terms	FDR corrected p - value
differentially expressed between parental species <i>C.rhenanus</i> and <i>C.perifretum</i>	3952 (2587)	none	
differentially expressed between invasive sculpins and both parental species	950 (611)	proteinaceous extracellular matrix (C) embryonic development (P) nucleic acid binding (F)	1.5 E-4 7.9 E-3 7.9 E-3
Transgressively over expressed in invasive sculpins: all genes	580 (387)	proteinaceous extracellular matrix (C) embryonic development (P) anatomical structure morphogenesis (P) transcription factor activity (F) nucleus(C) cellular component organization (P) DNA metabolic process (P) actin binding (F) motor activity (F) cell cycle (P)	4.0 E-7 3.0 E-6 7.2 E-5 3.9 E-3 4.1 E-3 4.1 E-3 4.1 E-3 4.1 E-3 4.5 E-3 4.6 E-3
Transgressively over expressed in invasive sculpins: genes with a static expression level	388 (265)	embryonic development (P) nucleus(C) transcription factor activity (F) cell cycle (P) DNA metabolic process (P) anatomical structure morphogenesis (P)	3.5 E-5 2.5 E-3 2.7 E-3 3.8 E-3 6.6 E-3 8.5 E-3
Transgressively over expressed in invasive sculpins: genes showing a plastic response	192 (122)	proteinaceous extracellular matrix (C) structural molecule activity (F) calcium ion binding (F) motor activity (F) actin binding (F)	6.8 E-9 1.6 E-3 2.3 E-3 3.9 E-3 7.9 E-3
Transgressively under expressed in invasive sculpins	120 (73)	none	
Expression in invasive sculpins matches <i>C.rhenanus</i> , but not <i>C.perifretum</i>	1203 (794)	none	
Expression in invasive sculpins matches <i>C.perifretum</i> , but not <i>C.rhenanus</i>	1229 (782)	none	

However, no enrichment of GO terms was detected for these genes. There are considerable numbers of features in which the invasive sculpins retain expression profiles from one parental species but not the other one (1203 and 1229 genes respectively) but none of these sets revealed overrepresented GO terms. The analysis revealed 950 genes in which the invasive sculpins differed significantly from both ancestral species and for this set we detected an overrepresentation of the GO-terms “proteinaceous extracellular matrix”, “embryonic development”, and “nucleic acid binding” (Table 2). For the genes, transgressively up regulated in the invasives, the same GO-terms were found to be enriched together with the additional terms “anatomical structure morphogenesis”, “transcription factor activity”, “nucleus”, “cellular component organization”, “DNA metabolic process”, “actin binding”, “motor activity” and “cell cycle” (Table 2). In contrast the genes transgressively down regulated in the invasive were not enriched in any GO term. Analogous analyses were carried out for each population and cross separately which corroborated the GO-terms and trends found with the inference based on pooled populations (not shown). The samples from the invasive lineage consisted in their majority (6 out of 8 individuals) from captures from the wild but contained 2 lab-raised fish. We visually inspected levels of gene expression (after correcting for the variance explained by PC 1) for all transgressively over expressed genes and separated them into two groups. The first group contained 388 genes for which expression values of individuals from the lab reared fish fell into the distribution of values of the wild caught fish. The second group consisted of 192 genes where the significant transgressive pattern was caused by the fishes from the wild, and where our lab reared fish showed a trend to be less differentiated from the parental populations. GO-term enrichment analysis was performed for these two sets (see Table 2). For the “genes with a static expression level” the GO-terms “embryonic development”, “nucleus”, “transcription factor activity”, “cell cycle”, “DNA metabolic process” and “anatomical structure morphogenesis” were found to be enriched. This recovers GO-terms already found during the GO-term analysis of all genes which are transgressively over expressed in invasive sculpins, though the FDR corrected p-values are slightly higher. This situation was mirrored for the “genes with a plastic response” for which the GO-terms “proteinaceous extracellular matrix”, “structural molecule activity”, “calcium ion binding”, “motor activity” and “actin binding” were enriched. With the exception of “calcium ion binding”, these terms were already found to be enriched in the set of all genes which are transgressively over expressed in the invasive sculpins.

Discussion

The theoretical appeal of hybridization lies in the fact that it may facilitate more complex evolutionary transitions than could be attained through the sequential accumulation of new mutations. A possible driver of such changes is given by the common process of transgressive segregation (Rieseberg *et al.* 2003a, Stelkens and Seehausen 2009, Stelkens *et al.* 2009), whereby recombined alleles in admixed individuals interact to create phenotypic values that lie outside the range observed in the ancestral species. Notably, transgressive patterns of gene expression are common in fish hybrids (Mavárez *et al.* 2009, Bougas *et al.* 2010, Normandeau *et al.* 2009, Renaut *et al.* 2009). This makes transcriptome analysis particularly useful to document the effects of hybridization but also poses questions related to the evolutionary significance of such patterns. In this study we document the emergence of novel patterns of gene expression in invasive *Cottus* from the river Rhine basin that is consistent with an evolutionary optimization process. Our results indicate that the majority of the distinguishing patterns of gene expression of invasive sculpins are evolutionarily modified after the initial admixture event rather than being an immediate feature of early generation hybrids. As for the technical aspects, we found nonlinear binding behavior and a wide range of dose-response slopes for the microarray probes that, unless taken into account, bias the inference of gene expression.

Microarray calibration

The microarray analysis described here followed established standard protocols with respect to the preparation of template, the hybridization and signal detection procedures. On the other hand it was complemented by performing a calibration of our custom array that assesses the relationship of signal intensity and transcript abundance to identify probes that produced a reliable signal. Since our calibration pool originated from the species analyzed here, the calibration is specifically tuned to the range of transcript concentrations encountered in this experiment. Even after excluding probes that were not responsive or did not have a linear dose response we observed a broad range of dose response behaviors that vary more than sixty-fold (Figure 1) although the dilution series was identical for all target sequences. It is likely that such distortions of the signal to gene expression ratio complicate comparisons of different approaches to study transcriptomes (Jeukens *et al.* 2010). This finding is in line with our previous results (Pozhitkov *et al.* 2006) and has prompted our claim that the diversity of binding-behaviors of probes needs to be considered in microarray experiments (Pozhitkov *et al.* in preparation). We have implemented a normalization procedure that establishes a common signal to fold-change

ratio for all probes and thus disentangles probe effects from changes in expression.

Hybrid origin and adaptive differentiation of invasive sculpins

The principal component analysis of the normalized expression values reveals a first major axis of variation that is correlated with the environment in which the fish were raised and possibly other factors (Figure 2). Ancestral species do not differentiate along this axis and it does not capture features that make the hybrid transcriptome different from what is observed in the parental species. In contrast, PC axis 2 clearly separated the ancestral species and indicates that the overall gene expression profile of the F₂ crosses and the invasive hybrid lineage is intermediate between the two parental species. Hybrid intermediacy is also reflected in the number of genes differentially expressed between the species (Figure 3). The number of significant genes between the *C. rhenanus* and *C. perifretum* is more than twice as high as the number of genes differentially expressed between the hybrid species and either parental species which is corroborated in the analysis at the population level. The relative proportions of genes differentially expressed in pairwise comparisons between species and populations agree with the current taxonomic distinction of the ancestral species (Freyhof *et al.* 2005) and levels of genetic differentiation among different populations observed at SNP or microsatellite loci (Nolte *et al.* 2005, Stemshorn *et al.* 2011). The number of differentially expressed genes and the divergence time (~2 Myr, Englbrecht *et al.* 2000) between *C. rhenanus* and *C. perifretum* are the largest in this analysis, but we have not found overrepresentation of GO terms that would indicate biological functions that play a special role in this divergence.

Although the invasive lineage of sculpins phenotypically resembles *C. perifretum*, it possesses a unique potential to invade summer warm river habitats (Nolte *et al.* 2005, 2006). While this inference is supported by general patterns of distribution and gene flow, specifically across hybrid zones, we have not identified the traits or precise ecological factors that cause the particular fitness properties of invasive sculpins. This study provides a first inventory of functional genetic variance that may contribute to novel adaptive phenotypes. The identification of the respective traits also represents a step towards understanding whether ecological changes in *Cottus* can be induced by the process of hybridization (Abbott *et al.* 2010, Nolte & Tautz 2011). In contrast to the within and between species comparisons, the invasive transcriptome is distinguished through differential expression of a set of genes that is significantly biased towards several functions (Table 2). This suggests that a significant proportion of the changes in gene expression is functionally related and therefore affects common, more complex,

phenotypes. While ideas about hybrid speciation center on an instrumental role of hybrid ancestry and the recombinant nature of novel adaptive traits (Mallet 2007, Abbott 2010) an alternative explanation for the presence of an adaptive potential relates to the evolution of traits in allopatry within the parental species. Such a preadaptation hypothesis posits that traits in invasive sculpins convey a fitness advantage without the genetic effects of admixture. Invasive sculpins shared expression patterns at 1203 and 1229 genes with *C. rhenanus* and *C. perifretum* respectively. While these genes could in principle represent characters that are retained from the ancestral species through a selective process, GO-term enrichment analysis could not detect any enrichment in these gene sets (see Table 2). Hence, the biological functions that are distinct and enriched in invasive sculpins are most likely not characters that were evolutionarily retained from one of the ancestors but have emerged in invasives. The fact that some transgressive patterns of gene expression of invasives could be recreated in F₂ crosses suggests that the process of hybridization has contributed to functional changes. The most significantly enriched biological process was “embryonic development”. This is surprising, since all fish in the analysis were juveniles transitioning to the adult stage. On the other hand, this can be explained because GO term analysis, though a valuable tool for meta-analysis, suffers from annotation biases due to idiosyncrasies of different research communities (Leong and Kipling 2009). However, screening sets of enriched genes in the light of evolutionary ecological hypothesis yields promising candidate genes. For example, invasive sculpins are found in parts of rivers that have a different temperature regime (Nolte *et al.* 2005, 2006). The influence of temperature on the distribution of species is described by the concept of oxygen and capacity limited thermal tolerance (OCLTT) (Pörtner and Knust 2007). It states for aquatic organisms that the capability to perform at high or low temperatures is determined by a mismatch between the demand for oxygen and the capacity to supply oxygen to the tissues. Accordingly, Eliason *et al.* (2011) showed that selection pressures associated with the temperature regime led to fast evolutionary modifications of the blood transportation system in salmonids. A reevaluation of the genes causing enrichment of the GO term “proteinaceous extracellular matrix” in invasive sculpins with respect to functions expected under the OCLTT concept yielded several genes (e.g. matrix metalloproteinase 2, transforming growth factor beta), that are known to play a role in formation of blood vessels (Fisher and Berger 2003) but were not annotated for the GO-term “angiogenesis”. Hence, it would be possible that an adaptive modification of the oxygen transportation system escapes GO term analyses.

From initial hybridization to the invasive phenotype

From an evolutionary ecological perspective an emerging hybrid lineage must separate from the parental species to avoid recurrent backcrossing and direct competition (Buerkle 2000, Barton 2001) and Seehausen (2004) and Stemshorn *et al.* (2011) discuss examples for how new adaptive peaks are exploited through the increased genetic variance present in a “hybrid swarm”. Barton (2001) suggested that the rise and fixation of newly adapted hybrid genotypes with a superior fitness will take some time and that a population of admixed, outbreeding hybrids will go through a lag phase before selection has optimized novel genotypic combinations. Although evidence that hybridization has played a key role in the evolution of hybrid species was obtained through experimental re-creation of hybrid phenotypes in the laboratory for sunflowers (Rieseberg *et al.* 2003b) and *Heliconius* butterflies (Mavárez *et al.* 2006) empirical studies that document the evolutionary transition from initial hybrids to an admixed lineage are still scarce. The sequence of events that occur in consecutive generations following initial hybridization was studied in *Senecio squalidus*, a hybrid species of plant that became invasive in the UK after being transplanted from a hybrid zone of its parental species, *Senecio aethnensis* and *Senecio chrysanthemifolius*. Hegarty *et al.* (2009) experimentally recreated crosses of the parental species and traced changes in gene expression from the F₁ to the F₅ generation. They detected a relatively high variance in gene expression and patterns of non-additive and transgressive gene expression in initial generations and found that these patterns were considerably reduced in the F₄ and F₅ generations post hybridization. Since seed germination rates sank from the F₁ to the F₃ generation, but increased later on, Hegarty *et al.* (2009) suspected that unintentional selection may have removed maladaptive patterns of gene expression. This indicates that phenotypic variability of admixed populations can be significantly altered after only a few generations in a particularly dynamic process (Nolte & Tautz 2010).

Although we have not analyzed consecutive generations of the artificial *Cottus* hybrids, an analysis can reveal to what extent patterns of gene expression in the natural hybrids are adumbrated in the lab reared hybrids. Invasive sculpin gene expression was less variable than that of the F₂ generation but was clearly more variable than in its parental species. This corresponds with a loss of heterozygosity across the invasive sculpin genome over time (Stemshorn *et al.* 2011) and agrees with the prediction that an initially maximal genetic diversity is reduced as maladaptive genetic variance is lost. Moreover, we have compared the numbers and types of transgressively over- and under- expressed genes between F₂ crosses and invasive sculpins to test how patterns that reflect initial admixture

may have been modified as invasive sculpins evolved. The numbers of transgressively up and down regulated genes in the F_2 crosses are approximately balanced but this pattern did not prevail in invasive *Cottus*. The number of transgressively down regulated genes dropped while the number of transgressively up regulated genes more than doubled from the F_2 generation to invasive *Cottus*. At the same time, many of the initial transgressive patterns were lost and new ones emerged. Interestingly, the set of transgressively up regulated genes in invasive sculpins turned out to be the set of genes that showed the strongest enrichment in our study (Table 2) while down regulated genes in invasives were not significantly enriched for any GO term. Enrichment analysis of the transgressively over expressed genes reproduced GO terms that were previously found to be characteristic for the whole transcriptome of invasive *Cottus*, albeit with higher significance levels (Table 2). This suggests that the functional enrichment detected in the differentially expressed genes is mainly caused by the genes which are transgressively over expressed in the invasive hybrid lineage and that this set represents a nonrandom set that is functionally linked. Intriguingly, transgressive overexpression seemed to be plastic for a large subset of these genes, which may suggest that phenotypic plasticity interacts with fixed traits to create evolutionary novelty (Moczek *et al.* 2011). This aspect remains unexplored for the origin of the evolutionary novelty and adaptation of the invasive *Cottus* hybrid lineage. To date, the set of transgressively overexpressed genes represents the best candidates to further study the adaptive advantage that invasive sculpins have over their ancestors in the lower River Rhine basin (Nolte *et al.* 2005, Stemshorn *et al.* 2011). Conversely, a significant number of transgressively down regulated genes in invasive sculpins were retained from the pool of genes that showed this pattern initially. We did not detect any enrichment of GO terms for this set of genes, but it is possible that selection has contributed to maintain these patterns and that they represent transgressive traits that have immediately contributed positively to the fitness of invasive sculpins.

