Assessing adaptive genetic variation for conservation and management of the European grayling (*Thymallus thymallus*)

J. V. Huml

PhD 2017

Assessing adaptive genetic variation for conservation and management of the European grayling (*Thymallus thymallus*)

Jana Vanessa Huml

A thesis submitted in partial fulfilment of the requirements of the Manchester Metropolitan University for the degree of Doctor of Philosophy

2017

Faculty of Science and Engineering

Manchester Metropolitan University

Abstract

In this PhD, functional genetic variation of European grayling (*Thymallus thymallus*) is assessed to inform conservation and management of the species. This study is the first to characterize immune variation at the Major Histocompatibility complex (MHC) in grayling. The MHC is a marker of high ecological relevance, because of the strong association between immunity and fitness. Taking advantage of advances in sequencing technology, an analytical pipeline optimized for high-throughput, efficient and accurate genotyping of multi-gene families in non-model species is presented. Immune genetic variation is compared to neutral marker data. Results confirm the hypothesis that neutral marker variation does not predict immune genetic variation. Further, the possible effect of supplementing wild populations with hatchery-reared fish on immune genetic variation is evaluated. Significantly lower estimates of heterozygosity were found in stocked than purely native populations. Lower differentiation at immune genes than at neutral markers are indicative of the effects of balancing selection acting upon the MHC, within purely native, but not stocked populations. Furthermore species distribution modelling is used to identify environmental parameters shaping the distribution of grayling. To evaluate risks imposed by climate change, the sensitivity of grayling to climatic variables and range changes under predicted future scenarios are assessed. Locally-optimised mitigation strategies are shown to increase habitat suitability estimates under conditions of climate change. Evolutionary dynamics between hosts and pathogens are important factors in determining an individual's susceptibility to disease. Studying the microbiome is therefore a promising tool to investigate the risk of disease-mediated extinctions, in relation to environmental conditions and host genetics. This study presents a preliminary analysis on the microbiome of grayling to inform experimental design for future large-scale studies. Altogether, data presented here contribute to improve the management of European grayling. More broadly, it informs conservation research in general, demonstrating the value of taking multiple approaches.

Acknowledgements

I want to thank the Grayling Research Trust and Manchester Metropolitan University for funding and Plymouth University for hosting me as a visiting student. Also, I thank Dr. Jonathan Ellis for proposing this PhD project. Further, I want to thank the Grayling Society, the Environmental Agency of the UK, the Salmon and Trout Conservation Trust and a number of local angling associations for distributing my request for sampling assistance. I want to thank all the anglers around the UK that have offered their help in sampling: Mike Hodell, Mike Pittendreigh, Munro Reid, Nick Chisholm, Paul Deaville, Pete Turner, Peter Bell, Richard Farrant, Rob Hartley, Roger Smith, Ryan Taylor, Simon Evans, Andrew Cartwright, Bob Male, Chris Backhouse, Christopher Hosker, Dave Hartley, David Baird, David Martin, David Rowley, Dr. N. Pringle, Gareth Evans, Mike Bosomworth, Gary Hyde, Greg Payne, James Munden, Jason Harper, John Goodliffe, John Shillcock, Keith Saunders, Mark Whitmarsh, Martin Bamford, Martin James, Michael Pattinson, Stephen & Helen Wilson, Steve Cheetham, Steve Rhodes, Steve Rimmer, Steve Skuce, Stuart Minnikin, Willie Miller, Robin Mulholland, Tony Barnett, Mark Varey, Alan Swann, David Tibbetts, Karl Jones, Stuart Wardle, Richard Cove, Dave Southall, Tim Jacklin, Ian Hawkins and Jeff Body.

Special thanks go to my supervisory team, Dr. Jonathan Ellis, Dr. Ed Harris, Dr. Martin Taylor and Dr. Robin Sen for their guidance, expertise and support. In particular, I want to thank Dr. Jonathan Ellis for his commitment, optimism and patience, which provided me with much support. A lot of helpful inputs were also provided by a number of PhD colleagues and staff both at MMU and Plymouth University, namely I want to thank Dr. Ana Rodiles for introducing me to the QIIME software and Dr. Selvino de Kort for providing me feedback on my progress during annual reviews.

Finally, I want to thank my family and friends for giving me confidence.

Contents

List of Figures

List of Tables

Chapter One: General Introduction

This chapter provides a broad context and review of relevant topics of research that will be addressed again through-out the following chapters of the thesis. After introducing relevant thematic and conceptual topics in turn, information on the biology and management of the study species, the European grayling (*Thymallus thymallus*), is given. Finally, the aims and objectives of each of the chapters are summarized.

Background

Global ecosystem alteration and species extinction make biodiversity maintenance a major challenge (Barnosky et al. 2012). Among the important drivers of anthropogenic biodiversity loss are habitat destruction, exploitation and pollution, which directly result from the growth of the human population (Butchart et al. 2010; Ceballos et al. 2015). Important ecosystem services affected by the loss of biodiversity include nutrient cycling, primary productivity, water supply, pollination and pest control (Keesing et al. 2010; Mace et al. 2012; Cardinale et al. 2012; Dirzo et al. 2014). The positive relationship between biodiversity and ecosystem function is clearly recognized, thus the loss of biodiversity is of immediate concern for human well-being (Cardinale et al. 2012; Ehrlich and Ehrlich 2013).

Biodiversity can be measured and defined at different levels: variation among genes, functional traits or species (Cardinale et al. 2012). There is a close relationship between biodiversity at these different levels of organization from genes to ecosystems (Whitham et al. 2006; Johnson and Stinchcombe 2007). For example local genetic variation within one species of the community can affect diversity and composition of other dependent community members or drive processes of co-adaptation or co-speciation (Wimp et al. 2004; Wade 2007). It is therefore generally acknowledged that the preservation of genetic diversity is a key parameter for maintaining biodiversity (Pertoldi et al. 2007; Sgrò et al. 2011).

There is evidence that the loss of genetic diversity has direct adverse effects on the viability of species and populations (Spielman et al. 2004; Frankham 2005a). In short time scales, the

loss of genetic variation due to inbreeding, and the reduction in fitness through an accumulation of recessive deleterious mutations, referred to as inbreeding depression, imposes a significant threat on natural populations (Charlesworth and Willis 2009). In long time scales, heritable genetic variation represents the basic resource required for adaptive evolution to occur (Chevin et al. 2010) and for populations to exhibit adaptive response to environmental change (Duplouy et al. 2013). The capacity of a population to thus adapt to environmental change directly relates to several factors, including genetic variation for traits under selection, the strength of selection and population size (Wright 1931; Falconer et al. 1996; Willi et al. 2006). Thus, the argument can be made that maintenance of genetic variation itself, in terms of evolutionary potential, becomes particularly important in the light of climate change (Hoffmann and Sgrò 2011).

The role of genetics in conservation biology

A major challenge for conservation biologists is to allocate limited resources in a way that maximises conservation impact (Bottrill et al. 2008). Genetic monitoring allows assessment of parameters such as effective population size and inbreeding, which are strong predictors for the probability of extinction of a population (Spielman et al. 2004; Frankham 2005a; Wright et al. 2008). Genetic monitoring is therefore a valuable tool for risk assessment and prioritization of populations for conservation efforts (Schwartz et al. 2007; Ottewell et al. 2016). Where species have become highly threatened in the wild, captive breeding programmes are initiated to prevent extinction (Frankham 2008). With the goal to support or reintroduce natural populations and recover self-sustainability, genetic considerations for the design of captive breeding programmes are crucial in order to minimize the loss of genetic diversity, inbreeding and adaptation to captivity (Frankham 2008; Fraser 2008; Griffiths and Pavajeau 2008; Robert 2009; Williams and Hoffman 2009).

Furthermore, assessment of gene flow between populations can resolve patterns of migration, which can inform conservation management by the identification and design of migration corridors (Epps et al. 2007; Manel and Holderegger 2013). Where natural migration cannot be restored, assisted migration or 'genetic rescue' through the artificial introduction of individuals (from a genetically similar stock) can restore genetic diversity in

highly inbred populations (Vilà et al. 2003; Frankham 2015; Whiteley et al. 2015). The success of an assisted migration strongly depends on the amount of genetic divergence between source and recipient population, which relates to the risk of disrupting local adaptations and outbreeding depression (Edmands 2007). However, the risk of this management option is high while the genetic basis of inbreeding and outbreeding depression remain poorly understood (Hedrick et al. 2014; Waller 2015; Hedrick and Garcia-Dorado 2016). The identification of genetic divergences between populations and the assignment of evolutionary significant units (ESU) for management is therefore one of the most important contributions of genetics to conservation biology (Allendorf et al. 2010; Frankham 2010).

Neutral versus functional genetic variation

Historically, there has been a focus on using selectively neutral genetic loci as a convenient tool to infer genetic variation in conservation biology (Sunnucks 2000; Kirk and Freeland 2011). Neutral markers putatively have no adaptive function, but empirically, there may be correlation between neutral marker heterozygosity and fitness (HFC). Positive HFCs would support the idea that neutral markers can be used to infer functional variation, for example by revealing loss of genetic variation across the genome through inbreeding (Szulkin et al. 2010).

Despite their widespread use and utility in revealing demographic processes, such as genetic drift, caution must be used when making conclusions from measurements of heterozygosity at neutral markers (Ding and Goudet 2005; Hansson and Westerberg 2008). The accuracy of estimating genome wide heterozygosity from a limited set of neutral markers may be poor (Hansson and Westerberg 2008; Szulkin et al. 2010). Thus, good experimental design should consider the amount of genome coverage required to provide a representative sample of the whole genome (DeWoody and DeWoody 2005; Kirk and Freeland 2011). For example, a population study of the endangered takahe (*Porphyrio hochstetteri*) could not accurately predict the inbreeding status of individuals using microsatellite markers (n=23 loci) as it was inferred from pedigree analysis (Grueber et al. 2011). Critically, neutral markers do not

typically provide direct information on selective processes and it has been shown that neutral and adaptive genetic variation can be affected differently during demographic processes like bottlenecks (Ejsmond and Radwan 2011; Oliver and Piertney 2012; Sutton et al. 2011,2015).

There are implications of making management decisions based on neutral marker data alone. When conservation priorities are based on genetic diversity surveys, the outcomes may differ significantly using neutral or adaptive markers (Bonin et al. 2007). Moreover, the scale and degree of population divergence measured for neutral markers does not necessarily represent adaptive population differentiation, which is of high relevance when management includes translocation (Johnson 2000). The maintenance of this adaptive population differentiation is of high importance in conservation biology, because it is thought to result in the highest diversity and opportunities for evolution in an uncertain future (Forest et al. 2007) and avoids genetic homogenization that reduces the spatial component of genetic variability among populations of a species (Olden et al. 2004).

Ecologically meaningful and subject to selection, functional markers are increasingly being used in conservation genetics (Hoffmann and Willi, 2008a; Kirk and Freeland, 2011; Piertney and Webster, 2010). Genes associated with the immune system make a good choice for such studies because of the strong association between immune competence and fitness, and the highly adaptive nature of pathogen recognition and elimination within the co-evolutionary dynamics between hosts and pathogens (Piertney and Webster 2010; Tellier et al. 2014; Sutton et al. 2015). Thus, because merging diseases are thought to be of major importance for biodiversity (Daszak et al. 2000) and extinction (Smith et al. 2006, 2009), genetic variation associated with pathogen resistance is an excellent model for conservation monitoring (Sommer 2005; Eyto et al. 2007; Dionne et al. 2009; Ujvari and Belov 2011).

The Major Histocompatibility Complex (MHC) in conservation genetics

Protein molecules encoded by the Major Histocompatibility Complex (MHC) bind specific antigens to their peptide binding region (PBR) and present them on cell surfaces to T-lymphocytes which need this stimulus to recognize antigens and to initialize the immune response cascade (Zinkernagel and Doherty 1974). What follows is a summary of MHC gene structure (see Jeffrey and Bangham 2000). Two classes of MHC genes can be defined: MHC class I molecules are involved in intracellular defense. They are expressed on the surface of all nucleic somatic cells and present antigens to CD 8+cytotoxic T-cells. Class I molecules are heterodimers defined by a transmembranic heavy chain and three extracellular domains (α1, α2, α3) and a β_2 microglobulin. The peptide-binding region is formed by the α1 and α2 domains of the heavy chain. MHC class II molecules can be found on activated T-cells, B-cells and macrophages and present antigens to CD 4 T-helper cells. They are also heterodimers consisting of a α- and a β-chain. The peptide binding region in the class II MHC is also formed by the α1 and β1 domain. MHC loci can be either of classical (a) or non-classical (b) type where the former encode functional MHC proteins as described above. Characteristics for the definition of classical loci are the conservation of residues that anchor the C-and N-termini of peptides, ubiquitous expression, selection on PBR and polymorphism (Saper et al. 1991; Miller et al. 2006). Non-classical MHC loci can have evolutionarily conserved recognition function (Hansen et al. 2007; Hofstetter et al. 2011) and several non-classical MHC genes are known to regulate innate immune response (Braud et al. 1999; Sullivan et al. 2006).

The genes of the MHC are the most variable coding loci known in vertebrates, and are a classic example of balancing selection, maintaining high levels of polymorphism because of selection for variation (Sommer 2005). Due to close interaction with pathogens, pathogen-mediated selection is thought to be the main mechanism to maintain the high diversity observed in this gene family (Eizaguirre et al. 2012). This is supported by a large body of empirical studies which show associations between MHC variants and pathogen susceptibility and resistance (e.g.: Miller et al. 2004; Meyer-Lucht and Sommer 2005; Zhang et al. 2006; Savage and Zamudio 2011). Balancing selection is thought to act via non-mutually exclusive mechanisms: heterozygote advantage, frequency dependent selection, and variation of selection in space and time (Piertney and Oliver 2005; Spurgin and

Richardson 2010). Additionally, sexual selection has been shown to play an important role in maintaining polymorphism through MHC-mediated mate choice in a range of taxa (Roberts and Gosling 2003; Consuegra and Leaniz 2008; Setchell et al. 2010; Strandh et al. 2012).

Many studies investigating MHC diversity in bottlenecked populations report a strong effect of genetic drift in eroding variation (Agudo et al. 2011; Miller et al. 2010; Miller and Lambert 2004; Schad et al. 2004), but the maintenance of MHC variation after a bottleneck has been documented several times (Aguilar et al. 2004; Oliver and Piertney 2012). It has been shown that the loss of genetic variation can be more pronounced for the MHC under the simultaneous effects of drift and balancing selection (Ejsmond and Radwan 2011; Sutton et al. 2011). On the other hand, variation can re-establish via mutation and selection again if the population expands (Ejsmond and Radwan 2011). Some natural populations have lost almost all MHC variation though, like Eurasian beaver (Babik et al. 2005) and cheetahs (Thalwitzer et al. 2010), but still show the capacity to combat infections. Thus, some species can compensate for a lack of variation at the MHC (Acevedo-Whitehouse and Cunningham 2006), but the mechanisms of selection are unknown. These examples of species with low MHC diversity, expanding and apparently coping well with immune challenges, may be rather exceptional though and it is possible that contrasting examples of species that have gone extinct are not documented (Frankham 2005a).

Because of the high sensitivity to population dynamics in addition to its functional importance, the MHC continues to be a top candidate to monitor adaptive genetic variation in a conservation context (Sommer 2005; Eyto et al. 2007; Piertney and Webster 2010; Ujvari and Belov 2011; Hansen et al. 2012; Marsden et al. 2013; Vásquez-Carrillo et al. 2014; Elbers and Taylor 2016). As host-pathogen interactions have been shown to be an important force acting upon the MHC, the simultaneous study of bacterial diversity and MHC variation further offers the possibility to gain valuable insights on evolutionary processes (Dionne et al. 2007, 2009b; Eizaguirre et al. 2012; Pavey et al. 2013).

Metagenomics of microbial communities

A field of research that has gathered much momentum through the development of new sequencing technologies is metagenomics of microbial communities (Hugenholtz and Tyson 2008). The application of culture-independent genomic tools has led to the discovery of vast diversity within and across natural microbial communities (Xu 2006). The identification of a high degree of association between eukaryotic organisms and microbial communities, including symbiotic, commensal and pathogenic relationships, has advanced our understanding of the importance of microorganisms in the evolution and functioning of higher organisms (Zilber-Rosenberg and Rosenberg 2008; Fraune and Bosch 2010; Fumagalli et al. 2011). The hologenome concept has emerged, where the individual organism is expanded by the totality of its microorganisms (microbiome) as a unit for selection, adaptation and evolution (Bosch and McFall-Ngai 2011; Gilbert et al. 2012; Rosenberg and Zilber-Rosenberg 2016).

There are a number of important contributions to conservation biology that microbiome research can make (Redford et al. 2012). There are several relevant conceptual areas of research: the threat of emerging infectious diseases (Altizer et al. 2003; Fisher et al. 2012); dynamics of pathogens and extinction in nature (Smith et al. 2009); and the interaction between stress, the microbiome and pathogens (Thurber et al. 2008; Verbrugghe et al. 2012; Amato et al. 2013; Boutin et al. 2013). It has therefore been suggested to use the microbiome to monitor the physiological states of individuals in a conservation context (Allendorf et al. 2010). Bioaugmentation of probiotic, protective bacteria has become a common practice in aqua culture as a strategy for disease mitigation (Verschuere et al. 2000; Lauzon et al. 2014; Newaj-Fyzul et al. 2014) and been proposed to be used in wildlife conservation as well (Bletz et al. 2013). Microbiome research is relevant to captive breeding, because shift in the composition of microbiomes in captivity (which may occur due to alteration of diet, the artificial environment, differences in social behaviour or medical treatment) could explain disease outbreaks or affect reintroduction success (Redford et al. 2012). There is evidence that incompatibility between co-adapted microbiomes can contribute to outbreeding depression (Brucker and Bordenstein 2013). Microbiome research is therefore highly relevant to translocation biology because it may enhance our ability to assess the risk of genetic rescue attempts of highly inbred populations.

Species distribution modelling as a tool in conservation biology

Every species can be thought of having a distinct, adaptive ecological niche, defined by Hutchinson (1957) as an n-dimensional space consisting of favourable environmental conditions. Across a hypothetical fitness landscape, adaptive peaks and valleys are predicted (Tilman 1980; Hirzel and Le Lay 2008). Species distribution modelling (SDM) builds upon the niche concept and a number of different techniques have been developed to predict the spatial distribution of species (Guisan and Thuiller 2005; Merow et al. 2013).

Understanding the environmental requirements of a species and its ecological niche properties, is of high relevance in conservation biology (Colwell and Rangel 2009; Guisan et al. 2013). Habitat suitability estimates are used for risk assessment and to prioritize areas for conservation (e.g.: Kremen et al. 2008; Wan et al. 2014), as environmental suitability has been shown to be a strong predictor of population persistence (Cabeza et al. 2004). Predicting the distribution of invasive species is used to allocate efforts to control expansion (Giljohann et al. 2011; Jiménez-Valverde et al. 2011). The distribution and habitat suitability of pathogens can be used to predict epidemic risk, to prioritize areas for research and management action (Puschendorf et al. 2009; Murray et al. 2011). By estimating sensitivity and tolerance limits for environmental parameters, predictions can be made about the impact of environmental change on species distribution (Dormann 2007; Thuiller et al. 2008). This has been used to make predictions about vulnerability to climate change (Thomas et al. 2004; Bellard et al. 2012). This approach can also be useful to estimate the sensitivity of species to other anthropogenic environmental alterations and to evaluate predictions of different management scenarios, in order to identify locally optimized conservation strategies (Rhodes et al. 2006; Bergström et al. 2013).

The European Grayling (Thymallus thymallus)

European grayling (*Thymallus thymallus*) provide an ideal system to study adaptive genetic variation in natural populations. Grayling have a wide distribution (see below) and exhibit pronounced population structure within small geographical scales and connected river systems, reflecting a high degree of habitat fidelity (Heggenes et al. 2006; Gum et al. 2009). The assessment of adaptive markers in grayling is important for the conservation of the species because management includes translocation of brood-stocks (Allendorf et al. 2001; Ayllon et al. 2006). Understanding the genetic basis of local adaptation can avoid adverse effects like outbreeding depression (Edmands 2007).

Grayling show high sensitivity to high temperature relative to other salmonids (Ibbotson et al. 2001; Jonsson and Jonsson 2009), and thus represent an ideal model species for studying the impact of climate change. Grayling also show high vulnerability to human impacts such as water quality alteration (Oberdorff et al. 2002) and pollution (Buhl and Hamilton 1990, 1991; Vuorinen et al. 1998; Uiblein et al. 2001) . Therefore, grayling are a suitable indicator species of habitat quality and a good system to study adverse effects of habitat degradation on evolutionary potential.

Phylogeography

The European Grayling belongs to the family of the Salmonidae, the only living family within the order of Salmoniformes in the class of Actinopterygii (Norden 1961). The distribution of European grayling ranges from France and Great Britain in the West until the Ural mountains in the East, the south of France and Montenegro on the Balkans in the South to Scandinavia in the north (Gum et al. 2009). Molecular genetic diversity between populations of the European grayling is high, supporting the hypothesis of multiple refugia during Pleistocene glacial periods and different routes of colonisation afterwards (Koskinen et al. 2000; Weiss et al. 2002; Gum et al. 2005). Mongolia has been proposed as the origin of grayling diversity (Schöffmann 2000) with the genus possibly expanding to European drainages via a dual corridor (Weiss et al. 2002). One refugium during glacial periods has been suggested close to the Caspian basin, from where north western Russia, Estonia and

Finland have been colonized, and another two in central Europe, possibly within ice-free parts of the Elbe or Vistula and the Main/Rhine drainage, providing the source population for expansions to Germany, Denmark, Poland, Sweden and parts of Norway (Koskinen et al. 2000; Gum et al. 2005). Three divergent clades are documented to originate from the Danube basin and other distinct lineages are found in the Adriatic and the Loire basin (Weiss et al. 2002; Meraner and Gandolfi 2012). The UK is thought to have been colonized from the Rhine system via the North Sea River, when Britain was still connected to the continent (Woolland 1986; Gum et al. 2005). The original distribution of grayling in the UK is uncertain, but is suggested to have concentrated on the River Ouse, Trent, Hampshire Avon, Severn, Wye, Thames, Ribble and Welsh Dee (Gardiner 2000) and were introduced to Scotland in the 1800s (Gardiner 1991).

Life History

The spawning season of grayling is in spring, induced by an increase in water temperature to 3 -11° C (Witkowski and Kowalewski 1988; Ibbotson et al. 2001). Grayling show upstream movements prior to spawning and distances migrated can be highly variable depending on the availability of suitable spawning habitat (Meyer 2001). Meyer (2001) report movements of up to 11.3 km in northern Germany (mean ± S.D.: 8.2 ± 2.8 km) whilst Ovidio et al. (2004) found movements between 70 and 4980 m (mean ± S.E.: 1234 ± 328 m) during pre-spawning migrations in a stream in Belgium. Sites suitable for spawning need to have a clean substratum of a size between 16–32 mm, a water depth between 30–40 cm, current velocity of 40–60 cm s^{-1} and a stable streambed (Beauchamp 1990; Sempeski and Gaudin 1995; Nykänen and Huusko 2002).

Males arrive at spawning grounds before females and dominant males defend suitable territories (Poncin 1996). When females approach, males display a courting behaviour, which leads, in case of success, to a joined release of eggs and milt to the gravel of the spawning ground (Fabricius and Gustafson 1955).The mating system of grayling is polygynandrous, where both sexes can mate with more than one partner over a reproductive season (Haddeland et al. 2015; Meraner et al. 2013). Reproductive success is greater for individuals mating with more than one partner and generally positively

correlated with body size for females (Northcote 1995; Haddeland et al. 2015). For males size has been shown to positively correlate with reproductive success in some cases, probably due to higher competitive abilities of larger males to secure high quality spawning territories (Haddeland et al. 2015), but was not the case under conditions of low density (Poncin 1996). After spending up to 32 days within the spawning grounds grayling return to the territories they inhabited prior to spawning, showing generally high site fidelity and homing success (Ibbotson et al. 2001; Meyer 2001; Ovidio et al. 2004).

The fertilized eggs hatch after about 177 degree days (Hulstere and Philippart 1982), which is calculated by summing up the average temperature of each day after the fertilization. Fry stay in the gravel for the first 4-5 days feeding on their yolk sac before they first emerge, having a fork length of about 1-1.9 cm (Scott 1985; Bardonnet *et* al. 1993). Grayling larvae (<3cm) are found within the river margins and dominantly feed on chironomid larvae (Sempeski et al. 1995; Sempeski and Gaudin 1995). Juveniles (> 3-4 cm) move to the river channel and start feeding on larger prey (Sempeski and Gaudin 1996). Sexual maturity is reached at an age of 3-5 years (Kristiansen and Dølving 1996; Ibbotson et al. 2001).

Conservation status and management

Decline in abundance of European grayling has been reported from several parts across its range (Northcote 1995; Persat 1996; Uiblein et al. 2001; Gum et al. 2003). The species is listed as protected in appendix II of the Bern convention (Swatdipong et al. 2010) and UK populations are considered endangered (Dawnay et al. 2011). The Environmental Agency in the UK lists habitat degradation, regulation of water and pollution as the main factors causing population declines (Ibbotson et al. 2001). An increase in predation pressure from piscivorous fish moreover constrains the recovery of populations (Gum 2007).

To compensate decreases in population density, stocking has been widely applied in various countries (Koskinen et al. 2002; Dawnay et al. 2011; Weiss et al. 2013; Persat et al. 2016). Fry used for stocking grayling were released without any knowledge about the genetic relationship between source and recipient population (Dawnay et al. 2011; Persat et al. 2016). This has led to the introgression of non-native gene pools and even the complete displacement of indigenous stocks in some places (Duftner et al. 2005; Meraner et al. 2014),

leading to genetic homogenization and the reduction of the spatial component of genetic variability among populations (Olden et al. 2004). In other cases stocked individuals have not contributed to the native population at all and no sign of introgression has been detected (Persat et al. 2016). Yet, in both these cases, the aim to maintain and support viable and diverse populations has not been successful.

The UK Environment Agency manages the hatchery programme for European grayling (*Thymallus thymallus*) in Calverton, Nottingham. Offspring of adults from a few source rivers are raised and introduced back into rivers within the management area of their parents, as they have been defined based on neutral genetic differentiation (Dawnay et al. 2011). Concerns about the effectiveness of such practice arise from findings of generally lower fitness of hatchery-reared fish compared to wild-reared individuals (Harbicht et al. 2014). Disruption of local adaptations and outbreeding effects are one reason for counterproductive effects (Satake and Araki 2012). However, adverse effects are possible even where the source and recipient population are the same. As fertilization happens artificially by mixing eggs and sperm of all adults that are used to create the hatchery offspring, natural mate choice is not allowed for. Because MHC-mediated mate choice is an important component of maintaining high levels of polymorphism, artificial fertilization can lead to substantial fitness reductions within the offspring generation compared to natural reproduction (Consuegra and Leaniz 2008). Moreover, through the relaxation of selection over just a period in life history in captivity, deleterious mutations are eliminated less and can substantially increase the genetic load of supplemented wild populations (Lynch and O'Hely 2001; Araki et al. 2009).

Study populations

Twelve populations of grayling within Great Britain were selected for assessment of immune genetic variation in this study (Figure 1.1). These populations had the advantage of available neutral genetic data, provided by Dawnay et al. (2011), for the same individuals studied here. The selected populations cover different management histories, with five non-stocked native populations (Dee, Severn, Ure, Wye and Wylye), four stocked native populations (Aire, Derbyshire Derwent, Dove and the Hampshire Avon) and three introduced populations (Clyde, Eden and Itchen).

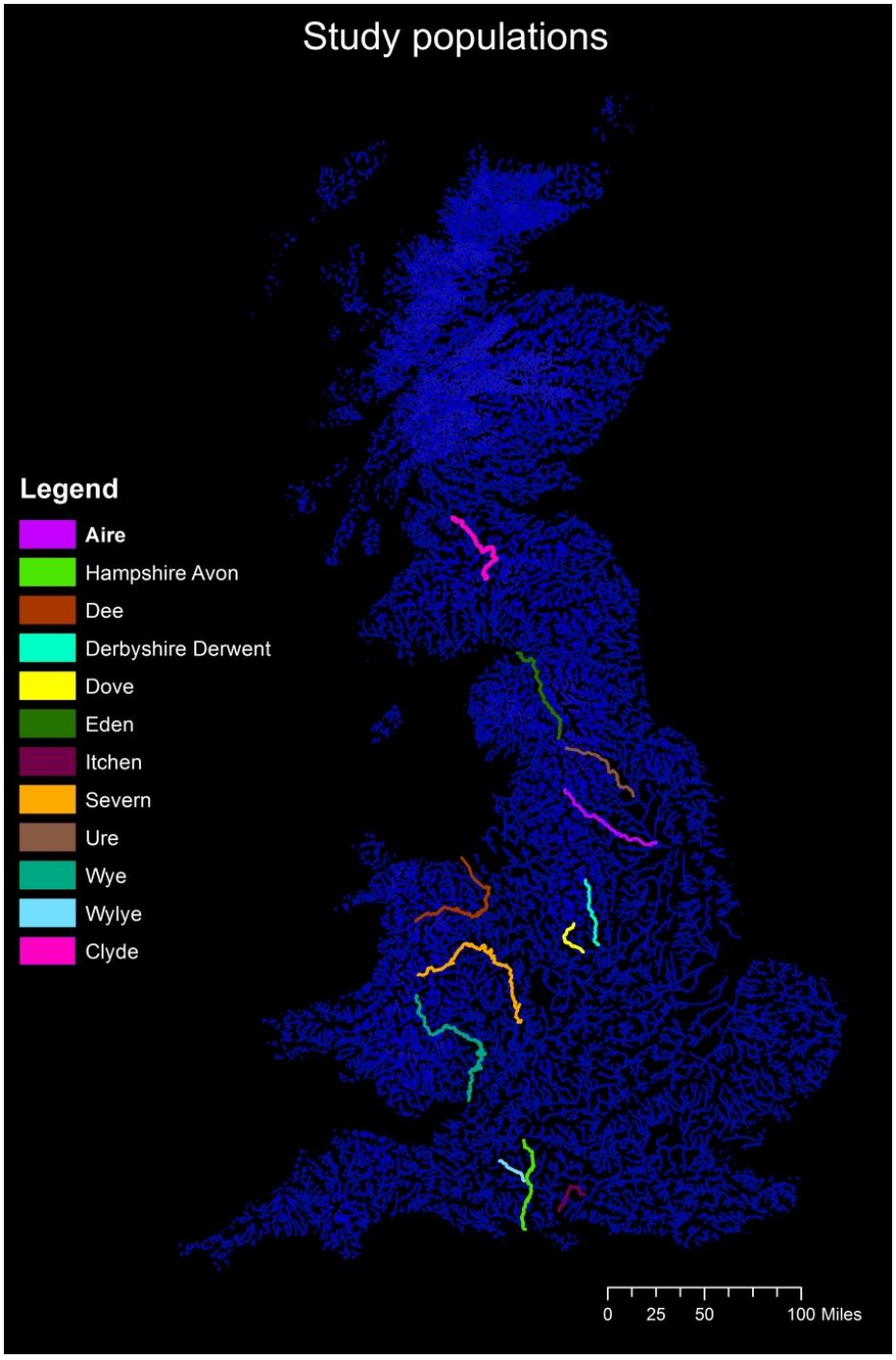


Figure 1.1: Locations of study populations

Aims and objectives

This study aims to characterise functional genetic variation at immune genes of MH loci in European grayling. Existing genetic data of related taxa are extended to this non-model species of conservation concern. Using next-generation sequencing technology, the PhD aims to investigate immunogenetic variation of natural populations in a conservation genetics framework, in comparison with existing neutral marker data. By doing so, the project develops an adaptive conservation genetics approach. In addition to conservation, evolutionary and population genetics approaches, the PhD aims to make the first assessment of the microbiome associated with grayling. This is of course important in relation to the immunogenetic motivation of the research, as well as being of wider interest. The PhD will also explore the impacts of environmental change on grayling populations by Maxent modelling approaches and examines interactions between Maxent models and estimates of genetic variation of grayling populations in immune markers.