One of the most conspicuous results of this study is that the gene expression features that distinguish invasive sculpins are mostly up regulated. Such a concerted change would not be expected if regulatory changes evolved independently to modify adaptive phenotypes as it was found by Lai *et al.* (2006). We consider it more likely that this result involves the exaggerated growth of some organs or tissues, which would result in a concerted rise of gene expression for genes specifically expressed in such tissues. On the one hand, this impairs the detection of the primary genetic factors that cause changes in gene expression. On the other hand it provides us with a more complex phenotype that can be studied experimentally. Besides mapping of the underlying genes (Cheng *et al.* in prep),

population genetic signatures of fixation in the invasive gene pool can be exploited to identify genomic regions (Stemshorn *et al.* 2011) that contribute to emerging phenotypes in the course of genomic admixture events.

Acknowledgements

We thank F. Volckaert, K. de Gelas, A. Kobler and L. Bervoets for their benevolent help in obtaining breeding stocks of *Cottus perifretum*. K. Lessenich and A. Mellin from Bezirksregierung Köln, Nordrhein - Westfalen and T. Heilbronner from the Siegfischereigenossenschaft have given permits to conduct the research. *Cottus* were kept in the laboratory of the MPI in Plön with permission from S. Hauschildt (Veterinaeramt in Plön, Schleswig – Holstein). E. Blohm-Sievers, 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, T. Domazet-Lošo 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.

Chapter 2

Temperature dependent gene expression and the invasion of new thermal environments in *Cottus*

Till Czypionka, Arne W. Nolte

Max Planck Institute for Evolutionary Biology, August-Thienemannstrasse 2, 24306 Plön, Germany.

Abstract: Phenotypic plasticity is considered to play an important role as a source of phenotypic novelty and changes in plastic traits may contribute to evolutionary change. Here, we test if the modification of pre-existing plastic responses contributed to the acclimatization and adaptation to different temperature regimes in freshwater sculpins of the genus *Cottus*. This is relevant to ectothermic fishes because water temperature is a fundamental ecological parameter that structures ecosystems along a river and critically determines physiological processes. While *Cottus* are typically known as stenotopic inhabitants of cold water bodies, a young invasive lineage of hybrid origin is currently colonizing summer-warm habitats. We have thus analyzed gene expression responses to changes in water temperature. Temperature specific expression profiles of four populations of the ancestral species and one population of the invasive lineage were assessed over a temperature range from 14°C to 25°C. This range was chosen to reflect a possible range of summer conditions in the originally occupied and newly invaded habitats. Contrary to our expectation that invasive *Cottus* show a marked response to raised water temperatures, the clearest expression profile differentiation between parent species and invasive *Cottus* was observed at temperatures < 21°C that are common in both habitats. Comparative analysis of population specific gene expression changes over temperature revealed plastic responses that are shared by all *Cottus* as well as changes in thermal plasticity between the ancestral species and the invasive lineage. In particular the increased plasticity of a mitochondrion related phenotype in the invasive lineage contributes to the functional differentiation from its ancestors. Our results demonstrate that an increase in plasticity of an already existing plastic phenotype played a role in phenotypic differentiation and diversification in *Cottus*.

Introduction

Genetic traits that cause adaptive phenotypic change play a key role in adaptive evolution and ecological speciation (Schluter 2000, 2009). However, many phenotypes are known to be not only determined by underlying genetics but to change plastically dependent on environmental conditions. Such 'phenotypic plasticity' can take the form of polyphenism (Stearns 1989), i.e. the expression of different discrete phenotypes based on the same genotype. An example for polyphenism is the expression of inducible defenses in response to kairomones of potential predators in *Daphnia cucullata* (Laforsch and Tollrian 2004). If the phenotypic response towards an environmental cue is continuous, the relationship is called a reaction norm (RN). Phenotypic plasticity is relevant for ecological speciation scenarios because differences in the environment to which organisms adapt can induce phenotypic change in plastic traits. Therefore, phenotypes should be measured under the conditions prevalent in the environment to which adaptation occurred (Barrett *et al.* 2009) to account for plastic components of a phenotype. While the realized phenotype in a certain environment change depends on both environmental and genetic components, the actual evolutionary change is happening at the genetic level only (Gienapp *et al.* 2007). This evolutionary change can be empirically assessed and separated from ancestral plastic responses, if adapted and the ancestral populations are compared under relevant environmental conditions (Parsons & Robinson 2006). Because of the contribution of environmental and genetic components to the realized phenotype, Schlichting & Pigluicci (1998; p. 79) propose the reaction norm itself should be interpreted as a trait under selection. How plastic traits and their RN evolve has been conceptualized with the most renowned of this processes being genetic assimilation. Genetic assimilation describes a process where plastic phenotypes lose or severely reduce their plasticity. This results in the expression of the phenotype independent of the formerly required environmental stimulus. The most famous example for this is a selection experiment by Waddington (1953) for a formerly plastic 'crossveinless' phenotype in *Drosophila* which became assimilated in ~23 generations. Thus evolution via genetic assimilation can occur rather fast. Conversely, based on a quantitative genetic model Lande (2009) suggests that abrupt adaptive reaction norm evolution towards a new environment starts with an increase in plasticity, i.e. a widening of the reaction norm, which allows for fast adaptation to a new environment. Genetic assimilation then occurs in a second phase (Lande 2009) when plasticity is reduced, while the expression of the new phenotype adaptive in the new environment is maintained. Irrespective of the details of these processes adaptation via RN evolution is characterized by two important properties. First, RN evolution is a

mechanism allowing for fast adaptation. Second, such fast adaptation can involve already existing complex phenotypes which, by being re-utilized under new environmental settings, can provide complex evolutionary innovations (Moczek *et al.* 2011). Accordingly, RN-evolution has been shown to play a key role in diversification and speciation (reviewed in Pfennig *et al.* 2010).

One of the environmental factors most frequently used to study phenotypic plasticity is temperature (e.g. Gracey *et al.* 2004, Eliason *et al.* 2011, Kingsolver *et al.* 2004). Temperature is also one of the key abiotic factors influencing species distribution (Angilletta 2009). It has been referred to as ‘the ecological master factor’ (Brett 1971) “because its profound effect on physiology of ectothermic animals” (Gamperl & Farrell 2004). Indeed, thermal adaptations have repeatedly been shown to allow organisms to extend their range and to lead to ecological speciation (reviewed in Keller & Seehausen 2012). Therefore, temperature is a suitable environmental factor to study the interplay of changes in phenotypic plasticity, adaptation and ecological speciation.

This interplay can be studied in European freshwater sculpins (*Cottus*), where an invasive lineage is colonizing a new habitat with a distinct temperature regime. Illies (1961) introduced a zonation system separating the lotic ecosystems into the regions of the krenal, the rhithral and the potamal. These different eco-regions are differentiated by abiotic factors like temperature, oxygen saturation flow velocity and contain different and characteristic biocenoses. *Cottus* are originally considered stenotopic for cold streams of the rhithral region (Kottelat & Freyhof 2007). The rhithral is characterized as summer-cold (~ max.20°C) and oxygen-rich (Illies 1961). In contrast, an invasive lineage of *Cottus* is currently colonizing the river Rhine and its tributaries belonging to the potamal region which was historically devoid of sculpins (Nolte *et al.* 2005). The potamal is differentiated from the rhithral by a higher maximal summer temperature (> 20°C) among other factors (Illies 1961). Indeed, the transition from rhithral to potamal circumscribes the distributional range of many species (Illies 1961).

The invasive lineage originated from hybridization of the species *Cottus rhenanus* and *Cottus perifretum*, whose habitats were connected due to antropogenic disturbances approximately 200 years ago (Nolte *et al.* 2005) and is genetically differentiated from its parent species (Stemshorn *et al.* 2011). During the upstream colonization of the River Rhine and its tributaries the invasive lineage established secondary contact with populations of its parent species *C. rhenanus* residing in rhithral confluents. Both readily hybridize and form very narrow hybrid zones (Nolte *et al.* 2006, 2009). The locations of

these hybrid zones coincide with the transition from rhithral to potamal which underlines differential adaptation of the different lineages of *Cottus* (Nolte *et al.* 2006). We have recently performed a gene expression screen comparing the invasive hybrid lineage to its parent species (Czypionka *et al.* 2012) and found the invasive lineage to be distinguished through a set of transgressively overexpressed and functionally related genes. Taken together, this evidence suggests that the situation in *Cottus* fits an ecological speciation scenario in which reproductive isolation is maintained by extrinsic factors which differ between the rhithral and the potamal and vice versa by genetically fixed adaptations of the parent species and the invasive hybrid lineage to their respective habitats.

In this paper we use the formation of a new hybrid lineage in *Cottus* (invasive *Cottus*) to explore phenotypic plasticity and the contribution of RN evolution to functional differentiation during ecological speciation. Since adaptation via RN evolution is supposed to be fast (see above), the young age of the invasive *Cottus* makes them especially useful for this analysis, because initial adaptive changes are not hidden behind secondary evolutionary changes (Nolte & Tautz 2010). As an environmental cue, we choose temperature because of its general relevance as an ecological factor and its specific relevance for the study system in analysis. Specifically, we focus on the temperature situation in summer, when maximal water temperatures which differentiate the two habitats (Illies 1961) are reached. *Cottus* are renowned as inhabitants of cold water habitats (Kottelat & Freyhof 2007), so we speculate that the invasion of the warmer waters of the potamal represents an important evolutionary change. Temperature data for the source habitats of some of the populations in analysis are available through environmental monitoring programs. This allows us to design our experiment to correspond to the situation in nature. As the phenotype to measure in relation to temperature we use gene expression which has been shown to change as an adaptation to habitats differentiated by their temperature (Whitehead and Crawford 2006). Analyzing the expression patterns of many genes in parallel provides an unbiased general pattern and bears the potential to identify hidden phenotypes of interest through the genes' annotation (Larsen *et al.* 2011). Specifically, we determine thermal transcriptional RNs (Aubin Horth and Renn 2009) using the same microarray and the same populations as in Czypionka *et al.* (2012).

We use our data to address the following four objectives. Firstly, we assess and characterize the amount of thermal plasticity in gene expression in *Cottus*. We compare the results obtained for the different populations and species on a functional level with a focus on the differences between invasive *Cottus* and the parent species. Thereby we test the hypothesis that *invasive Cottus* are distinguished from the parent species through the

amount of phenotypic plasticity in certain functional relevant phenotypes. Secondly, we specifically identify genes, for which genetic assimilation or RN widening have occurred during the formation of invasive *Cottus*. These candidate genes are employed to test the hypothesis that the proposed processes of RN evolution also occurred in the ecological speciation in the *Cottus* system. Thirdly, we determine how the expression profiles of the populations and species are differentiated at the individual temperatures. This allows us to identify temperatures at which the invasive *Cottus* are functionally differentiated from the parent species. Here, our hypothesis is that the strongest signature of functional differentiation is found at the higher temperatures characteristic for the potamal habitat, which invasive *Cottus*, but not its parent species, were able to colonize. Fourth, we test the hypothesis that RN evolution (i.e. genetic assimilation and RN widening) contributed to this functional differentiation of invasive *Cottus* from its ancestral parent species.

Methods

Study populations in analysis and fish care

The populations analyzed in this study are identical to those already used in Czypionka *et al.* (2012), consisting of two populations of *C. rhenanus* (Broel & Naaf), two populations of *C. perifretum* (Witte Nete (WN) & Laarse Beek (LB)) and one population of invasive *Cottus* (Sieg). All fish used in this experiment were born and raised under controlled conditions in our fish facility. During winter 2010/2011, water temperature was lowered to 4°C. When the temperature was raised to 8-10°C (April 2011) spawning readily occurred in artificial shelters partially buried in sand. Fish larvae hatched during May 2011. Fish were initially raised with artemia and later live and frozen invertebrates. Water temperature and photoperiod in the fish facility were adjusted over seasons to mimic the natural conditions in central Europe. At the time of transfer to the temperature gradient (end of September 2011; see below), fish weighed approx. between 0.2 and 0.4 g and were kept at 17°C. Thus the fish used in this work resemble juveniles in their first summer. Since the conditions under which *Cottus* were raised were identical for all fishes entering analysis, the analysis of plastic changes refers to the concept of phenotypic plasticity where environment influences the realized phenotype, but not to the concept of developmental plasticity, where environmental influences during embryonic development causes subsequent phenotypic differentiation (Crispo 2007). Hence, RNs reported in this work do not constitute 'developmental RNs' but so called 'labile RNs' (Lynch & Walsh 1998 p.659).

Temperature treatment

Temperature regime in nature

In order to be able to relate our experiment to the conditions in nature we compared water temperature data from an environmental monitoring program for the river Sieg and the stream Broel. The lower Sieg river is a potamal river inhabited by invasive *Cottus*. The Broel belongs to the rhithral and is inhabited by *C. rhenanus*. Populations from both water bodies are analyzed in this paper. Water temperature data were provided by the “Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen” under <http://luadb.lids.nrw.de/LUA/wiski/> (assessed 11/22/2012; Sieg: station „Menden-S“ 50°48'N 7°9'E , Broel: station „Broel“ 50°50'N 7°25'E). The environmental monitoring stations were in close proximity to the original sampling sites from which our fish facility stock populations for the Broel population (*C. rhenanus*) and Sieg population (invasive *Cottus*) were originally derived. Consequently, the temperature data were used to choose ecological relevant parameters for the temperature treatment.

Temperature gradient experiment

A gradient flow through system consisting of 4 chambers (volume 27,6 l each) was set up to expose the populations in analysis to different temperatures. The system was set up in a climate room with 14°C air temperature. Water with a temperature of 14°C entered the left chamber (0,32 l /min) (sketch in supplementary file 1). Chambers were connected at the bottom allowing for a moderate water exchange. This led to a water flow from the left chamber to the right chamber that had an effluent. Temperatures in the chambers were adjusted to 14°C, 17°C, 21°C and 25°C by heating elements (power: 150-200 W). Air stones were used to aerate and circulate the water within the chambers, thereby providing homogenous conditions.