CHAPTER TWO aims to develop a novel bioinformatics pipeline for the analysis of NGS amplicon datasets, where loci exhibit a complex architecture. Class I and class II MH loci are optimised for genotyping by developing a synthesis of existing genotyping methods. A comprehensive dataset, including the peptide binding regions (PBR) of the MHC class I α2 chain (UBA), both class II chains (DAA, DAB) as well as the α1 chain of the unclassical ZE locus, is compiled for twelve populations within the UK to inform about adaptive genetic variation of the species. An analytical pipeline is developed for high throughput Illumina sequencing data that allows for high confidence genotyping of multi-gene families like the MHC. Genotyping repeatability is tested for different analytical methodologies and an integrated approach is suggested to increase accuracy and economy.

CHAPTER THREE examines MH loci within an evolutionary genetics context. Hypotheses to be examined include:

(a) Is a single classical locus for both the class I and class II regions, with high allelic divergence in the class I (Shum et al. 2001) and similar levels of diversity within both class II chains (Gómez et al. 2010) a general pattern across salmonids and also found within grayling?

(b) Is balancing selection a main force in shaping diversity at the MHC and can be detected within the PBR for the classical loci?

(c) Is balancing selection also acting on the non-classical ZE lineage?

CHAPTER FOUR will investigate the population genetics of European grayling in a conservation context using existing neutral marker data (Dawnay et al. 2011) and novel MH data generated as part of this PhD. In doing so, the relative role of selection versus demography and genetic drift in shaping ecologically meaningful genetic variation can be examined. This comparison allows to test the following hypothesis and predictions:

(a) Neutral marker variation cannot predict variation at the MHC, because selection is acting as a major force on the MHC

(b) Current risk assessment of population viability and the definition of management units based on neutral marker diversity is not consistent with results for immune genetic markers involved in evolutionary processes

(c) Stocking history affects genetic variation at MH loci more than at neutral sites, because of the inference with natural evolutionary processes

CHAPTER FIVE will investigate environmental parameters that shape the distribution of European grayling. To do so SDM tools are used. After identification of the dominant climatic factors that define limitations to grayling distribution within its entire range, the UK is chosen as a subset area to take advantage of long-term monitoring data sets, such as water chemistry and current velocity (flow), in order to test the sensitivity of grayling to non- climatic parameters, which potentially affect habitat suitability.

CHAPTER SIX aims to further develop this work to assess how future climate change scenarios potentially affect grayling. MAXENT models are also used to investigate how locally optimized strategies for habitat improvement might affect future predictions of habitat suitability as a mitigation strategy under climate change conditions (chapter 6).

CHAPTER SEVEN takes a NGS approach to investigate and characterise the microbiome of grayling in natural populations, for the first time. Because of the important role of pathogen-mediated selection, the study of microbial communities that shape selective

regimes across different environments has the potential to enhance our understanding of the adaptive value of observed immune genetic variation. We therefore compare different methodological approaches to characterise the microbiome associated with individual grayling to inform future large-scale studies. In particular, we aim to inform about the resolution to dissolve bacterial phylogeny targeting different regions of the bacterial 16S gene, assess the relative difference of the microbiome associated with the mouth or the gills and the effect of sample treatment.

Finally, **CHAPTER EIGHT** considers the findings of each chapter together and discusses these in the context of the current literature. Further, implications for management of grayling are suggested and directions for future research proposed.

Chapter 2: Novel NGS analysis pipelines for the rapid assessment of highly variable multi-gene families

Abstract

Advances in sequencing technologies now allow large-scale studies of functional genetic variation in a conservation context. These advances have the potential to provide valuable insights into the process of genetic adaptation, highly warranted in a time characterised by high-rates of species extinction in consequence of rapid environmental change and anthropogenic ecosystem alteration. The high sensitivity of next-generation sequencing (NGS) tools can make the distinction between true variants and artefacts challenging, however. This is particularly true for multi-gene families, like the major histocompatibility complex (MHC), due to the common presence of multiple loci, the similarity of true recombinant alleles and artificial chimers or individual copy-number variation. A golden standard, regarding sample replication, sequencing-coverage and the inclusion of negative controls, to assure high confidence genotyping, whilst minimizing sequencing effort, has not yet been established. Here it is shown, that using current approaches genotyping repeatability without sample replication can be low and a combination of two different analytical approaches to overcome this without the need for sample replication is presented.

Introduction

Human ecosystem alteration is a major threat to biodiversity (Barnosky et al. 2012). This has led to the advent of conservation biology as an independent discipline in its own right. Within this, the study of adaptive potential and how it is affected by demographic and environmental factors is at the core of conservation genetic research (Willi et al. 2006; Ouborg et al. 2010). Because neutral marker diversity does not reflect variation at genes under selection (Sutton et al. 2011, 2015), the aim of adaptive conservation genetics is to directly target variable genetic regions that impact fitness (Ouborg et al. 2010). New genotyping-by-sequencing (GBS) techniques, such as restriction-site-associated DNA sequencing (RADseq) or transcriptome sequencing allow tens of thousands of markers to be genotyped simultaneously without prior genetic data of the study species (Davey et al. 2011). These approaches are promising to advance knowledge about adaptive genes (Allendorf et al. 2010). For example, genomic regions potentially under positive or balancing selection can be identified through the statistical association of outlier loci, or correlations of allele frequencies with important ecological variables (Coop et al. 2010; Hohenlohe et al. 2010). However, providing evidence for the adaptive nature of genetic variation at candidate regions is often challenging due to confounding factors like linkage, background selection and heterogeneous recombination rates (Cruickshank and Hahn 2014; Shafer et al. 2015). Further investigation using quantitative trait loci (QTL) mapping, controlled crosses, or genome-wide-association studies (GWAS) are necessary to link genetic variation to phenotypic traits and fitness (Narum et al. 2013; Vitti et al. 2013). In the meantime, applying NGS approaches to already well-established candidate loci where there is prior knowledge of their function and ecological importance is fruitful (Piertney 2010). This approach is particularly suitable for population genetic surveys as the number of individuals and populations can be increased by reducing the number of target loci with the same sequencing effort (Ekblom and Galindo 2011).

Recent advances in DNA sequencing technology have revolutionized the field of genetic research (Mardis 2008; Zhang et al. 2011; Koboldt et al. 2013). Referred to as next-generation sequencing (NGS), these methods are characterized by massively parallel sequencing for data collection of several millions of base pair reads in a single run (Metzker

2010; Mardis 2013). The increasing cost-effectiveness of this methodology to generate large amounts of sequence data, and the applicability to non-model species, make next-generation sequencing a highly relevant approach for conservation genetics (Avise 2009; Allendorf et al. 2010; Ekblom and Galindo 2011).

The Major Histocompatibility Complex (MHC) is a good example of this case. The Major Histocompatibility Complex (MHC) is widely implicated as a region of rapid, adaptive evolution based on the functional significance it has for immune response in vertebrates (Sommer 2005). Numerous studies have shown a direct link between pathogen resistance and MHC variation (e.g.: Miller et al. 2004; Meyer-Lucht and Sommer 2005; Zhang et al. 2006; Savage and Zamudio 2011). The MHC standing genetic variation has the potential to buffer populations against novel or emerging diseases (Sommer 2005; Dionne et al. 2009; Ujvari and Belov 2011), which are thought to be an increasing threat under climate change (Harvell 2002). Thus, because of the high degree of variation and adaptive importance, the MHC is a good candidate to study the role of drift and selection in natural populations (Evans et al. 2010; McClelland et al. 2013).

Using NGS technologies for a candidate gene approach greatly facilitates the process as no cloning step is required to sequence separate alleles. Indexing techniques allow multiplexing of many individuals in a single sequencing run. However, there are some down sides of NGS technologies. Sequencing error rates can be substantially higher for NGS technologies than for traditional Sanger sequencing (Kircher and Kelso 2010). Also, PCR errors, chimeric sequences and sample cross contamination can achieve high read numbers using NGS, imposing a challenge on the correct distinction between artefacts and true variants (Li and Stoneking 2012; Sommer et al. 2013; Lighten et al. 2014a; Robasky et al. 2014; Tosar et al. 2014). Genotyping highly variable multi-gene families, like the MHC, has some additional intrinsic challenges (Babik 2010; Lighten et al. 2014a). For example, the evolutionary history of multi-gene families is often characterized by events of gene duplication (Okita et al. 1989; Klein et al. 1998; Nei and Rooney 2005). As a consequence, multiple co-amplifying loci are common and the presence of pseudo-genes, inter-locus gene conversions and copy-number variation between individuals can make an assignment of variants to a particular locus more difficult (Hess and Edwards 2002; Reusch et al. 2004; Babik 2010; Cheng et al. 2012). Natural recombination within or between loci can be frequent and mimic artificial chimers (Cullen et

al. 2002). A recent review summarizes the most important experimental and bioinformatics approaches taken to overcome these challenges, but concludes that a golden standard has not yet been established (Lighten et al. 2014a).

Babik et al. (2009) were the first to genotype the MHC using NGS technology. These authors suggested the application of allele validation thresholds (AVTs), both within and among amplicons (amplicons here defined as all reads obtained from one ID tag, typically one PCR), to separate genuine alleles from artefacts. Under this approach, to classify a variant as a putative allele it needs to be observed in at least two independent PCRs with a minimum of two reads in one of them or in three independent PCRs as a single copy (Babik et al. 2009). Moreover, they empirically identified an average per-amplicon frequency threshold of 3% as suitable to differentiate between artefacts and genuine alleles. This builds on the assumption that errors happen randomly and artefacts exhibit lower frequencies than true variants. Their approach was modified in a number of subsequent studies (e.g.: Zagalska-Neubauer et al. 2010; Huchard et al. 2012; Sepil et al. 2012). However, it has been criticised that circumstances under which artefacts may exhibit higher frequencies than genuine alleles are neglected (Sommer et al. 2013; Lighten et al. 2014a). Biased errors can be introduced through the sequencing technology (Gomez- Alvarez et al. 2009; Gilles et al. 2011) or cross-contamination and can be amplified at higher levels than random artefacts leading to false positives (Lighten et al. 2014a). Amplification efficiencies of true variants can be variable and increase the rate of false negatives for those alleles with lower than average efficiency (Sommer et al. 2013).

Sommer et al. (2013) addressed the issue of variation in allele amplification efficiency in MHC genotyping. Differences in the amplification efficiency of alleles can arise through primer mismatches, differential denaturation caused by significantly higher GC contents in one allele or preferential amplification of shorter alleles (Walsh et al. 1992). An important step in their experimental design is to include an independent replicate of each sample, which reduces the risk of both false positive and false negative findings compared to other methods. In their dataset, allele amplification efficiencies varied up to 12-fold between variants. Taking into account estimated allele efficiencies they developed a method to assess the confidence level of each genotype based on Galan's T1 threshold for minimum read numbers (Galan et al. 2010; Sommer et al. 2013). One drawback in their method to

estimate allele amplification efficiencies across more than one locus is that copy number variation is not considered. The relative amplification efficiency of alleles, frequently found in homozygous states can be inflated, whilst the relative efficiency of alleles found only in single copies can be underestimated. As an alternative Lighten et al. (2014b) suggest a method to estimate copy numbers per individual by estimating the sum of squares between observed and expected number of reads for each copy number scenario and selecting the best fitting model. However, this approach does not take unequal amplification efficiencies across variants into account. Thus, while neither of these methods is perfect, both methods represent valuable improvements on frequency-based thresholds.

Salmonid MHC class I and class II genes both contain one classical gene and are found in different linkage groups (Sato et al. 2000). The α1 and α2 domains of class I molecules are encoded by the linked exon 2 (~261 bp) and exon 3 (~279 bp) of the UBA gene, separated by an intron of ~18kb (Shiina et al. 2005; Lukacs et al. 2007). Divergence between lineages of the salmonid UBA gene is high, with sequence identity among lineages as low as 40% (Shum et al. 2001; Miller et al. 2006; McClelland et al. 2013). Therefore, a variety of primers are needed to amplify allelic variants at this locus (Miller et al. 2006; McClelland et al. 2011). Duplicated regions of the salmonid MHC class I can be found containing so-called non-classical loci (Shiina et al. 2005; Lukacs et al. 2007, 2010). The characteristics to distinguish a fully functional classical locus from a non-classical locus are the conservation of eight functionally important residues that anchor the C-and N-termini of peptides (Brown et al. 1993; Kaufman et al. 1994), ubiquitous expression, selection on peptide binding residues (PBR) and polymorphism (Miller et al. 2006). Salmonid gene duplication events have arisen due to genomic tetraploidy in the evolutionary history of the group (Allendorf and Thorgaard 1984). Duplications of functional genes may be released from functional constraints, becoming selectively neutral, gaining new functionality or turning into pseudogenes (Raes and Van de Peer 2003; Magadum et al. 2013). Gene duplication is therefore an interesting source of genetic variation contributing to evolutionary potential (Taylor and Raes 2004). Associations of pathogen resistance with variation at non-classical class I loci has been described in salmonids (Miller et al. 2004; Johnson et al. 2008). Non-classical loci in salmonids can be classified into the S-, Z-, U- and L- lineages, which are distributed over five different linkage groups in Atlantic salmon (Lukacs et al. 2010). Non-

classical loci of the Z-lineage have been found in several teleost species (Okamura et al. 1993; Erp et al. 1996; Kruiswijk et al. 2002; Miller et al. 2006; Lukacs et al. 2010). The ZE locus exhibits characteristics that make its assignation as a non-classical locus questionable (Stet et al. 2003). The presence of conserved potential peptide anchoring residues, exon-intron organization, ubiquitous expression, and polymorphism support a classical nature of this locus in several teleost species (Kruiswijk et al. 2002; Miller et al. 2006; Lukacs et al. 2010). MHC class II molecules are also heterodimers consisting of a α and a β chain. The peptide binding region is formed by the α1 and β1 domain with the DAA gene encoding for the α and the DBB gene for the β chain. European grayling make a case-study of the extension of genetic data to a non-model species of conservation concern since there is no prior knowledge of their MHC loci and they are closely related to well-studied salmonids like Atlantic salmon (Davidson et al. 2010; Vincent et al. 2013; Houston et al. 2014). , .

Here a novel NGS pipeline is synthesized and documented, which was optimized to genotype highly variable multi-gene families in the salmonid MHC. In particular, MHC variation in grayling is analysed in the context of existing data in closely-related taxa, for both classical loci and poorly studied non-classical loci. Because conservation geneticists often work within budget constraints, competing genotyping methods are also evaluated in scenarios where sequencing coverage is low. Specifically, the aims are to: (a) Compare the repeatability of genotyping methods for measuring genetic variation in MHC in grayling; (b) Critically evaluate competing genotyping analysis pipelines for NGS data; (c) Document a novel analysis pipeline, which integrates the methods of Sommer et al. (2013) and Lighten et al. (2014b), whilst overcoming the drawbacks of each (replicated samples and high sequencing coverage), designed to optimise accuracy and economy. The results are critically discussed in the context of modern methods to evaluate rapidly evolving, multi-locus genotypes and the analysis pipeline is documented as a reference for future work.

Materials and Methods

Primer Design

Known primers were used to target alleles of the most common L3 I lineage of the MHC class I UBA α 2 exon 3 (Miller et al. 2006; Table 2.1). Grimholt et al. (1997) published one single primer pair for the MHC class I UBA α 2 exon 3 used in Atlantic salmon (*Salmo salar)* and an adaptation of this primer sequence was used for amplifying other UBA α 2 exon 3 lineages (Table 2.1). Primer sequences for the DAA exon were based on published primers developed for brown trout (*Salmo trutta*) with matching sequences in Atlantic salmon (*S. salar*) and rainbow trout (*Oncorhynchus mykiss*) (Stet et al. 2002). Primers for the β1 domain encoded by exon 2 of the DBB gene, involved in peptide binding, have been described by Pavey et al. (2011) for salmonids. However, it was found that the reverse primer could act as a forward primer at the keratin II locus and was re-designed (Table 2.1). Primer sequences for the non-classical ZE exon 3 were developed from alignments with *O. mykiss* and *S. salar* within putatively conserved regions (Table 2.1). To test amplification of target loci, cloning was done using an *Escherichia coli* (DHα5) Pjet cloning kit (Fisher) and subsequent Sanger sequencing was done on an ABI3130 (Applied Biosystems) capillary instrument following standard protocols, using Big-Dye v.3.

Table 2.1 Primer sequences for target MHC loci

Target	name	Forward primer	Reverse primer	Product size (bp)
UBA α2 exon 3	L3-I	CCAGGTGATGTATGGATGTGA	GATACTTCTTAAGCCAATCAATGCA	~233
UBA α2 exon 3	L3-II	GTGAGTGGGATGATGAGACTG	CTTCCCATAGTCCACATACTTCTT	~230
MHCII α1 exon 2	DAA	GGATGCAGTGATTCAGATGGA	CTCTTCTGGGTTCTTGTAAGCTTTT	~213
MHCII β1 exon 2	DAB	ATGTTTTCCTTTTAGATGGATATTTT	GTCTTATCCAGTACGACAC	~286
ZE α1 exon 3	ZE	CCTGGTATCCATGAGTTCAC	ACTTTGAACCACTGTTCCTTG	~176

Primer sequences were obtained from the literature for amplification in salmonids; modifications were applied to optimize amplification in grayling

DNA samples and NGS library preparation

Previously collected fin clip samples of grayling were used (see Dawnay et al. 2011). DNA was extracted using a modified salt procedure (Dawnay et al. 2011). Between 37 and 40 individuals were selected from each of twelve populations (Dawnay et al. 2011). A replicate was made for each sample along with ten randomly distributed negative controls for each locus.

An individually tagged sequencing approach was used in a nested PCR design, with the outer primer containing the Illumina adapter sequences and tags and the inner primer the target specific sequence (after Lange et al. 2014; Figure 2.1). This approach allows different inner primers to be used with the same set of tagged outer primers and is therefore flexible and cost-efficient. The assay was designed as a one-step PCR on a Fluidigm Access Array microfluidic chip (Lange et al. 2014; Moonsamy et al. 2013) but was modified here to a two-step PCR approach on conventional thermocyclers.

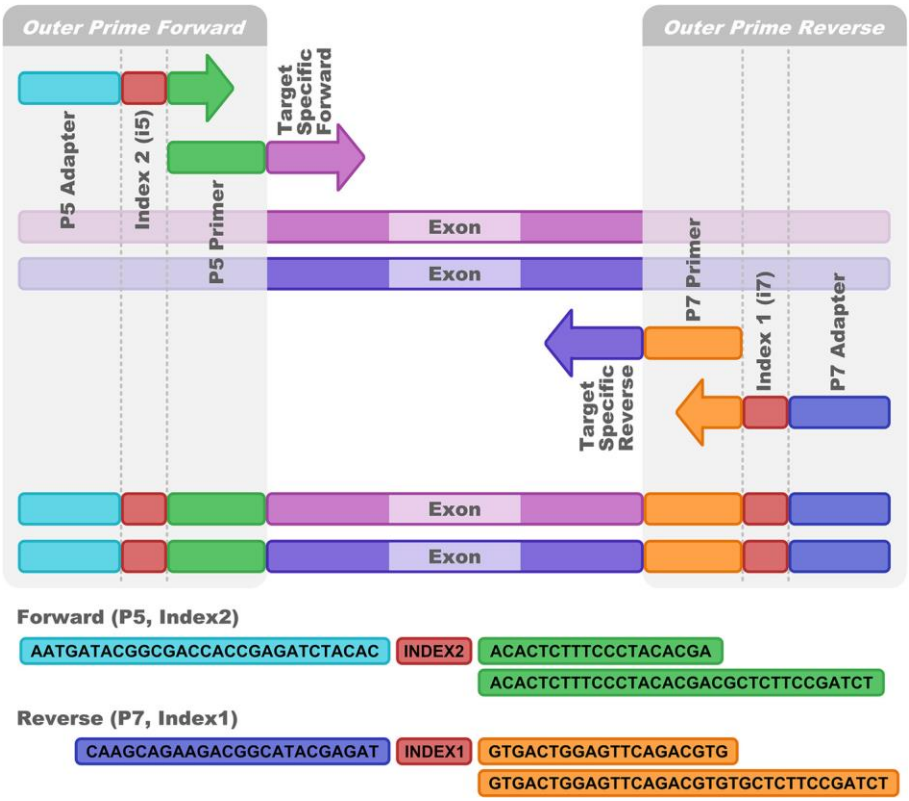


Figure 2.1 nested PCR-design for amplicon-library preparation*: outer primer contain the Illumina adapter sequence, an index sequence and a 19 bp long sequence in common with the inner primer, which contains the target specific sequence; from Lange et al. (2014)*

Library preparation using PCR was done as follows. Inner target specific PCR was performed in total volumes of 6 µl containing 3.75 mM $MgCl_{2,}$, 0.2 mM each dNTP, 4% DMSO, 0.2 µM of each target specific primer, FastStart High Fidelity Reaction Buffer and 0.15 U of FastStart High Fidelity Blend Enzyme (Roche/Sigma Aldrich). PCRs were performed in Prime (Bibby) PCR cyclers or in a ABI 1 PCR cycler with following thermal profile: 95°C for 10 min, followed by 15 cycles at 95°C for 25 s, target specific temperature and annealing time and 72°C for 90 s, and a final extension at 72°C for 5 min. Target specific temperatures and annealing times were 59°C for 60 s (DAA locus), 60°C for 45 s (DAB and L3II locus), 58°C for 45 s (Ze locus) and 55°C for 90 s (L3I locus). PCR products were then diluted (1:20) in H2O and 3 µl used as template in the second outer PCR reaction which was carried out in a total volume of 7 µl containing 3.75 mM $MgCl_{2,}$, 0.2 mM each dNTP, 4% DMSO, 0.1 µM of each outer primer, FastStart High Fidelity Reaction Buffer and 0.25 U of FastStart High Fidelity Blend Enzyme (Roche/Sigma Aldrich). The thermal profile of the second PCR was 95°C for 10 min, followed

by 27 cycles at 95°C for 25 s, 60°C for 60 s and 72°C for 90 s, and a finishing step at 72°C for 5 min. Of all samples, 20% were run out on 1% agarose gels to check amplification success. All PCRs were pooled per locus for each population prior to purification using AmpureXP (Beckmann and Coulter) following manufacturers guidelines. These pools were quantified using Qubit 3.0 Fluorometer (Thermo Fisher Scientific). All populations were pooled for each locus (equimolar concentrations). These pools were then run on a bioanalyzer chip 2100 (Agilent) to check product size and that unincorporated adaptors and primers were removed. Samples were then pooled in equal concentrations across loci and sequenced using a Miseq Nano Run.

Data analysis

For quality control, all reads with a quality score below 20 in more than 90 percent of the sequence were filtered using Filter by Quality tool on Galaxy Server (Goecks et al. 2010). Only sequences with both paired-end reads of sufficient quality were retained and aligned using Mothur (Schloss et al. 2009). Primer mismatches (>1bp) and frame-shifts were filtered and examined for repeated sequences that could be derived from co-amplifying pseudo-genes. Read counts were adjusted if a variant was present in a negative control. In this case, the highest read count of the variant observed in a control was subtracted from all amplicons where this variant was detected. Genotyping methods (described below) were applied if amplicons had a minimum of twenty reads, for the DAA and DAB locus, where preliminary work suggested a single amplifying locus and forty for the L3I, L3II and ZE locus, where more than a single locus was amplified.

Method (1) modified 'Sommer pipeline' (Sommer et al. 2013)

Genotyping was performed using methods described in Sommer et al. (2013), with the following modifications. Treatment of variants classified as 1-2 bp different to a more common cluster within the amplicon was modified as follows (Figure 2.2 step IIa). The Sommer pipeline usually classifies variants of this category present in an amplicon but not the technical replicate as artefacts. However here, if a variant was encountered that was not

present in the technical replicate, but had been classified as a putative allele in step I in another amplicon (and hence was observed as the most frequent cluster in another amplicon), it was not assigned as an artefact, but as 'unclassified'. This approach was taken to account for genuine alleles with high similarity. On the other hand, if the most frequent variant within one amplicon was not present within the technical replicate, an assignment error was assumed and the individual excluded from the analysis.

Using all variants classified as putative alleles, relative allele amplification efficiencies and Galan's T1 threshold for sequencing coverage corrected for minimum allele efficiency, were calculated in R following the script provided by Sommer et al. (2013). Unclassified variants identified at the end of the implementation of the 'Sommer pipeline' were re-evaluated based on their relative efficiency. These were scored as putative alleles, if after dividing read count by efficiency, the modified value was higher than the frequency of any putative artefact. The rationale for this approach is that biologically real alleles may be misclassified as putative artefacts if their efficiency happens to be low. Thus this method reduces the likelihood of undetected heterozygosity due to the loss of low efficiency alleles being misclassified as artefacts (thus accounting for type II genotyping error in the sense of Lighten et al. (2014a)).

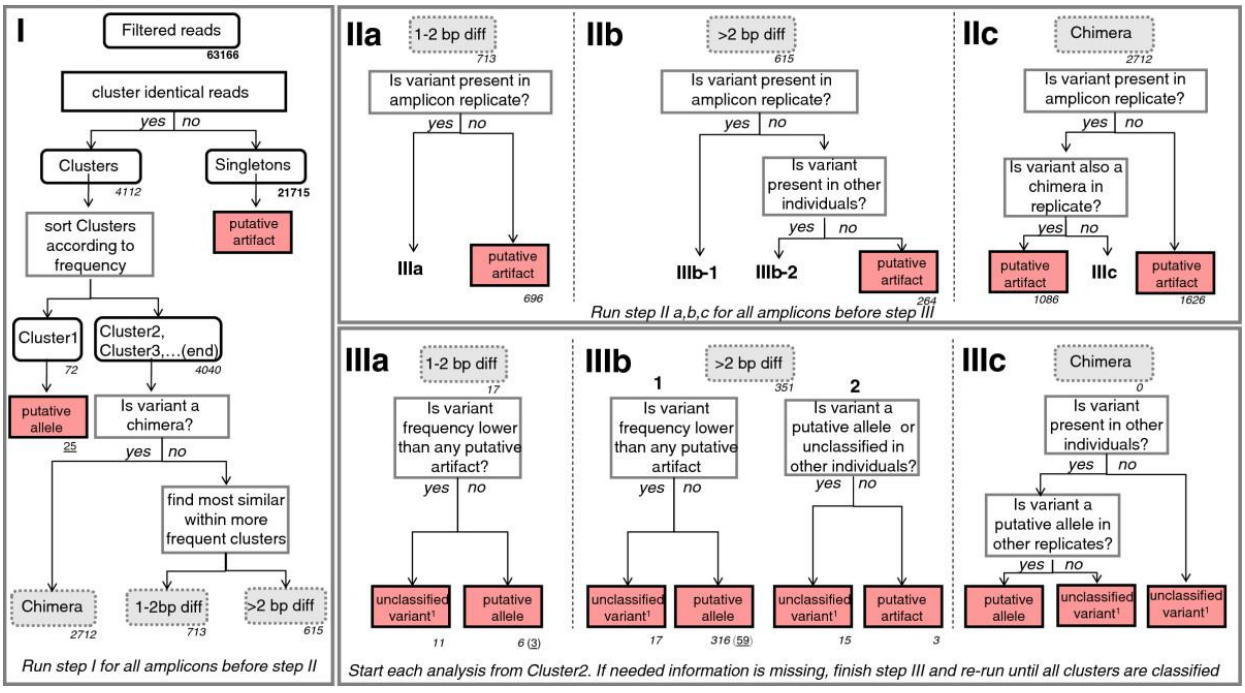


Figure 2.2 Analysis pipeline suggested by Sommer et al. (2013)

Method (2): modified ‘Lighten pipeline’ (Lighten et al. 2014b)

As a second approach, all amplicons were evaluated using the methodology as suggested by Lighten et al. (2014b), but testing the modification of incorporating classifications made using the Sommer methods and disregarding read counts of variants classified as artefacts in improving repeatability of estimates between replicates. The methodology proposed by Lighten et al. (2014b) includes two different approaches based on copy number-variation (CNV) and degree of change (DOC). The CNV-method estimates the copy number of variants by comparing observed and expected number of reads under a range of alternative genetic models and identifies the simplest model that fits the data best. This method was applied to compare the effect of control read subtraction on the overall fit of the data to specific copy number scenarios. In this case up to five loci were considered and an F-ratio test was used to test if control read subtraction resulted in significantly lower variance and better fit.

In order to identify the copy number scenario that best fit the data for each locus and individual, the sum of squares of the genetic model of each scenario was calculated for each individual and an F-ratio test was applied between the two top models. Scenarios considered were based on the highest number of alleles observed as the highest number of possible loci. Independently, the DOC was calculated around each variant and the variant with the highest DOC classified as the last putative allele. The rationale for DOC is that genuine alleles should be sequenced with greater depth than artefacts (Lighten et al. 2014b).