The temperatures in this gradient were chosen to reflect low summer temperatures in the rhithral (14°C), high summer temperatures in the rhithral and low summer temperatures in the potamal (17°C), high summer temperatures in the potamal (21°C) (see supplementary file 2). The temperature of 25°C was chosen as a representative of high temperatures reached in the potamal in very hot summers (invasive *Cottus* were found in shallow water microhabitats on shallow gravel banks within the river Sieg at 28°C. Arne Nolte, personal observation). Each temperature chamber was equipped with 5 net cages (~ 17 cm x 12 cm x 13.5 cm; JBL NBox (JBL, Neuhofen, Germany) and Marina Fish net breeder (Hagen, Holm, Germany). Temperature and oxygen saturation were continuously monitored

throughout the course of the experiment using the *Multi 350i* multi-parameter system equipped with a *ConOx* probe (WTW GmbH, Weilheim, Germany) to confirm conditions were comparable among the net cages from one chamber. The system was monitored for 5 days (22.9.11 to 27.9.2011) to assure conditions were stable (maximal standard deviation per net cage: 0.44 °C). At the time the experiment was set up, fish in our fish facility were acclimatized to 17°C mimicking summer conditions (supplementary file 2). Twenty fish from each of the populations in analysis were transferred to the net cages at the 17°C chamber of the gradient flow through system. After an acclimatization period of 24 h, 10 fish from each population were transferred to net cages at the 14°C and the 21°C chambers. Again 20 fish from each population were transferred from the fish facility stock to the net cages at the 17°C chamber. Following an additional 24 h acclimatization period, fishes from the 21°C chamber were transferred to the net cages of the 25°C chamber and 10 fishes from each population were transferred from the 17°C chamber to the 21°C chamber. As a result, fish were acclimatized stepwise to the different temperatures in the gradient.

Following this acclimatization procedure, fish were exposed to the different temperatures for 14 days (29.9.2011 – 12.10.2011). This time period was chosen based on the observation that substantial changes in temperature (see supplementary file 2) already occur over such short periods in natural environments. Also, temperatures above 21°C were rarely maintained for longer time periods during the summers of 2011/2012 for which water temperature data are available (supplementary file 2). Fish were fed to saturation with frozen and live insect larvae. Net cages were checked to contain food four times a day to prevent starvation induced artifacts due to higher metabolic rates at higher temperatures (Guderley 2004). Sampling was carried out in the morning of day 15 (13.10.2011). Net cages containing fish were removed from the flow through system and immediately submerged into a bath of carbonated water thereby anaesthetizing all fish in the net cage simultaneously. Carbonated water was adjusted to the temperature of the chamber the fish were taken from to avoid a temperature shock. Overanesthetized fish were killed by immersion into RNA_{later}. Fish were kept in RNA_{later} at -20°C until RNA extraction.

Analysis of gene expression

Total RNA was extracted from whole fish using a Trizol protocol and RNA concentrations were determined using the Agilent RNA 6000 Nano kit (Agilent manual number G2938-90034). For each net cage, two pools were prepared from equimolar amounts of total RNA

of 4 fish per pool. The number of 4 instead of 5 fish per pool was chosen due to the death of one fish in the course of the experiment (LB-population @ 21°C; 30.9.2011) in order to maintain a balanced design. Fish used for RNA extraction were selected to represent average sizes. We have not observed conspicuously sick individuals. Labeled cRNA was produced from total RNA pools using the Agilent Low Input Quick Amp Labeling kit. Labeled cRNA was then hybridized to a microarray especially designed for the *Cottus* populations analyzed here (Czypionka *et al.* 2012) following the Agilent protocol for One-Color Microarray-Based Gene Expression Analysis (version 6.5, May 2010). Microarrays were scanned with an Agilent Microarray Scanner type C and feature extraction was carried out using the Agilent Feature Extraction software (version 10.7.3.1). The 'gprocessed signal' was taken as raw expression level. Signals flagged for bad quality or saturated intensity were interpreted as missing data. For the *Cottus* microarray used here, we have developed a calibration and normalization procedure described in Czypionka *et al.* (2012). Briefly, the calibration consists in an assessment of the target-binding behavior of the probes on the microarray by hybridizing different concentrations of a pool of labeled cRNA to the microarray. Based on this calibration, we a) excluded probes with a non-linear and unfavorable binding behavior and b) normalized raw-expression levels for differences in probe-binding behavior so that normalized expression levels indeed reflect differences in transcript abundance (details given in Czypionka *et al.* 2012). We normalized raw expression levels from this experiment by the same method based on the calibration from Czypionka *et al.* (2012). Between-microarray normalization was carried out by dividing the normalized signal intensities of each microarray by their 75%-tile as in Czypionka *et al.* (2012). Microarray data normalization and all other statistical analysis were carried out in R (R Development Core Team 2010). Populations were assessed by two replicates at each temperature. One of the replicates of the Laarse Beek population maintained at 14°C had to be removed because of insufficient incorporation of labeled nucleotides during synthesis of labeled cRNA which proved to affect data quality. Therefore, the measurements for 14°C treatment of the LB population were based on a single replicate. For the other treatments expression levels (ELs) were averaged over the replicates, providing a single EL for each population at each temperature. Genes still containing missing data after the averaging step were excluded from the analysis. This step is essential considering plasticity assessment for each population was based on the maximum EL range over temperatures (see below). Additionally, genes having no variance in their ELs over temperatures were excluded from analysis. These were genes expressed very low or not at all and whose EL

was set to 0 by our normalization procedure.

Plasticity in gene expression

Plasticity was detected at the population level as a substantial change in EL over temperature. EL change was assessed as the difference between the maximum and the minimum EL, i.e. the population specific EL range over temperature. To decide, if this EL change was substantial or just resulted from sources of variability unrelated to temperature, we calculated the gene specific standard deviations (SD) of gene expression from the replicates. Since these values were only based on two measurements each, the average of these SDs over all populations and temperatures was used as a gene specific measure for temperature unrelated variability in gene expression. Genes were identified as plastic in their gene expression if their EL range was larger than 4 times their averaged SD. While this threshold is arbitrary, it is motivated by the fact that under a normal distribution with its centre exactly in the middle between the maximum and the minimum EL, 4 SDs would correspond to the 95 % confidence interval. Plasticity was assessed for each gene at the population level. Plasticity of genes was further characterized, by fitting a linear function between temperature and ELs for each population. The slope of this function was used as a proxy for an increase or decrease of expression level with temperature. Genes with a negative slope were characterized as having a “negative RN” in the analyzed population. Correspondingly, a positive slope indicated a “positive RN”. In order to determine which plastic responses are common in *Cottus* and which are specific to certain species or the invasive *Cottus*, genes were categorized according to the population(s) in which this plasticity was detected. This analysis was carried out for the population comparisons Broel-WN-Sieg and Naaf-LB-Sieg. Thereby both comparisons contained a population from each parent species and the invasive *Cottus*. Plastic genes were divided in seven different subsets according to the detection of thermal plasticity in their expression: (1) only in the *C. rhenanus* population, (2) only in the *C. perifretum* population, (3) only in the invasive Sieg population, (4) in all three populations (5) only in the two parent species populations, (6) only in the *C. rhenanus* population and the invasive Sieg population or (7) only in the *C. perifretum* population and the invasive Sieg population. This analysis was carried out separately for negative or positive RNs. Venn diagrams were prepared to illustrate the distribution of these gene sets between populations using the limma package (Smyth 2005).

Signatures of RN-evolution

The characterization of plastic responses on the population level only revealed a

difference in the strength of plastic response if this difference involves a change in the categorization as plastic or not plastic.

In order to identify the contribution of RN evolution to the functional differentiation of invasive *Cottus* an analysis independent of this categorization was performed. We identified genes for which genetic assimilation (a loss of plasticity of gene expression) or respectively a widening of the RN (an increase in plasticity) had occurred in invasive *Cottus* respectively to the populations of its parent species. Genes where the minimal or the maximal EL-range was found in the invasive Sieg population were taken as candidates for RN-evolution. A one-tailed one-sample t-test was carried out, to test if the invasive EL-range was significantly ($p < 0.01$) different from the EL range in the parent populations. Additionally, the genes for which significant reaction norm widening was detected in invasive *Cottus* were split into genes with a positive RN in the invasive and genes with a negative RN in the invasive. The resulting two gene sets were also functionally characterized by GO term enrichment.

Differentiation of gene expression at individual temperatures

The low coverage of population specific gene expression at individual temperatures ($n = 2$) impedes a statistical sound assessment of differential expression on a per gene basis. Instead, we assess differentiation of the entire gene expression profiles between the populations at the individual temperatures by the calculation of temperature specific distance matrices between the transcription profiles using the amap package (Lucas 2010) in R. These matrices were based on the Euclidian distance measures between population specific expression values scaled by their standard deviation over populations of all genes in analysis. While distance measures for individual genes might also be affected by low replication, such random effects cancel each other out in the calculation of the distance matrices from many genes. Based on the temperature specific distance matrices, neighbor joining trees were build using the ape package (Paradis *et al.* 2004) in R. The same method was also used to assess the effects of genetic assimilation and RN-widening on gene expression differentiation at individual temperatures. In this case, distance matrices were calculated based on the expression values of sets of identified candidate genes.

Functional characterization by GO term enrichment

Functional characterization of different subsets of genes was carried out by GO term enrichment analysis using the software Blast2GO (Conesa *et al.* 2005). GO term

annotation for the genes on the microarray is described in Czypionka *et al.* (2012). A one-tailed Fisher's exact test with false discovery rate (FDR)-correction implemented in Blast2GO was employed to detect significant (FDR corrected p-value < 0.05) GO term enrichment in gene sets against the background of the genes entering the analysis after data filtering.

Results

Temperatures in natural habitat

The seasonal change in water temperature in potamal habitat of the invasive *Cottus* and the rhithral habitat of the parent species *C. rhenanus* was compared by contrasting temperature monitoring data from the habitats of two populations analyzed in this study (supplementary file 2). These were the Broel stream inhabited by *C. rhenanus* and the river Sieg inhabited by invasive *Cottus*.

During the summer (~June to October), the temperature in the river Sieg is most of the time up to 4°C warmer than in the stream Broel. Water temperatures in both locations are fluctuating with changes of up to ~10°C in only 2 weeks (6/2012). These fluctuating changes in water temperature are correlated between river and stream. In winter temperatures in both habitats decrease to 3°C and less. Unfortunately, no water temperature data could be retrieved for the Broel stream for the period from 1/2011 to 4/2011 and for the River Sieg for the period from 1/2012 to the first half of 4/2012. Therefore, it is not possible to contrast winter temperatures between Broel and Sieg as clearly as summer temperatures.

Plastic patterns of gene expression

The microarray used in this analysis contained probes for 13329 genes. The calibration and normalization procedure developed by Czypionka *et al.* (2012) lead to the exclusion of 3924 genes because of improper probe binding behaviors. From the remaining 9405 genes, 1325 were removed from analysis because of missing data. 33 genes were excluded from analysis because they had no variance in their expression over temperatures. Thus, in total 8047 genes entered analysis.

Plasticity in gene expression of the individual populations

The number plastically expressed genes ranged from 1697 to 2683 between populations. When separating RNs into “positive RNs” and “negative RNs” as a response to the temperature gradient, it was found that except for the Naaf population, more genes with negative than with positive RNs were found. Supplementary file 3 contains a table specifying for each gene and population, which genes were identified to have a positive or negative RN. The count of plastic genes per population is provided here in the format (total| with negative RN| with positive RN) for each population: Broel (1697| 1068| 629); Naaf (1895| 788| 1107); Laarse Beek (2683| 1604| 1079); Witte Nete (2080| 1139| 941); Sieg (2159| 1177| 982).

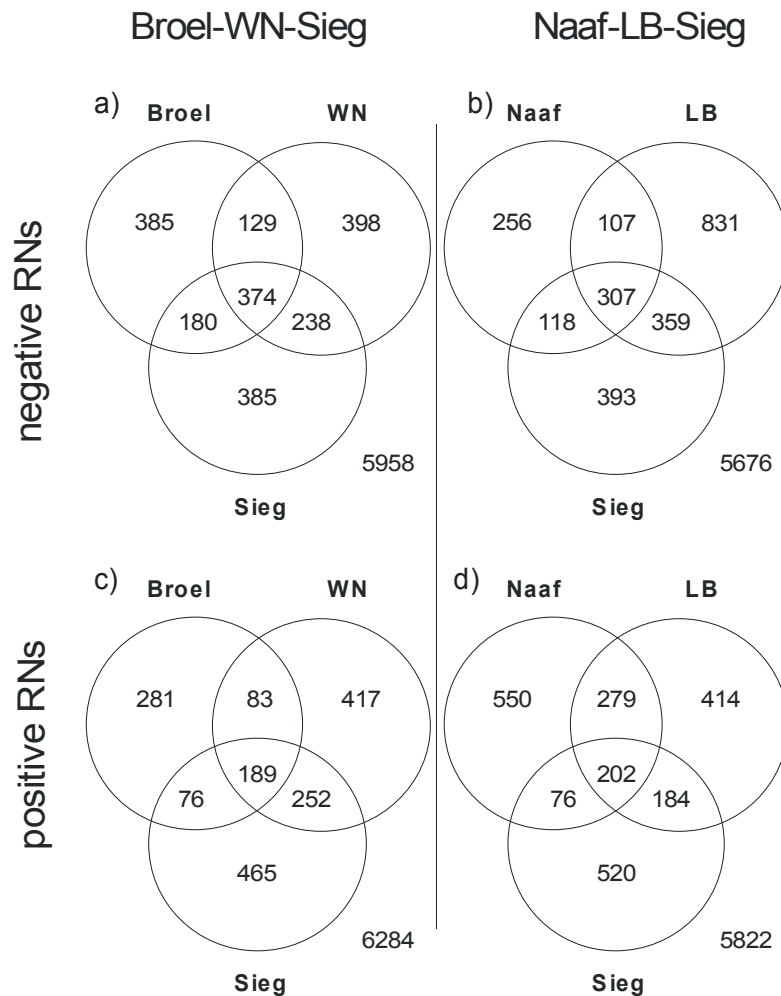


Fig 1: Venn Diagrams indicating numbers of genes showing thermal plasticity in expression in two population comparisons each containing one population of *C. rhenanus* (Broel, Naaf), *C. perifretum* (LB, WN) and invasive *Cottus* (Sieg). (a & b): Genes with a negative RN. (c & d): Genes with a positive RN. (a & c) Broel-WN-Sieg comparison. (b & d): Naaf-LB-Sieg comparison.

Thus, the invasive Sieg population is not distinguished from the parental populations through any conspicuous trend for higher or lower number of plastically expressed genes.

Shared plastic responses and their functional characterization

In order to identify plastic responses, which distinguish invasive *Cottus* and its parent species, genes were categorized according to prevalence of expression level plasticity in populations of *C. rhenanus*, *C. perifretum* and invasive *Cottus*. This categorization was carried out as for the comparisons Broel-WN-Sieg and Naaf-LB-Sieg separately (Fig 1).