Lighten et al. (2014b) assume that real alleles contribute a relatively higher amplicon read depth than artefacts within an amplicon. This assumption is not necessarily met when amplification efficiencies are uneven across alleles or when low efficiency alleles are present at lower frequency than artefacts (Sommer et al. 2013). In such cases, amplicons may not show an inflection point (highest DOC) or may have conflicting CNV and DOC estimates and be discarded due to insufficient data quality (Lighten et al. 2014b). Thus, information on low efficiency alleles might be lost, particularly if coverage is not very high. Consequently, an alternative approach was also employed to correct total read numbers of putative alleles with low efficiency: total read count was divided by relative amplification efficiency. CNV

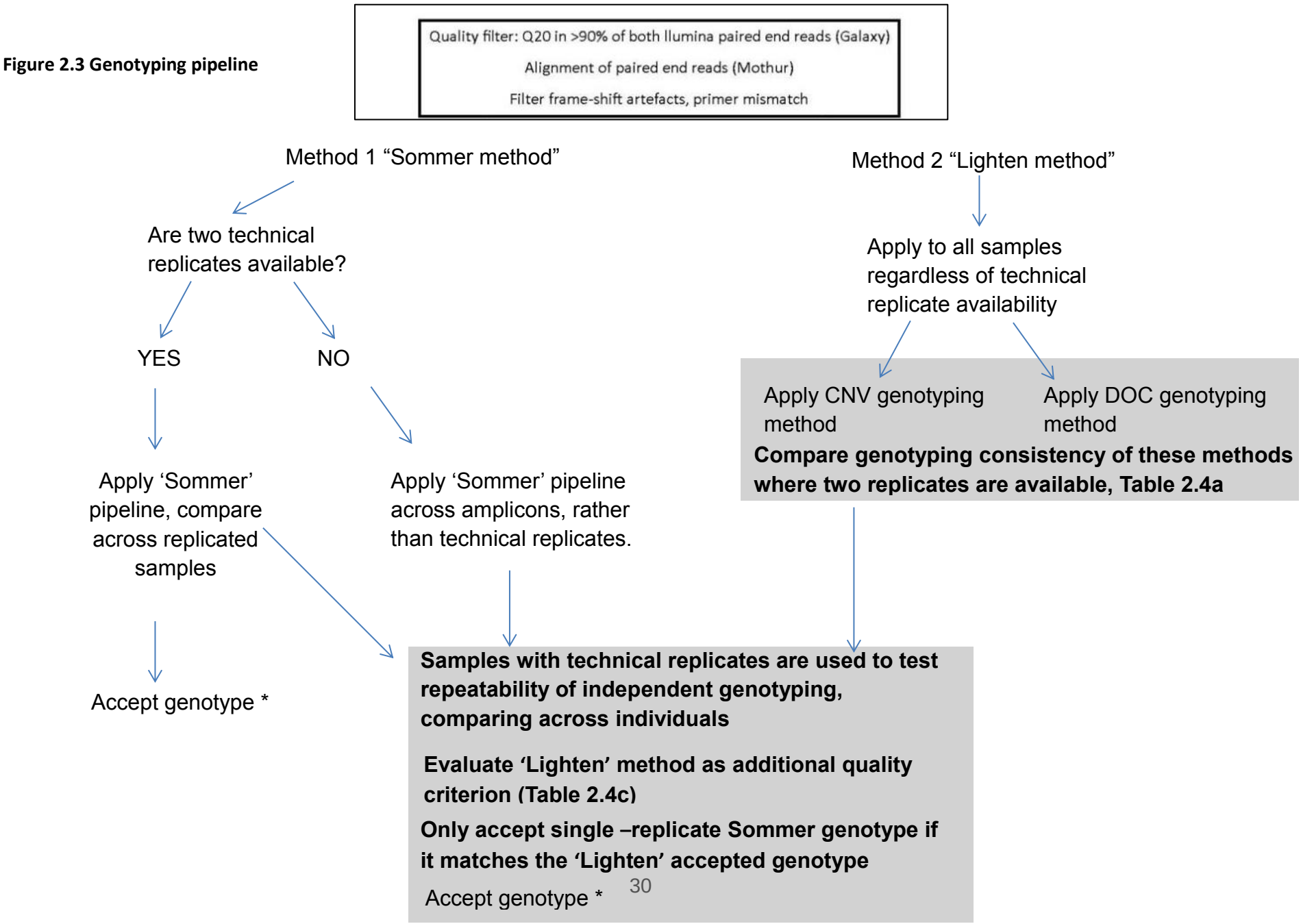
and DOC estimates for raw read counts and efficiency corrected read counts were thus compared in their reliability to make consistent estimates between replicates.

These methods were synthesized into a new approach (see Results). Consistency of genotyping methods was compared between the Sommer method with replicates available to the methods of Lighten. Further, repeatability of genotype estimates obtained from the Lighten methods was compared between replicates. Finally, the repeatability of genotype estimates from the Sommer method when using only a single replicate was compared to estimates obtained from both replicates. The criterion of a matching estimate provided by the Lighten method was evaluated to enhance accuracy of single replicate estimates, when using the Sommer methods (see Results, Figure 2.3).

Results

The Illumina Nano run reported here resulted in n=1,227,780 reads (Table 2.2). A small number of reads were observed in negative controls: the mean reads for these were 10, 1.8, 0.2, 3.9 and 23.8 for the DAA, DAB, L3I, L3II and ZE loci, respectively. The fit of the overall dataset to specific copy number scenarios was significantly better after control read subtraction (F-test: F=0.67, p<.0005). Mean coverage (per amplicon after control read subtraction), sample size, percentage of replicated samples, relative amplification efficiency and the calculated threshold T1 to obtain a minimum of two reads are shown in Table 2.3 for each locus.

Figure 2.3 Genotyping pipeline



* Consistency between replicated Sommer genotypes with Lighten genotypes is assessed, Table 2.4b

Table 2.2 Filtering steps of sequence reads obtained from an Illumina Miseq Nano run

Illumina Nano Run (all reads: 1227780)						
	Target	**Both primers**	**Tag, full length**	**Reading frame**	**Quality filter (above Q20 in ≥ 90%) of both paired reads**	**After singleton filter**
DA	268121	244191	225992	209902	171244	169596
DB	203094	171962	149336	138036	98488	98978
ZE	222510	224422	217555	188613	106259	165092
UBA	390325	163792	149143	228351	238568	163973
sum	1084050	923342	844896	764902	614559	597639
%	**88**	**75**	**69**	**62**	**50**	**49**

For each locus the number of sequences retained after each filtering step is given for the Illumina Miseq Nano run on a total of twelve populations; for the UBA locus the results are summarized for the two different primer pairs used (L3I and L3II)

Table 2.3 Summary statistics of analysed loci

	DAA	**DAB**	**UBA (L3I)**	**UBA (L3II)**	**ZE**
Mean coverage	143	89	96	55	106
Percentage of replicated samples (total number of samples)	72 (442)	49 (378)	24 (362)	51 (406)	53 (426)
Highest allele amplification efficiency	1.8	1.6	3.3	3.4	2.7
Lowest allele amplification efficiency	0.9	0.4	0.13	0.4	0.76
T1 (corrected for minimum efficiency)	20	20	146	63	40

Mean coverage, percentage of replicated samples, estimates of minimum and maximum allele amplification efficiency and Galan's T1 threshold corrected for minimum efficiency per locus

Several samples were excluded from the analysis due to a potential assignment error (DAA n=7; DAB n=1; L3II n=7, ZE n=2), where the most frequent variant within one amplicon was not present in the replicate. For the DAA and DAB locus the genotypes of most individuals were consistent with the single classical class II locus system found within other salmonids (Stet et al. 2002). However, individuals for the DAA (n=4) and DAB (n=1) loci exhibited three alleles and were excluded from subsequent analysis. Highest allele count for L3I locus was n=6 and n=5 for the L3II and ZE loci.

The repeatability of Lighten genotype estimates increased in all cases by incorporating classifications of the Sommer method and disregarding artefact read counts (CNV method 8% and DOC method 17% of mean increase in repeatability across loci), which was mainly driven by the exclusion of chimeric sequences. This was therefore implemented as a general approach. The repeatability of genotype estimates form the Lighten methods on 'raw' reads, which had not been corrected for amplification efficiency, between replicates ranged from 21% to 100% depending on the method applied and the locus (Table 2.4A). The most stringent method of genotyping, using only samples with matching CNV and DOC estimates, generally performed best (Table 2.4A). The consistency of genotype estimates of the Lighten methods on 'raw' reads with genotype estimates of the Sommer method ranged from 38% to 98% depending on the method applied and the locus (Table 2.4B).

The correction of raw read counts for low amplification efficiency increased the repeatability of genotype estimates between replicates, applying the Lighten methods in most cases for the DAA and DAB, where efficiency estimates are not confounded by copy number variation (Table 2.4A). Consistency of genotype estimates between the Sommer and Lighten methods increased for the DAA and DAB loci, when correcting raw read counts for low amplification efficiency (Table 2.4B). This was not the case for the more than one locus systems L3I, L3II and ZE, where correcting for low allele amplification efficiency did not enhance repeatability and consistency.

Independent assessment of amplicons between replicates (i.e. comparing genotypes assigned by the Sommer Method) variable: 86% for DAA and DAB, 67% for L3I, 34% for L3II and 26% for ZE. When using the Lighten methods in conjunction with the Sommer methods for genotype calling, repeatability was enhanced in all cases (Table 2.4C). For DAA, DAB and

L3I loci, genotyping confidence of ≥95% was achieved using matching genotype estimate provided by the Lighten methods as an additional quality criteria for genotyping not replicated samples (Table 2.4C). This threshold was not achieved for L3II or ZE loci, which also showed lowest repeatability between Lighten estimates (Table 2.4C). In the case of the ZE locus where the mean amplicon coverage was higher, an increase in consistency between Lighten estimates was observed with increasing amplicon coverage (Figure 2.4). The same was true for genotype estimates of replicated samples when evaluated independently. Here, a matching CNV-DOC estimate in addition to a coverage above 200 resulted in 100% congruency with estimates made with both replicates available (Figure 2.4). However, only one non- replicated sample fulfilled the criteria of having a high enough coverage above 200 and a matching CNV-DOC estimate, so that these criteria could not be effectively applied to our data set.

Quality criteria to include a genotype were set so that for a genotype to be assigned at least one replicated sample had to be above Galan's T1 threshold (corrected for minimum allele efficiency). Samples without a successful replicate had to be above Galan's T1 threshold and have a matching genotype using the Lighten methods (Figure 2.3). For the DAA, DAB (CNV or DOC match) and L3I locus (CNV=DOC match), high reproducibility was shown with this additional selection criterion (Table 2.4). For the L3II and ZE locus only replicated samples were considered for down-stream analysis.

Fifteen alleles were identified for the DAA and L3I locus, ten for the DAB, thirty for the L3II and twelve for the ZE locus. Of all individuals 8% for the DAA locus, 5% for the DAB locus, 20% for the L3I locus, 60% for the L3II locus and 40% for the ZE locus exhibited low efficiency alleles, which were found at frequencies lower than the most frequent artefact. Two alleles for the DAA, L3I, L3II and ZE locus respectively, and four alleles of the DAB locus had been previously identified by Sanger sequencing and genotyped in all individuals, where they had been observed.

Table 2.4 Evaluation and comparison of genotyping pipelines

		DAA			DAB			UBA (L3I)			UBA (L3II)			ZE		
		DOC=CNV (%)	CNV (%)	DOC (%)	DOC=CNV (%)	CNV (%)	DOC (%)	DOC=CNV (%)	CNV (%)	DOC (%)	DOC=CNV (%)	CNV (%)	DOC (%)	DOC=CNV (%)	CNV (%)	DOC (%)
A	**raw**	93	88	72	100	90	76	86	65	52	59	41	21	47	48	22
	corrected	95	92	75	93	88	88	73	55	39	53	43	19	48	42	19
B	**raw**	96	98	91	93	93	84	79	82	78	44	44	38	72	72	56
	corrected	97	98	94	99	99	98	79	83	71	43	40	37	66	72	52
C	**raw**	96	95	95	98	98	95	95	87	88	64	54	64	65	60	53
	corrected	97	96	96	98	98	95	88	83	82	62	61	69	65	58	53

A) Repeatability of Lighten genotype estimates between replicates

B) Consistency of Lighten with 'Sommer' genotype estimates

C) Repeatability of genotype estimates using single replicates independently, when using different Lighten estimates as additional quality criterion

Consistency of genotype estimates for different selection criteria between the methods as described by Lighten et al. (2014) and by Sommer et al. (2013) per locus; the methods presented by Lighten are evaluated for copy number estimate of variants (CNV), degree of change (DOC) estimate and matching CNV and DOC estimate (CNV=DOC)

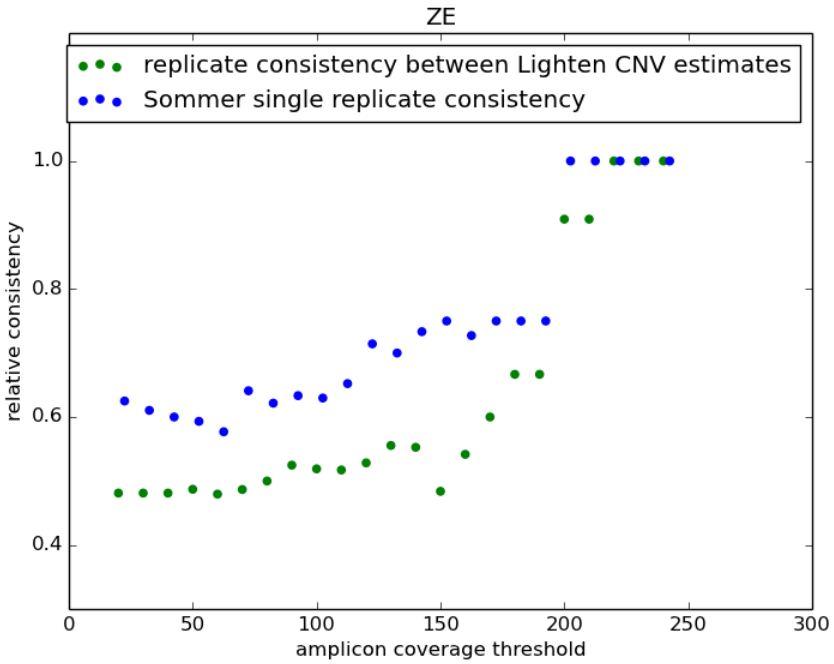


Figure 2.4 Relationship between per amplicon coverage and consistency*: of Lighten estimates and of single sample estimates using the Sommer method using the Lighten estimates as additional selection criterion*

Discussion

There is a growing trend to study functional instead of neutral genetic variation in conservation genetics (Hoffmann and Willi 2008; Kirk and Freeland 2011). Here, competing approaches for candidate gene studies have been evaluated to improve experimental design, data-through-put, cost-efficiency and genotyping confidence in future studies.

Sample replication (e.g. in the Sommer Pipeline; Sommer et al. 2013), is an ideal approach to measure functional genetic variation using NGS, because it allows for rigorous quality control. This is particularly true where cross-contamination is a problem. As the sensitivity of NGS methods is much higher than traditional methods, cross-contamination might generally be underestimated (Tosar et al. 2014) and the an inclusion of negative controls in the sequencing run is recommend even in the absence of bands on an agarose gel as an important measure to indicate the level of noise present in a sample due to cross contamination. To account for background noise, subtraction of the highest read counts in negative controls (Nguyen et al. 2015) could significantly improve the fit of the overall data

set to specific copy number scenarios. Additionally, the inclusion of a replicate of each sample is still the best way to distinguish between cross-contamination (or artefacts) and true variants (Sommer et al. 2013). This is evident from the moderate to low repeatability of genotypes from independent amplicons compared to genotypes obtained through replicate comparison. When comparing variants across all individuals within the dataset, artefacts were identified less efficiently. This may be because more than half of all artefacts are found in more than one amplicon. Such repeatable errors have been identified in previous studies (Gomez- Alvarez et al. 2009; Gilles et al. 2011; Lighten et al. 2014a) and make their distinction as artefacts without the comparison against a sample replicate more difficult. Additionally, finding alleles that exhibit good efficiency in only one replicate can be used to identify cross-contamination. Whilst increasing automation in the library preparation process could help to improve cross-contamination risk, this might still not be available for many conservation genetics studies. Therefore, an efficient informatics approach to identify dubious samples is of great benefit, without the need of replicating every sample. Here, it was shown that including a modified Lighten methodology can result in a minimum of 95 % repeatability for replicated amplicons compared to genotypes assigned using both replicates for the DAA, DAB and L3I loci. This highlights the benefits of using more than a single method for genotyping (Lighten et al. 2014a).

For L3II and ZE loci, the results indicate that genotyping confidence can be improved through increasing sequencing coverage. These loci have more than a single locus copy and thus showed the highest number of alleles per individual. These loci therefore showed high number of individuals which exhibited low frequency alleles, lower than any artefact, so that 60% and 40% of genotypes changed by accounting for low efficiency. If prior knowledge about the expected highest number of alleles exists for a locus, this can be adjusted for by increasing sequencing depth for these loci. Here, for a maximum of six alleles per individual, a sequencing coverage of 200 would reduce the variance in Lighten estimates between replicates.

Because the Lighten methods do not account for unequal allele amplification efficiencies, incorporating efficiency estimates derived from the Sommer method was proposed here, enhancing consistency for the one locus DAA and DAB system. However, where more than a single locus is amplified, estimation of efficiency is hampered by the fact that efficiency and

copy number cannot easily be distinguished. Therefore no improvement in repeatability between replicates or consistency between methods was achieved by correcting raw reads by amplification efficiency. Further improvement of allele efficiency with respect to copy number variation would be beneficial.

Here the Sommer and Lighten NGS methods were synthesized so the drawbacks of each method (requirement of replicates or high coverage, respectively) may be overcome. Together, it is suggested that these methods increase sequencing accuracy and, consequently, efficiency. The standard use of the Sommer methods is recommended with the addition of replicated sub-samples in order to formally evaluate repeatability. Further, the addition of a matching genotyping estimate adopted from the Lighten methods is recommended as an additional quality criterion.

Chapter 3: Evolutionary genetics of European grayling (*Thymallus thymallus*) Major Histocompatibility (MH) genes

Abstract

Immune genes of the major histocompatibility complex (MHC) are among the most polymorphic genes known in vertebrates. This high degree of variation is thought to result from dynamic evolutionary processes between host and pathogens. Here, variation at major histocompatibility class I and class II genes within twelve populations of European grayling (*Thymallus thymallus*) is determined, extending existing knowledge on other salmonid taxa. The results suggest a single classical locus for the class II (with DAA and DAB chains) and high allelic divergence within the classical class I (UBA), confirming the generality of this pattern among salmonids. On the other hand, alleles derived of the classical class I UBA locus and the non-classical ULA could not be distinguished due to high similarity. Phylogenetic comparison of variants described for the U-lineage in grayling, suggest the presence of non-classical UCA, UFA and UHA loci. High allelic variation is described for the non-classical ZE lineage and evidence for balancing selection is given for sites outside the peptide binding region (PBR) as it is described in humans. Evidence for balancing selection was also found for the class II DAA and DAB loci. Moreover, evidence for recombination at both classes confirms the important role of this evolutionary force in maintaining high degrees of polymorphism at the MHC. Together this study adds to the current knowledge on immune genetic variation in salmonids and the patterns of selection acting upon it.

Introduction

The evolutionary history of the major histocompatibility complex (MHC) dates back ~460 million years to the emergence of jawed vertebrates (gnathostomes), with cartilaginous fish (sharks) as the oldest group (Flajnik et al. 1999). It likely originates from a proto-MHC region present in an ancestor of all bilaterian species (Danchin and Pontarotti 2004). The evolution of the adaptive immune system in jawed vertebrates, with the MHC as a central component, is believed to have been facilitated by two genome duplications, hypothesised to have happened once before the emergence of jawless vertebrates (agnathans) and again in a common ancestor of jawed vertebrates (Flajnik and Kasahara 2001). The relaxation of functional constraints in paralogous regions may then have allowed rapid evolution of new functions giving rise to the vertebrate MHC and adaptive immune system (Rached et al. 1999).

Fully-functional (classical) MHC genes, important for the initialization of the adaptive immune response and the discrimination of self and non-self (Chaplin 2010) have been found in all jawed vertebrates studied (Kulski et al. 2002). However, the architecture of the MHC and the number of classical loci in the class I and class II genes can be variable across species (Aoyagi et al. 2002; Shiina et al. 2004; Traherne 2008; Star et al. 2011). Although an increased number of MHC loci could be advantageous for the recognition of pathogens, trade-offs limit the number of MHC loci: with an increased number of MHC molecules, more self-peptides are likely to bind to them, which may induce negative selection on T-cells in the thymus and therefore reduce overall T-cell repertoire (Takahata 1995). Gene expansion and contraction at the MHC is therefore a dynamic evolutionary process (Piontkivska and Nei 2003) with expansions especially occurring during times of species radiation (Klein et al. 1993a; O'Connor et al. 2016). During times of MHC contraction formerly classical genes can degrade to non-classical genes with limited expression and functionality (Miller et al. 2002). Relieved from functional constraints, non-classical genes can gain new functionality or turn into pseudogenes (Raes and Van de Peer 2003; Magadum et al. 2013).

The genes of the MHC are the most variable expressed loci known in vertebrates (Hedrick 1999). However, mutation rates are not higher at the MHC than for other parts of the

genome (Klein et al. 1993b) and inter- and intra-locus recombination and gene conversion are thought to play important roles in the generation of the high levels of variation observed (Zangenberg et al. 1995; Parham and Ohta 1996; Reusch and Langefors 2005; von Salomé and Kukkonen 2008). Non-classical MHC genes are an interesting example of standing genetic variation in this context, as variation at classical loci can be increased substantially by inter-locus recombination (Reusch and Langefors 2005) and additionally because of possibilities for different adaptation and more flexible expression patterns (Taylor and Raes 2004).

Different to many other vertebrate species the MHC class I and class II genes are found in different linkage groups in teleost (bony fish) (Bingulac-Popovic et al. 1997; Sato et al. 2000) and are therefore simply assigned as MH (Stet et al. 2003). This may have arisen by maintaining different paralogues of the MHC functional and results in high flexibility in the combination of class I and class II alleles in different haplotypes, which allows selection to act independently on each class (Shum et al. 2001; Stet et al. 2003). Though differences in evolutionary mode and history have been also shown in vertebrate species where both classes are in linkage (Boyson et al. 1996; Bonneaud et al. 2004), the pattern observed in mammals is reversed in teleosts (Ohta et al. 2000). In mammals, class I genes show species-specific lineages (Adams and Parham 2001), whilst in teleost fish the class I is characterized by highly divergent allelic lineages, which are shared across species and maintained over long evolutionary time scales (Shum et al. 2001; Miller et al. 2006; McClelland et al. 2013). In contrast the class II evolves in a species-specific manner in teleosts and does not show divergence in different allelic lineages, as is found in primates (Bontrop et al. 1999). The increased flexibility of unlinked MH classes may compensate for the low number of classical loci found in salmonids, where the MH class I and class II genes both contain a single classical locus (Aoyagi et al. 2002; Grimholt et al. 2003).

In salmonids, duplicated regions of the MH class I can be found in different linkage groups, containing various non-classical loci (Shiina et al. 2005; Lukacs et al. 2007, 2010). Five different class I lineages have been classified as U, Z-, S-, L- and P-lineages and described in detail by Grimholt (2016), summarized as follows. The U lineage contains the classical locus, defined as UBA, fulfilling all characteristics for this designation. These are the conservation of eight functionally important residues that anchor the C-and N-termini of peptides (Brown

et al. 1993; Kaufman et al. 1994), ubiquitous expression, selection on peptide binding residues (PBR) and polymorphism (Miller et al. 2006). Other non-classical loci of the U-lineage have been described as UCA, UDA, UEA, UFA, UGA, UHA and ULA. The UFA lineage is a pseudogene in salmonids (Miller et al. 2006), whilst the others show variable degrees of conservation of peptide anchor residues, expression and polymorphism. The ZE locus exhibits various characteristics that make its assignation as a non-classical locus questionable (Stet et al. 2003). The presence of conserved peptide anchoring residues, exon-intron organization, ubiquitous expression, and polymorphism support the classical nature of this locus in several teleost species (Kruiswijk et al. 2002; Miller et al. 2006; Lukacs et al. 2010). The sites of polymorphism observed at the ZE locus are, however, not located within the PBR (where the PBR is deduced from alignment with the human HLA peptide), which is in contrast highly conserved (Grimholt et al. 2015). It has therefore been suggested that residues involved in peptide binding are not consistent with those inferred from the human HLA (Kruiswijk et al. 2002) and this could explain the maintenance of the observed polymorphisms (Dirscherl and Yoder 2014). However, the high degree of functional conservation and the presence of at least one expressed Z-locus in every teleost studied, suggest an important functionality of this locus (Grimholt et al. 2015). Upregulation of ZE-expression after vaccine immunization with live attenuated *Edwardsiella tarda* in zebrafish suggests that it has an immune function (Yang et al. 2012). The L-, S- and P-lineages do not show conserved peptide anchoring residues and are not predicted to have peptide binding properties (Grimholt et al. 2015).

Non-classsical MHC genes have been shown to have a variety of functions, mainly investigated in human or other mammalian model systems (Hofstetter et al. 2011). These include various forms of pattern recognition, binding peptides or lipids, as well as natural killer cell or T cell receptors, suggesting important immune regulatory functions of non-classical MHC genes, bridging the innate and adaptive immune response (Rodgers and Cook 2005). Associations of pathogen resistance with variation at non-classical class I loci has been also documented in salmonids, in particular with the UCA/UDA or UEA loci (Miller et al. 2004; Ozaki et al. 2007; Johnson et al. 2008).

The overall aim of this study is to investigate extension of immunogenetic surveys in wild salmonids beyond MH class II loci. Evolutionary genetic surveys of MHC class I UBA α2, class

II DAA α1and DAB β1 regions, non-classical ZE locus and other non-classical U-lineages in European grayling are presented. In this study:

(i) predictions of a single classical locus for both the class I and class II regions, with high allelic divergence in the class I (Shum et al. 2001) and similar levels of diversity within both class II chains (Gómez et al. 2010), as found in other salmonid taxa, are tested.

(ii) diverse allelic variation at the classical UBA locus/loci (Miller et al. 2006) is predicted and the U-lineages within the α2 region are broadly targeted to investigate this.

(iii) evolutionary genetic variation of non-classical loci of both the U and Z-lineage present in grayling is quantified, hypothesising to find polymorphic regions of adaptive nature. This hypothesis is tested by investigating if selection is acting on non-classical genes across a large number of individuals (96-223).

Materials and Methods

Sequencing of 37-40 individuals from each of twelve populations on an Illumina Miseq and subsequent genotyping was performed as described in Chapter 2.

Prior to estimation of evolutionary genetic parameters, amino acid (AA) sequences were derived for the class I α2 region by the identification of the reading frame within the overlap of the sequences obtained from two different primers used to target the UBA locus (see previous chapter). Multiple class I lineages/loci amplified (see Results) so assignment to particular lineages and loci was necessary. To do this, amplicon sequences were aligned with published sequences of the U- and Z lineage from closely related salmonid taxa as described by Shum et al. 2001, Consuegra et al. 2005b, Kiryu et al. 2005, Shiina et al. 2005, Miller et al. 2006 and McClelland et al. 2011 in *Oncorhynchus mykiss*, *Oncorhynchus nerka* and *Salmo salar* using ClustalW (Larkin et al. 2007). The best model of evolution was then inferred using ProTest 2.4 (Abascal et al. 2005) with AICc as the model selection criterion. The parameters of the inferred model were implemented in PhyML 3.1 (Guindon et al. 2010),

which estimates maximum likelihood phylogenies. Tree stability was estimated over 100 bootstrap cycles, and consensus trees were built in Phylip (Felsenstein 1993).

For all loci (class I and class II) the evolutionary history of the observed MH sequences and the PBR was then investigated based on alignments of salmonid MH sequences to human HLA sequences (Stet et al. 2002; Grimholt et al. 2015). Residues important for peptide anchoring of classical class I genes were inferred from alignment to human HLA sequences (Grimholt et al. 2015) and compared with substitutions observed in different classical and non-classical lineages in other salmonid taxa. Sites within the six peptide binding pockets of the MH class I sequences were assigned as presented by Grimholt et al. (2015) and properties inferred based on Saper et al. (1991) and Hashimoto et al. (1999). Peptide binding sites of MH class II sequences were identified based on human HLA crystal structures (Brown et al. 1993). The following analysis were done only for the MH I ZE and the MH II loci, where an unambiguous locus assignment was possible. Mean nucleotide diversity (Kimura 2-parameter model) was calculated for both PBR and non-PBR sites in MEGA 6.0 (Tamura et al. 2013). The rates of synonymous and non-synonymous substitutions were calculated using the Nei & Gojobori method (Nei and Gojobori 1986) applying the Jukes Cantor correction (Jukes and Cantor 1969) for multiple hits in MEGA (Tamura et al. 2013), using a Z-test for positive selection (Nei and Kumar 2000). To identify sites under selection a suite of tools available on the “datamonkey” webserver (Delport et al. 2010a) were implemented. First the Genetic Algorithm Recombination Detection (GARD) tool (Pond et al. 2006) was used to test for recombination within sequences, which could affect the inference of phylogenetic trees and find possible break points. To detect sites under episodic diversifying selection a mixed effects model of evolution (MEME) was used (Murrell et al. 2012). The Fast Unconstrained Bayesian AppRoximation (FUBAR) method was used to detect sites under pervasive diversifying or purifying selection (Murrell et al. 2013). A neighbour joining phylogenetic tree was used as input for both methods. Where evidence for recombination was given, a consensus phylogenetic tree was built using separate neighbour joining trees for each segment and used as an input for the inference of selection. A codon test was implemented to identify the best fitting model of nucleotide substitution, which was used for all analysis (Delport et al. 2010b). Results at a 0.1 level of significance were also considered as evidence for selection, using the FUBAR and MEME

methods. In order to detect non-random amnio acid replacements as a consequence of selection for particular physiochemical properties, TREESAAP 3.2 (McClellan and Ellison 2010) was implemented, using a sliding window of 15 AAs. Here, amino acid substitutions are ranked according to the resulting change in physio chemical properties into categories 1-8 with increasing magnitude of change. Only changes of categories seven and eight, representing most drastic changes, were considered as evidence for selection (McClellan and Ellison 2010). Alleles were named using the nomenclature described by Shum et al. (2001).

Results

Evolutionary genetics of MH class I α2

For the UBA locus, a total of 35 unique sequences were identified, all of which translated to different alleles based on AA identity. Eight sequences were obtained from both primer pairs, eight only from the L3I primer and 19 only from the L3II primer. For 97 individuals, genotypes were obtained for both primers (see previous chapter). The majority of individuals showed between four and eight alleles (Figure 3.1), suggesting that generally a minimum of four loci were amplifying.

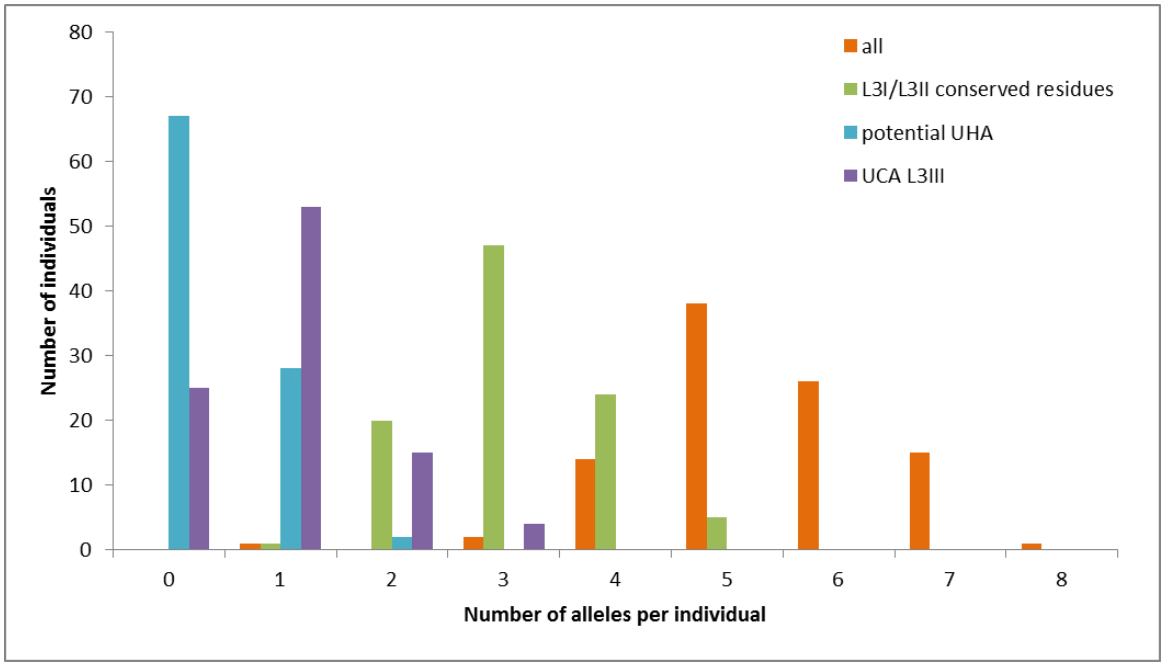


Figure 3.1 Alleles observed per individual for the MH class I α2: *only given for individuals genotyped for both primer pairs; allele counts are shown for all alleles, for only those with conserved anchor residues of the L3I and L3II lineage, the potential UHA alleles and the UCA L3II lineage*

Evidence for the amplification of multiple class I U lineages

Further evidence for the amplificiation of multiple U lineage loci using primers targetting the UBA locus is demonstrated by clustering of lineages in the phylogenetic analysis (Figure 3.2). ProTest inferred as the best fitting model of evolution in both cases Blosum62 (Henikoff and Henikoff 1992) with a gamma distribution of rates between sites of 4, an alpha value of 0.82 for the UBA and deducing equilibrium AA frequencies from the alignment (Chao 1984). Analysis of all obtained sequences with MH class I sequences of other salmonid taxa, clustered twelve sequences within the L3III lineage of the UCA locus (Figure 3.2). All of them deviated from the conserved pattern described for residues involved in peptide anchoring at position 40 (W/L), corresponding to our sequence alignment, which has been observed for the UBA L3II and Z-lineage in other salmonids (Table 3.1). Two of them had an additional K/N substitution at position 39 (Table 3.1). Up to three alleles of the UCA cluster were observed per individual (Figure 3.1). No sequences clustered closely to the classical UBA L3III lineage (Figure 3.2). Two obtained sequences were observed closest to the UHA L3VI lineage of other salmonid taxa (Figure 3.2). They showed conservation of peptide anchoring

residues for all of the five positions our sequences covered, which is true for the UHA lineage in addition to all UBA and the ZE lineages (Table 3.1). The majority of individuals (67) did not have one of the potential UHA alleles (Figure 3.1). Of the remaining 21 alleles four showed substitutions within the peptide anchoring residues (Table 3.1). Three of them exhibited a T/S substitution at position 36, which has also been observed for alleles of the UFA pseudogene and the UGA lineage in other salmonids, one with an additional K/N substitution at position 39 and one with an additional W/L sustitution at position 40 (Table 3.1). The other allele that deviated from the conserved pattern of peptide anchoring residues had a W/C substitution at position 40, not described for any salmonid class I lineage, but found within the human non classical HLA-G lineage (Grimholt et al. 2015) (Table 3.1) . This allele had the highest overall frequency and was observed in a total of 75 individuals. The remaining 17 alleles showed complete conservation of peptide anchoring residues, or a W/L substitution at position 40, which is also observed in classical alleles. They were mainly associated with the L3I and L3II lineage, though some alleles clustered close to the UEA lineage but did not show the T/I substitution described as characteristic for this lineage at position 36. Because of the high similarity between the classical UBA L3I lineage and the non classical ULA L3I lineage, sequences derived from either of these loci could not be distinguished. The majority of individuals showed between two and four alleles of the conserved L3I/L3II category at the class I α2 region, five individuals exceeded this and showed five alleles and one individual had only one allele (Figure 3.1).

Table 3.1 Conserved peptide anchoring residues at the C and N peptide terminus covered by sequencing grayling MH class I α2

	Peptide C terminus				Peptide N terminus	
	Y/F 6/123	**T 36/143**	**K 39/146**	**W 40/147**	**Y 53/159**	**No of sequences**
A) Observed						
	*	*	*	*	*	15
	*	*	*	L	*	14 (10 L3 III)
	*	S	*	*	*	1
	*	S	*	L	*	1
	*	S	N	*	*	1
	*	*	*	C	*	1
	*	*	N	L	*	2 (L3 III)
B) Described						
UBA L3I	*	*	*	*	*	
UBA L3III	*	*	*/R	*	*	
ULA L3I	*	*	*	*	*	
UBA L3II	*	*	*	*/L	*	
UFA (pseudo) L3II	*	S	*	*	*	
UCA/UDA L3II	*	*/M	*/E	*	*	
UEA L3IV	*	I	*	*	*	
UGA L3V	L	S	*	*	*	
UHA L3VI	*	*	*	*	*	
ZE L3VII	*	*	*	*/L	*	
L-lineage	A	Q	D	G	V	
S-lineage	*	Y	*	R	V	
P-lineage	*	*	D	*	F	

Positions are given for the alignment in this study and the according position in the full length alignment; (A) Alleles observed in grayling in this study, (B) Alleles described for different lineages in other salmonids (Kiryu et al. 2005; Miller et al. 2006; Lukacs et al. 2010; Grimholt et al. 2015)

Figure 3.2 Phylogenetic tree of salmonid MH class I α2*: All grayling (Thth) sequences were observed in this study; # at the end of the sequence indicates that all peptide anchoring residues are conserved and #L indicates a W/L substitution at position 40; genbank accession numbers are given at the start of sequence names of other taxa; Bootstrap values obtained over 100 cycles are given;*

Evolutionary genetics of ZE α1

Twelve alleles were identified for the ZE α1 regions, with ten of them corresponding to different AA translations (Figure 3.3a). There was no variation at the nucleotide and protein level within the PBR as inferred from the human HLA (Figure 3.3a). For one of the sites suggested by Kruiswijk et al. (2002) as an alternative PBR, evidence of pervasive diversifying selection at a posterior probability >0.9 (CI 0-3) was given by the FUBAR method, but no variation was found within the other two suggested sites, which were covered by the obtained sequences (Figure 3.3a). FUBAR identified an additional other site with posterior probability >0.95 (CI 0-2) to be under balancing selection (Figure 3.3a), whilst the MEME analysis did not give any evidence for selection. For one site FUBAR gave evidence for purifying selection with a probability >0.9 (CI 0-3), one position before a residue involved in salt-bridges (Figure 3.3a) (Stet et al. 2003). Mean nucleotide distance within the suggested PBR was 3 times higher (0.09) than for the rest of the sequence (0.03). However a z-test for positive selection was not significant ($p = 0.14$, $Z = 1.06$). GARD analysis did not give evidence for recombination at this locus. TREESAAP gave evidence for positive destabilizing selection on amino acid properties acting upon the last two thirds of the sequence (Figure 3.4). The tendency to form an alpha helix was the only positively selected amino acid property of category 8 ($p < 0.001$), representing most drastic changes. Identified selection on amino acid properties of category 7 included turn tendencies ($p < 0.01$), power to be at the N-terminal ($p < 0.05$) and compressibility ($p < 0.05$). Increasing selection for all of these properties was found after AA position 17 (Figure 3.4).

Evolutionary genetics of MH class II α1 and β1

Individual allele counts for the MH class II DAA and DAB locus provided evidence for a single classical class II locus in grayling, as has been described for other salmonid taxa (Grimholt et al. 2003). For the DAA and DAB locus 15 and 10 alleles were identified respectively, all of which encoded different protein sequences. Evidence for historical positive selection was identified for both the PBR region ($p = 0.0003$, $Z = 3.51$) and non-PBR region ($p = 0.02$, $Z = 2.06$) at the DAA locus. For the DAB locus evidence for positive selection was identified only for the PBR region ($p = 0.005$, $Z = 3.39$) and not for the non-PBR region ($p = 0.09$, $Z = 1.35$).

Average pairwise nucleotide distance was considerably higher (0.15, 0.21) for the PBR than for non-PBR regions (0.03, 0.05) for the DAA and DAB locus respectively. The F81 substitution model had highest support (lowest AIC) for both class II loci. GARD analysis gave no evidence for recombination events at the DAA locus and a neighbour-joining tree was used to identify sites under selection. Two breakpoints were identified for the DAB locus, at position 56 and 166 of the amino acid alignment, which was confirmed by significant topological incongruence using the KH test ($p < 0.05$), giving evidence for recombination at this locus. A consensus of the neighbour-joining trees for each segment was used to identify sites under selection. MEME inferred seven and six sites under episodic diversifying selection ($P<0.1$), with six and four of them significant at the 0.05 level for the DAA and DAB loci, respectively (Figure 3.3b, 3.3c). The FUBAR method detected 15 and 18 sites at a posterior probability >0.9 (CI 0-3) with evidence of pervasive diversifying selection, with 10 and 8 sites having a posterior probability >0.95 (CI 0-2) for the DAA and DAB locus respectively (Figure 3.3b, 3.3c). No evidence was found for purifying selection at the DAA locus. For the DAB locus, evidence for pervasive purifying selection was found with a posterior probability >0.9 (CI 0-3) for one site, next to the N-linked glycosylation site suggested for salmonids (Stet et al. 2002) and a posterior probability >0.95 (CI 0-2) next to the N-linked glycosylation site as described in humans (Figure 3.3c). From a total of 16 putative peptide binding sites, 10 had evidence for diversifying selection at the DAA locus (Figure 3.3b). For 5 other sites evidence for diversifying selection was found, three of them a conserved residue in the human HLA sequence. For the DAB locus, 11 of 18 putative peptide binding residues had evidence for diversifying selection (Figure 3.3c). Another 6 sites showed evidence for diversifying selection, one of them a conserved residue in the human HLA sequence. For the DAA gene TREESAAP identified one positively selected amino acid property at category 8, buriedness ($p < 0.05$), which is increased within the first quarter of the sequence (Figure 3.4). Thermodynamic transfer hydrophobicity was another selected amino acid property of category 7 ($p < 0.01$), which is increased at the end of the sequence (Figure 4). For the DAB gene a total of three amino acid properties had evidence for positive selection. These were compressibility ($p < 0.01$) within category 8 and power to be at the middle of an alpha-helix ($p < 0.001$) and isoelectric point ($p < 0.01$) within category 7 (Figure 3.4).

Comparison of diversity at the different class I and class II loci

The largest number of alleles and highest mean nucleotide distance were identified for the L3I/II lineages of the class I α2 domain containing the classical UBA locus (Table 3.2). However, this includes additional alleles of the non-classical ULA locus. The largest number of sites with evidence for balancing selection was identified for the class II loci. The DAB locus also showed the highest number of significant breakpoints as a result of recombination (Table 3.2). Evidence for recombination events was given for the class I α2 region as well, with both the L3I/L3II lineages and the L3III lineage showing one breakpoint. The class I α1 ZE locus had the lowest number of sites with evidence for balancing selection, but exceeded the non-classical UCA locus in mean nucleotide distance for the inferred PBR (Table 3.2). Evidence for purifying selection was found at one site at this locus as well, which was only also observed at the class II DAB locus at two sites (Table 3.2).

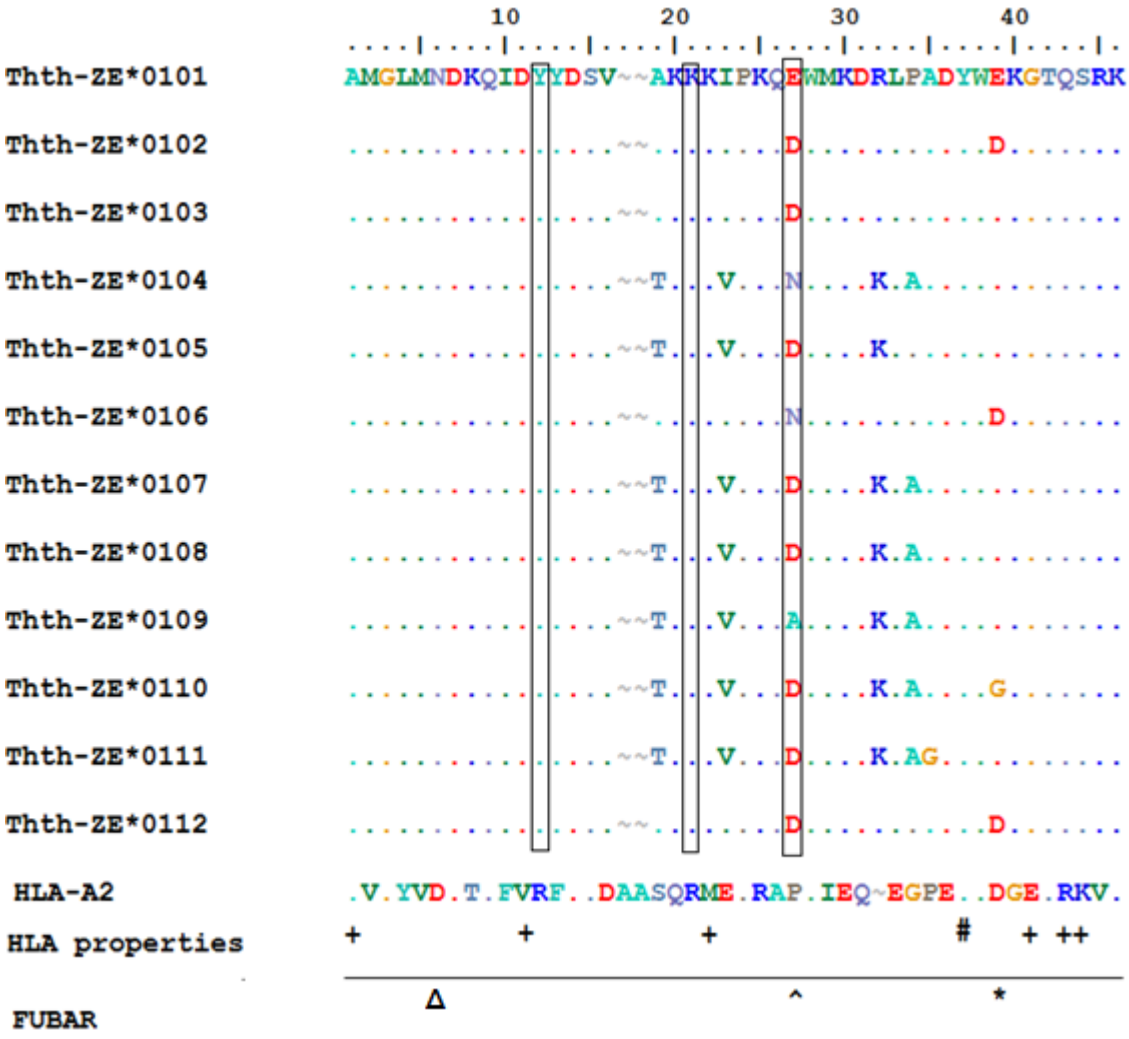


Figure 3.3a Alignment of MH class I ZE α1 sequences obtained in this study with the human HLA: *(+) indicate potential contact residues and # potential conserved residues of the human HLA (=) indicate N-linked glycosylation sites in the HLA sequence*; ** indicates diversifying selection was detected at a 0.05 significance level and ^ detection at a 0.1 significance level for the MEME and FUBAR method respectively; triangles indicate detection of purifying selection at a 0.1 significance level and filled triangles purifying selection at a 0.05 significance level*; *alternative PBR suggested by Kruiswijk et al. (2002)are boxed;*

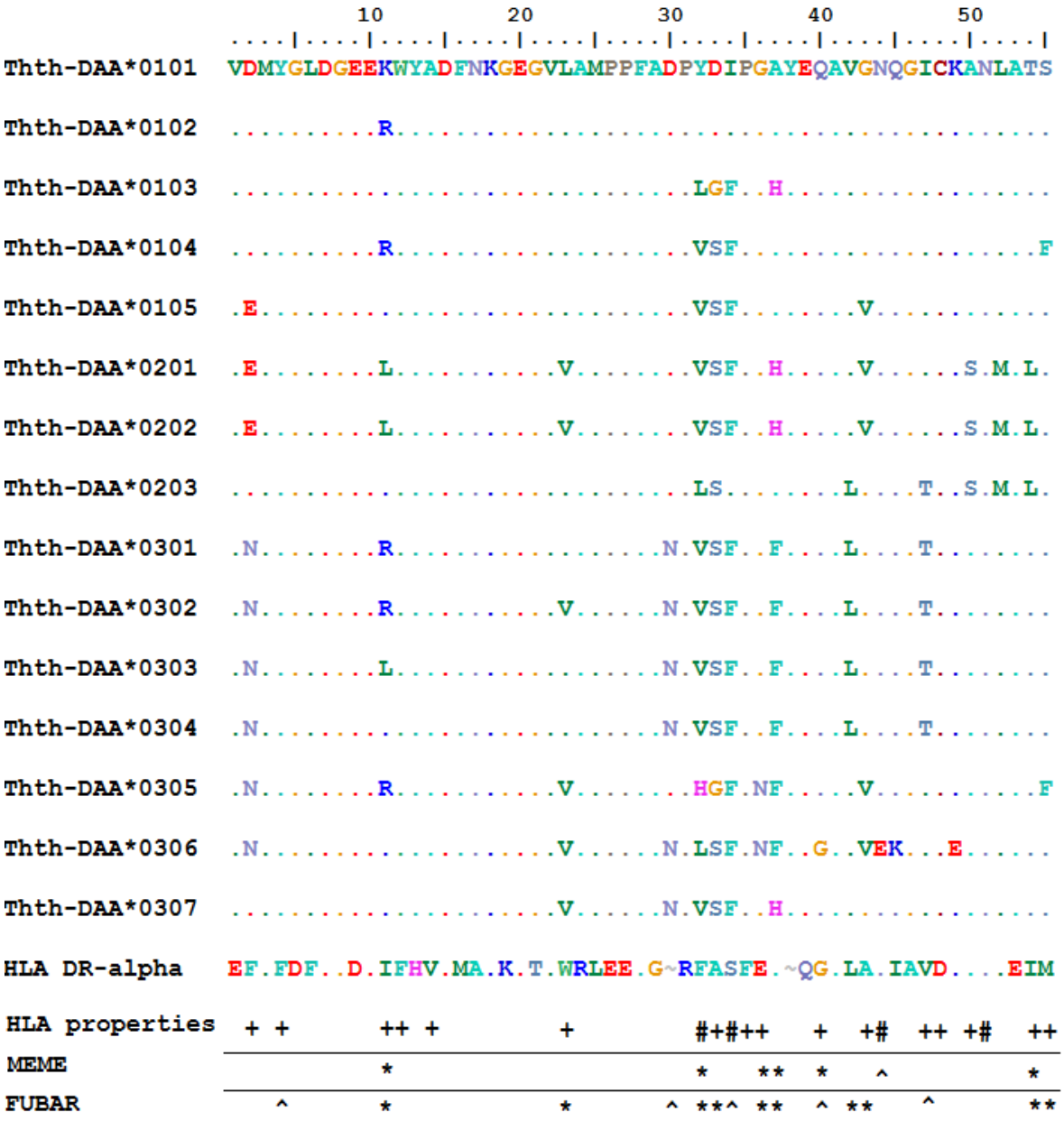


Figure 3.3b Alignment of MH class II α1 sequences obtained in this study with the human HLA: *(+) indicate potential contact residues and # potential conserved residues of the human HLA (=) indicate N-linked glycosylation sites in the HLA sequence*; ** indicates diversifying selection was detected at a 0.05 significance level and ^ detection at a 0.1 significance level for the MEME and FUBAR method respectively; triangles indicate detection of purifying selection at a 0.1 significance level and filled triangles purifying selection at a 0.05 significance level*;

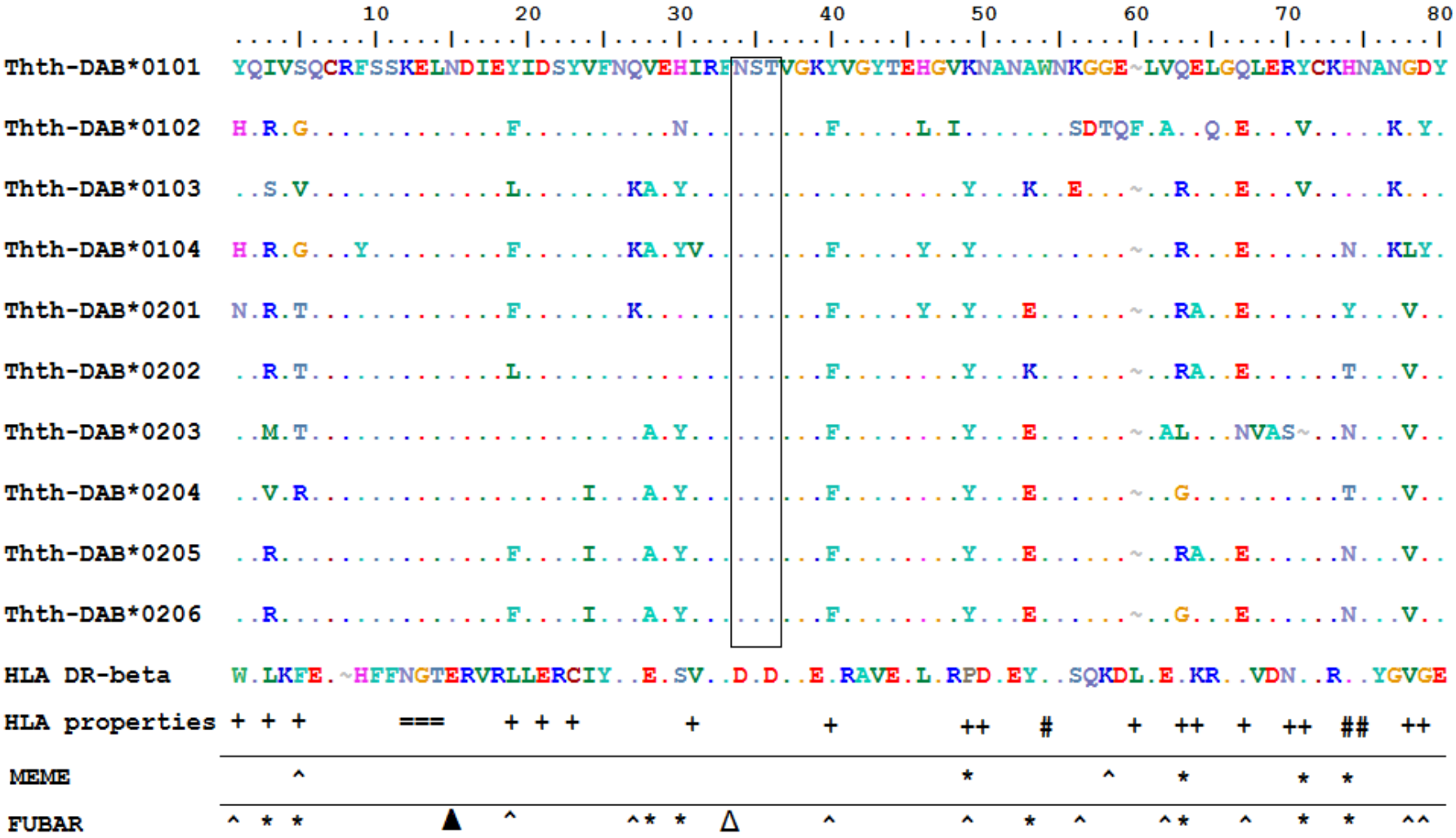
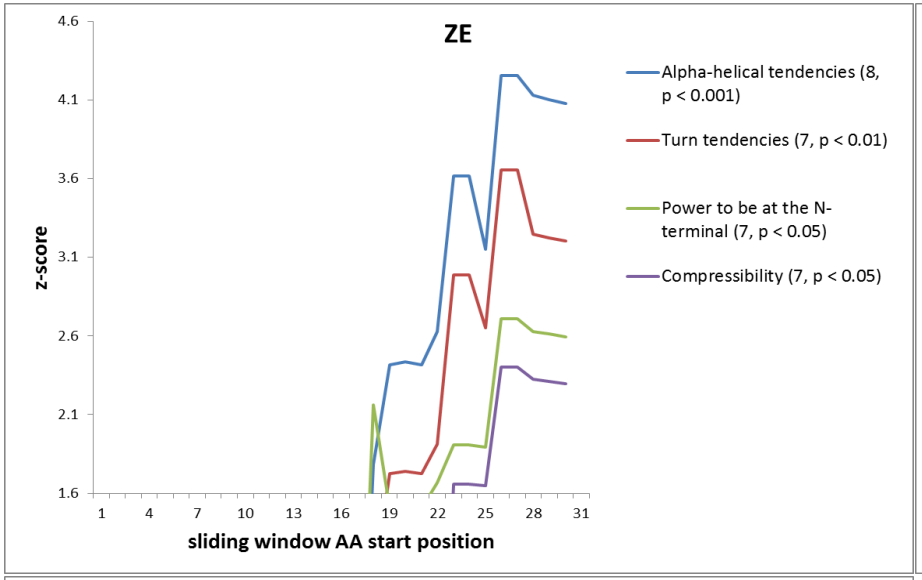
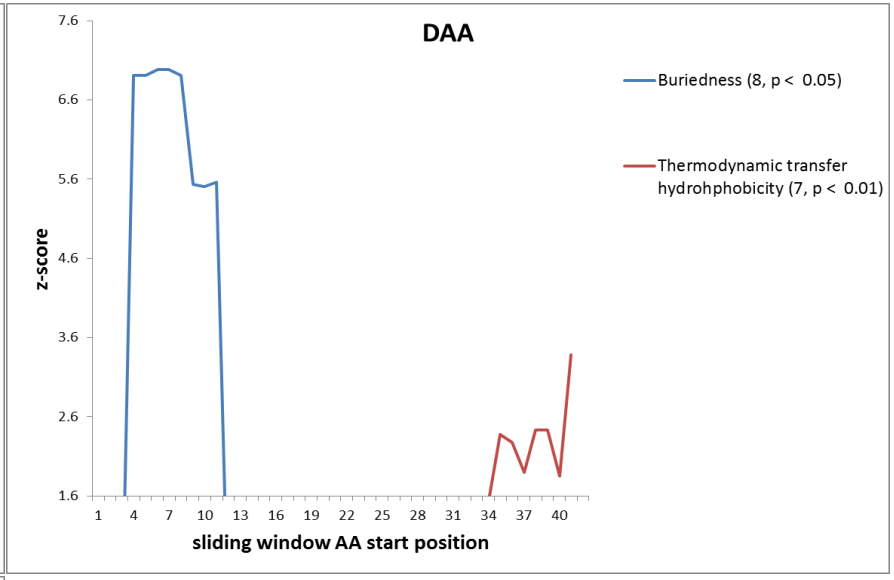


Figure 3.3c Alignment of MH class II β1 sequences obtained in this study with the human HLA; *(+) indicate potential contact residues and # potential conserved residues of the human HLA (=) indicate N-linked glycosylation sites in the HLA sequence*; ** indicates diversifying selection was detected at a 0.05 significance level and ^ detection at a 0.1 significance level for the MEME and FUBAR method respectively; triangles indicate detection of purifying selection at a 0.1 significance level and filled triangles purifying selection at a 0.05 significance level*; *potential N-linked glycolisation sites in salmonids are boxed (Stet et al. 2002);*





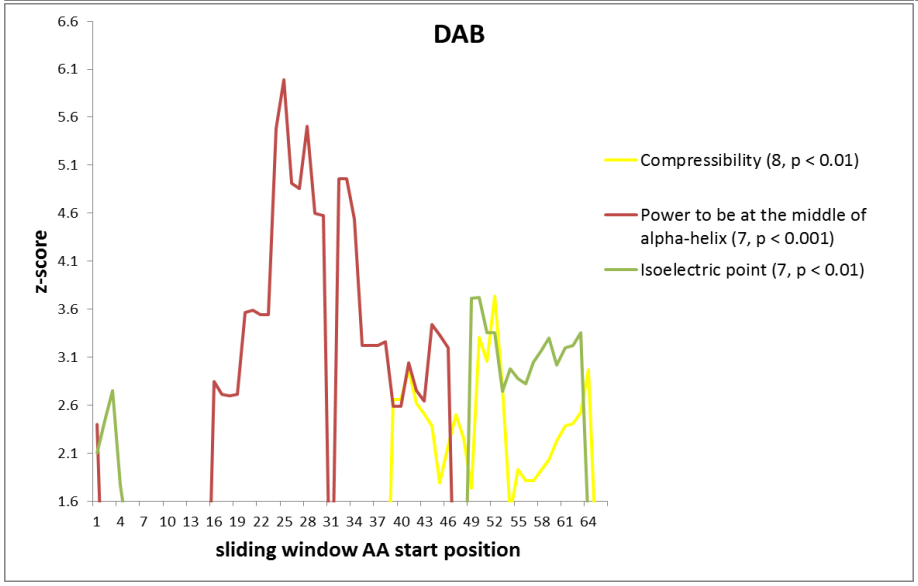


Figure 3.4 TREESAAP for MH class I ZE and class II DAA and DAB: *category of change and significance across the sequence are shown in brackets; only z-scores above 1.6, equivalent to p < 0.05 are shown;*

Table 3.2 Comparison of variation observed for the different MH regions studied

	class I			class II	
chain	**α2**	**α2**	**α1**	**α1**	**β1**
Locus /lineage	**UBA, ULA (L3I/II)**	**UCA (L3III)**	**ZE**	**DAA**	**DAB**
No. alleles (AA identity)	17	12	10	15	10
Mean nucleotide distance (PBR)	0.45	0.03	0.09*	0.15	0.21
Mean nucleotide distance (not PBR)	0.05	0.08	0.03*	0.03	0.05
Recombination (GARD)	1	1	0	0	2
No. sites under balancing selection (MEME+FUBAR)	9	3	2	16	19
No. sites under purifying selection (FUBAR)	0	0	1	0	2

**for the ZE locus the PBR was defined as suggested by Kruiswijk et al. (2002)*

Discussion

This study provides further evidence for the generality of patterns of diversity within the MH across the salmonid genus. This includes a single classical locus for the class II and high allelic divergence within the classical class I (Grimholt et al. 2003). Further, phylogenetic comparison of grayling U-lineage sequences with sequences of other salmonid taxa suggests the presence of the non-classical UCA, UFA and UHA loci in grayling. Strong evidence for balancing selection was found for the classical class II loci, and specifically within the PBR. Additionally, evidence for balancing selection was also given for the non-classical ZE lineage and an adaptive nature of polymorphism is suggested.

Some ambiguity remains, regarding a single classical locus for the class I, because no classical L3III lineage alleles have been identified and the alleles derived from the unclassical ULA locus and the UBA classical L3I lineage could not be distinguished, because of the high similarity. The classical class I locus is characterized by high allelic divergence reflected by mean nucleotide distance twice as high as for the class II loci in this study (Table 3.2) (Shum et al. 2001). Two classical UBA α2 lineages were identified in grayling (L3I, L3II), whilst only non-classical L3III lineage alleles were identified based on phylogenetic analysis (Figure 3.2). Evidence for balancing selection acting on non-classical UCA and ZE loci has been given, supporting both functionality and adaptive polymorphism. We did find some non-classical alleles that could not be assigned to particular loci, showing some unique residues at peptide anchoring sites, not observed in other salmonids, which merits further research in the future.