The invasive Sieg population is not distinguished from the parental populations by any conspicuous trend for a higher or lower number of genes with a plastic response (Fig 1). Functional characterization via GO-term enrichment revealed common plastic responses for *Cottus*, but also plastic responses which were characteristic for invasive *Cottus*, as detailed in the following. GO-term enrichment results for all 28 gene sets depicted in Fig 1 given in supplementary file 4. A common plastic response for *Cottus* was detected through concordant results of GO term enrichment analysis between the both population comparisons: genes with a negative RN in all populations were enriched for the GO-terms 'proteinaceous ECM', 'embryo development' (respectively its child term 'multicellular organismal development'), 'ion transport', 'structural molecule activity' and 'generation of precursor metabolites and energy'. GO term enrichment also revealed characteristic plastic responses of invasive *Cottus*. Genes with a negative RN exclusively in the invasive Sieg population were enriched for the GO terms 'generation of precursor metabolites and energy' and 'mitochondrion'. The GO-term 'DNA metabolic process' was enriched in genes with a positive RN in the invasive Sieg population. These GO terms were also enriched in other sets of genes with a plastic response shared between invasive Sieg population and single populations of the parent species (supplementary file 4). This suggests that these GO terms describe a plastic response especially pronounced in invasive *Cottus*. Genes annotated for the GO-terms discussed in this paragraph are also reported in supplementary file 3.

Evidence for proposed modes of RN evolution

We identified 255 genes with a pattern consistent with genetic assimilation and 735 genes consistent with a RN widening scenario in invasive *Cottus* (Table 1 a; affected genes are also listed in supplementary file 3). Genes showing a signature of RN widening in invasive *Cottus* were enriched in the GO terms "cellular metabolic process" and "generation of precursor metabolites and energy" (Table 1a). In order to refine this functional characterization, genes with a widened RN were separated according to the slope of their RN in the invasive *Cottus* (415 with a positive RN, 320 with a negative RN). Both sets were functionally characterized through GO-term enrichment (Table 1b). Genes with a widened negative RN were enriched for the GO terms 'generation of precursor metabolites and energy' and 'mitochondrion' among others. Genes with widened positive RNs in invasive *Cottus* were enriched in the GO terms "nucleoplasm" and "nucleobase-containing compound metabolic process".

Table 1: GO term enrichment analysis of genes being subject to RN evolution in invasive *Cottus*. For gene sets showing significant enrichment (FDR corrected p-value < 0.05), the GO term is reported along with the FDR corrected significance level (FDR) and the number of genes annotated for the respective GO term in the analyzed gene set (# in subset). (a) Analysis for genes for which genetic assimilation or respectively RN widening has occurred in invasive *Cottus*. (b) Analysis for genes for which RN widening towards a RN with a negative or respectively positive slope in invasive *Cottus* has occurred.

a) change of RN in invasive <i>Cottus</i>	# of genes	enriched for GO term	FDR	# in subset
genetic assimilation	255	-		
RN widening	735	cellular metabolic process	3.1 E-4	177
		generation of precursor metabolites and energy	5.3 E-4	32
b) RN widening	# of genes	enriched for GO term	FDR	# in subset
positive RN in invasive <i>Cottus</i>	415	nucleoplasm	3.5 E-3	35
		nucleobase-containing compound metabolic process	6.5 E-3	55
negative RN in invasive <i>Cottus</i>	320	generation of precursor metabolites and energy	1.1 E-8	27
		mitochondrion	1.0 E-5	52
		electron carrier activity	4.0 E-2	48
		Ion channel activity	4.0 E-2	74

Differentiation of gene expression at different temperatures

While the above analysis of plasticity looked at the change of gene expression over temperatures, it does not tell us about expression profile differentiation at discrete temperatures. This differentiation is illustrated by a neighbor joining tree calculated on population transcription profiles in Figure 2a. At 14, 17 and 21°C gene expression differentiation reflects the taxonomic relationship with the populations of *C. rhenanus* (Broel and Naaf) and *C. perifretum* (LB and WN) clustering together. The Sieg population representing invasive *Cottus* takes an intermediate position between the populations of the parent species in accordance with its hybrid origin. The branches separating the species are much shorter than the branches of the individual populations. Branches separating the species become shorter at 21°C, indicating a reduction in the expression

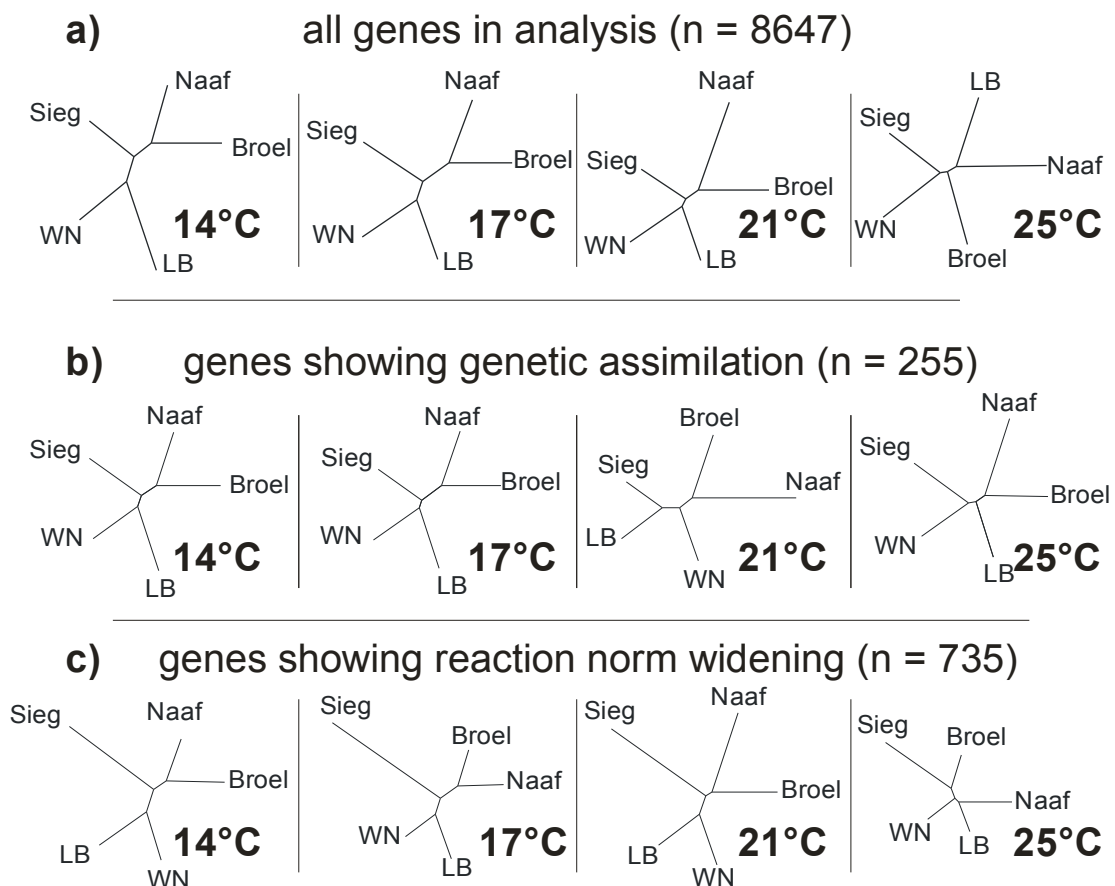


Fig 2: Neighbor-joining trees indicating the differentiation of gene expression patterns between the populations Broel, Naaf (*C. rhenanus*), LB, WN (*C. perifretum*) and Sieg (invasive *Cottus*) at the analyzed temperatures. (a) Expression profile differentiation for all 8647 genes in analysis. Except at 25°C, expression profile differentiation is in agreement with taxonomy. Populations of *C. rhenanus* (Broel, Naaf) and *C. perifretum* (LB, WN) cluster together while the invasive Sieg population takes an intermediate position in agreement with its hybrid origin. (b) Differentiation of gene expression patterns of 255 genes for which genetic assimilation of expression levels occurred in invasive *Cottus*. The differentiation at 14°C and 17°C is similar to expression profile differentiation for all 8647 genes in analysis. At 21°C and 25°C the invasive Sieg populations clusters with the *C. perifretum* populations. (c) Differentiation of the gene expression patterns of 735 genes for which genetic assimilation of the expression levels occurred in invasive *Cottus*. At 14°C, 17°C and 21°C gene expression differentiation is in agreement with taxonomy. The branch separating the invasive Sieg population from the populations of the parent species is much longer than in part (a) and (b) of this figure. At 25°C, differentiation is reduced compared to lower temperatures and does not reflect taxonomic relationship anymore.

profile differentiation observed at lower temperatures. This effect is also observed at 25°C where gene expression differentiation does not reflect taxonomic relationship anymore. Thus the most pronounced expression profile differentiation is observed at temperatures $\geq 14^\circ\text{C}$ and $< 21^\circ\text{C}$. According to the environmental monitoring data (supplementary file 2), such conditions occur in both habitats during large parts of the summer (June to August), but also during late spring (Mai) and early autumn (September). In order to identify how RN evolution contributed to the observed expression profile differentiation between species and populations, neighbor joining trees were prepared from the sets of candidate genes for genetic assimilation or RN widening in the invasive (Table 1). Figure 2b illustrates the expression profile differentiation of the 255 genes for which genetic assimilation has occurred in the invasive *Cottus* (Table 1a). At 14°C and 17°C neighbor joining trees are in agreement with taxonomy. At 21°C and 25°C the invasive Sieg populations branches from the populations of *C. perifretum*. Figure 2c displays the expression profile differentiation of the 735 genes for subject to RN widening. Here at all temperatures the branches of the invasive Sieg population are much longer than in Figure 2a as well as in Figure 2b. This indicates that genes for which RN widening has occurred differentiate the expression profile of the invasive *Cottus* from the parent species more than average (Figure 2a) or genetically assimilated genes (Figure 2b). Genes subject to RN widening were found to be enriched for different GO terms depending if they had a positive or a negative RN in invasive *Cottus* (Table 1b). This suggests that these two gene sets reflect different processes. Therefore, we refined our analysis by preparing separate neighbor joining trees based on these two subsets of genes with RN widening from Table 1b (supplementary file 5). This indicates that for the 320 genes with negative RNs in invasive *Cottus*, expression profile differentiation of the invasive *Cottus* is most pronounced at 14°C and 17°C. For the 415 genes with positive RNs in invasive *Cottus*, expression profile differentiation is observed at higher temperatures (17°C to 25°C) (supplementary file 5).

Discussion

Phenotypic plasticity has been repeatedly considered a factor that permits fast adaptation and promotes diversification (Pfennig *et al.* 2010, Moczek *et al.* 2011). In this study, we assessed thermal plasticity of gene expression in the context of an ongoing ecological speciation event in *Cottus*. We identified plastic responses in relation to temperature that are shared between all populations in analysis. Functional characterization of the respective genes related them to the processes of cold induced angiogenesis. Our study

also revealed a list of genes with expression patterns that were in line with the hypothesis that genetic assimilation and RN widening occurred during the evolution of the invasive *Cottus*. In contrast to our original expectation, the clearest expression profile differentiation was found at the temperatures $< 21^{\circ}\text{C}$, and thus at temperatures observed in the newly invaded and the originally occupied habitat. For this temperature range, RN widening of a mitochondrion related phenotype contributed to the phenotypic differentiation of invasive *Cottus* from the parent species. Our data thereby support the notion that RN-evolution can contribute to phenotypic novelty of potential adaptive relevance. Most studies of transcriptional RNs analyze a reduced set of candidate genes under two environmental conditions and use a linear model to describe expression levels as a function of genotype and environment (e.g. Côté *et al.* 2007, Evans & Bernatchez 2012). In contrast, this study was highly explorative with candidate genes and relevant temperatures not exactly known beforehand. Extending a modeling approach to our study, covering 8047 genes and four temperatures, would have required an increase in replication (van de Pol 2012) beyond what is justifiable for such an exploratory study. Instead we used simple statistics, which while being less quantitative than a modeling approach archived a successful analysis of gene expression plasticity differentiation. Since the recent increase in interest into evolvability of thermal tolerance in the context of climate change (Chevin *et al.* 2010, Somero 2010, Hoffmann and Sgrò 2011) the feasibility and simplicity of our approach might be also of interest beyond the field of speciation research.

Characterization of thermal plasticity in *Cottus*

Invasive *Cottus* are not distinguished from the parental populations by the number of genes showing thermal plasticity (Fig 1). In contrast, functional characterization of the common and population specific plastic responses (supplementary file 4) provided evidence for differences in the plastic responses of invasive *Cottus* and its parent species. For the subsets of genes with positive RNs (Fig 1 c and d), GO term enrichment for the GO term "DNA metabolic process" revealed a shared functional link between those genes with positive RN in the invasive Sieg population, which distinguishes it from the population of its parent species except the WN population.

A plastic response shared by all *Cottus* was identified through GO term enrichment of the terms 'proteinaceous ECM', 'embryo development' and 'structural molecule activity' for the genes with negative RNs in the all three populations in both populations comparisons (Fig 1 a and b). Czypionka *et al.* (2012) found enrichment for the same GO terms in a set of

candidate genes which were up regulated in *Cottus* in response to environmental factors and speculated that this GO term enrichment might capture signal from blood-vessel formation. Indeed, a review of the genes underlying this GO term enrichment in the temperature gradient experiment included several genes (for example *hypoxia-inducible factor*, *vascular endothelial growth factor* and several angiopoetins; see supplementary file 3) with a known role in promoting angiogenesis (Pugh & Ratcliffe 2003). All these genes were characterized to have a negative RN, i.e. to be higher expressed at the lower temperatures in analysis. This finding is therefore in agreement with the scenario that these GO terms capture signal from the process of “cold induced angiogenesis” (Egginton and Cordiner 1997). Briefly, this process consists in the increase of the number and density of blood vessels, which allows for an improved supply of blood, and thereby of oxygen and other metabolites to the tissues. This thermal compensation is therefore thought to be adaptive by maintaining the aerobic scope over different temperatures (Egginton 2002).