Both classical class II chains have been shown to be diverse, with more allelic variation present at the DAA chain and higher allelic divergence at the DAB chain. Evidence for recombination has been provided for both classes highlighting the importance of this evolutionary force in shaping variation at the MH genes (Zangenberg et al. 1995; Parham and Ohta 1996; Reusch and Langefors 2005; von Salomé and Kukkonen 2008). The highest rate of recombination was found here at the MH class II DAB locus. Comparatively higher

rates of recombination for the class II than for the class I has also been documented by Consuegra et al. (2005a).

The presence of a classical L3III lineage in grayling, which has not been identified in this study, seems probable as these ancient UBA lineages are generally shared across species and maintained over long evolutionary time scales (Shum et al. 2001; Miller et al. 2006; McClelland et al. 2013). However, it is also possible that alleles defined as the UCA lineage are actually classical alleles of the L3III lineage, as the assignment was purely based on phylogenetic relationships here. The pattern observed for peptide anchoring residues for these alleles, a Leucin residue at position 40 in all alleles is not typical for the UCA lineage in other salmonids, but rather associated with the L3II or L3VII lineage (Table 3.1). Two of these alleles had an additional K/N substitution, which has to our knowledge not been observed in other salmonids, but has been observed in spotted gar (*Lepisosteus oculatus*), a sister group of teleosts, within the U lineage (Grimholt et al. 2015, additional file 11). This substitution was also observed in another allele, in combination with a T/S substitution at position 36, which is associated with the UFA and UGA lineages in other salmonids (Table 3.1). Further work would be needed to elucidate the origin of these alleles. The presence of up to three alleles for the UCA lineage is consistent with this locus being duplicated in other salmonids (Miller et al. 2006; Lukacs et al. 2010). However, not all individuals exhibited alleles of this lineage, indicating that more allelic variation is present, which remained uncovered in this study. In Atlantic salmon the UCA loci have been described as pseudogenes (Lukacs et al. 2010). There was no indication of this in the current study.

An unusual deviation from the conserved pattern of peptide anchoring residues was observed in one allele (ID 16465_1415), which showed a high overall frequency (0.77), showing a W/C substitution at position 40 (Table 3.1). This substitution has been described as a unique characteristic of the human non-classical HLA-G, important for its function in inhibiting natural killer (NK) cells (Gonen-Gross et al. 2003). Up-regulation of HLA-G in fetus cells during mammalian pregnancy protects them from the attack of NK cells in the absence of maternal MHC I motives (Pazmany et al. 1996). Interestingly, this substitution has been also observed in *Latimeria chalumnae (*Coelacanthiformes) (Stet et al. 1996) a sister group of all ray-finned fish (Actinopterygii), which is viviparous (Wourms et al. 1991). Although this

allele is clearly not an adaptation to viviparity in grayling, a functional role in NK cell mediation seems possible. The allele clustered within the L3I lineages, but its classification as such would suggest a third locus containing alleles of this lineage as indicated by allele counts per individual (Table 3.1). The investigation of the source and role of this allele merits further research.

The ZE α1 region showed a high degree of conservation within the residues associated with the PBR in humans, also observed for other teleost species (Grimholt 2016). Nevertheless, there was a relatively high degree of polymorphism and ten distinct alleles (AA identity) were observed, comparable to the class II DAB locus (Table 3.2). For two of the polymorphic sites evidence for balancing selection was given by FUBAR analysis, suggesting an adaptive nature of the maintenance of polymorphism at this locus, as proposed by Kruiswijk et al. (2002). The authors have also suggested alternative sites for the PBR at this locus and here evidence for selection was given in grayling for one of these sites (Figure 3.3b), but not for another two. TREESAP also gave evidence for selection on physiochemical properties of the ZE protein, which is indicative of its functionality. Evidence for functionality in the immune response has been given in zebrafish (*Danio rerio),* where an upregulation of the ZE-expression was observed after vaccine immunization with live attenuated *Edwardsiella tarda* (Yang et al. 2012). Further research is recommended to elucidate the functional properties of the ZE gene products.

Significantly higher rates of non-synonymous to synonymous substitutions and higher nucleotide divergence within the PBR were observed for both MH II genes, giving evidence for historical pathogen driven balancing selection. Higher rates of non-synonymous to synonymous substitutions were observed for the DAA gene in the non-PBR as well, but significance was more than 60 times higher for the PBR. For both genes some evidence for diversifying selection was given for residues conserved in the human HLA gene. However, the inference of functionality of the protein residues are currently derived from human and mouse models and have not been developed for teleost or salmonids (Gómez et al. 2010). Although the pattern seems to broadly follow that observed in human and mouse (Stet et al. 2002), some incongruence cannot be excluded, yet Gómez et al. (2010) found the same residue conserved in human (H82 in Figure 6) under diversifying selection in *Oncorhynchus kisutch*, *Oncorhynchus mykiss* and *Salmo salar* as observed for grayling here.

TREESAAP also gave evidence for selection for fuctional properties on the MH class II genes. These included properties that affect secondary protein structure, such as buriedness for parts of the DAA gene and alpha helical properties and compressibility for parts of the DAB gene. Evidence for selection on physiochemical properties, like increased hydrophobicity (DAA) or selection for a zero electrical net charge (DAB) was also given. It is likely that these alterations affect pathogen binding and represent adaptive changes.

This study is another example of high immune genetic variation maintained by diverse patterns of selection in natural populations. It further extends on the current knowledge of immune genetic diversity in salmonids and highlights potential areas for future research. Now, that historic patterns of selection on the grayling MH have been elucidated, further population genetic analysis will be used to investigate how selection is shaping current within population diversity in relation to demographic forces.

Chapter 4: Neutral variation does not predict immunogenetic variation in the European grayling, *Thymallus thymallus*

Abstract

In the light of current rates of environmental change induced by human ecosystem alteration and climate change, the assessment and preservation of evolutionary potential is of high relevance in conservation biology. Neutral genetic markers continue to be the method of choice in many population genetic surveys that aim to inform species conservation, based on the rationale that they reflect genome-wide genetic variation and can be used as a surrogate for adaptive genetic variation and evolutionary potential. The aim of this study was to compare the performance of neutral markers with functionally important immune genes of the Major Histocompatibility Complex (MHC), underlying evolutionary processes, to inform decisions for management of European grayling (*Thymallus thymallus*). As a species of great socioeconomic value, it is primarily managed through stocking wild populations with hatchery-reared fish. This study is the first to (i) characterize the MH class II α-chain (DAA) and β-chain (DAB) and (ii) to assess the effect of stocking on immune genetic variation in grayling. High throughput Illumina sequencing was done for twelve populations: three where grayling have been introduced (Clyde, Eden and Itchen), four described as 'native stocked' (Aire, Derbyshire Derwent, Dove and the Hampshire Avon) and five native populations (Dee, Severn, Ure, Wye and Wylye). Both MH II genes showed evidence for balancing selection based on a Ewen-Watterson test. No significant correlations between MH and neutral marker levels of heterozygosity, allelic richness or inbreeding were found. Significantly lower levels of heterozygosity for both 'introduced' and 'native stocked' populations in comparison to 'native' populations were only evident from MH markers. Patterns of population structure were similar between neutral and MH II markers.. Significantly lower differentiation at the MH II than for microsatellites was only apparent when considering purely native populations. This study adds to the increasing demand of including ecologically meaningful genetic markers for the information and assessment of management decisions.

Introduction

Current rates of species extinction resulting from large-scale human ecosystem alterations are comparable to those of historic events of mass extinctions (Ceballos et al. 2015). In order to develop sound conservation strategies to maintain biodiversity against this background, it is generally acknowledged that taking account of genetic factors is crucial (Pertoldi et al. 2007; Sgrò et al. 2011). In the short term, the loss of genetic variation directly impacts on population viability due to negative effects associated with inbreeding and inbreeding depression (Spielman et al. 2004; Charlesworth and Willis 2009). In the long term, populations can only persist in a changing environment if they harbour heritable functional genetic variation, which is the prerequisite for adaptive evolution to occur (e.g. Duplouy et al. 2013).

Based on the assumption that neutral genetic markers, which do not encode functional proteins, represent genome-wide levels of variation (Kirk and Freeland 2011), they continue to be the method of choice in many conservation genetic surveys. Although this method has a wide range of useful applications for conservation biology (e.g. measuring isolation and gene-flow), adaptive genetic variation (defined as functional genetic variation that translates to differences in individual fitness) is not measured using neutral marker analysis (Hedrick 2001; Kirk and Freeland 2011). Since adaptive genetic variation may be affected differently than neutral variation by demographic processes like bottlenecks (because of the additional effects of selection which remain through the bottleneck (Ejsmond and Radwan 2011; Oliver and Piertney 2012; Sutton et al. 2011, 2015)), conservation decisions based only on neutral markers may be insufficiently informed (Bonin et al. 2007). Consequently, there is an increasing demand to include functional genetic markers, which represent ecologically meaningful variation, in conservation genetic studies (Hoffmann and Willi 2008; Piertney and Webster 2010; Kirk and Freeland 2011). This inclusion of evolutionary dynamics into diversity assessment is thought to greatly improve ecological monitoring and conservation planning (Brodersen and Seehausen 2014).

Genetic markers associated with immunity are top candidates to monitor adaptive genetic variation for conservation purposes because they have known associations with fitness and

evolutionary dynamics (Sommer 2005; Eyto et al. 2007; Dionne et al. 2009; Ujvari and Belov 2011). As diseases are thought to be a major threat to biodiversity (Daszak et al. 2000) and species persistence (Smith et al. 2006, 2009), genetic variation associated with immune competence is an excellent model for conservation monitoring (Sommer 2005; Eyto et al. 2007; Dionne et al. 2009; Ujvari and Belov 2011).

The assessment of adaptive genetic variation is particularly warranted in the case of European grayling (*Thymallus thymallus*, Salmonidae). Management for conservation purposes includes supplementary breeding and translocation of brood-stocks (Allendorf et al. 2001; Ayllon et al. 2006). In response to declines in abundance recorded in many parts of its range (Persat 1996; Uiblein et al. 2001; Gum et al. 2003; Dawnay et al. 2011), supplementation with hatchery reared fish (stocking) has been widely applied (Koskinen et al. 2002; Dawnay et al. 2011; Weiss et al. 2013; Persat et al. 2016). However, the genetic relationships between source and recipient population were largely unknown (Dawnay et al. 2011; Persat et al. 2016), with the consequence of the complete displacement of indigenous stocks in some places, as revealed by neutral genetic data (Duftner et al. 2005; Meraner et al. 2014). In other cases, released fish have not contributed to the native population at all, which is attributed to their very low survival rates (Persat et al. 2016).

Genetic maladaptation to local conditions of introduced fish is one reason for the generally observed lower fitness compared to native stocks (Araki et al. 2007; Araki and Schmid 2010; Williamson et al. 2010; Harbicht et al. 2014). This is of great concern where stocked and native individuals do interbreed, as disruption of local adaptations and outbreeding depression can have long-term impacts on fitness of local stocks (Allendorf et al. 2001; Lorenzen et al. 2012; Satake and Araki 2012). Adverse effects on fitness are possible even where the source and recipient population are the same. The genetic load of supplemented wild populations can increase substantially if deleterious mutations are eliminated less efficiently through the relaxation of selection over a period of early life-history in captivity (Lynch and O'Hely 2001, Araki et al. 2009). Also, natural mate choice is thought to be an important component of selection on immune genetic variation (Consuegra and Leaniz 2008; Milinski 2006), which is not allowed for in current supplementary breeding programmes of grayling. Altogether, there are a number of mechanisms how stocking could directly interfere with natural selection and evolutionary processes. However, no genotyping of

functional genetic loci has been conducted for European grayling in order to assess the effects of the widely applied stocking practice on adaptive genetic variation.

The first larger genetic survey of grayling in the UK was done by Dawnay et al. (2011) using neutral genetic markers. They found high levels of differentiation between populations, which has also been as reported from continental Europe (Gross et al. 2001; Koskinen et al. 2002; Gum et al. 2005). Four demographic clusters were identified (Dawnay et al. 2011). Stocking policies were altered in response to these data. This is based on the rationale that the maintenance of genetic integrity and distinctiveness in conservation efforts, is thought to result in the highest diversity and opportunities for evolution in an uncertain future (Forest et al. 2007).

Here, immune variation in European grayling at the Major Histocompatibility (MH) class II genes is examined, representing highly diverse, non-neutral markers. In teleost fish, unlike other vertebrates, the class I and class II genes are not within one complex and hence they are only designated as MH (Stet 2003). In particular, variation at the MH class II α-chain (DAA) and β-chain (DAB), covering most of the class II peptide binding region (PBR) is assessed. Variation at the MH genes is compared to neutral genetic marker variation, which has been assessed for the same individuals by Dawnay et al. (2011). Comparison between neutral and adaptive marker variation allows the distinction between the relative role of selection versus demography and genetic drift.

The hypothesis that microsatellite diversity does not reflect MH diversity is tested, as selection acts as a major force on MH but not microsatellite loci. Current risk assessment of population viability and the definition of management units based on neutral marker diversity is therefore predicted not to be consistent with results for immune genetic markers involved in evolutionary processes. Stocking history is also predicted to affect genetic variation at MH loci more than at neutral sites, because of the inference with natural evolutionary processes.

Materials and Methods

Samples

Between 37 and 40 individuals were selected from each of twelve previously sampled populations (Dawnay et al. 2011) covering different management classes. These consisted of five native populations (Dee, Severn, Ure, Wye and Wylye), four 'native stocked' (Aire, Derbyshire Derwent, Dove and the Hampshire Avon) and three introduced populations (Clyde, Eden and Itchen) (these designations follow Dawnay et al. 2011). Illumina sequencing and genotyping was done as described in chapter two.

Sequence and diversity analysis

First, summary statistics of genetic diversity were calculated for all populations. Conformity with Hardy-Weinberg equilibrium was investigated using Fisher's exact tests implemented within Genepop (Rousset 2008) using 10000 dememorizations, 100 batches and 10000 iterations per batch. Observed and unbiased expected heterozygosity were calculated in GenAlex 6.5 (Peakall and Smouse 2012). Weir and Cockerham's (1984) F-statistics and allelic richness were calculated using Fstat (Goudet 2001). Tests of significant differences of F_{IS} estimates and of F_{ST} estimates were based on 24000 randomisations and 66000 permutations respectively. Allele frequency differences among populations were also investigated using Fisher's exact tests implemented in Genepop (dememorization = 1000, batches =100, iterations per batch = 1000) (Rousset 2008). For all loci amino acid diversity was calculated for the peptide binding region (PBR) as p-distance within and across populations in MEGA 7.021 (Kumar et al. 2016).

To assess the potential effect of stocking on adaptive genetic variation, differences in observed and expected heterozygosity, inbreeding coefficient F_{IS} and allelic richness between native, native stocked and introduced populations were tested using Kruskal Wallis tests and pairwise comparisons were performed using Mann-Whitney-Wilcoxon tests in R (R Core Team 2015).

Microsatellite marker data from Dawnay et al. (2011) was tested for deviations from neutral expectations, performing an F_{ST} outlier test in BayeScan 2.1 (Foll and Gaggiotti 2008).with a

prior odd ratio of 10 and 20 pilot runs of 5000 iterations and an additional 50000 iterations burn in. MH locus data were compared to neutral microsatellite based diversity for the same populations to evaluate how well the latter reflects ecologically meaningful genetic variation. Additionally pairwise F_{ST} estimates were compared using Mann-Whitney-Wilcoxon tests between MH II and microsatellite loci, using (i) all populations, (ii) using only native and native stocked or (iii) only purely native populations. Introduced populations were not separately considered in this comparison because results are expected to be less biologically meaningful, due to founder effects resulting from the translocation event. To assess whether population structure reflected by neutral markers is supported by adaptive genetic differentiation, a neighbour-joining phylogenetic tree was built based on Nei's genetic distance (Nei 1972) in PHYLIP as a consensus of 2000 bootstrapped replications (Felsenstein 1993) for all genes studied.

Finally, the more recent effects of selection on each population and on each gene were evaluated. A Ewen-Watterson homozygosity test in its original definition (Ewens 1972; Watterson 1978) and a slightly modified version by Slatkin (1994) were conducted in ARLEQUIN 3.5 (Excoffier and Lischer 2010). The Ewen-Watterson test compares allele frequencies observed within each population to those expected under neutrality for populations at mutation-drift equilibrium. The test assumes population equilibrium and is sensitive to demographic changes. During population bottlenecks low frequency alleles are lost at a higher rate, producing allele frequencies that are more even than expected under neutrality (Watterson 1986). Similarly population expansion leads to an increase in low frequency alleles and lower heterozygosity than expected under neutral-equilibrium (Meyer et al. 2006). In order to distinguish demographic and selective forces and their effect on allele frequency changes a Ewen Watterson test was also performed on the microsatellite data from Dawnay et al. (2011) for all populations studied. Because of the limited statistical power of the Ewen Watterson test in detecting weak or moderate selection, significance without correcting for multiple testing or marginally significant results ($p < 0.1$, $p > 0.9$) are generally also considered as evidence (Ewens 1972; Larson et al. 2014).

Results

Genetic diversity

Total number of alleles per population ranged from 2-7 for the DAA locus and 2-6 for the DAB locus. Two populations Aire (AIR) and Severn (SEV) showed significant heterozygote deficits (Table 4.1) and departure from HWE for the DAB gene. F_{IS} estimates ranged from -0.22 to 0.33 for the DAA locus and from -0.2 to 0.6 for the DAB locus. Average AA diversity within the PBR was 0.11 for DAA and 0.41 for DAB across all populations. There was greater within population than between population AA diversity for all but the Eden population for the DAB locus (Table 4.1).

Expected heterozygosity and allelic richness for DAA and DAB differed significantly across introduced, native stocked and native populations (Pairwise Mann-Whitney-Wilcoxon test, Figure 4.1). 'Native stocked' populations showed significantly lower expected heterozygosity than native populations (Pairwise Mann-Whitney-Wilcoxon; p = 0.046). Between native and introduced populations, significant differences were observed for expected and observed heterozygosity and allelic richness (Pairwise Mann-Whitney-Wilcoxon; p = 0.02, p = 0.001, p = 0.003). No significant differences were observed between groups for F_{IS} values or effective population size which was inferred from microsatellites (Dawnay et al. 2011). For microsatellites only allelic richness followed the same pattern as observed for the MH (Figure 4.1).

Comparison of microsatellite and MH diversity

F_{ST} outlier tests gave no indication for selection on microsatellite loci. No significant correlations were observed between microsatellite and MH expected heterozygosity, observed heterozygosity, allelic richness or F_{IS} (Figure 4.2). Expected heterozygosity, allelic richness and inbreeding coefficients (F_{IS}) were significantly higher for MHII genes than for microsatellites (Pairwise Mann-Whitney-Wilcoxon; p = 0.007, p = 0.003, p = 0.01) (Figure 4.3). Populations for which a past bottleneck had been detected using neutral data (Dawnay et al. 2011) did not differ significantly from those populations where no evidence for past bottlenecks was found in any comparison.

Population-differentiation

A Mantel test showed significant correlations between MH and microsatellite pairwise F_{ST} estimates for all genes (Mantel test DAA: P = 0.001, r= 0.55; DAB: p = 0.02, r =0.38) (Figure 4.4). The Wylye population, which showed significant intra-population differentiation for microsatellites and was split in an A and B subpopulation in Dawnay et al. (2011) did not show significant differentiation for the MH genes and was treated as one population. Considering all or only native and native stocked populations in a pairwise comparison of MH II to microsatellite F_{ST} no significant differences were observed (Figure 4.6). However, considering only purely native populations resulted in significantly lower differentiation for both MH II genes than for microsatellites (Pairwise Mann-Whitney-Wilcoxon; DAA: p = 0.009, DAB: p = 0.02) (Figure 4.6).

Pairwise F_{ST} estimates significantly greater than zero were found for most population pairs for both genes (Table 4.2). Un-rooted neighbour-joining phylogenetic trees suggest a similar pattern of population sub-groups for MH II genes as for neutral markers (Figure 4.7). The Dee population groups with cluster C rather than A and the Derbyshire Derwent with A rather than D, where they were grouped for neutral markers (Figure 4.7).

Table 4.1 Sample locations with population classification and summary of genetic diversity for microsatellite markers (from Dawnay et al.(2011)), MH class II α (DAA) and β chain (DAB)

		Microsatellites (from Dawnay et al.(2011))							DAA						DAB					
pop	**class**	**N**	**Ne**	**Na**	**He**	**Ho**	**Fis**	**Bottle necked**	**N**	**Na**	**He**	**Ho**	**Fis**	**Mean AA diversity PBR**	**N**	**Na**	**He**	**Ho**	**Fis**	**Mean AA diversity PBR**
AIR	N S	39	63.9	3	0.45	0.44	0.02		31	4.0	0.59	0.42	0.30	0.23	29	4.0	0.54	0.28	**0.5***	0.42
CLD	I	64	68.6	2.3	0.39	0.37	0.05	*	40	2.0	0.31	0.38	-0.22	0.44	37	2.0	0.29	0.35	-0.20	0.5
DEE	N	52	43.2	3.5	0.54	0.51	0.04	*	27	6.7	0.80	0.74	0.07	0.3	26	5.9	0.77	0.54	0.30	0.42
DBD	N S	39	36.9	2.7	0.42	0.45	0.07		35	3.4	0.54	0.57	-0.06	0.32	36	2.0	0.50	0.58	-0.18	0.5
DOV	N S	50	64.4	2.6	0.35	0.32	0.04		26	2.8	0.30	0.23	0.23	0.33	35	2.9	0.30	0.26	0.14	0.51
EDN	I	45	48.7	2.5	0.4	0.38	0.04	*	33	3.0	0.36	0.24	0.33	0.29	36	3.5	0.34	0.25	0.26	0.4
HAV	N S	58	32.5	2.5	0.42	0.45	0.07	*	37	4.9	0.73	0.70	0.04	0.34	33	4.9	0.75	0.73	0.03	0.46
ITH	I	50	86.6	2.5	0.39	0.38	0.02		34	2.0	0.42	0.29	0.31	0.25	20	2.0	0.41	0.45	-0.10	0.65
SEV	N	39	40.8	2.8	0.42	0.41	0.03		31	3.0	0.53	0.39	0.27	0.29	30	3.9	0.57	0.23	**0.6***	0.42
URE	N	58	62.5	2.9	0.35	0.32	**0.09**		31	6.0	0.80	0.58	0.28	0.27	30	5.0	0.59	0.40	0.32	0.47
WYE	N	55	121	3	0.4	0.4	0		30	3.7	0.64	0.63	0.02	0.28	22	3.0	0.65	0.55	0.16	0.47
WLA/B	N	48/51	33.5	2.2	0.34	0.32	0.05	*	34	4.0	0.73	0.68	0.07	0.29	25	4.0	0.75	0.56	0.26	0.43

Population classifications are given for native (N), native stocked (NS) and introduced (I) populations; number of genotyped samples (N) and where available allelic richness (Na), expected heterozygosity (He), observed heterozygosity (Ho), inbreeding coefficient (F_{IS}), with values significant in bold and significant deviation from Hardy-Weinberg equilibrium after Hochberg-Bonferroni correction indicated (*), are given; for microsatellite markers estimated effective population size (Ne) and detection of a bottleneck (*) is given; population abbreviations are followed as in Dawnay et al. (2011); NA indicates that a measurement is not available;

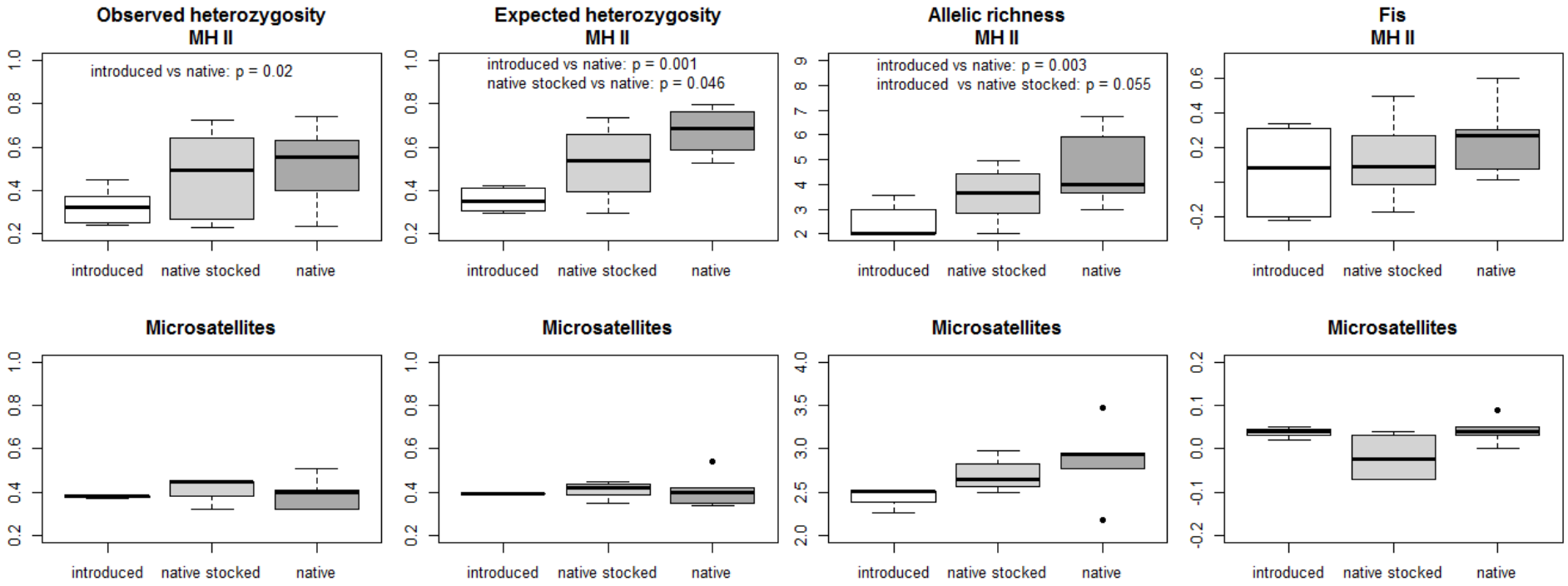


Figure 4.1 ***Comparison of genetic diversity at the MH II and microsatellites within native, native stocked and introduced populations:*** *p-values for pairwise comparisons are derived from Mann-Whitney-Wilcoxon tests and only reported for significant or marginally significant differences;*

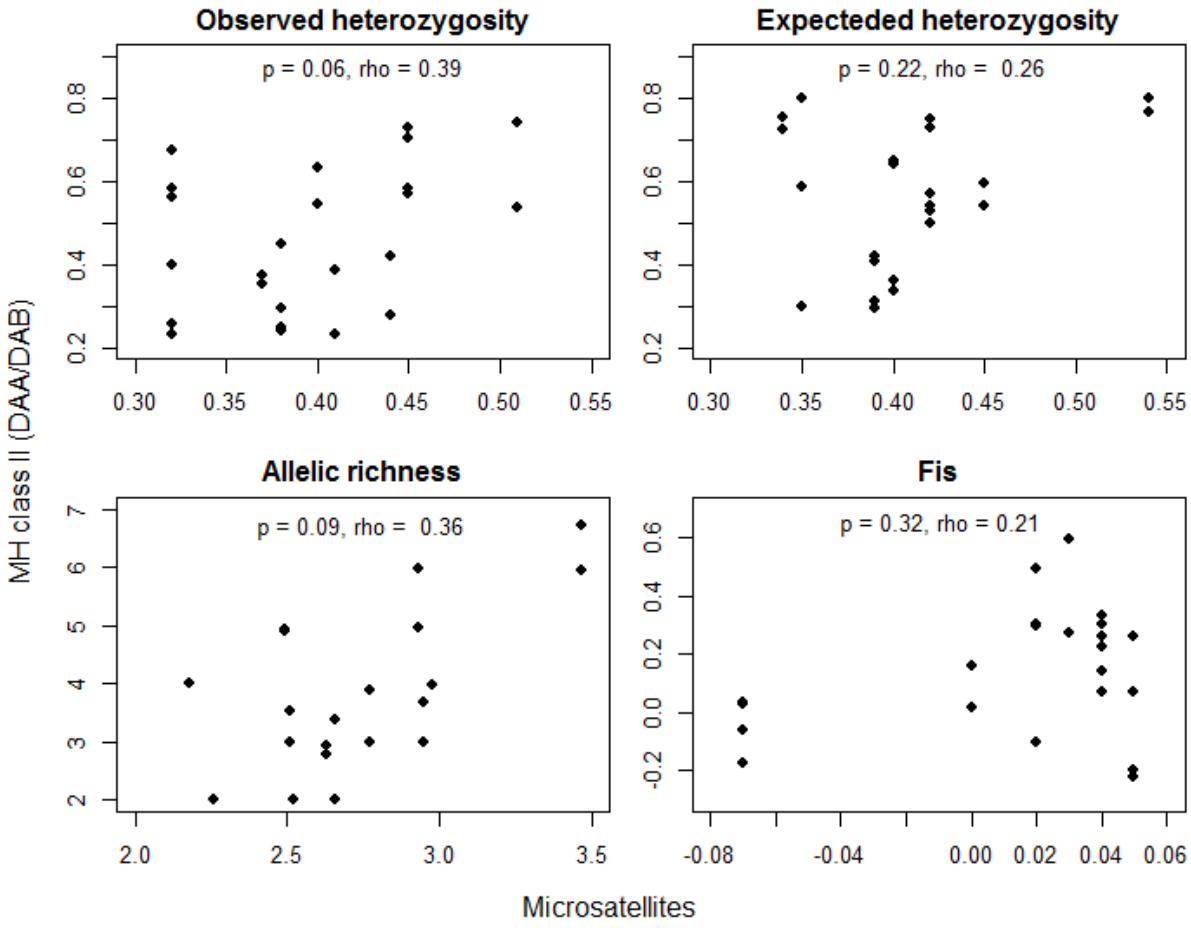


Figure 4.2 ***Relationship between microsatellite and MH II diversity measurements:*** *Spearman correlation coefficients and p-values are given for comparisons of microsatellite and MH II diversity measurements*

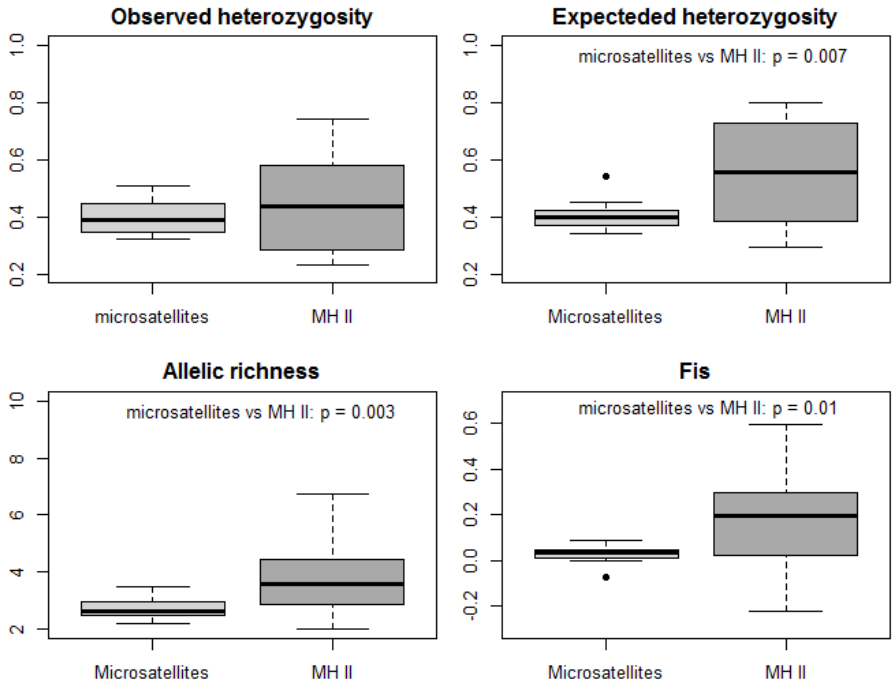


Figure 4.3 ***Relationship between microsatellite and MH II diversity measurements:*** *P values of Mann-Whitney-Wilcoxon tests with significant differences between markers are given*

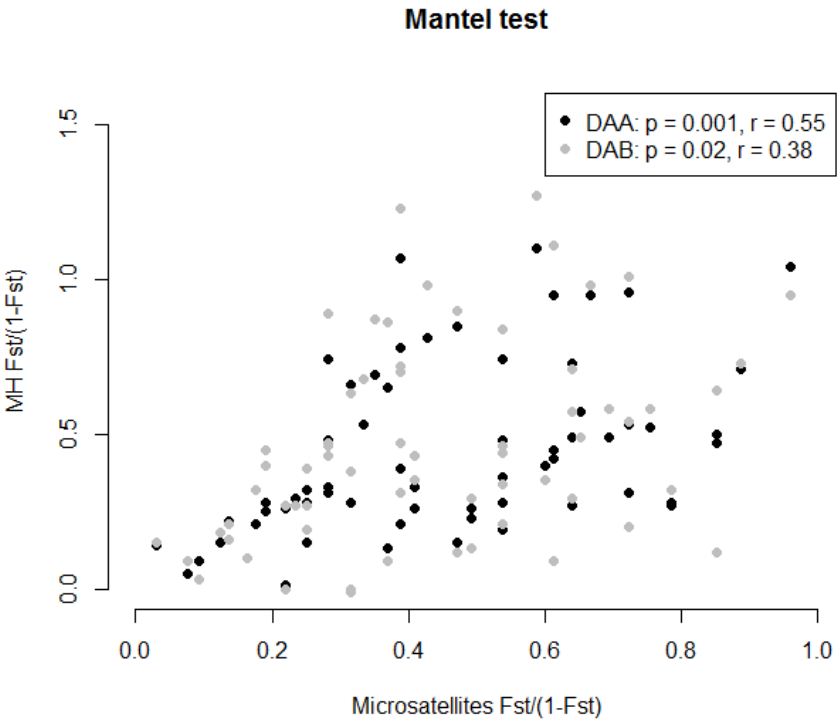


Figure 4.4 Relationship of pairwise F_{ST} /(1- F_{ST}) of microsatellites and MH markers studied: *P values are given with Spearman correlation coefficients in significant cases;*

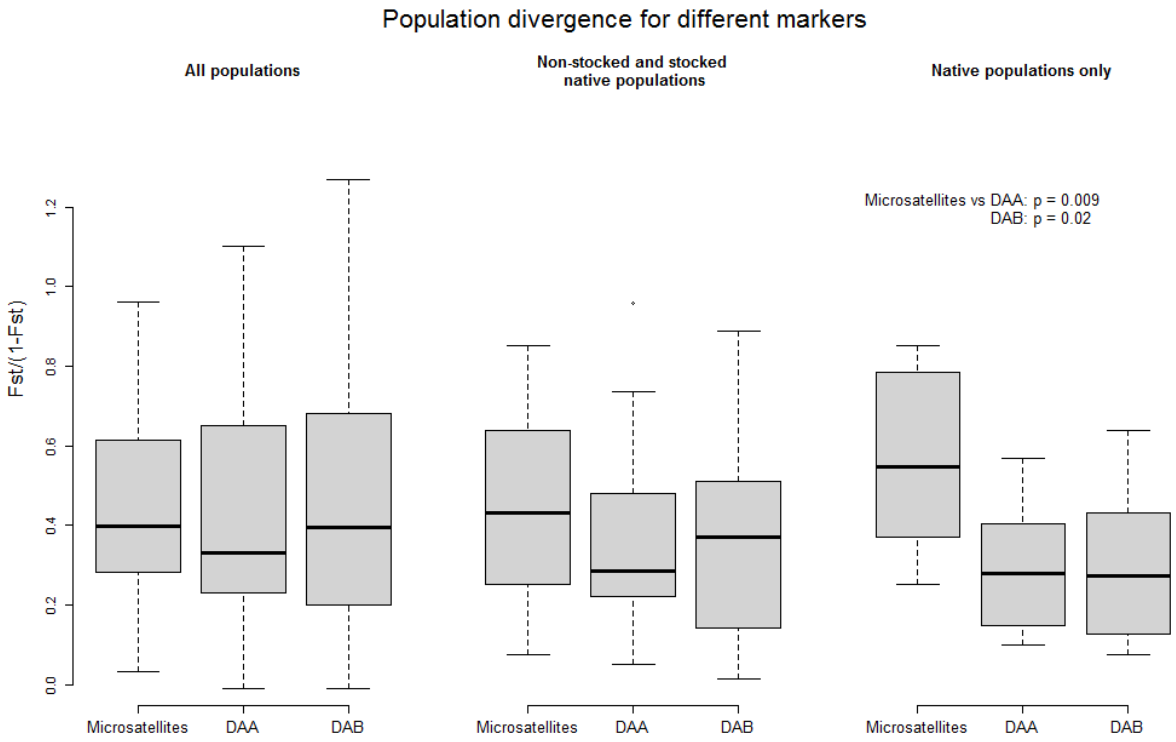


Figure 4.5 Population divergence measured as pairwise F_{ST} /(1- F_{ST}) for microsatellites and MH II markers: *for all (left), only native and native stocked populations (middle) and purely native populations (right): P values of Mann-Whitney-Wilcoxon tests with significant differences between markers are given*

Table 4.2 Pairwise linear F_{st} (F_{st}/(1-F_{st})) between all populations for each gene individually

DAA	**AIR**	**CLD**	**DEE**	**DBD**	**DOV**	**EDN**	**HAV**	**ITH**	**SEV**	**URE**	**WYE**
CLD	**0.81**										
DEE	**0.28**	**0.29**									
DBD	**0.53**	**0.14**	**0.16**								
DOV	**0.74**	-0.01	**0.26**	0.15							
EDN	**0.69**	0.01	**0.22**	**0.15**	-0.01						
HAV	**0.33**	**0.31**	**0.05**	**0.21**	**0.28**	**0.25**					
ITH	**0.65**	**1.10**	**0.26**	**0.74**	**1.07**	**0.95**	**0.33**				
SEV	**0.73**	**0.95**	**0.32**	**0.28**	**0.96**	**0.85**	**0.48**	**1.04**			
URE	**0.28**	**0.50**	**0.15**	**0.31**	**0.45**	**0.42**	**0.19**	**0.53**	**0.47**		
WYE	**0.49**	**0.52**	**0.13**	**0.27**	**0.49**	**0.36**	**0.26**	**0.71**	**0.23**	**0.28**	
WLA/B	**0.39**	**0.78**	**0.10**	**0.48**	**0.72**	**0.66**	**0.09**	**0.21**	**0.57**	**0.27**	**0.40**
DAB	**AIR**	**CLD**	**DEE**	**DBD**	**DOV**	**EDN**	**HAV**	**ITH**	**SEV**	**URE**	**WYE**
CLD	**0.99**										
DEE	**0.45**	**0.26**									
DBD	**0.68**	0.10	**0.14**								
DOV	**0.66**	0.00	**0.17**	0.07							
EDN	**1.03**	-0.01	**0.27**	**0.15**	0.01						
HAV	**0.50**	**0.37**	**0.09**	**0.25**	**0.26**	**0.39**					
ITH	**0.69**	**0.87**	**0.27**	**0.53**	**0.59**	**0.91**	**0.29**				
SEV	**0.89**	**0.98**	**0.26**	**0.37**	**0.74**	**1.04**	**0.50**	**0.71**			
URE	**0.43**	**0.12**	**0.13**	**0.15**	**0.06**	**0.12**	**0.19**	**0.39**	**0.64**		
WYE	**0.72**	**0.76**	**0.13**	**0.37**	**0.55**	**0.76**	**0.37**	**0.54**	0.09	**0.43**	
WLA/B	**0.52**	**0.63**	0.07	**0.39**	**0.43**	**0.67**	0.01	**0.28**	**0.52**	**0.28**	**0.38**

Microsatellites

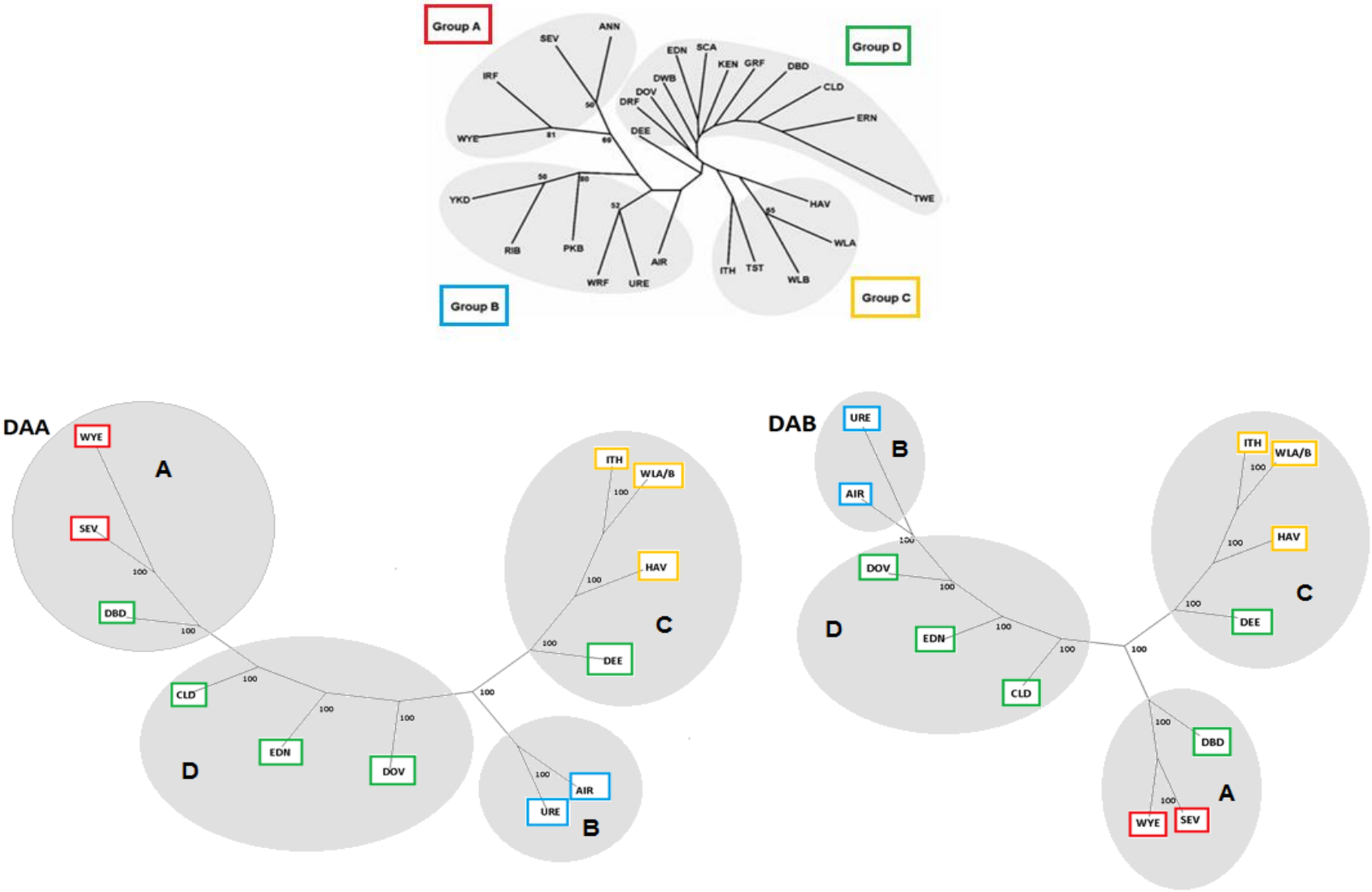


Figure 4.6: ***Unrooted phylogenetic trees based on Nei's genetic distance for microsatellites (from Dawnay et al. 2011), DAA and DAB of the MH class II:*** *Bootstrap support based on 2000 replicates is given; colour codes reflect the assignation to clusters based on neutral markers in Dawnay et al. (2011), with red for cluster A, blue for cluster B, yellow for cluster C and green for cluster D*

Selection

For the Aire (AIR), Clyde (CLD) and Dove (DOV) population no evidence for selection was given for any MH gene and microsatellite (Appendix 2) results gave no indication for recent population decline . For the Dee (DEE), Eden (EDN), the Hampshire Avon (HAV) , Severn (SEV) and the Wylye (WLA/B) populations allele frequencies were significantly more even than expected for both MH genes and microsatellites, but in each case the difference between observed and expected allele frequencies was greater for microsatellites, indicating a dominant effect of a recent population decline (Appendix 2). Populations that did not show larger significant differences between observed and expected allele frequencies for microsatellites than for MH genes, were the Ure population for the DAA (p = 0.01) and the DAB locus for the Wye population (p = 0.05) (Figure 4.8).

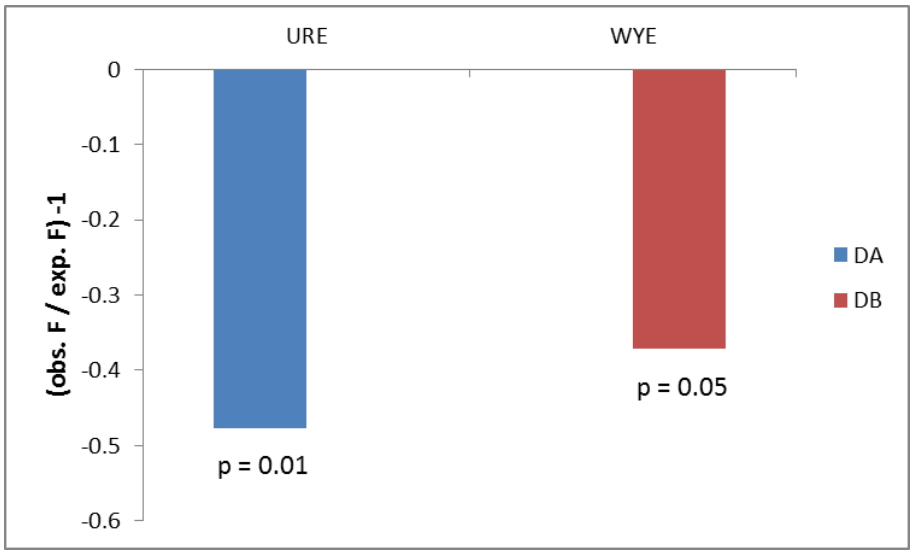


Figure 4.8 Results of Ewen Watterson tests of neutrality: *Ratios of observed to expected F values are standardized to zero by subtracting 1; only significant (p < 0.05, p>0.95) or marginally significant (p < 0.1, p>0.9) results are shown for MH markers and for populations where standardised F values of significant microsatellite results (in grey) were not smaller; negative values indicate allele frequencies more even than expected under neutrality, with balancing selection or population decrease as possible explanations; positive values indicate a higher amount of low frequency alleles than expected under neutrality, with directional selection or population increase as possible explanations;*

Discussion

The results clearly highlight the need for adaptive genetic markers in addition to neutral markers to inform conservation decisions (Piertney and Webster 2008; Sutton et al. 2011). No measurement of genetic diversity correlated between MH II and neutral markers, confirming the hypothesis that diversity at the MH II is strongly affected by selection and is not accurately reflected by neutral marker diversity. This is important as neutral marker surveys are still the method of choice for many conservation genetic studies building upon the implicit assumption that they can be used as a proxy for genome-wide variation (Kohn et al. 2006; Kirk and Freeland 2011). If populations were to be prioritized for management based on levels of heterozygosity, quantified as the population average, the top three identified for MH II and microsatellites would have only a single matching population. As loss of genetic diversity can occur more quickly under the simultaneous effects of drift and balancing selection during early stages of a population bottleneck on one hand, and diversity be re-established quicker after a population bottleneck, on the other hand, the MH is a more sensitive marker to detect vulnerable populations (Ejsmond and Radwan 2011; Sutton et al. 2011; Oliver and Piertney 2012).

Most strikingly, significant differences in heterozygosity between management classes observed for the MH II genes were not present within the microsatellite data at all. The only significant difference Dawnay et al. (2011) found across the twenty-seven populations they studied was higher allelic richness for native than for introduced populations. This can be explained by the effects of founder individuals not being genetically diverse or in low numbers, so that generally lower numbers of alleles are present in introduced than in native populations (Mayr 1954) or that individuals were stocked from an already bottlenecked population (Dawnay et al. 2011). For two of the introduced populations studied here, a genetic bottleneck was identified based on microsatellites. Nevertheless, average observed and expected heterozygosity did not differ for neutral data between any of the classes, whilst there was a clear difference for MH II markers. Clear signatures of balancing selection acting on the MH II are shown, both historically (see previous chapter) as well as more recently. This was indicated from significantly higher levels of expected heterozygosity and

allelic richness in comparison to neutral markers, higher within than across population AA diversity and significantly more even allele frequencies than expected under neutrality in populations where this was not observed for microsatellites (Figure 4.8). There was also no indication of an effect of effective population size, which did not differ significantly across the management classes. Hence, significantly lower expected heterozygosity in stocked than in native populations for the MH, but not for microsatellites is an indication for potential negative effects of the stocking process itself on MH II variability. As mortality of juveniles is exceedingly high within the first few months in the wild (>90% for salmon) this life stage would be most likely to experience disease-mediated selection (de Eyto et al. 2011). Through relaxation of selection in a hatchery environment, evolutionary processes mediated through heterozygote advantage or frequency dependent selection could be disrupted, diluting differences in survival based on immune competence during critical early life stages. The possibility of selection to act through natural mate choice is removed through artificial fertilization within a hatchery programme as well. It has been shown, that MH mediated mate choice, increasing parasite resistance, can be important in maintaining MH diversity in salmonids (Consuegra and Leaniz 2008). Additionally the artificially selected individuals used for stocking will have a disproportional reproductive success. As there is a bias in individuals used to produce hatchery stocks towards those individuals that are caught more easily, there might be also an unintentional selection for reduced predator avoidance.

Further evidence for the disruption of selective processes and relative increase of drift through stocking is given by significantly lower levels of differentiation between MH II and microsatellites being only apparent for purely native populations (Figure 4.6). Pairwise F_{ST} values were not significantly different between MH II and microsatellites across all populations and patterns of population structure were mainly consistent between both markers, though two populations did not cluster within the same groups for the MH II as they did for microsatellites. This suggests similar selection pressures on the MH II across the study area.

It is important to consider the intrinsic differences between the types of markers compared here. Microsatellites generally show higher mutation rates than MH markers, which can result in higher average heterozygosity and stronger patterns of differentiation (Hedrick 1999). However, here heterozygosity estimates were significantly higher for MH markers,

indicating the importance of balancing selection in generating high levels of diversity and pairwise F_{ST} estimates were not significantly different between MH and microsatellites markers across all populations.

Inbreeding coefficients F_{IS} were significantly higher for the MH II than for neutral markers (Figure 4.3) and two populations, the Aire and the Severn, significantly departed from HWE due to heterozygote deficiency at the DAB locus (Table 4.1). Whilst a technical cause, like an allelic drop-out at the DAB locus cannot be completely ruled out, elevated (but not significant) F_{IS} vales have also been found for these populations at the DAA locus, which support a biological cause for the observed pattern, as biased kin-sampling or a Wahlund effect (Wahlund 1928) would also be reflected by microsatellites (considering relatively lower or equal levels of differentiation between MH and microsatellites). Higher inbreeding coefficients (F_{IS}) for genes under balancing selection than for neutral markers seem counter intuitive. However, a higher loss of diversity under the simultaneous effects of selection and drift has been shown both empirically and by simulation models (Ejsmond and Radwan 2011; Sutton et al. 2011). Additionally, MH mediated mate choice is not necessarily dissortative, seeking highest offspring dissimilarity, but assortative, where particular haplotypes confer highest resistance, e.g. under frequency dependent selection (Eizaguirre et al. 2009). Highly unbalanced reproductive success for particular MH genotypes resulting in elevated inbreeding might be expected in particular when spawning grounds are scarce (Castric et al. 2002), which has been documented for the Severn population (Lewis 2006).

Conclusions

All MH genes studied showed evidence for recent balancing selection, indicated by significant results of a Ewen Watterson test independent from microsatellites. No measurement of diversity correlated between MH II and microsatellites, highlighting the need of adaptive genetic markers for comprehensive evaluation of genetic diversity (Piertney and Webster 2008; Kirk and Freeland 2011). Additionally, significant differences in genetic diversity were found between management classes, not reflected by neutral markers, with introduced populations showing generally lowest diversity. Significantly lower expected heterozygosity in stocked than in native populations raises further doubt on the

efficiency of this management strategy in supporting long term viable populations with high adaptive potential.

Chapter 5: The role of climate and other environmental factors in defining habitat suitability of European grayling

Abstract

Species distribution modelling tools are commonly applied to assess sensitivity of species to environmental parameters and estimate their distributional range. Whilst climate parameters are thought to be sufficient to predict range limits, it is recognised that the inclusion of other relevant parameters, such as land cover, can increase the accuracy of suitability predictions, particularly at smaller scales. The identification of non-climate related parameters that impact on habitat suitability is particularly important for identifying priorities for habitat improvement. Identification of these parameters may also be used to identify mitigation strategies that may reduce environmental stress under climate change conditions. Here, the sensitivity of European grayling, a salmonid freshwater species of conservation concern, to non-climate related and potentially relevant environmental variables is tested. The contribution of non-climatic parameters to the fine-tuned final model was 49%. Land cover, current velocity (flow), water nutrient, calcium and copper concentration as well as predator density (*Phalacrococax carbo*) were identified as important parameters defining habitat suitability for grayling. Omission errors were more than twice as high for the climate only model.

Introduction

Human ecosystem alterations such as habitat loss and degradation, invasive species, and overharvesting can cause environmental stress (Crain et al. 2008; Carilli et al. 2009; Brown et al. 2013). Environmental stress, defined as any environmental factor that causes a change in a biological system with potentially harmful effects (Hoffmann et al. 1991), can directly impair the adaptive potential of populations and increase vulnerability to extinction (Hoffmann and Hercus 2000; Charmantier and Garant 2005; Frankham 2005a). Together, such stressors impact biodiversity in a way comparable to historic mass extinction events (Barnosky et al. 2011; Ceballos et al. 2015). Climate change can exacerbate other threats to biodiversity, such as habitat degradation, through cumulative or synergistic effects (Brook et al. 2008). Extinctions may then happen due to other proximate causes long before physiological tolerances to high temperatures become predominant (Cahill et al. 2013). It is therefore vital to explore interactions between climate related parameters and other environmental factors, which affect habitat suitability to manage biodiversity effectively.

Species distribution modelling (SDM) techniques offer valuable tools to determining the key suite of environmental parameters, that condition a species' distribution, by combining observations of species occurrences with environmental estimates (Elith and Leathwick 2009). The implementation of SDMs has been successful in conservation and reserve planning (Esselman and Allan 2011; Rodríguez et al. 2007), invasive species management (Jiménez-Valverde et al. 2011), epidemiology research (Puschendorf et al. 2009) and predicting potential effects of climate change on biodiversity (Bellard et al. 2012). However, there are some intrinsic challenges of SDMs (Araújo and Guisan 2006) including bias in species records (Beck et al. 2014), selection of predictor variables (Synes and Osborne 2011), extrapolation into new environmental conditions not included in model calibration (Pearson et al. 2006) and unrealistic assumptions such as constancy of limiting factors and no evolutionary adaptation (Dormann 2007).

In response to these challenges, significant methodological improvements have been developed. The widely used Maxent program (Phillips et al. 2004) for presence-only species

distribution modelling has been continuously improved to enhance model performance (Phillips and Dudík 2008). Various methods were introduced to account for sampling bias, like the selection of background points from a target group of related species (Phillips et al. 2009) or reducing the spatial density of occurrences (Kramer-Schadt et al. 2013). Processes of species-specific tuning of model regularization and variable selection have been suggested to enhance model performance (Warren and Seifert 2010; Anderson and Gonzalez Jr. 2011; Warren et al. 2014) and methods to increase the independence of test data sets have been developed (Radosavljevic and Anderson 2014). Additionally, the acknowledgement of the limitations and uncertainties of models, which always represent simplifications, help users to better match their outcomes to specific applications (Araújo and Peterson 2012; Guillera-Arroita et al. 2015). Whilst techniques to improve critical evaluation and model performance continue to be developed (Renner and Warton 2013), studies that link model outputs of suitability to field measurements of fitness like crop yield or reproductive output are encouraging examples of the ecological relevance of SDMs (Brambilla and Ficetola 2012; Estes et al. 2013).

The classical approach of SDMs is to model 'bioclimatic envelopes' (Pearson and Dawson 2003). This is because climate is generally the dominant factor in determining species ranges (Araújo and Peterson 2012) and climatic variables are therefore considered as sufficient to describe main changes in distribution (Araújo and Guisan 2006; Bucklin et al. 2015). As a first approximation this approach can show climatic requirements and indicate where predicted tolerance limits are dramatically exceeded under conditions of climate change (Pearson and Dawson 2003). However, improvements in model performance by the addition of non-climate related parameters can be observed, particularly at high resolutions, where other local factors can become dominant (Pearson and Dawson 2003; Pearson et al. 2004; Heikkinen et al. 2007; Luoto et al. 2007; Stanton et al. 2012). This means shifting the focus from species distribution models towards Environmental Niche models, which includes mechanistic analysis (ENMs, Kearney 2006) and habitat suitability models (HSMs, Hirzel and Le Lay 2008) as resolution gets higher and differences in local conditions are aimed to be resolved (although the terms are often used interchangeably in the literature (Miller 2010; Araújo and Peterson 2012)). This approach gives the opportunity to pinpoint local climate refugia (Austin and Van Niel 2011) and design climate change mitigation strategies through

the identification of other non-climate related drivers of biodiversity change (Pereira et al. 2010). Stanton et al. (2012) show that the increase in predictive performance is highest when there are interactions between climatic and other variables. These authors highlight the importance of including other relevant variables, such as land cover, for studying their effect in combination with climate (Stanton et al. 2012). This is suggested to result in more realistic predictions of changes in distribution as a consequence of climate change.

European grayling (Salmonidae, *Thymallus thymallus*) are a protected species of conservation and economic importance that show high sensitivity to pollution (Buhl and Hamilton 1991, 1990; Vuorinen PJ *et* al. 1998; Uiblein et al. 2001) and tolerate water quality alterations only within a narrow range (Oberdorff et al. 2002). As such, grayling make an ideal indicator species of habitat quality and climate change response. This study aims to investigate the influence of climate and other habitat quality-related parameters on the distribution of this species. First, a climate-only model is developed to identify climatic parameters restricting distribution. Second, the sensitivity of grayling to additional parameters (e.g. velocity of the stream current (flow), land use, water chemistry and predator density) is tested in order to investigate their relative impact on habitat suitability for grayling. The results are discussed in the context of grayling conservation management.

Materials and Methods

The first part of the study was conducted across the whole range of the distribution of European grayling, developing a climate-only model with the aim to identify restrictive climatic parameters. In the second part, the focus is on a subset study area, for which environmental data, such as water chemistry parameters, was available from long term monitoring surveys or simulations (stream velocity/flow) at high spatial density. The aim of the subset model was to assess the relative importance of other parameters that potentially impact on habitat suitability for grayling, in addition to climate.

1. Climate distribution modelling of European grayling

All available occurrence records of European grayling (*Thymallus thymallus*) were obtained from the Global Biodiversity Information Facility (GBIF; www.gbif.org). These covered the whole latitudinal range of the species, but not all of its longitudinal range, which ranges until the Ural mountains in the East. Model training and evaluation was restricted to parts with available records, which was only the western part of the distribution of grayling (Figure 5.1). The model was then projected to the whole part of the distribution. A map of main European watercourses was obtained from the European Environmental Agency to define freshwater habitats. The study area was limited within 1km of these river lines (Figure 5.1). In order to account for sampling bias the records were spatially rarefied, by retaining only one occurrence record per 30 km^2, using the SDM toolbox (Brown 2014) for Arc GIS 10.3 (ESRI). The remaining records (292) (Figure 5.1), were used for species distribution modelling in Maxent (Phillips et al. 2004).

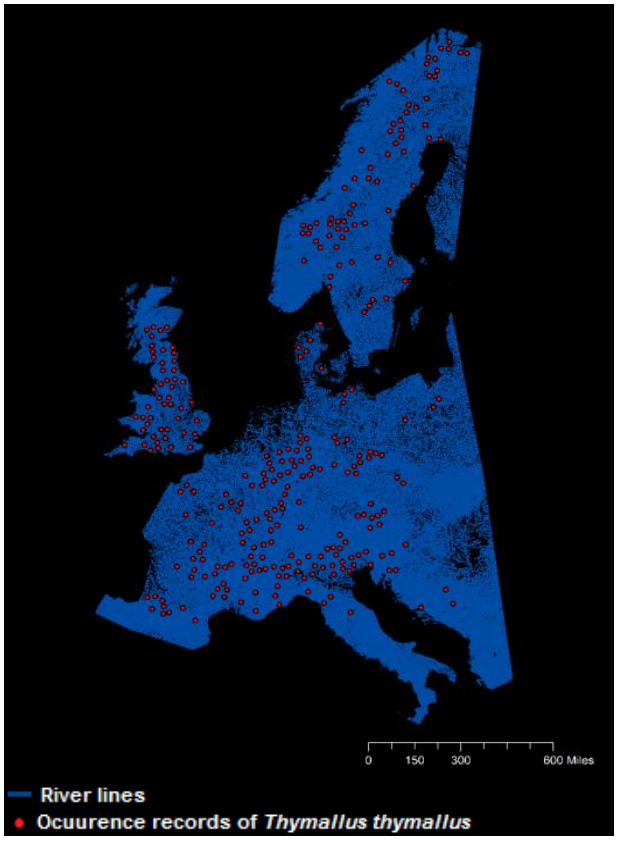


Figure 5.1 Unique occurrence records of European grayling across the entire range spatially rarefied at a scale of 30 km^2*: in red; in blue river lines of the total study area;*

Current climate data was obtained from the WorldClim database at a resolution of 30'' (available as bioclimatic parameters 1-19) (Hijmans et al., 2005). All layers were projected in ETRS 89 and tested for correlations (Appendix 3.1) across the study area, which was used in the variable selection process (see below). Variable selection, model calibration and evaluation was done as described below.