Another plastic response towards temperature common to all *Cottus* populations was indicated by enrichment of the GO term “generation of precursor metabolites and energy” in the genes sharing negative RN in all populations and in both population comparisons (Fig 1 a and b). Enrichment of this GO term was mainly caused by genes forming a part of the respiratory chain (e.g. NADH dehydrogenase, cytochrome c, isocitrate dehydrogenase, various ATPsynthase subunits; see supplementary file 3). Significant enrichment for the GO term “generation of precursor metabolites and energy” and “mitochondrion” was also detected for the genes being expressed exclusively in the invasive Sieg population, but also in the sets of genes with negative RNs only in Sieg and Broel (Fig 1 a) or in Sieg and LB (Fig 1 b). We interpret this as the manifestation of a single mitochondrion related phenotype with variation in the amount of plastic response between different populations. Interestingly, all the gene sets with the corresponding GO term enrichment consist of genes having a negative RN in the invasive Sieg population, (see supplementary file 4). Plasticity for this mitochondria related phenotype therefore seems to be most pronounced in the invasive Sieg population. Because RNA was extracted from the entire fish, the changes in the amount of mitochondria specific transcripts could be caused by an increase of expression of these genes or an increase in the number of mitochondria. Both, an increase in mitochondrial volume density per tissue or the compensatory modification of mitochondrial properties have been reported to be induced by lower temperatures in fishes (reviewed in Guderley 2004). These thermal compensations are thought to modify aerobic metabolic capacity (O’Brien 2011). This

finding therefore hints to an interesting difference in plasticity for a higher order phenotype, which is of relevance for the thermal ranges in which ectotherms and especially fishes can thrive (Fangue *et al.* 2009, Pörtner *et al.* 2010).

Evidence for proposed modes of RN evolution

The second objective of this work was to assess whether proposed modes of RN evolution, i.e. genetic assimilation or respectively RN widening, occurred during the formation of the invasive *Cottus* from its ancestral parent species. Indeed, a number of genes for which genetic assimilation or RN widening had occurred in invasive *Cottus* were identified (Table 1a). RN widening seems to be the more significant of the two processes, not only because of the higher number of genes, but also because GO term enrichment suggested a functional cohesion between the genes with a widened RN in invasive *Cottus* (Table 1a). The separate GO term enrichment analysis of genes with a widened positive RN and a widened negative RN in invasive *Cottus* (Table 1b) recaptured GO terms discussed above as characteristic for specific plastic responses of invasive *Cottus* (supplementary file 4). The GO term “nucleobase containing compound metabolic process” enriched in the genes with a widened positive RN in invasive *Cottus* (Table 1b) is a parent term of the GO term “DNA metabolic process” and therefore most likely represents the same phenotype. GO terms enriched in those genes with widened negative RN in invasive *Cottus* (Table 1b) contained many genes of the respiratory chain (supplementary file 4) and refer to the mitochondrion related phenotype discussed in the previous paragraph.

The classification of genes as plastic in Fig 1 depended on the expression level change over temperature being larger than a gene specific threshold. This approach thus only permitted to detect RN widening if the increase in plasticity entailed a change in categorization. Such a situation is indeed observed for the phenotypes identified with a widened RN in the invasive (Table 1b). The corresponding GO terms are found enriched in genes displaying transcriptional plasticity exclusively in the invasive Sieg population or in the invasive Sieg population and a parent species population (supplementary file 4). Like also discussed above, this suggests a more pronounced plasticity for these traits in the invasive *Cottus* compared to its parent species. In summary, there is evidence from two independent analyses, the population specific categorization of genes as plastic (supplementary file 4) and specific detection of RN widening (Table 1b), indicating that RN widening occurred in invasive *Cottus* for a mitochondrion related phenotype and a phenotype causing functional enrichment for GO terms like “DNA metabolic process” and

“nucleus”.

Expression profile differentiation over temperature

In this experiment, we assessed gene expression under typical summer temperatures, which differ between rhithral and potamal. As explained in the introduction, we expect a functional differentiation between invasive *Cottus* and its ancestral parent species especially at those temperatures found exclusively in the potamal habitat. The temperature range covered by us ranged from 14°C, a temperature commonly exceeded in summer in both rhithral and potamal to 25°C, a temperature which can be observed in the potamal (see supplementary file 2). Since we covered this range by four different temperatures, we can narrow down the temperature range in which expression profile differentiation between the *Cottus* parent species and invasive *Cottus* is most pronounced. Temperature specific expression profile differentiation between populations is illustrated in Fig 2a. The distance matrix on which the neighbor joining tree in Fig 2a is built is calculated from distance measurement between populations over all genes in analysis, wherefore it is relatively robust to experimental noise. Indeed, in Fig 2a populations maintained at 14°C to 21°C are separated according to their taxonomic relationship, with populations of the same species being more similar in their expression profile than populations from the other species and with the hybrid species taking an intermediate position. The situation at the lower range of analyzed temperatures are therefore in agreement with the results from Czypionka *et al.* 2012, who also observed overall expression profile differentiation in agreement with taxonomy. For the temperatures above 20°C, which capture the situation in the potamal (Illies 1961), this situation changes. At 21°C, branches separating the species are already much shorter than at the lower temperatures, while at 25°C (Fig 2a) the expression profile differentiation between temperatures does not recapture their taxonomic relationships. Accordingly, populations lose their species specific gene expression profiles. Such a pattern is in agreement with a scenario of phenotypic decanalization and the release of cryptic variation due to environmental perturbation (Masel & Siegel 2009, Rutherford 2000). Vice versa, shared species specific expression profiles at lower temperatures could be interpreted as shared adaptations to past selection pressures (Ghalambor *et al.* 2007). While species specific expression patterns can also result from non-adaptive processes like drift and this situation might also apply in *Cottus* (Czypionka *et al.* 2012), such non-adaptive phenotypic divergence should not be correlated with certain temperatures corresponding to ancestral experienced environments of *Cottus*. Therefore we propose

that adaptive relevant functional differentiation is found at $< 21^{\circ}\text{C}$. We can of course not exclude the possibility that relevant functional differentiation not captured in the overall expression profile also occurred at temperatures $> 21^{\circ}\text{C}$. Still our conclusion makes sense considering the temperature data for the potamal Sieg river (supplementary file 2), where temperatures fluctuate most of the time between 16 and 21°C and temperatures higher than 20°C occurred for maximal periods of 5 days. During our experiment, *Cottus* from all populations in analysis survived a 2 week exposure to 21°C (except for one individual of the LB population) and 25°C and did not show any signs of distress. It is therefore plausible that short periods of higher temperatures are tolerated by the ancestral and the invasive *Cottus* and that the actual functional differentiation and potential adaptation of the invasive *Cottus* to the potamal habitat occurs mostly at temperatures $< 21^{\circ}\text{C}$ commonly encountered in both rhithral and potamal. Based on this finding we reject our third hypothesis that the strongest signature of functional differentiation is found at the higher temperatures characteristic for the potamal habitat during summer. Instead, given the environmental monitoring data (supplementary file 2), the functional differentiation between invasive *Cottus* and its parent species should not only play a role under conditions observed during large parts of the summer (June to August), but already during late spring (\sim Mai) and early autumn (\sim September).

To test the fourth hypothesis whether proposed modes of RN-evolution contribute to this differentiation we calculated neighbor-joining trees illustrating the expression differentiation based on the genes for which genetic assimilation or RN widening was detected (Table 1a). Figure 2b illustrates the expression level differentiation for the 255 genes for which genetic assimilation was detected in invasive *Cottus*. At lower temperatures (14°C and 17°C) expression level differentiation recaptures taxonomy as seen in Fig 2a. At higher temperatures (21°C and 25°C) invasive *Cottus* cluster with *C. perifretum*. However, the contribution of genetic assimilation (Fig 2b) to the overall gene expression differentiation of the invasive Sieg population from its parent species seems marginal when compared to the corresponding figure for 735 genes showing RN widening in the invasive *Cottus* (Fig 2c). In Fig 2c, the branches separating the invasive Sieg population from the populations of the parent species are also much longer than in Figure 2a. This holds true for the temperatures from 14°C to 21°C but not at 25°C , where overall expression differentiation in these genes is reduced (Figure 2c). Bearing in mind that we identified expression profile differentiation to be most pronounced at temperatures $< \sim 21^{\circ}\text{C}$, we accept the hypothesis that a RN evolution, especially RN widening, contributed to the differentiation of invasive *Cottus* from its ancestral parent species. It is encouraging,

that the 735 genes for which RN widening was detected are in part functionally correlated as evidenced by GO term enrichment (Table 1a), suggesting also a phenotypic relevance of this pattern.

The distinction of these 735 genes into genes, which have a negative or a positive RN, refined the results of the GO term enrichment (Table 1b, discussed in the previous paragraph). Therefore neighbor-joining trees analogous to Fig 2 were also built for the two subsets from Table 1b. The corresponding figure (supplementary file 5) shows that for the 415 genes with a widened RN and a positive slope in invasive *Cottus*, differentiation is most pronounced at the temperatures from 17°C towards 25°C. In contrast, the 320 with a widened RN and a negative slope in invasive *Cottus* are most differentiated at 14°C and 17°C. These are the temperatures at which we propose that functional differentiation and potential adaptation of invasive *Cottus* occurred (note this also may apply to temperatures >17°C and < 21°C, which were not analysed). GO term enrichment analysis for these 320 genes of potential adaptive relevance (Table 1b) related this pattern to the previously mentioned mitochondrion related phenotype.

Higher order phenotypes and future directions

In this paper, we assessed thermal plasticity in gene expression and identified population- and species specific as well as shared plastic responses in *Cottus*. We also could narrow down the temperature range under which expression profile differentiation between invasive *Cottus* and its ancestral parent species occurred. Our analysis further suggests that proposed models of RN-evolution, especially RN-widening (Lande 2009) do apply here and contribute to this expression profile differentiation. By reviewing the annotation and meta-annotation (i.e. GO terms) of the genes in question, we are able propose higher order phenotypes, i.e. phenotypes above the transcriptome level of potential adaptive value. The identification of such phenotypes would provide an important step to further our understanding about the role of plasticity evolution in ecological speciation (Moczek *et al.* 2011). As a next step, these findings have to be verified and characterized in more detail by follow-up experiments to clearly understand their function and mechanistic role in adaptation (Dalziel *et al.* 2009) to an environment with a new thermal regime. The influence of temperature on the distribution of ectotherms is described by the concept of oxygen and capacity limited thermal tolerance (OCLTT) (Pörtner & Knust 2007), which states that the key factor determining a species' thermal range is its capability to supply sufficient oxygen to its tissues at the respective temperatures. The OCLTT might apply especially at the transition from rhithral to potamal which is not only characterized by

differences in temperature, but also in oxygen saturation, which while being always high in the rhithral streams can be reduced at the bottom layer of potamal rivers (Illies 1961). Above we suggested that the process of 'cold induced angiogenesis' (Egginton and Cordiner 1997) plays a role in *Cottus*. Also, we found evidence that an increase in plasticity of a higher order phenotype related to mitochondria occurred in invasive *Cottus* and contributes to its phenotypic divergence from its parent species. Indeed, both suggested phenotypes are known as compensatory adaptations to deal with changes in environmental temperature or oxygen concentration (Storz *et al.* 2010). Increased capillary growth, described above for the process of cold induced angiogenesis, displays a mean to increase O₂ diffusion capacity of the tissue (Storz *et al.* 2010). This could be experimentally verified by comparing the capillary to fiber ratio of different tissues (Egginton 2002) between invasive and ancestral *Cottus*. Mitochondria show many different plastic compensatory modifications (reviewed Guderley 2004). These include mitochondrial volume density in muscle fibers, which is found to be especially plastic in moderately active demersal fish like *Cottus* (Guderley 2004) or other adaptations like changes in mitochondrial capacity itself, which were observed for short-horned sculpin *Myoxocephalus scorpius* (Guderley and Johnston 1996). Further experiments, including the measurement of physiological parameters and structural changes in tissues should address, which of the manifold known compensatory adaptations (O'Brien 2011, Seebacher *et al.* 2010) are realized in the *Cottus* system. It is reassuring that Dhillon und Schulte (2011) observed intraspecific variation in the thermal plasticity of mitochondria in killifish, which they relate to ecological and especially temperature differences of their respective habitats. This illustrates that such plasticity in mitochondrial amount can evolve (fast) within species in relation to environmental conditions, making our inferences for the *Cottus* system plausible.

Acknowledgements

We thank F. Volckaert, K. de Gelas, A. Kobler and L. Bervoets for their benevolent help in obtaining breeding stocks of *Cottus perifretum*. K. Lessenich and A. Mellin from Bezirksregierung Köln, Nordrhein - Westfalen and T. Heilbronner from the Siegfischereigenossenschaft have given permits to conduct the research. *Cottus* were kept in the laboratory of the MPI in Plön with permission from S. Hauschildt (Veterinaeramt in Plön, Schleswig – Holstein). Harald Deiwick constructed the gradient aquarium. E. Blohm-Sievers, E. Bustorf, S. Frese, H. Harre, G. Augustin, D. Mertens have contributed to the laboratory work and fish care. We thank 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.

Conclusion

In this thesis the role of hybridization and the special role of changes in phenotypic plasticity in the formation of an invasive lineage of *Cottus* were investigated.

In the first chapter, I presented evidence which suggests that hybridization induced phenotypic variation, in particular transgressive expression patterns, partially explain the characteristic phenotype of the invasive hybrid lineage. In the second chapter, I showed that RN-widening (*sensu* Lande 2009) of a pre-existing plastic phenotype has contributed to the phenotypic differentiation of the invasive hybrid lineage from its parent species. The conclusions which can be derived from these findings individually were already discussed in the two chapters themselves. Here, I extend these insights by setting the separate findings gained in both chapters in relation to each other and to the current state of research on the *Cottus* system.

Relating the results of the two chapters

Given that the two chapters of this thesis were designed to address different aspects of the transcriptome differentiation in *Cottus*, it is reassuring that important results from both chapters corroborate each other. For example, in the first chapter a set of 580 genes was shown to be transgressively overexpressed in the invasive lineage. For 192 of those genes, evidence suggested that transgressivity was due to environmental effects not removed by our normalization approach. These 192 genes with a suggested plastic component in their expression patterns were enriched in the GO-term 'proteinaceous extracellular matrix' among others. In chapter 2, enrichment for the same GO-term was indeed found to be characteristic for a plastic response in *Cottus*, namely for those genes with a negative RN in all populations in analysis. On the other hand, in the second chapter a mitochondria related phenotype was identified as a main candidate for the differentiation between the invasive hybrid lineage and its parent species. This phenotype was not detected in the analysis of the first chapter. Six of the eight invasive *Cottus* analyzed in the first paper were derived from the wild. These fish were acclimatized to 17°C in our aquaria for one week. Still, like seen for the 192 genes discussed above, acclimatization and normalization did not remove all effects of environment. We can therefore not exclude the possibility that environmental effects, for which our normalization procedure did not account, also impaired our the detection of the mitochondrion related phenotype in the first chapter. In conclusion, this underlines the need to account for the effects of environment in experimental design. Specifically, environmental conditions, under which

the relevant phenotypic differentiation between the invasive lineage and the parent species occurs, need to be identified.