2. Habitat suitability model of European Grayling for the UK

The subset study area was limited to only a part of the total distribution of European grayling, which was the UK without Scotland (Figure 5.2). The reason for this was to take advantage of multiple extensive datasets available for this area (such as water chemistry) with the aim to identify other non-climate related parameters that affect the distribution of

grayling on smaller scales, as well as direct measurements of river surface water temperature (Orr et al. 2010).

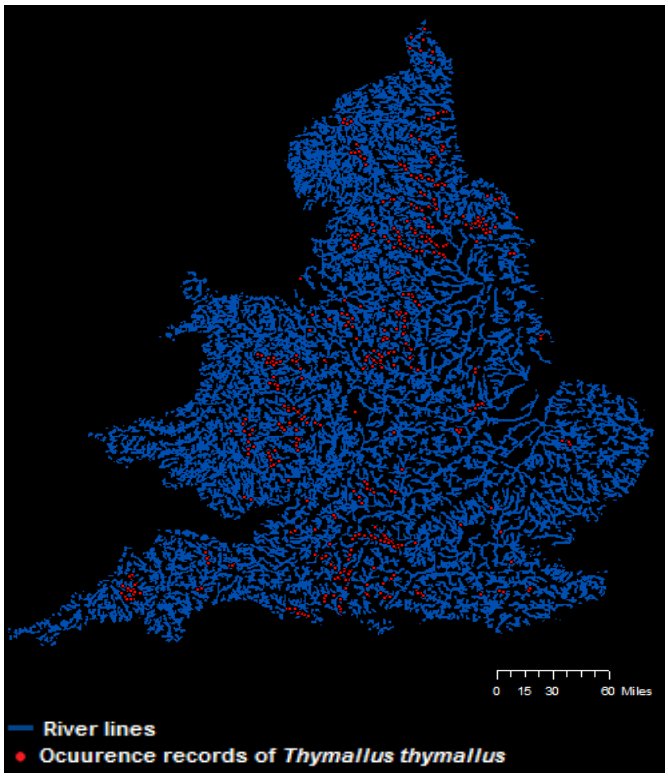


Figure 5.2 Unique occurrence records of *Thymallus thymallus* in the UK spatially rarefied at a scale of 1km^2

Unique occurrences points (693) that were recorded since 1990 for European Grayling (*Thymallus thymallus*) were obtained from the GBIF database (www.gbif.org) for the final study area (Figure 5.2). This time-frame was selected as there is a lack of recently updated records in GBIF (pers. obs.), so that using only records made after 2000 would significantly underestimate the current distribution. To account for uneven sampling effort and reduce spatial clustering, subsampling of occurrence records was done. Records were spatially rarefied at a scale of 1 km^2 for the UK, using the SDM toolbox (Brown 2014) for Arc GIS 10.2 (ESRI), leaving 441 records in total. The scale is the same as the resolution of environmental layers and was chosen for the generally high site-fidelity of grayling, aiming for a maximal resolution of small-scale differences in habitat preference (Nykänen et al. 2004). Only occurrence points which had records for all environmental parameters were used for model

tuning and evaluation. These were 118 occurrence points for the entire set of parameters and 272 for the final reduced model.

A subset of eleven bioclimatic parameters were selected (Hijmans et al. 2005), having high importance in predicting the global distribution of grayling (Table 5.1). Land cover data for 2000 was retrieved from the European Environmental Agency at a resolution of 100 m. Predation by the Great Cormorant, *Phalacrocorax carbo*, on inland fish populations has increased significantly over the past thirty years (Callaghan et al. 1998; Vetemaa et al. 2010; Winfield et al. 2003), having severe local impact in some cases (Winfield et al. 2003; Vetemaa et al. 2010). To investigate the impact of this emerging predator, GBIF records of Great Cormorant that were made after 1990 were used to make a Gaussian Kernel density layer, representing the individual counts per 10 km^2 using the SDM toolbox (Brown 2014) for Arc GIS 10.3 (ESRI). Measurement of biochemical oxygen demand (BOD), calcium, chloride, copper, nitrate, nitrite, orthophosphates, pH, suspended solids and total ammonium were obtained from long-term surveys conducted by the Environment Agency of the UK. These measurements fulfilled the requirement to have a good spatial coverage of measurements (minimum of 6000 sites), necessary to acquire data from a large number of occurrence and background points, after filtering using the following approach. For each site, all records since 2000 were averaged and sites were then filtered if either the mean or the standard deviation exceeded a modified z-score of 3.5, to remove outlier values with erroneous database entries or high fluctuation over the past 15 years (Iglewicz and Hoaglin 1993). Water temperature measurements were derived from the River Surface Water temperature database as described in Orr et al. (2010). For each region, all records since 1985 were extracted and filtered if measurements exceeded a modified z-score of 3.5. Of the remaining records, averages were calculated per site, separately for each season. The time-period was chosen to maximize the records per site in order to resolve local differences despite strong fluctuations. River flow data were obtained from simulations which were done for the Environment Agency of the UK (Prudhomme and Davies 2009a). Average annual flow rate and the 5^{th} (Q95) and 90^{th} (Q10) percentile flow of annual flow were used. The Q95 flow value, exceeded 95% of the time, was used to represent low flow, whilst the Q10 flow value, exceeded 10% of the time, was used to reflect high flow. For all

point measurements a buffer of 2km was created and raster layers were created at a resolution of 30'' (0.93 x 0.93 = 0.86 km^2 at the equator), assigning the value closest to the centre of the cell using ArcGIS 10.3 (ESRI). All layers were projected in ETRS 1989 and tested for correlation (Appendix 3.2), which was used in the variable selection process (see below). Table 5.2 lists variables tested for relative impact on habitat suitability for grayling.

We tested the importance of each of the 30 parameters in predicting grayling distribution, including climate, water chemistry, land cover and predator density data. Variable selection, fine-tuning of regularization and model evaluation was done as described below. To compare model performance when using only climate data, but fine-tuning on the same scale, a subset model was additionally applied using only the predefined bioclimatic variables (Table 5.1).

Table 5.1 Parameter contribution and permutation importance for the reduced fine-tuned Maxent model of the distribution of European grayling across its' global range

Bioclimatic parameter	Contribution	Permutation importance
Mean temperature of the coldest quarter (Bio 11)	11.16	53.66
Precipitation of the warmest quarter (Bio 18)	21.74	12.82
Mean temperature of the driest quarter (Bio 9)	12.04	12.30
Precipitation of the wettest quarter (Bio16)	12.90	7.91
Minimum temperature of the coldest month (Bio6)	4.81	6.96
Precipitation of the driest month (Bio 14)	0.37	2.23
Temperature seasonality (standard deviation *100) (Bio 4)	10.00	1.29
Maximum temperature of the warmest month (Bio 5)	2.91	1.15
Mean temperature of the wettest quarter (Bio8)	2.18	0.90
Isothermality (Bio 2 / Bio 7) (* 100) (Bio 3)	8.36	0.57
Precipitation seasonality (Coefficient of Variation) (Bio 15)	13.53	0.21

Variables are ranked by permutation importance

Table 5.2 Initial variable set used in the variable selection process of the habitat suitability model

Climate	Human Habitat Modification				Biotic interactions
Bio Climatic variables	**Water Temperature**	**Flow**	**Landuse**	**Water Chemistry**	**Predation**
Isothermality (Bio 2 / Bio 7) (* 100) (Bio 3)	Autumn Mean	Flow average	**Landcover 2000**	Biological Oxygen Demand (BOD)	***Phalacrocorax carbo* Density**
Temperature Seasonality (standard deviation *100) (Bio 4)	Spring Mean	**Flow Q10**		**Calcium**	
Max Temperature of Warmest Month (Bio 5)	**Summer Mean**	**Flow Q95**		Chloride	
Min Temperature of Coldest Month (Bio6)	Winter Mean			**Copper**	
Mean Temperature of Wettest Quarter (Bio8)				Nitrate	
Mean Temperature of Driest Quarter (Bio 9)				**Nitrite**	
Mean Temperature of Coldest Quarter (Bio 11)				**Orthophosphates**	
Precipitation of Driest Month (Bio 14)				Ph	
Precipitation Seasonality (Coefficient of Variation) (Bio 15)				Suspended Solids	
Precipitation of Wettest Quarter (Bio16)				Total Ammonium	
Precipitation of Warmest Quarter (Bio 18)					

parameters retained in the final model are in bold

Variable selection, model calibration and evaluation

As models with many variables and high complexity tend to be too closely fitted to the training data and perform poorly when transferred (Warren and Seifert 2011) a process to select only the most important variables and to optimize regularization was implemented (Warren et al. 2014). Warren and Seifert (2011) demonstrated that using Akaike's information criteria, corrected for sample size (AICc), for model selection has significant advantages over other methods like AUC, which has been criticized by others (Lobo et al. 2008; Jiménez-Valverde 2012), resulting in more accurate inference of variable importance and higher model transferability. Following the procedure described in Warren et al. (2014), 74 models were generated with random multiplier values ranging from 0.2 to 15 with 0.2 increment steps, in order to tune species-specific regularisation and identify the value with best fit to the data set (Merow et al. 2013). Only hinge features, which combine linear and step functions (Elith et al. 2011) were used as they have been found to best discriminate the importance of different predictors (Polce et al. 2013) and produce smooth models suitable for projection (Elith et al. 2010). The model with the lowest AICc was selected and all variables with >5% contribution were retained. All variables with 0% contributions and permutation importance were discarded. All other variables were added in the order of their contribution, so long as they did not show correlations above 0.7 to any other variable in the model (otherwise removed). For the selected reduced-variable set 74 models were again generated with random multiplier values ranging from 0.2 to 15 with 0.2 increment steps and hinge features and the model with the lowest AICc was selected. To test the transferability of the final model a 'masked geographically structured' evaluation approach (Radosavljevic and Anderson 2014) was implemented. In this approach, the study area is divided into k geographic blocks and evaluation is done on each block in turn. Background sampling is performed on all areas apart from the test block to enhance independency of the test data. This was done in the ENMeval R package (Muscarella et al. 2014), where this approach is available as 'block' method with k=4. To facilitate model interpretation and make comparisons between areas of high and low suitability, the continuous model outputs were classified using two different thresholds. As a conservative approach, the Minimum Training Presence threshold (MTP) was selected (omission error = zero) in order to identify areas of least suitability (Pearson et al. 2007). Thresholds based on model sensitivity and

specificity have been shown to give most accurate predictions about habitat suitability (Liu et al. 2005; Jiménez-Valverde and Lobo 2007; Bean et al. 2012). The Equal Sensitivity and Specificity (ESS) threshold, where specificity equals sensitivity, and the probability of false positives is the same as for false negatives, was selected to further classify between low/medium and high suitability habitat.

Results

1. Climate-only modelling

Eleven bioclimatic variables were retained in the reduced model (Table 5.1). These were, in the order of their relative permutation importance: mean temperature of the coldest quarter (Bio11), precipitation of the warmest quarter (Bio18), mean temperature of the driest quarter (Bio9), precipitation of the wettest quarter (Bio16), minimum temperature of the coldest month (Bio6), precipitation of the driest month (Bio14), temperature seasonality (Bio4), maximum temperature of the warmest month (Bio5), mean temperature of the wettest quarter (Bio8), isothermality (Bio3) and precipitation seasonality (Bio 15).

The regularization multiplier of the model with lowest AICc was 2.4. Evaluating the model using the geographically masked approach, where each quarter of the study area is used in turn for testing, resulted in a mean test AUC of 0.57 ± 0.12. The 10^{th} percentile test omission using this approach was 0.34 ± 0.08. Using a less stringent evaluation approach of bootstrapping over a random test percentage of 25 % of the occurrence points, not limiting background selection for training, resulted in an average test AUC of 0.68 ± 0.03 and a 10^{th} percentile test omission of 0.14. Following the ranking of mean variable contribution measurements, measures of precipitation have highest importance, including precipitation of the warmest quarter (Bio18), precipitation of the wettest quarter (Bio16) and precipitation seasonality (coefficient of variation) (Bio15) (Table 5.1). Highest permutation importance shows the most important parameter to be mean temperature of coldest quarter (Bio11) with over 50%, followed by precipitation of the warmest quarter (Bio18) and mean temperature of driest quarter (Bio 9) (Table 5.1). Model predictions of habitat suitability are shown in Figure 5.3.

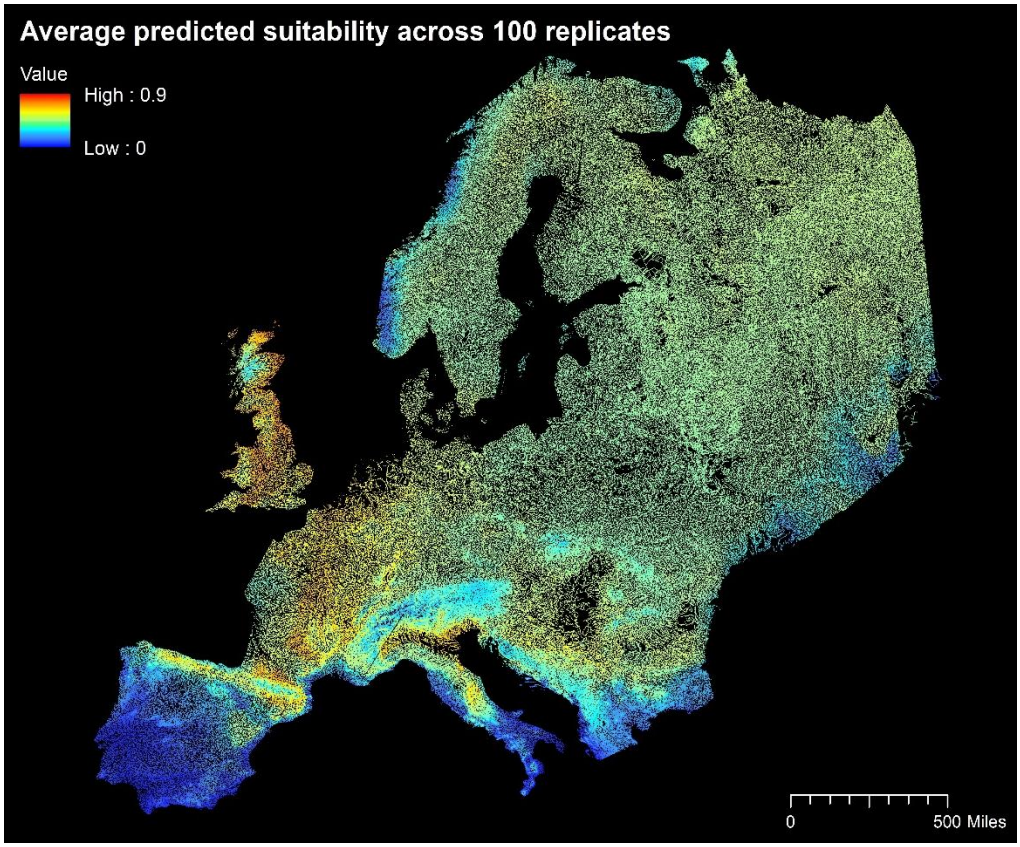


Figure 5.3 Average predictions of habitat suitability for European grayling across 100 replicates: *the mean standard deviation was 0.04 and the maximum 0.11*

2. Climate and Habitat distribution modelling

The random multiplier of the full model with lowest AICc was 4.2. Twelve parameters of the initial 30 were retained in the fine-tuned Full model, which had a random multiplier value of 2.2. The average test AUC across evaluation blocks, using the 'masked geographically structured' approach was 0.76 ± 0.002. Average minimum test omission rate was 0.004± 0.00005 and average 10^{th} percentile omission rate 0.12 ± 0.002. The predicted habitat suitability across the study area is shown in Figure 5.4.

Following the same approach for variable selection and model tuning using the bioclimatic subset only (Table 5.2), resulted in an AICc of the best fitting model substantially higher (AICc = 9548) than for the model including other environmental variables (AICc = 5211). The mean test AUC across evaluation blocks was 0.71 ± 0.05, the average minimum test omission rate was 0.02 ± 0.00007 and average the 10^{th} percentile omission rate 0.26 ± 0.04. Details on model predictions and variable contributions can be found in Appendix 3.3.

Minimum Training Presence (MTP) and Equal Sensitivity and Specificity thresholds showed 11% of the area was classified as unsuitable (below MTP), 61% as low-medium suitability (below ESS) and 28% as high suitability (above ESS; Figure 5.4). Climatic environmental parameters contributed 51% of the final set of variables used to build the model and had a permutation importance of 60% (Table 5.3). These were precipitation of the wettest quarter (Bio16), isothermality (Bio3), described as the ratio of the mean diurnal range (Bio2) to the mean annual range (Bio7), maximum temperature in the warmest month (Bio5), mean water temperature in summer and the 90^{th} (Q10) percentile and 5^{th} (Q95) of annual flow, representing highest and lowest flow rates respectively. Water chemistry parameters had a total contribution of 37% to the final model and a permutation importance of 26% (Table 5.3). These parameters were the concentrations of calcium, dissolved copper, nitrite and orthophosphates. All parameters, which are affected by anthropogenic activity had a permutation importance of 41%. These included copper, nitrite and orthophosphate concentration, land-use and flow rate, which is affected by stream regulation, water abstraction and dam constructions. Additionally the density of the predator *Phalacrocorax carbo* had a permutation importance of 1% (Table 5.3).

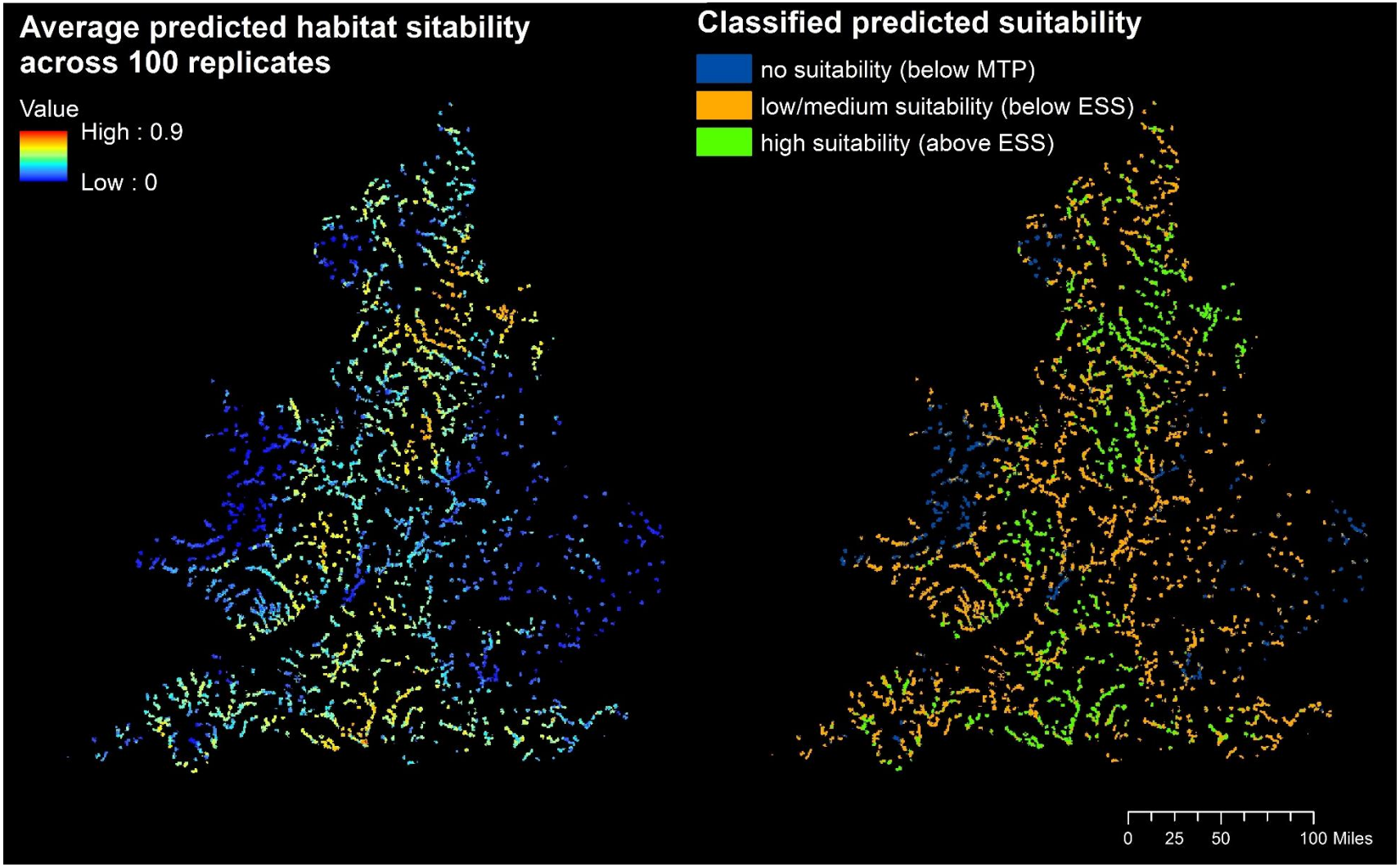


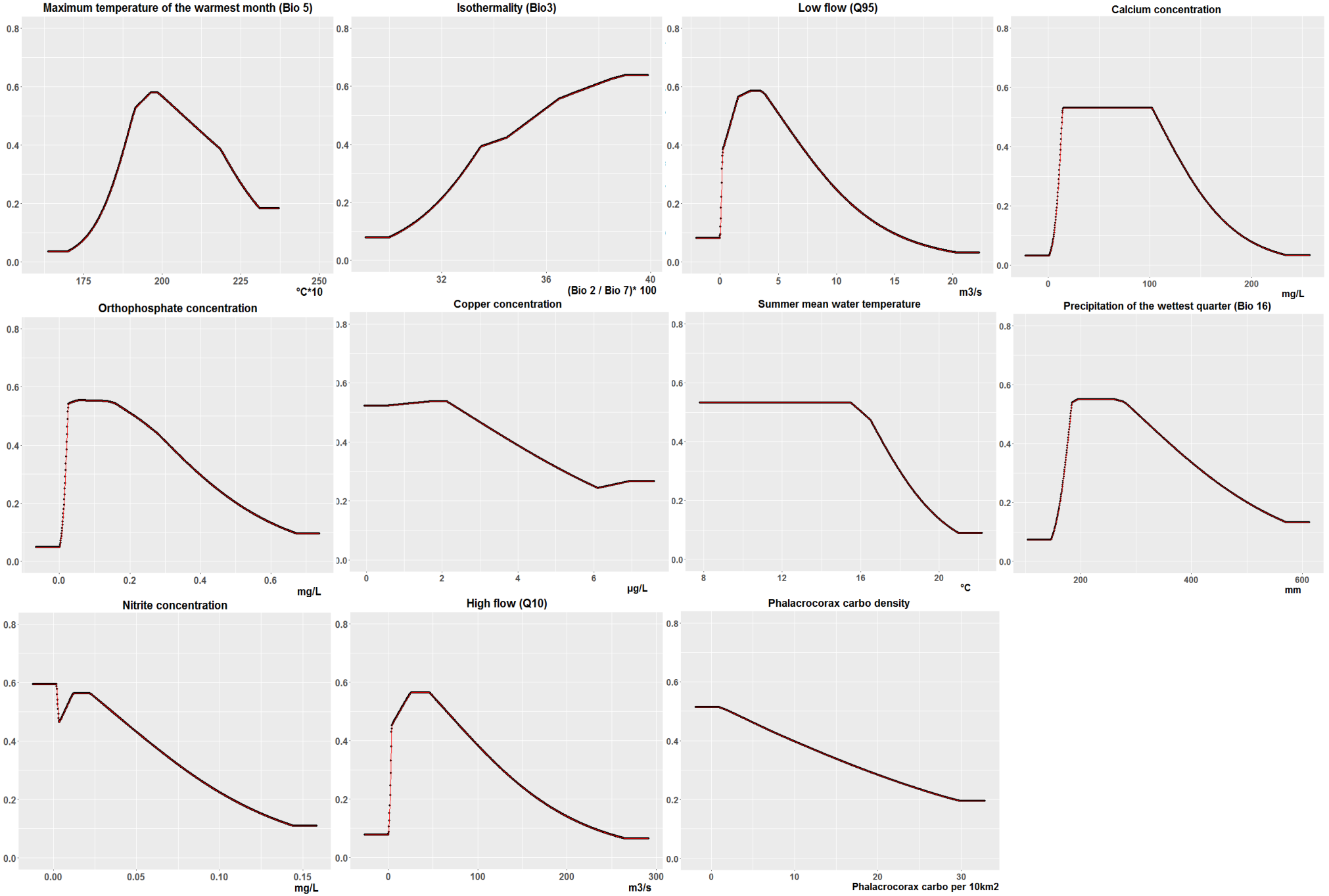
Figure 5.4 Maxent estimates of habitat suitability for *Thymallus thymallus*: *the mean standard deviation was 0.01 across the study area*

Table 5.3 Relative contribution and permutation importance for all variables included in the habitat suitability model

Parameter description	Contribution	Permutation importance
Maximum temperature of the warmest month (Bio 5)	8.02	24.31
Isothermality (Bio 2 / Bio 7) (* 100) (Bio 3)	10.54	15.27
Land cover	12.06	12.83
Low Flow (95th Quantile of Annual Flow)	10.35	10.45
Calcium Concentration	20.67	9.67
Orthophosphate concentration	12.15	7.95
Dissolved Copper Concentration	2.18	5.36
Summer Mean Water Temperature	6.55	4.89
Precipitation of the wettest quarter (Bio 16)	9.10	4.14
Nitrite concentration	2.26	2.88
High Flow (10th Quantile of annual Flow)	5.94	1.21
Phalacrocorax carbo density	0.17	1.05

Parameters are ranked by permutation importance

In the order of their permutation importance Figure 5.5 shows the response of grayling to the environmental parameters studied. A probability of presence above 50% was observed for maximum temperatures in the warmest month (Bio5) between 19 and 20.8 °C, isothermality values above 36, low flow(Q95) between 1 to just over 5 m^3/s, calcium concentrations between 18 and ~105 mg/L, orthophosphate concentration between 0.02 and 0.2 mg/L, copper concentrations between 0 and 2.6 µg/L, mean water temperature in summer between 8 and 16 °C, precipitation in the wettest quarter between 182 and 306 mm, nitrite concentrations between ~0.01 and ~0.04 mg/L, high flow (Q10) between 10 and 70 m^3/s and no *Phalacrocorax carbo* count within 10 km^2 (Figure 5.5). Highest suitability in regard to land use was observed for fruit trees and berry plantations, inland marshes and broad-leaf forest. Land use classifications with a probability of presence below 50% were observed for urban fabric and industrial or commercial areas, natural grasslands and moors and heathlands.


Maximum temperature of the warmest month (Bio 5)
°C*10
Isothermality (Bio3)
(Bio 2 / Bio 7)* 100
Low flow (Q95)
m3/s
Calcium concentration
mg/L
Orthophosphate concentration
mg/L
Copper concentration
µg/L
Summer mean water temperature
°C
Precipitation of the wettest quarter (Bio 16)
mm
Nitrite concentration
mg/L
High flow (Q10)
m3/s
Phalacrocorax carbo density
Phalacrocorax carbo per 10km2

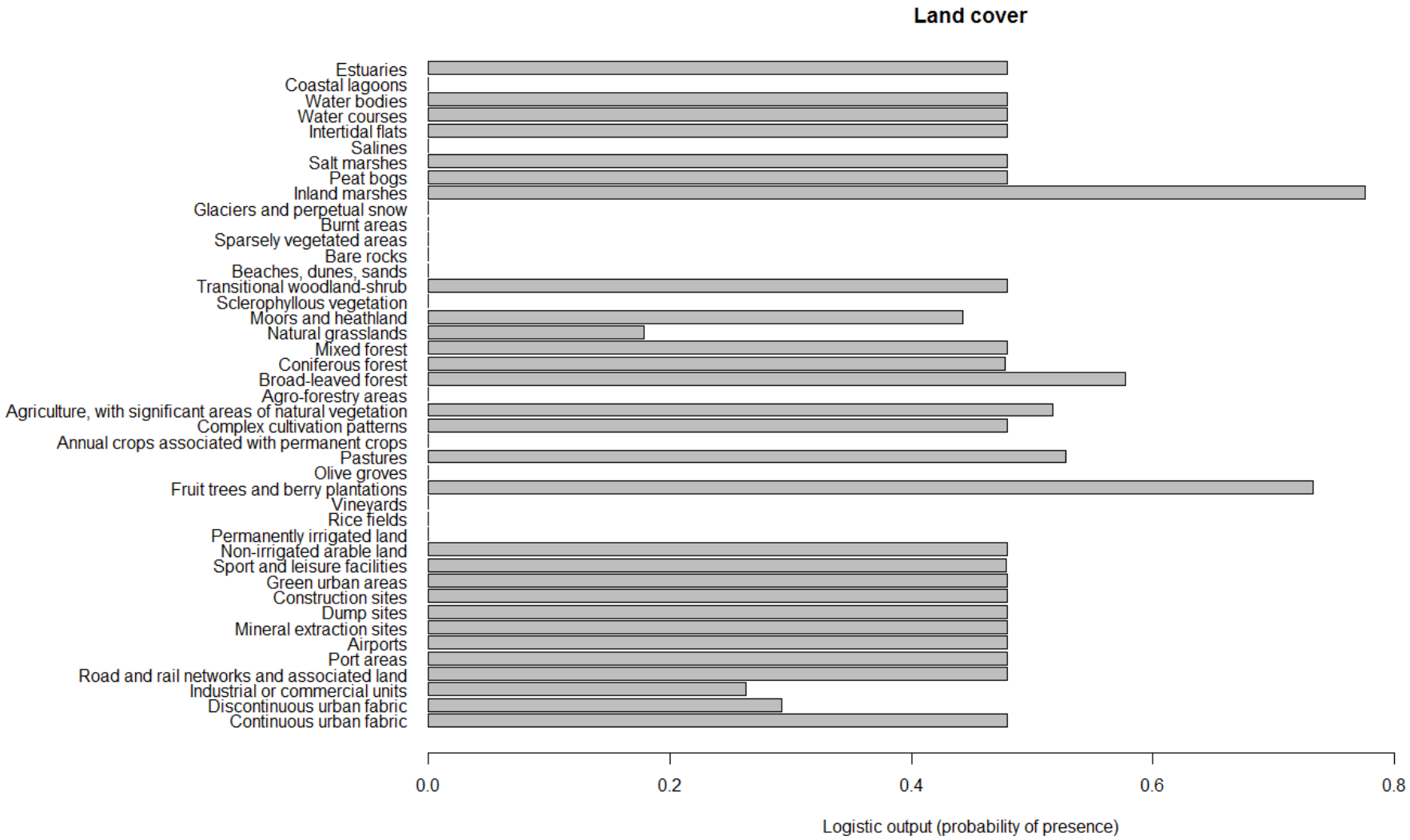


Figure 5.5 Maxent response curves for all parameters of the final model: *parameter responses were modelled excluding all other variables and represented as the mean across ten replicates; the order reflects permutation importance, except for land cover which is shown at last;*

Discussion

The climate only model achieved substantially lower discriminative power compared to the model including other environmental parameters and showed a 10^{th} percentile omission rate more than twice as high, evaluated for the subset study area. Non-climate related parameters had an overall permutation importance of 49% in the final model, highlighting their importance in addition to climate to define habitat quality for European grayling. A particular importance of flow parameters was indicated by the high contribution of precipitation parameters in the climate only model, showing the highest power to predict grayling distribution across its whole range. Whilst mean temperature of the coldest quarter also had high prediction importance, reflecting a sharp tolerance limit below minimal temperatures, highest temperatures show much less importance. Precipitation of the warmest quarter (Bio18) has a higher contribution in defining the range limits of European grayling, suggesting that droughts and low flow regimes are more important in limiting the distribution than highest temperature. Flow has also been identified as a critical driver of habitat suitability for salmonids by Wenger et al. (2011), highlighting the importance of including hydrological data on flow regimes, which can only to some extent be derived from precipitation and climate data alone. This can partly explain the low discriminative power of the climate only model. Furthermore, land use changes and pollution are leading to habitat loss and impairing the natural distribution of freshwater species (Foley et al. 2005; Weijters et al. 2009; Alabaster and Lloyd 2013). It is therefore conceivable that climate data alone is not sufficient to achieve high discriminative power to model habitat suitability (Pearson et al. 2004; Stanton et al. 2012). However, there are some technical limitations of the presented model. The distribution could not be estimated from the entire natural range for a lack of data along the longitudinal gradient in the East. Additionally, migration was not considered in order to identify and exclude not accessible areas.