Further characterization of relevant environmental conditions

The need to identify environmental relevant parameters was already addressed in the second chapter, where evidence suggesting that phenotypic differentiation occurs especially at temperatures below 21°C was presented. Still, as already discussed in the same chapter, our analysis could have missed relevant phenotypic differentiation which did not manifest in the overall transcriptome differentiation between the species. This for example could be phenotypic differences dependent on expression level changes of only very few genes, or phenotypic differences, which do not result from changes in expression level but from differences in the coding sequence. As a next step, the results from chapter two should be complemented by other methods than expression profile analysis. A set of appropriate methods was presented in Walsh *et al.* (1997). These authors analyzed thermal tolerance and metabolism (in particular oxygen consumption) in various species of North American fresh water sculpins. Their methods were specifically tailored to address differences in the distribution of various species of *Cottus* between spring and stream habitats. Thus they should be applicable for our study system. Like discussed in chapter two of this thesis, the ability to at least passively tolerate the imposed temperature treatments for the period of the experiments was observed in all *Cottus* populations in analysis. However, this does not mean that all populations can cope with thermal challenges equally well. Analysis of differences in thermal tolerance should be extended to incorporate a measure of organismal performance. A widely used biochemical index for growth but also for general health condition is given by the RNA-DNA-ratio (Grant 1996). This ratio has been shown to reflect limitations in growth and metabolism due to temperature in rainbow trout (Mathers *et al.* 1993) and also was shown to work for many other freshwater fishes including *Cottus* (Bennett & Janz 2007). The RNA-DNA-ratio therefore displays an appropriate measure to address subtle differences in thermal tolerance between the invasive hybrid lineage and its parent species.

Integration of results in the Cottus research program

This thesis constitutes the first systematic phenotypic assessment of the invasive hybrid lineage, its parent species, but also of the initial stages of a hybrid swarm which were assessed through F₂ crosses between parent species in the first chapter. It therefore closes an important gap in the research program on *Cottus*, which so far was almost

exclusively based on the analysis of neutral markers. Both chapters revealed sets of candidate genes whose expression patterns differentiated the invasive hybrid lineage from its parent species. Functional analysis of these candidate genes suggested several higher order phenotypes, like e.g. the mitochondrial related phenotype in chapter two, which are of potential adaptive relevance. These suggested higher order phenotypes should be confirmed by follow up experiments like discussed in chapter two of this thesis. The successful conformation and characterization of such higher order phenotypes would be a further step towards the integration of various layers of analysis (genome, transcriptome, higher order phenotypes). Our lab has developed expertise in quantitative trait locus (QTL) mapping (Cheng & Nolte in preparation). This technique could be used to reveal the genetic architecture controlling these traits, but also to map the loci causing the underlying changes in gene expression, i.e. identify relevant expression QTLs (Whiteley *et al.* 2008).

Using a set of 81 ancestry informative markers Stemshorn *et al.* (2011) estimated that approximately 2.5 % of the genome of the invasive *Cottus* and required to maintain the hybrid lineage as a whole. Meanwhile, with the *Cottus* genome being sequenced for several individuals of the invasive lineage, corresponding ancestry informative screens will enable to characterize such regions in even greater detail (Nolte in preparation). As a long term perspective, the results of QTL mapping for certain phenotypes of potential adaptive value should be integrated with information of loci of fixed ancestry in the invasive hybrid lineage, whose retention is suggested to result from selection (Stemshorn *et al.* 2011). These data would allow to explicitly test for a role of complementary gene action (Rieseberg *et al.* 2003a) in the emergence of transgressive phenotypes of adaptive value in the invasive lineage. The data on phenotypic differentiation between the invasive hybrid lineage and its parent species presented in this thesis provide an important step towards this goal.

Bibliography

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Väinölä R, Wolf JBW, Zinner D (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.
- Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC (2010) Homoploid hybrid speciation in action. *Taxon*, **59**, 1375–1386.
- Ainouche ML, Jenczewski E (2010) Focus on polyploidy. *New Phytologist*, **186**, 1–4.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.
- Angilletta MJJ (2009) *Thermal Adaptation: A Theoretical and Empirical Synthesis*. Oxford University Press, USA.
- Arnold ML (2007) *Evolution Through Genetic Exchange*, 1st ed. Oxford University Press, USA.
- Aubin-Horth N, Renn SCP (2009) Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. *Molecular Ecology*, **18**, 3763–3780.
- Barrett RDH, Rogers SM, Schluter D (2009) Environment specific pleiotropy facilitates divergence at the Ectodysplasin locus in threespine stickleback. *Evolution; International Journal of Organic Evolution*, **63**, 2831–2837.
- Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.
- Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.
- Barton NH, Briggs DEG, Eisen JA (2007) *Evolution*. Cold Spring Harbor Laboratory.
- Bell G (2012) Evolutionary rescue and the limits of adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**.
- Bell G, Gonzalez A (2011) Adaptation and Evolutionary Rescue in Metapopulations Experiencing Environmental Deterioration. *Science*, **332**, 1327–1330.
- Bennett PM, Janz DM (2007) Seasonal changes in morphometric and biochemical endpoints

- in northern pike (*Esox lucius*), burbot (*Lota lota*) and slimy sculpin (*Cottus cognatus*). *Freshwater Biology*, **52**, 2056–2072.
- Bernatchez L, Renaut S, Whiteley AR, Derome N, Jeukens J, Landry L, Lu G, Nolte AW, Østbye K, Rogers SM, St-Cyr J (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1783–1800.
- Bougas B, Granier S, Audet C, Bernatchez L (2010) The transcriptional landscape of cross-specific hybrids and its possible link with growth in Brook Charr (*Salvelinus fontinalis* Mitchill). *Genetics*, **186**, 97–107.
- Brett JR (1971) Energetic Responses of Salmon to Temperature. A Study of Some Thermal Relations in the Physiology and Freshwater Ecology of Sockeye Salmon (*Oncorhynchus nerka*). *American Zoologist*, **11**, 99–113.
- Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity*, **84**, 441–451.
- Carroll SP, Hendry AP, Reznick DN, Fox CW (2007) Evolution on ecological time-scales. *Functional Ecology*, **21**, 387–393.
- Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, Myers RM, Petrov D, Jónsson B, Schluter D, Bell MA, Kingsley DM (2010) Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science*, **327**, 302–305.
- Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. *Nature Reviews Genetics*, **10**, 783–796.
- Chevin L-M, Lande R, Mace GM (2010) Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. *PLoS Biol*, **8**, e1000357.
- Chevireon ZA, Whitehead A, Brumfield RT (2008) Transcriptomic variation and plasticity in rufous-collared sparrows (*Zonotrichia capensis*) along an altitudinal gradient. *Molecular Ecology*, **17**, 4556–4569.
- Comai L, Madlung A, Josefsson C, Tyagi A (2003) Do the different parental “heteromes” cause genomic shock in newly formed allopolyploids? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **358**, 1149–1155.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

- Côté G, Perry G, Blier P, Bernatchez L (2007) The influence of gene-environment interactions on GHR and IGF-1 expression and their association with growth in brook charr, *Salvelinus fontinalis* (Mitchill). *BMC Genetics*, **8**, 87.
- Coyne J, Orr H (2004) *Speciation*. Sinauer Associates Inc., USA
- Crispo E (2007) The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution*, **61**, 2469–2479.
- Czypionka T, Cheng J, Pozhitkov A, Nolte AW (2012) Transcriptome changes after genome-wide admixture in invasive sculpins (*Cottus*). *Molecular Ecology*, **21**, 4797–4810.
- Dalziel AC, Rogers SM, Schulte PM (2009) Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Molecular Ecology*, **18**, 4997–5017.
- David JR, Gibert P, Moreteau B (2004) Evolution of reaction norms *In Phenotypic Plasticity: Functional and Conceptual Approaches* (DeWitt & Scheiner eds.) pp. 50–63. Oxford University Press.
- DeWitt TJ, Scheiner SM (2004) *Phenotypic Plasticity: Functional and Conceptual Approaches*. Oxford University Press.
- Dhillon RS, Schulte PM (2011) Intraspecific variation in the thermal plasticity of mitochondria in killifish. *The Journal of Experimental Biology*, **214**, 3639–3648.
- Dieckmann U, Doebeli M, Metz JAJ, Tautz D (2004) *Adaptive Speciation*. Cambridge University Press.
- Domazet-Loso T, Tautz D (2010) A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*, **468**, 815–818.
- van Dyken JD, Wade MJ (2009) Quantifying the Evolutionary Consequences of Conditional Gene Expression in Time and Space. *Genetics*, **184**, 439–453.
- Egginton S (2002) Temperature and angiogenesis: the possible role of mechanical factors in capillary growth. *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology*, **132**, 773–787.
- Egginton S, Cordiner S (1997) Cold-Induced Angiogenesis in Seasonally Acclimatized Rainbow Trout (*Oncorhynchus Mykiss*). *Journal of Experimental Biology*, **200**, 2263–2268.
- Eliason EJ, Clark TD, Hague MJ, Hanson LM, Gallagher ZS, Jeffries KM, Gale MK, Patterson DA, Hinch SG, Farrell AP (2011) Differences in thermal tolerance among Sockeye Salmon populations. *Science*, **332**, 109–112.
- Englbrecht CC, Freyhof J, Nolte AW, Rassmann K, Schliewen U, 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.
- Evans ML, Bernatchez L (2012) Oxidative phosphorylation gene transcription in whitefish species pairs reveals patterns of parallel and nonparallel physiological divergence. *Journal of Evolutionary Biology*, **25**, 1823–1834.
- Fangue NA, Richards JG, Schulte PM (2009) Do mitochondrial properties explain intraspecific variation in thermal tolerance? *Journal of Experimental Biology*, **212**, 514–522.
- Fisher WE, Berger DH (2003) Angiogenesis and antiangiogenic strategies in pancreatic cancer. *International Journal of Gastrointestinal Cancer*, **33**, 79–88.
- Fontdevila A (2005) Hybrid genome evolution by transposition. *Cytogenetic and Genome Research*, **110**, 49–55.
- Freyhof J, Kottelat M, Nolte AW (2005) Taxonomic diversity of European *Cottus* with description of eight new species (Teleostei: Cottidae). *Ichthyological Exploration of Freshwaters*, **16**, 107–172.
- Gamperl AK, Farrell AP (2004) Cardiac plasticity in fishes: environmental influences and intraspecific differences. *The Journal of Experimental Biology*, **207**, 2539–2550.
- Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, **21**, 394–407.
- Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*, **17**, 167–178.
- Gracey AY, Fraser EJ, Li W, Fang Y, Taylor RR, Rogers J, Brass A, Cossins AR (2004) Coping with cold: An integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 16970–16975.
- Grant GC (1996) RNA-DNA ratios in white muscle tissue biopsies reflect recent growth rates of adult brown trout. *Journal of Fish Biology*, **48**, 1223–1230.
- Guderley H (2004) Metabolic responses to low temperature in fish muscle. *Biological reviews of the Cambridge Philosophical Society*, **79**, 409–427.
- Guderley H, Johnston I (1996) Plasticity of fish muscle mitochondria with thermal acclimation. *Journal of Experimental Biology*, **199**, 1311–1317.

- Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ (2009) Extreme changes to gene expression associated with homoploid hybrid speciation. *Molecular Ecology*, **18**, 877–889.
- Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ (2006) Transcriptome Shock after Interspecific Hybridization in *Senecio* Is Ameliorated by Genome Duplication. *Current Biology*, **16**, 1652–1659.
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP (2006) A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science*, **313**, 101–104.
- Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–485.
- Illies J (1961) Versuch einer allgemeinen biozönotischen Gliederung der Fließgewässer. *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, **46**, 205–213.
- Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L (2010) The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Molecular Ecology*, **19**, 5389–5403.
- Kawakami T, Strakosh SC, Zhen Y, Ungerer MC (2010) Different scales of Ty1/copia-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species. *Heredity*, **104**, 341–350.
- Kearney M (2005) Hybridization, glaciation and geographical parthenogenesis. *Trends in Ecology & Evolution*, **20**, 495–502.
- Keller I, Seehausen O (2012) Thermal adaptation and ecological speciation. *Molecular Ecology*, **21**, 782–799.
- Kellermann V, Heerwaarden B van, Sgrò CM, Hoffmann AA (2009) Fundamental Evolutionary Limits in Ecological Traits Drive *Drosophila* Species Distributions. *Science*, **325**, 1244–1246.
- Kingsolver JG, Izem R, Ragland GJ (2004) Plasticity of Size and Growth in Fluctuating Thermal Environments: Comparing Reaction Norms and Performance Curves. *Integrative and Comparative Biology*, **44**, 450–460.
- Knapen D, Knaepens G, Bervoets L, Taylor MI, Eens M, Verheyen E (2003) Conservation units based on mitochondrial and nuclear DNA variation among European bullhead populations (*Cottus gobio* L., 1758) from Flanders, Belgium. *Conservation Genetics*, **4**, 129–140.

- Kottelat M, Freyhof J (2007) *Handbook of European Freshwater Fishes*. Publications Kottelat.
- Laforsch C, Tollrian R (2004) Inducible defenses in multipredator environments: cyclomorphosis in *Daphnia cuculatta*. *Ecology*, **85**, 2302–2311.
- Lai Z, Gross BL, Zou Y, Andrews J, Rieseberg LH (2006) Microarray analysis reveals differential gene expression in hybrid sunflower species. *Molecular Ecology*, **15**, 1213–1227.
- Lai Z, Nakazato T, Salmaso M, Burke JM, Tang S, Knapp SJ, Rieseberg LH (2005) Extensive Chromosomal Repatterning and the Evolution of Sterility Barriers in Hybrid Sunflower Species. *Genetics*, **171**, 291–303.
- Lamatsch DK, Stöck M (2009) Sperm-Dependent Parthenogenesis and Hybridogenesis in Teleost Fishes. *Lost Sex* (eds I. Schön, K. Martens & P. Dijk), pp. 399–432. Springer Netherlands.
- Lande R (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology*, **22**, 1435–1446.
- Landry CR, Hartl DL, Ranz JM (2007) Genome clashes in hybrids: insights from gene expression. *Heredity*, **99**, 483–493.
- Lang M, Murat S, Clark AG, Gouppil G, Blais C, Matzkin LM, Guittard É, Yoshiyama-Yanagawa T, Kataoka H, Niwa R, Lafont R, Dauphin-Villemant C, Orgogozo V (2012) Mutations in the neverland Gene Turned *Drosophila pachea* into an Obligate Specialist Species. *Science*, **337**, 1658–1661.
- Larsen PF, Schulte PM, Nielsen EE (2011) Gene expression analysis for the identification of selection and local adaptation in fishes. *Journal of Fish Biology*, **78**, 1–22.
- Leong HS, Kipling D (2009) Text-based over-representation analysis of microarray gene lists with annotation bias. *Nucleic Acids Research*, **37**, e79.
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Inc., USA
- Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279–283.
- Masel J, Siegal ML (2009) Robustness: mechanisms and consequences. *Trends in Genetics*, **25**, 395–403.
- Mathers EM, Houlihan DF, McCarthy ID, Burren LJ (1993) Rates of growth and protein synthesis correlated with nucleic acid content in fry of rainbow trout, *Oncorhynchus mykiss*: effects of age and temperature. *Journal of Fish Biology*, **43**, 245–263.

- Mavárez J, Audet C, Bernatchez L (2009) Major disruption of gene expression in hybrids between young sympatric anadromous and resident populations of brook charr (*Salvelinus fontinalis* Mitchell). *Journal of Evolutionary Biology*, **22**, 1708–1720.
- Mavárez J, Linares M (2008) Homoploid hybrid speciation in animals. *Molecular Ecology*, **17**, 4181–4185.
- Mavárez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M (2006) Speciation by hybridization in Heliconius butterflies. *Nature*, **441**, 868–871.
- de Meester L, Gómez A, Okamura B, Schwenk K (2002) The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica*, **23**, 121–135.
- Meyer A, Salzburger W, Schartl M (2006) Hybrid origin of a swordtail species (Teleostei: Xiphophorus clemenciae) driven by sexual selection. *Molecular Ecology*, **15**, 721–730.
- Michalak P, Noor MAF (2003) Genome-wide patterns of expression in Drosophila pure species and hybrid males. *Molecular Biology and Evolution*, **20**, 1070–1076.
- Michalak P, Noor MAF (2004) Association of Misexpression with Sterility in Hybrids of Drosophila simulans and D. mauritiana. *Journal of Molecular Evolution*, **59**, 277–282.
- Miller-Pierce MR, Preisser EL (2012) Asymmetric priority effects influence the success of invasive forest insects. *Ecological Entomology*, **37**, 350–358.
- Moczek AP, Sultan S, Foster S, Ledón-Rettig C, Dworkin I, Nijhout HF, Abouheif E, Pfennig DW (2011) The role of developmental plasticity in evolutionary innovation. *Proceedings of the Royal Society B: Biological Sciences*.
- Lucas A (2010) amap: Another Multidimensional Analysis Package. R package version 0.8-5. <http://CRAN.R-project.org/package=amap>
- Nolte AW, Freyhof J, Stemshorn KC, Tautz D (2005) An invasive lineage of sculpins, Cottus sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2379–2387.
- Nolte AW, Freyhof J, Tautz D (2006) When invaders meet locally adapted types: rapid moulding of hybrid zones between sculpins (Cottus, Pisces) in the Rhine system. *Molecular Ecology*, **15**, 1983–1993.
- Nolte AW, Gompert Z, Buerkle CA (2009a) Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.

- Nolte AW, Renaut S, Bernatchez L (2009b) Divergence in gene regulation at young life history stages of whitefish (*Coregonus* sp.) and the emergence of genomic isolation. *BMC Evolutionary Biology*, **9**, 59.
- Nolte AW, Tautz D (2010) Understanding the onset of hybrid speciation. *Trends in Genetics*, **26**, 54–58.
- Normandeau E, Hutchings JA, Fraser DJ, Bernatchez L (2009) Population-specific gene expression responses to hybridization between farm and wild Atlantic salmon. *Evolutionary Applications*, **2**, 489–503.
- O'Brien KM (2011) Mitochondrial biogenesis in cold-bodied fishes. *The Journal of Experimental Biology*, **214**, 275–285.
- Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.
- Parsons KJ, Robinson BW (2006) Replicated evolution of integrated plastic responses during early adaptive divergence. *Evolution; international journal of organic evolution*, **60**, 801–813.
- Pavey SA, Collin H, Nosil P, Rogers SM (2010) The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, **1206**, 110–129.
- Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, **25**, 459–467.
- Phillips PC (1996) Waiting for a compensatory mutation: phase zero of the shifting-balance process. *Genetics Research*, **67**, 271–283.
- Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras J-B, Stephens M, Gilad Y, Pritchard JK (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, **464**, 768–772.
- van de Pol M (2012) Quantifying individual variation in reaction norms: how study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution*, **3**, 268–280.
- Pörtner HO, Knust R (2007) Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science*, **315**, 95–97.
- Pörtner HO, Schulte PM, Wood CM, Schiemer F (2010) Niche dimensions in fishes: an integrative view. *Physiological and Biochemical Zoology: PBZ*, **83**, 808–826.

- Pozhitkov A, Noble PA, Domazet-Lošo T, Nolte AW, Sonnenberg R, Staehler P, Beier M, Tautz D (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Research*, **34**, e66.
- Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends in Plant Science*, **13**, 288–294.
- Pugh CW, Ratcliffe PJ (2003) Regulation of angiogenesis by hypoxia: role of the HIF system. *Nature Medicine*, **9**, 677–684.
- Renaut S, Nolte AW, Bernatchez L (2009) Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Biology and Evolution*, **26**, 925–936.
- Rieseberg LH (1991) Homoploid Reticulate Evolution in *Helianthus* (Asteraceae): Evidence from Ribosomal Genes. *American Journal of Botany*, **78**, 1218–1237.
- Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. *Heredity*, **83**, 363–372.
- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C (2003a) Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. *Science*, **301**, 1211–1216.
- Rieseberg LH, Widmer A, Arntz AM, Burke B (2003b) The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **358**, 1141–1147.
- Rosenthal DM, Rieseberg LH, Donovan LA (2005) Re-creating Ancient Hybrid Species' Complex Phenotypes from Early-Generation Synthetic Hybrids: Three Examples Using Wild Sunflowers. *The American Naturalist*, **166**, 26–41.
- Rosenthal DM, Schwarzbach AE, Donovan LA, Raymond O, Rieseberg LH (2002) Phenotypic Differentiation between Three Ancient Hybrid Taxa and Their Parental Species. *International Journal of Plant Sciences*, **163**, 387–398.
- Rutherford SL (2000) From genotype to phenotype: buffering mechanisms and the storage of genetic information. *BioEssays*, **22**, 1095–1105.
- Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. *Nature*, **396**, 336–342.
- Schlichting C, Pigliucci M (1998) *Phenotypic Evolution: a Reaction Norm Perspective*. Sinauer.

- Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press.
- Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. *Science*, **323**, 737–741.
- Seebacher F, Brand MD, Else PL, Guderley H, Hulbert AJ, Moyes CD (2010) Plasticity of oxidative metabolism in variable climates: molecular mechanisms. *Physiological and Biochemical Zoology: PBZ*, **83**, 721–732.
- Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.
- Seehausen O, Van Alphen JJM, Witte F (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science*, **277**, 1808–1811.
- Shulman MJ, Ogden JC, Ebersole JP, McFarland WN, Miller SL, Wolf NG (1983) Priority Effects in the Recruitment of Juvenile Coral Reef Fishes. *Ecology*, **64**, 1508–1513.
- Smyth G (2005) limma: Linear Models for Microarray Data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* Statistics for Biology and Health. (eds R. Gentleman, V.J. Carey, W. Huber, R.A. Irizarry, S. Dudoit, M. Gail, K. Krickeberg, J. Sarnet, A. Tsiatis & W. Wong), pp. 397–420. Springer New York.
- Somero GN (2010) The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine “winners” and “losers”. *The Journal of Experimental Biology*, **213**, 912–920.
- Stearns SC (1989) The Evolutionary Significance of Phenotypic Plasticity. *BioScience*, **39**, 436–445.
- Stelkens RB, Schmid C, Selz O, Seehausen O (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evolutionary Biology*, **9**, 283.
- Stelkens R, Seehausen O (2009) Genetic distance between species predicts novel trait expression in their hybrids. *Evolution*, **63**, 884–897.
- Stemshorn KC, Reed FA, Nolte AW, Tautz D (2011) Rapid formation of distinct hybrid lineages after secondary contact of two fish species (*Cottus* sp.). *Molecular Ecology*, **20**, 1475–1491.
- Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, **64**, 479–498.
- Storz JF, Scott GR, Cheviron ZA (2010) Phenotypic plasticity and genetic adaptation to high-

- altitude hypoxia in vertebrates. *The Journal of Experimental Biology*, **213**, 4125–4136.
- Taylor EB, Boughman JW, Groenenboom M, Sniatynski M, Schluter D, Gow JL (2006) Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, **15**, 343–355.
- Ungerer MC, Strakosh SC, Zhen Y (2006) Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Current Biology*, **16**, R872–R873.
- Urban MC, de Meester L (2009) Community monopolization: local adaptation enhances priority effects in an evolving metacommunity. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 4129–4138.
- Visser ME (2008) Keeping up with a warming world; assessing the rate of adaptation to climate change. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 649–659.
- Volckaert F, Hänfling B, Hellemanns B, Carvalho G (2002) Timing of the population dynamics of bullhead *Cottus gobio* (Teleostei: Cottidae) during the Pleistocene. *Journal of Evolutionary Biology*, **15**, 930–944.
- Waddington CH (1953) Genetic Assimilation of an Acquired Character. *Evolution*, **7**, 118–126.
- Waddington CH (1956) Genetic Assimilation of the Bithorax Phenotype. *Evolution*, **10**, 1–13.
- Walsh SJ, Haney DC, Timmerman CM (1997) Variation in thermal tolerance and routine metabolism among spring- and stream dwelling freshwater sculpins (Teleostei: Cottidae) of the southeastern United States. *Ecology of Freshwater Fish*, **6**, 84–94.
- Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences*, **103**, 5425–5430.
- Whiteley AR, Derome N, Rogers SM, St-Cyr J, Laroche J, Labbe A, Nolte A, Renaut S, Jeukens J, Bernatchez L (2008) The Phenomics and Expression Quantitative Trait Locus Mapping of Brain Transcriptomes Regulating Adaptive Divergence in Lake Whitefish Species Pairs (*Coregonus* sp.). *Genetics*, **180**, 147–164.
- Wu H (2010). *maanova*: Tools for analyzing microarray experiments. R package version 1.20.0. Modified by Yang H, Sheppard K, with ideas from Churchill G, Kerr K and Cui X. <http://CRAN.R-project.org/package=maanova>
- Xia Q, Cheng D, Duan J, Wang G, Cheng T, Zha X, Liu C, Zhao P, Dai F, Zhang Z, He N, Zhang L, Xiang Z (2007) Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, *Bombyx mori*. *Genome Biology*, **8**, R162.

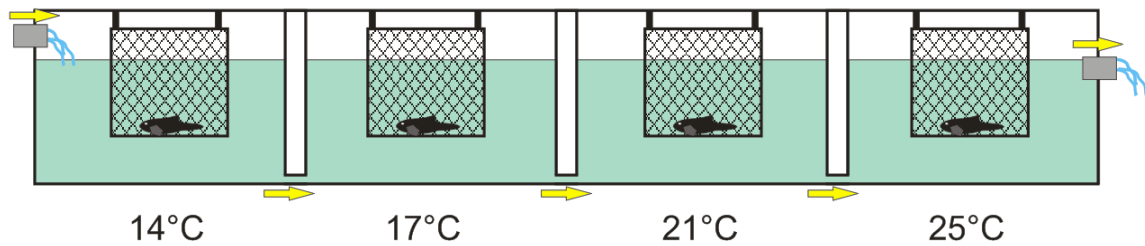
Appendix A – supplementary material from Chapter 1

Supplementary Table 1: Target transcripts, microarray probes and differences in gene expression between *C. rhenanus*, *C. perfretum*, invasive sculpins and laboratory-bred F₂ crosses.

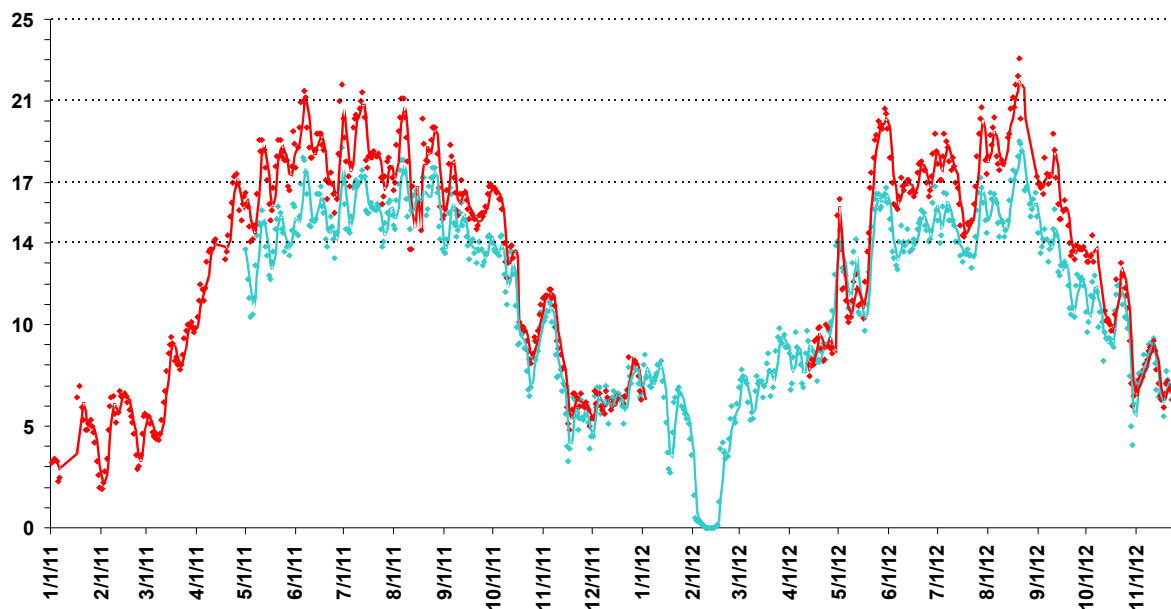
This supplementary table contains 7481 lines and 25 columns and therefore is not suitable for presentation in print. It can be accessed online as supplementary material from Czypionka *et al.* (2012) under:

<http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2012.05645.x/supinfo>

Appendix B – supplementary material from Chapter 2



Supplementary file 1: Sketch of the temperature gradient. The temperature gradient consists of 4 chambers of plexiglass (length: 35.7 cm, depth 38.7 cm, height 24 cm), which are separated by an air-filled plexiglass wall (2.6 cm) for insulation. At the bottom chambers are connected by 3 small tunnels (area ~ 6 cm²). Water (14°C, 0,32 l/min) enters the system at the left chamber. Water passes through the tunnels to the other chambers and leaves the system through an effluent at the chamber at the right. The direction of water flow is indicated by yellow arrows. Tunnels are loosely stuffed with slices of aquarium cleaning sponge material (not indicated in the sketch) in order to slow down the velocity of water exchange and allow the building up of homogenous conditions in the respective chambers. Within the chambers, air stones (not indicated in the sketch) are used to aerate and circulate the water within the chambers. Heating elements (power: 150 W to 200 W; not indicated in the sketch) lying at the bottom of the three chambers on the right are used to adjust the temperature within the chambers to 17°C, 21°C and 25°C respectively. In each chamber, 5 net cages are submerged (only one net cage per chamber is indicated in the sketch). Small (~7 cm * 4 mm) bars of glass were placed in the net cages as a weight to avoid that air bubbles get trapped under the net cages. Fish were transferred and acclimatized to the net cages as described in the methods section.



Supplementary file 2: Water temperature data for the rhithral stream Broel (blue) and the potamal river Sieg (red). Points represent daily average temperatures, lines were calculated as the moving average over 4 days. During summer, temperatures in the Sieg exceed those in the Broel. For the river Sieg, no temperature data were available for the period from 1/1/2012 to 4/11/2012. For the stream Broel, no temperature data were available for the time before 5/1/2011. Therefore, a meaningful comparison for winter temperatures is not possible. Note that the temperature drop to 0°C in the first half of February 2012 corresponds to the freezing point of fresh water and therefore most likely does not represent temperature data for flowing water in the Broel, but a local freeze of the temperature sensor.

Supplementary file 3 is a table with 8047 lines and 19 columns and therefore is not suitable for presentation in print. It is planned to publish chapter 2 in a scientific journal, with supplementary file 3 being made available as online supporting material through the homepage of the journal.

For the purpose of this thesis, supplementary file 3 can be accessed online under:

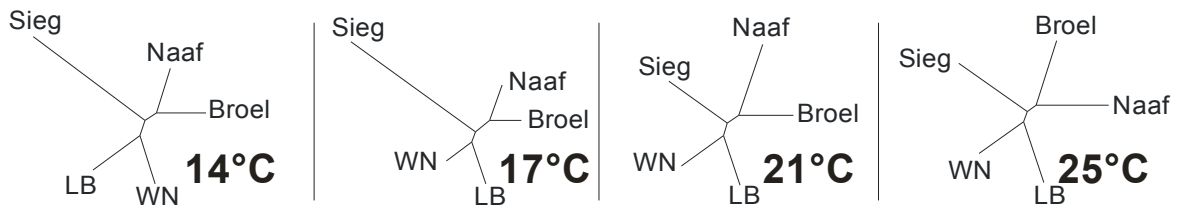
www.evolbio.mpg.de/~czypionka/supplementary_file_thesis/

Supplementary File 3: Overview of those genes for which thermal plasticity in expression level was detected. The first column “ID” is just a numerical denominator. The column “gene_name” provides a unique name for each gene. This information is included so that the results can be related to expression data from Czepionka *et al.* (2012) which was uploaded at NCBI’s Gene Expression Omnibus (accession number GSE36755). The column “probe sequence” provides the sequences of the probes for the individual genes printed on the microarray. The column “target sequence” provides the sequence of the supposed target transcript inferred by RNA sequencing as described in Czepionka *et al.* (2012). The column “best_blast_hit” provides the most significant blast hit of the target sequence during annotation with Blast2GO as described in Czepionka *et al.* (2012). The column “sequence description” contains an easy to read description of the gene that was compiled by Blast2Go based on all the blast hits of the target sequence. The 5 columns “plasticity_Broel”, “plasticity_Naaf”, “plasticity_Sieg”, “plasticity_LB” and “plasticity_WN” provide information about the characterization of transcriptional plasticity and correspond to the situation also illustrated in Fig 1. For each population in analysis, genes with a significant plastic change in gene expression over temperature are reported as “positive RN” if the slope of the RN is positive or respectively as “negative RN”. The column “RN_evolution_Sieg” identifies genes for which a change in plasticity between invasive *Cottus* and the populations of its parent species has been identified and for which GO term enrichment results are reported in Table 1. Genes are either labeled as “RN widening” or “genetic assimilation” depending on which process is observed. Enrichment of different candidate gene sets for the GO terms “DNA metabolic process”, “embryo development”, “generation_of_precursor_metabolites_and_energy”, “mitochondrion”, “ion_transport”, “proteinaceous extracellular matrix (ECM)” and “structural_molecule_activity” played a central role in the analysis. To be able to identify the genes underlying this enrichment, annotation for these GO terms is marked in the last seven columns for the corresponding genes.

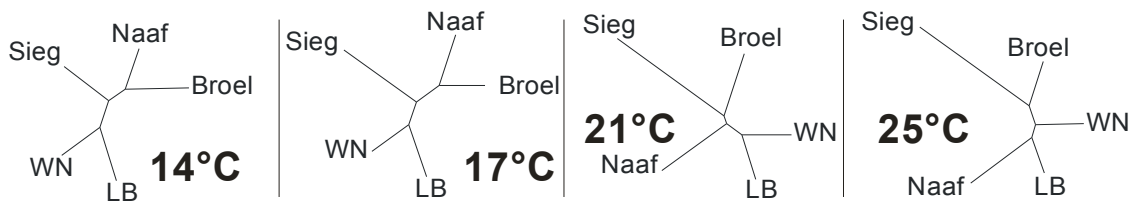
	plasticity detected negative RNs	GO terms enriched with FDR < 0.05	FDR	
Broel-WN-Sieg	in all three populations	proteinaceous extracellular matrix generation of precursor metabolites and energy structural molecule activity ion transport embryonic development mitochondrion extracellular space calcium ion binding ribosome carbohydrate binding	5.2 E-9 1.4 E-4 2.0 E-3 1.4 E-2 1.9 E-2 2.5 E-2 2.7 E-2 2.7 E-2 4.6 E-2 4.6 E-2	
	only in the <i>C.rhenanus</i> population	behavior ion transport cellular homeostasis	4.4 E-2 4.4 E-2 4.4 E-2	
	only in the <i>C.perifretum</i> population	-	-	
	only in the invasive Sieg population	mitochondrion generation of precursor metabolites and energy	4.4 E-6 1.8 E-4	
	only in the <i>C.perifretum</i> population and the invasive Sieg population	cytoplasm	3.0 E-2	
	only in the <i>C.rhenanus</i> population and the invasive Sieg population	ion channel activity generation of precursor metabolites and energy ion transport regulation of biological quality	1.5 E-3 6.9 E-3 1.2 E-2 3.8 E-2	
	only in the two parent species populations	proteinaceous extracellular matrix structural molecule activity	2.9 E-11 2.5 E-2	
	positive RNs			
	in all three populations	-	-	-
	only in the <i>C.rhenanus</i> population	-	-	-
	only in the <i>C.perifretum</i> population	-	-	-
	exclusively in Sieg	-	-	-
only in the <i>C.perifretum</i> population and the invasive Sieg population	DNA metabolic process	4.1 E-2		
only in the <i>C.rhenanus</i> population and the invasive Sieg population	-	-	-	
only in the two parent species populations	-	-	-	
Naaf-LB-Sieg	negative RNs			
	in all three populations	proteinaceous extracellular matrix multicellular organismal development ion transport structural molecule activity generation of precursor metabolites and energy receptor binding	1.7 E-5 3.1 E-3 4.2 E-3 1.8 E-2 3.6 E-2 4.8 E-2	
	only in the <i>C.rhenanus</i> population	-	-	
	only in the <i>C.perifretum</i> population	endoplasmatic reticulum Golgi apparatus	2.0 E-4 3.9 E-2	
	only in the invasive Sieg population	generation of precursor metabolites and energy cell-cell signalling mitochondrion	3.5 E-2 3.5 E-2 3.8 E-2	
	only in the <i>C.perifretum</i> population and the invasive Sieg population	generation of precursor metabolites and energy mitochondrion	6.2 E-10 1.1 E-4	
	only in the <i>C.rhenanus</i> population and the invasive Sieg population	-	-	
	only in the two parent species populations	proteinaceous extracellular matrix	7.1 E-3	
	positive RNs			
	in all three populations	-	-	-
	only in the <i>C.rhenanus</i> population	-	-	-
	only in the <i>C.perifretum</i> population	-	-	-
only in the invasive Sieg population	DNA metabolic process nucleus	2.2 E-3 1.5 E-2		
only in the <i>C.perifretum</i> population and the invasive Sieg population	-	-	-	
only in the <i>C.rhenanus</i> population and the invasive Sieg population	-	-	-	
only in the two parent species populations	-	-	-	

supplementaryfile 4: Results from the GO term enrichment analysis of the different subsets of the Venn diagrams in Fig 1. The table at the top summarizes the results from the population comparison Broel-WN-Sieg (Fig 1a and c). The table at the bottom summarizes the results from the population comparison Naaf-LB-Sieg (Fig 1b and d). The first column (“plasticity detected”) identifies which are the gene sets in analysis. In the column “GO terms enriched with FDR < 0.05”, the significantly enriched (FDR-corrected p-value < 0.05) GO terms (if any) are reported for the respective gene sets. The column FDR provides the significance level (FDR-corrected p-value) for the significant GO term enrichment.

RN widening with invasive negative RN (n = 320)



RN widening with invasive positive RN (n = 415)



Supplementary file 5: Neighbor-joining trees indicating the differentiation of the gene expression patterns between the populations Broel, Naaf (*C. rhenanus*), LB, WN (*C. perifretum*) and Sieg (invasive *Cottus*) at the analyzed temperatures. (a) For 320 genes showing a RN widening in invasive *Cottus* with a negative RN in the invasive population. Invasive *Cottus* are especially differentiated from its parent species at 14°C and 17°C. Expression level differentiation reflects taxonomy at all temperatures. (b) For 415 genes showing a RN widening in invasive *Cottus* with a positive RN in the invasive population. Expression level differentiation reflects taxonomy only at 14°C and 17°C. Differentiation is most pronounced at the higher temperatures (21°C and 25°C) but already starts to increase at 17°C.

Declaration of Author's Contribution

In this thesis, I present the results of my doctoral research. The experimental design was conceived by Dr. Arne Nolte and myself. Practical laboratory work as well as data analysis were carried out mainly by me, but also with help from collaborators. Sonja Dembeck, Heike Harre and Elke Bustorf helped in maintaining the fish stocks. Specific contributions are detailed below for the two chapters.

Chapter 1

This chapter was published in *Molecular Ecology* (Czypionka *et al.* 2012).

F2 crosses were prepared and bred in the lab together with Jie Cheng and Arne Nolte. 454 transcriptome sequencing performed together with Jie Cheng and with help by Heinke Buhtz. Raw sequence analysis of the transcriptome sequencing data was carried out together with Jie Cheng. The microarray design was done by me. Microarray calibration and normalization was implemented and carried out by me, based on concepts developed by and with support from Alexander Pozhitkov. Microarray labwork was done together with Elke Blohm-Sievers. Data analysis was performed by me. Arne Nolte and I wrote the manuscript together.

Chapter 2

The temperature gradient experiment was carried out by me. Total RNA extraction following the temperature experiment was carried out by me together with Elke Bustorf and Sarah Frehse. Microarray labwork was carried out together with Elke Blohm-Sievers. Data analysis was performed by me. The manuscript was written together with Arne Nolte.

Till Czypionka

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 first chapter of this dissertation has been published as an 'Original Article' in *Molecular Ecology* (Czypionka *et al.* 2012). The publication of chapter 2 in a scientific journal is planned as well. This work, neither in parts and nor as a whole, has not been subject to an examination procedure at another university before.

This work has been undertaken in compliance with the German Research Foundations (Deutsche Forschungsgemeinschaft) rules of good academic practice.

Plön, 20th of March 2013

Till Czypionka

Curriculum vitae

Till Czypionka

9th of september 1983

born in Weilburg (Germany)

school

1990 – 1994

primary school "Christian Spielmann Schule", Weilburg

1994 – 1998

gymnasium "Gymnasium Philipinum", Weilburg

1998 – 2002

gymnasium "Jugenddorf-Christopherusschule", Braunschweig (Germany), finishing with Abitur

university

10/2003 – 9/2006

study of biological sciences at the University of Constance (Germany)
degree "Bachelor of Science in Biological Sciences"

10/2006 – 1/2007

Study in the International Study Programme "BioOcean" at the IfM-Geomar in Kiel (Germany)

2/2007 – 6/2007

Study in the International Study Programme "BioOcean" at the *Syddansk Universitet* in Odense (Denmark)

10/2007 - 9/2008

Master of Science thesis at the Ecosystem Functioning lab (Dr.Cristian Vargas), Environmental Sciences Center EULA Chile, *Universidad de Concepción* in Concepción (Chile)

postgraduate education

4/2009 - 6/2009

Internship about hybrid zones of house mouse at the Max Planck Institute for Evolutionary Biology in Plön (Germany)

since 7/2009

Ph.D. - thesis "Evolutionary analysis of gene expression in hybrid sculpins" at the Evolutionary Genetics of Fishes lab (Dr. Arne W. Nolte), Max Planck Institute for Evolutionary Biology in Plön

Acknowledgements

This thesis could not have been finished without the support of many people. Some acknowledgements have already given in the specific chapters. Here, I would like to complement this with some more *global* acknowledgments.

First of all, I would like to thank my supervisor Dr. Arne Nolte. Arne is always willing to share his enthusiasm for evolutionary biology, fish and fishing. He is an outstanding supervisor, who is very reliable and always available for his students.

I also would like to thank Prof. Diethard Tautz. Diethard did not only took the time to patiently listen to the ideas of a fresh-from-university biological oceanographer about how evolution works, but who also gave me the possibility to get started in evolutionary research.

I also thank Prof. Hinrich Schulenburg for his willingness to evaluate this thesis.

Besides acknowledging the bigwigs, I would emphasize and acknowledge the exceptionally good and collegial working atmosphere in the *Cottus* group. Especially I would like to thank Cheng Jie, who in theory would have started his Ph.D together with me. In practice, while I started with a six week sick leave, he started with working for two. Elke Blohm-Sievers, who is not an official part of the *Cottus* group, helped with the microarray analysis, which was not only helpful but indeed a lot of fun.

Many people made my time in Plön a good one. I do not intend to name all of them here. An exception is only granted for the core of the lunch group (Weini & Jens) who provided me not only with friendship, but with key logistic support (food, bed, transport etc.).

Finally, I would like to thank my parents Bodo und Cornelia and my sister Anna.

Maria Czypionka (most likely unintentionally) provided emotional support during the final phase of the thesis writing.