Climate parameters had an overall contribution to current-day suitability of 51% and permutation importance of 60% using the habitat suitability modelling approach (Figure 5.6). Suitability regarding the highest temperature in the warmest month showed highest permutation importance and ranged from 19 °C to 21 °C, in accordance to what has been

described as the upper limit at which growth is still possible in grayling at 21 °C (Mallet et al. 1999). For the mean water temperature in summer, suitability started decreasing above 16 °C, which is lower than has been described as the optimal temperature for grayling (between 17°C and 18°C,Persat and Pattee 1981; Northcote 1995). Isothermality (Bio3), defined by the ratio of the diurnal to the annual temperature range followed next in permutation importance . Though diurnal temperature fluctuations are associated with stress in salmonids (Imholt et al. 2011), the effects of extreme temperatures within an increased annual temperature range would be worse under conditions of chronic exposure and little diurnal fluctuation (Wehrly et al. 2007). Additionally, the opportunities for behavioural adaptations might be enhanced under higher diurnal ranges if local climate refuges exist so that activity and movement patterns can be adjusted in order to avoid exposure to lethal temperatures (Fraser et al. 1993; Sutton et al. 2007; Malcolm et al. 2008). Precipitation of the wettest quarter, which is closely related to flow, was also retained in the habitat suitability model, but as lowest (Q95) and highest (Q10) flow were included, it is likely that the additional information gained form peak precipitation relates to the frequency and severity of flood events. Flood events can increase mortality within the 0+ age class and damage prey invertebrate populations (Elwood and Waters 1969; Jensen and Johnsen 1999). Additionally pollutants bound within sedimentary particles can be redistributed and become bioavailable during heavy flood events resulting in toxic effects on fish (Hollert et al. 2000; Hilscherova et al. 2007). Our results confirm a sensitivity of grayling to low flows (Q95) and probability of presence fell below 50% if flow was lower than 1 m^3/s in more than 5% of the year and sharply drops for values below 0.5 m^3/s. Low flow regimes have been shown to increase mortality in the 0+ age class of grayling potentially because marginal stream areas are lost, which represent their preferred habitat, so that exposure to predators is increased (Riley et al. 2009). Additionally increased temperature and lower concentrations of dissolved oxygen are associated with low flow regimes (Solomon and Sambrook 2004). High Flow (Q10) had a lower contribution and showed a decrease in suitability if flow exceeded 70 m^3/s in over 10% of the time.

Water chemistry parameters all over contributed to habitat suitability with 37%, as measured by permutation importance (Figure 5.6). The measurements of water chemistry are naturally fluctuating and not static as represented here. Averaging measurements across

years was done to incorporate the effects of more extreme years or seasons. Calcium concentration showed highest importance among the water chemistry parameters. Calcium concentration is directly associated with the natural buffer capacity of freshwater systems against acidification (Schneider 2011). It reduces both a drop in pH and toxicity of elevated aluminium concentrations which are the two main negative effects on salmonids associated with acidification (Kroglund et al. 2007; Malcolm et al. 2014). Calcium reduces the toxicity of metals through competing for binding sites at the gills of fish and reducing the total uptake (Hunn 1985; Playle 1998; Chowdhury et al. 2016). Metals are a significant pollutant in the UK due to its' strong, mostly historic mining activities (Lewin and Macklin 1987), which continue to affect river sediments (Macklin et al. 1997). A probability of presence for grayling at old or active mining sites over 50% as revealed from the land use layer shows that populations are exposed to metals to some extent (Figure 5.7). Next highest importance among the water chemistry parameters was attributed to orthophosphate concentration. This is in agreement with the conception that eutrophication is the most important anthropogenic pressure on freshwater systems in Europe (Birk et al. 2012). Studies showing that assemblages of fish species and their population parameters closely reflect eutrophication intensities confirm the impact on fish distributions (Pont et al. 2007; Blabolil et al. 2016). High phosphate concentrations are the most common cause for eutrophication in fresh water systems (Correll 1998). Fish are affected indirectly through an increased growth of phytoplankton, which reduce oxygen levels in the water through their respiratory activities. This effect is further enhanced by increased respiratory activities of bacteria decomposing accumulating biomass leading to hypoxia and anoxia (Diaz and Breitburg 2009). Particularly fish eggs and embryos are highly vulnerable to hypoxia (Elshout et al. 2013). 0.15 mg /L is defined as the maximum phosphate concentration to avoid effects of eutrophication (van Dijk et al. 1994) and identified as the threshold after which habitat suitability decreases for Grayling (Figure 5.7). In a warming climate eutrophication is expected to become accelerated through higher productivity, whilst solubility of oxygen in water decreases with temperature, magnifying harmful effects of hypoxia (Ficke et al. 2007). Copper was the only metal with sufficient data locations to be included in this analysis and had 2.2% contribution to the model. The negative effects of copper on growth, reproduction, behaviour and survival of fish are well documented (McKim and Benoit 1971; Buckley et al. 1982; Ayllon et al. 2006; McIntyre et al. 2012; Wang et al. 2013; Sovová et al. 2014). Similar

to other metals the mechanism of toxicity is mainly related to a disruption of ion regulation and total loss of ions in particular (Laurén and McDonald 1985; Alsop and Wood 2011; Chowdhury et al. 2016). MacRae et al. (1999) found that rainbow trout did already show slight mortality at only 1 µg/L and over 25% mortality at 10 ug/L free (uncomplexed) copper. Using absolute copper concentrations rather than bioavailability is leading to some imprecision, but was done for a lack of sufficient data to calculate bioavailability across all sites. A decrease in suitability with absolute copper concentration above 2 µg/L is in general agreement with what MacRae et al. (1999) found for the minimum concentration affecting rainbow trout assuming that some part will be found in complex and not bioavailable. That is however magnitudes lower than the 40 µg/L (0.04 mg/L) stated in the European Directive 2006/44/EC 'On the quality of fresh waters needing protection or improvement in order to support fish life' as the minimum acceptable concentration of total dissolved copper to support salmonids. This huge discrepancy highlights the need to focus on bioavailability rather than absolute concentrations to obtain precise estimates of environmental risk (Reiley 2007; Comber et al. 2008). It is probable that other metals like zinc and aluminium, that were not included, have a similar relevance as copper. Nitrite had the lowest contribution of the water chemistry parameters to overall habitat suitability (Figure 5.6). Elevated concentrations of organic nitrogen like ammonium result in high bacterial nitrification activities resulting in the accumulation of nitrite. The toxicity of nitrite to fish is well studied and salmonids show particularly high sensitivity (Palachek and Tomasso 1984; Tomasso 1986; Kroupova et al. 2005, 2008). Wastewaters are a major source of anthropogenic pollution of freshwater systems with organic nitrogen, but major improvements in wastewater treatment can achieve significant reductions (Jorgensen and Weatherley 2003; Aissa-Grouz et al. 2015). The estimated maximum concentration which did not affect habitat suitability was estimated as 0.02 mg/L in this study, which is twice of what is stated in the European Directive 2006/44/EC 'On the quality of fresh waters needing protection or improvement in order to support fish life' as maximum tolerable concentration of nitrite for salmonids.

Land cover was among the parameters with highest contribution to habitat suitability (Figure 5.6). Generally, higher suitability was inferred for extensive agriculture like fruit and berry plantations or agricultural land with substantial natural vegetation (Figure 5.7).

Broadleaf forest showed good suitability as well, potentially due to higher amounts of terrestrial invertebrates available as prey associated with broadleaf trees within the riparian zone, compared to grasslands for example (Tyler and Ormerod 1991; Kawaguchi and Nakano 2001; Ormerod et al. 2004). This could be one explanation why natural grasslands showed lower suitability. Additionally this type of vegetation has a dense canopy which creates a good cover, with the potential to create cooler microhabitats (Malcolm et al. 2008). Inland marshes showed considerably high suitability as well (Figure 5.7). Marshes provide a good protection against erosion, attenuate floods and support summer base flows (Giles et al. 2004).

The results suggest that the increased predation of *Phalacrocorax carbo* on inland freshwaters (Callaghan et al. 1998) does impact on habitat suitability of grayling. Locally severe impacts on populations have been documented in some cases (Vetemaa et al. 2010; Winfield et al. 2003). However, studies on the effects on population dynamics and potential over-exploitation through *Phalacrocorax carbo* could not support long term effects on sustainability of fish populations, which were either not affected in their population dynamics (Pienkowski et al. 1998) or compensated losses with increased growth rate and lower age at maturity (Britton *et* al. 2002). It is therefore rather in the combination with other habitat related factors that increase the risk for predation that result in a threat to population sustainability (Frenz et al. 1997; Uiblein et al. 2001).

Conclusions

This study highlights the importance of assessing not climate related parameters and their potential impact in addition to climate on habitat suitability. As the quality of current habitat depended to almost 50% on not climate related parameters, the identification of other than climate related parameters offers the opportunity to develop locally optimized mitigation strategies in order to reduce environmental stress. Habitat suitability models could be shown here to be a valuable tool to aid the development of conservation strategies in this respect.

Chapter 6: Climate change: Risk and mitigation strategies for European grayling

Abstract

Climate change is expected to become a major threat to biodiversity and potentially exacerbate other environmental stressors. This study evaluates the risk climate change imposes on European grayling (*Thymallus thymallus*), a salmonid freshwater species, and tests the suitability of different habitat improvement strategies in mitigating negative effects, using Maxent models. It can be shown that locally optimized habitat improvement can increase habitat suitability estimates by up to 31% and suggestions on local priorities on habitat improvement are made, including the reduction of eutrophication and metal pollution. This study aims to give specific management advice for conservation of grayling in the face of climate change.

Introduction

Climate change is expected to become a major threat to biodiversity over the next decades (Thomas et al. 2004; Bellard et al. 2012). The integration of climate change predictions into current conservation and biodiversity planning is therefore essential (Araújo et al. 2004; Heller and Zavaleta 2009). This requires knowledge on sensitivities of species to various climatic parameters and the potential impact of climate change on habitat suitability and species distribution (Hulme 2005). This allows conservation management to allocate limited resources in a way that maximises conservation returns (Bottrill et al. 2008). Vulnerability assessment is used to prioritize conservation effort, incorporating exposure and sensitivity to climate change as well as adaptive capacity (Williams et al. 2008; Dawson et al. 2011). Species distribution models (SDMs) are a powerful tool to assess specific sensitivity to such change and make predictions on the influence of climate on species distribution (Thomas et al. 2004). Whilst there are number of uncertainties, e.g. regarding the accuracy of predictions of future change and the projections of species distributions into new environmental conditions (Pearson et al. 2006; Dormann 2007), with a careful modelling design and acknowledgment of the limitations of the outputs (Araújo and Peterson 2012), SDMs are important in revealing the sensitivity and exposure of species to climate change (Pearson and Dawson 2003; Araújo et al. 2006).

Whilst estimates of species responses to climatic parameters are essential to assess the impact of climate change, the integration of other non-climate related drivers of biodiversity change (Pereira et al. 2010) and studying their effect in combination with climate is thought to result in more realistic predictions (Stanton et al. 2012). In chapter 5 it was shown that a model that included non-climate related parameters, such as water chemistry and land cover, to estimate habitat suitability for European grayling within the UK, had a substantially better fit than the climate only model. In this context this study investigates the risk of climate change to grayling and evaluates possible habitat improvement strategies to enhance adaptability and the probability of persistence of grayling populations within the UK. Relative to other salmonids grayling show high sensitivity to high temperature (Ibbotson et al. 2001; Jonsson and Jonsson 2009), which makes them a suitable target for the

development of climate change mitigation strategies, of which many other community members can benefit.

Materials and Methods

Data for future climate scenarios were downloaded from the WorldClim database (Hijmans et al. 2005) for the General Circulation Model HadGEM2-AO (UK Meteorological Office, United Kingdom) for each of the representative concentration pathways (r.c.p.) assuming minimal change (r.c.p. 2.6) and maximal change (r.c.p. 8.5) for 2050. Projections of future mean water temperatures in summer were done by calculating the percentage degree of change for mean air temperatures in the warmest quarter (Bio10) under each scenario and applying the same percentage degree of change to current mean water temperatures in summer. Eleven future flow scenarios were obtained from the Centre for Ecology and Hydrology (CEH) for 2050 (Prudhomme and Davies 2009b). Model inputs and parameters were set as described for the fine-tuned subset model in chapter 5. Future projections were made using 100 replicates. Standard deviations between flow scenarios were low (see Results) and therefore predictions were averaged across scenarios. A multivariate environmental similarity matrix (MESS) was created in Maxent (Phillips and Dudík 2008), as described in (Elith et al. 2010), and the variable that is subject to greatest change in future scenarios was identified. Exposure to change of this variable was assessed in the context of grayling sensitivity. To evaluate the effect of habitat improvement as a mitigation strategy under climate change conditions, future projections were additionally done under scenarios where either orthophosphate, nitrite or copper concentration was artificially shifted to the maximum values that had been shown not to impact on habitat suitability for grayling (chapter 5). These variables were selected, because they were shown to affect habitat suitability for grayling (chapter 5) and could be targeted through management within a habitat improvement context. Maximum concentrations of orthophosphates, nitrite and copper concentration that did not impact on habitat suitability were estimated as 0.15 mg/L, 0.02 mg/L and 2 µg/L respectively (chapter 5).

Results

Future projections

Under conditions of climate change, the projections for 2050 predict predominantly a significant loss of high suitability habitat (Figure 6.1; Table 6.1). This was estimated as a reduction in highly suitable areas of 20% for the r.c.p. 2.6 and the r.c.p. 8.5 scenario of change (Table 6.1). An increase in the area that becomes unsuitable is more pronounced in the r.c.p. 2.6 projection of change, where Isothermality (Bio2/Bio7) decreases more, particularly in the South West, due to a greater increase in the annual temperature range (Bio7) in relation to the diurnal range (Bio2). For the r.c.p. 8.5 projections, the diurnal range increases more proportionally, so that the decrease in isothermality is less pronounced (Figure 6.1). Large parts of the study area exhibit environmental conditions under climate change scenarios not included in training (Figure 6.2), which was mainly driven by the increase in maximum temperature of the warmest month (Bio5). The optimal temperature of the warmest month (Bio5) is exceeded in large parts within the study area with a maximum of 8.2°C (Figure 6.3 (left)). Figure 6.3 (right) shows the according predicted increase in mean water temperatures in summer in relation to grayling sensitivity (habitat suitability is estimated to be impacted above 16°C, chapter 5).

Table 6.1 Relative percentage change in habitat classification for 2050

	current	2050 r.c.p. 2.6	2050 r.c.p. 8.5
no suitability (below MTP)	13	35	21
low-medium suitability (below ESS)	59	57	73
high suitability (above ESS)	27	7	7

Relative area in percent under the following classification: no suitability (below MTP), low-medium suitability (below ESS) or high suitability (above ESS) for the current habitat suitability estimate and as projected for 2050 under the r.c.p. 2.6 and 8.5 respectively;

An overall recovery of suitable habitat of up to 10% and of highly suitable habitat of 4%, relative to the total study area, is predicted to be possible by making locally optimal improvements in either orthophosphate, nitrite or copper concentration (Table 6.2). Possible improvements would greatly affect areas in which grayling is currently present, as highlighted by the number of known current occurrences of grayling, which would be predicted to benefit (Figure 6.4). The highest net total gain in estimated suitability that can be achieved reveals locally optimal strategies for improvement, targeting either orthophosphate, nitrite or copper concentration (Figure 6.5 & 6.6 for r.c.p. 2.6 and 8.5 respectively). Highest local increases in habitat suitability of 31% were observed for reductions in orthophosphate concentrations (Figure 6.5 & 6.6). For other locations, a reduction in nitrite or metal pollution had higher effects on suitability achieving relative maximum improvements of 17% and 19% under the r.p.m. 2.6 scenario of change and 14% and 16% under the r.p.m. 8.5 scenario of change respectively (Figure 6.5 & 6.6).

Within the global distribution of European grayling, the UK is expected to be within an area of least change of maximum temperatures, as indicated by the maximum temperature of the warmest month (Bio5) in relation to grayling sensitivity (Figure 6.7).

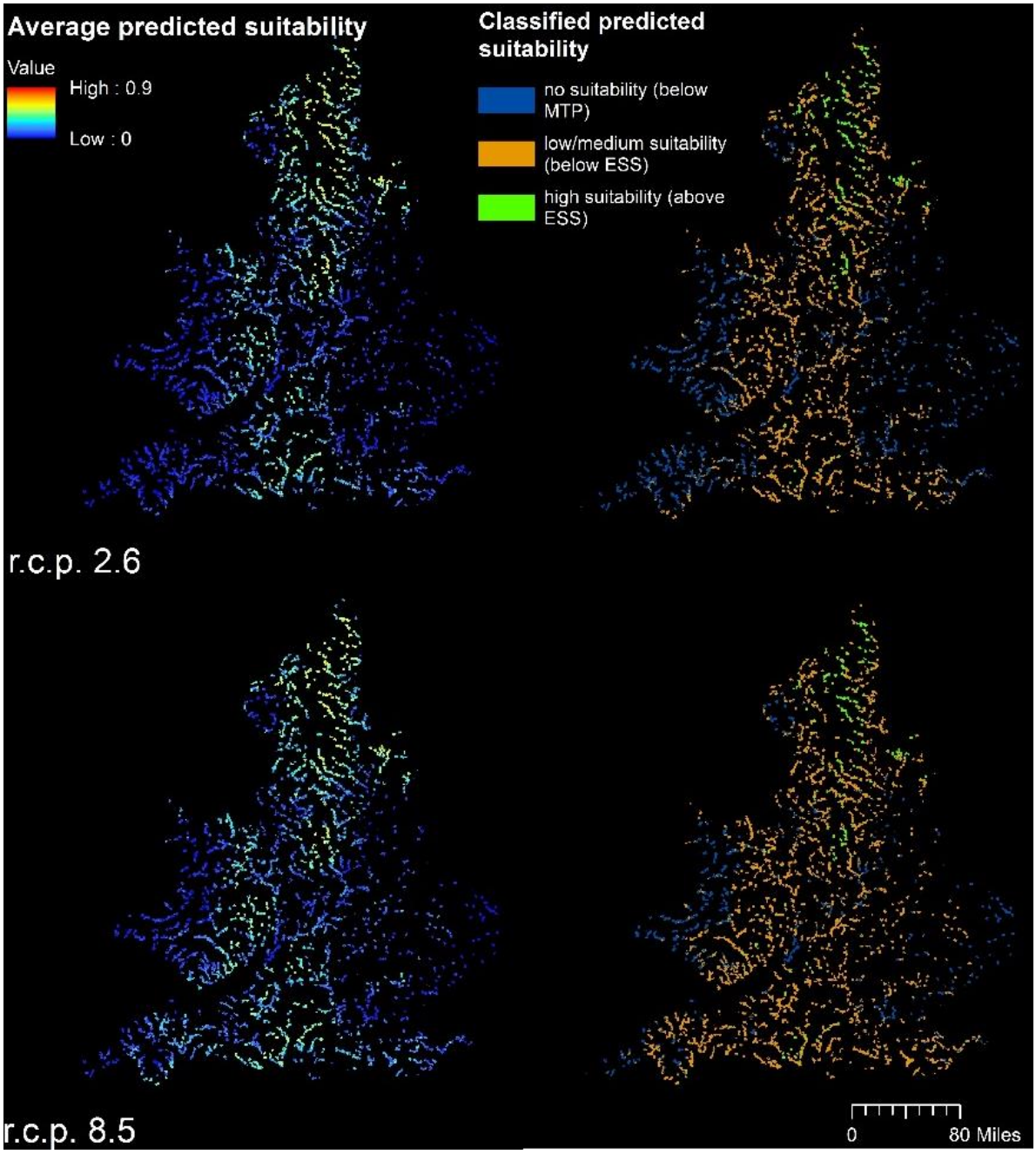


Figure 6.1 Projections of habitat suitability estimates for 2050 for r.c.p. 2.6 (above) and 8.5 (below): *Averages across 100 replicates are shown (the mean standard devaition was 0.06 for both the 2.6 scenario and the 8.5 scenario of change across the study area)(left), with warmer colors indicating higher suitability; Habitat classification were done using the Minimum Training Presence Threshold (MTP) and Equal Training Sensitivity and Specificity (ESS) threshold (right)*

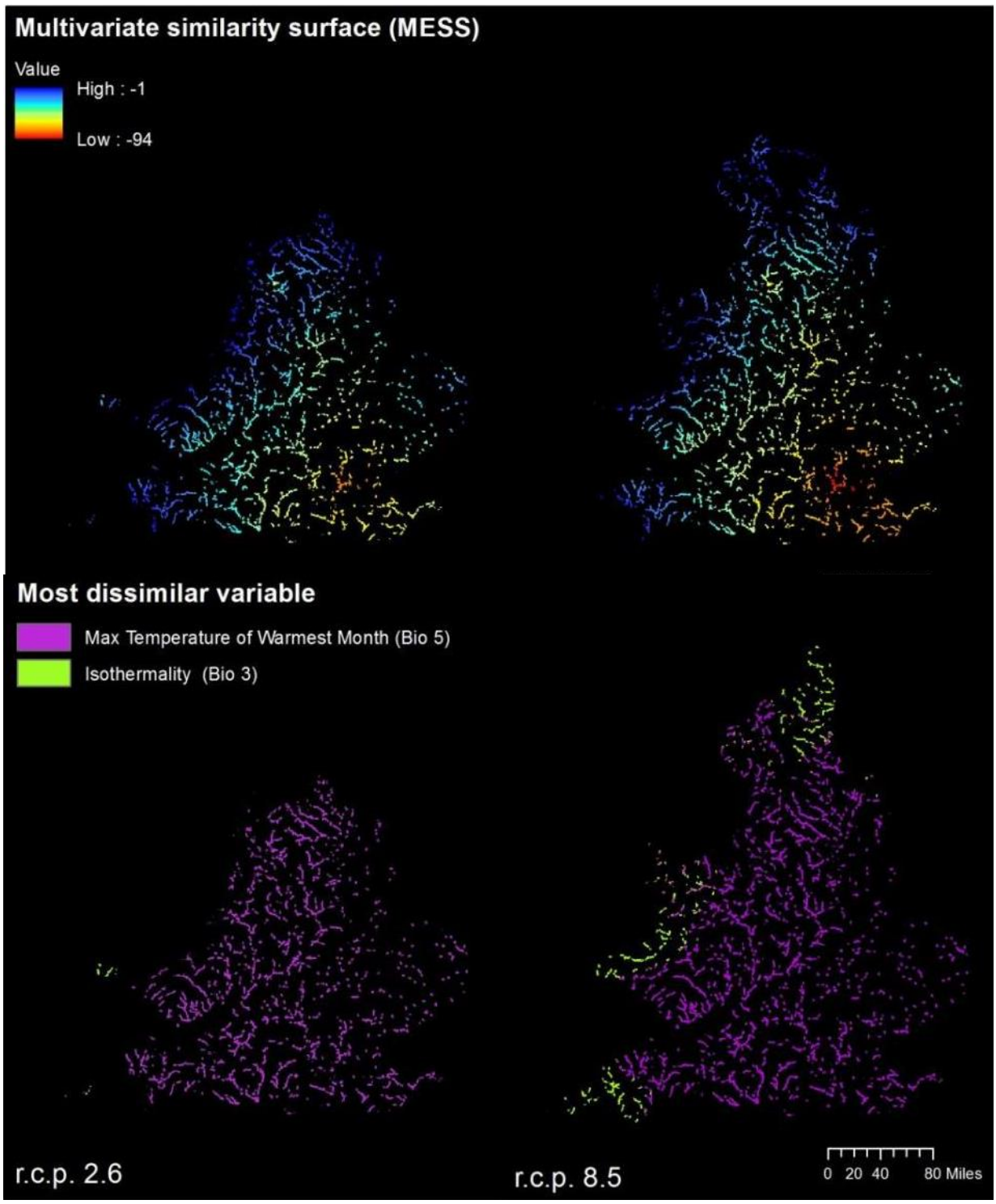


Figure 6.2 Multivariate environmental similarity surface: *top: areas with dissimiliarity to training data are shown, represented as negative values (highest dissimilarity in red); bottom: variable of highest dissimilarity for each point across the study area*

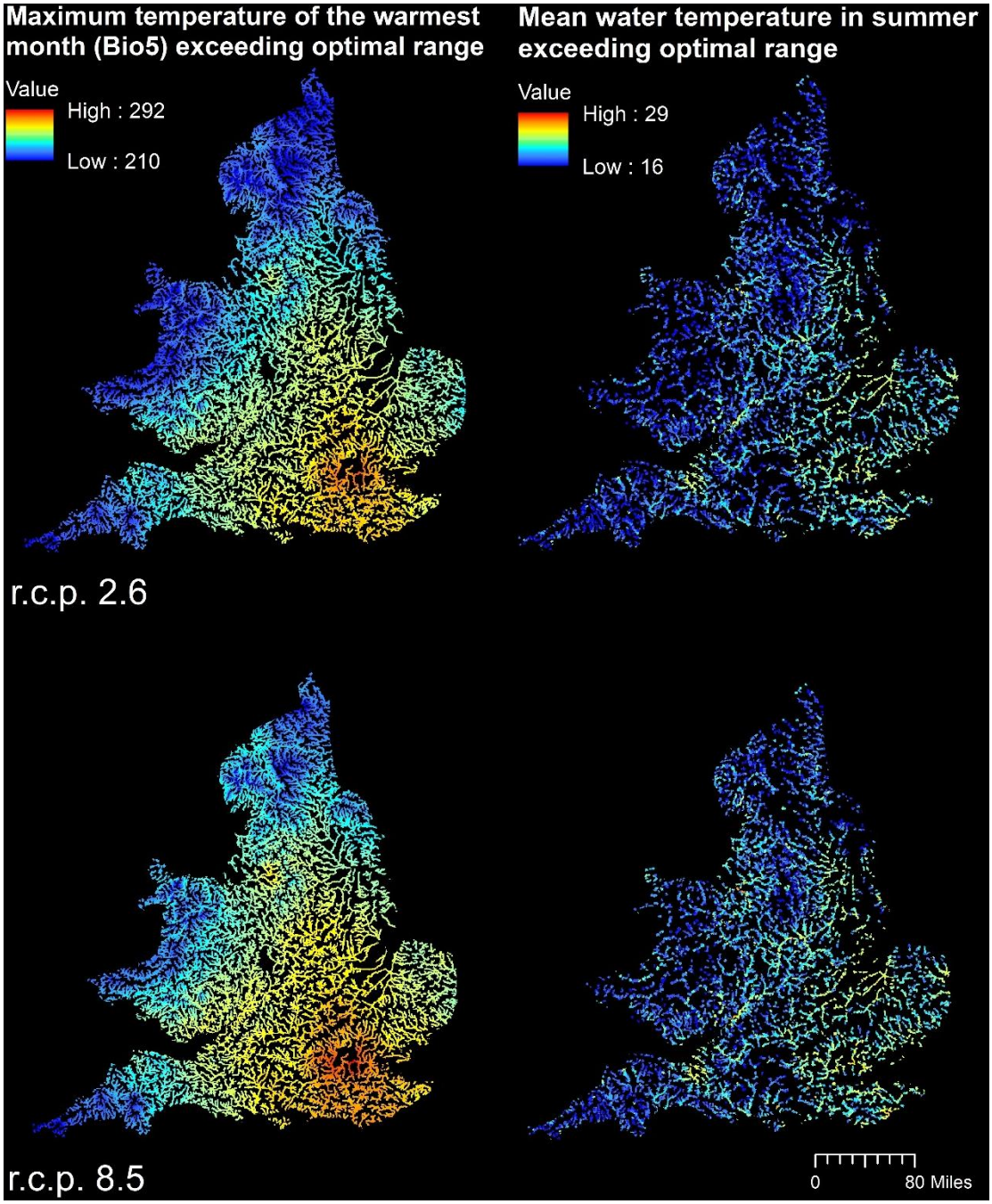


Figure 6.3 Maximum temperature of the warmest month (Bio5) (°C*10) and mean water temperature (°C) in summer is shown where it exceeds the optimal range for grayling: *habitat suitability is estimated to be impacted above 21°C for maximum temperature of the warmest month and above 16°C for mean water temperature in summer (chapter 5);*

Table 6.2 Relative percentage change in habitat classification for 2050, when Orthophosphate, Nitrite or Copper concentration is reduced to the maximum estimated as tolerable

		Orthophoshpate (0.15mg/L)	Nitrite (0.02 mg/L)	Copper (2 µg/L)	All (max)
r.c.p. 2.6	**increase above MTP**	8	5	6	10
	increase above ESS	2	2	1	4
r.c.p. 8.5	**increase above MTP**	5	1	2	6
	increase above ESS	1	0	1	2

Relative increase of suitable area in percent under the following classification: no suitability (below MTP), low-medium suitability (below ESS) or high suitability (above ESS) as projected for 2050 under the r.c.p. 2.6 and 8.5 respectively when either Orthophoshpate, Nitrite or Copper concentration is reduced to the maximum that was estimated tolerable or the parameter which maximizes suitability for each cell (All)

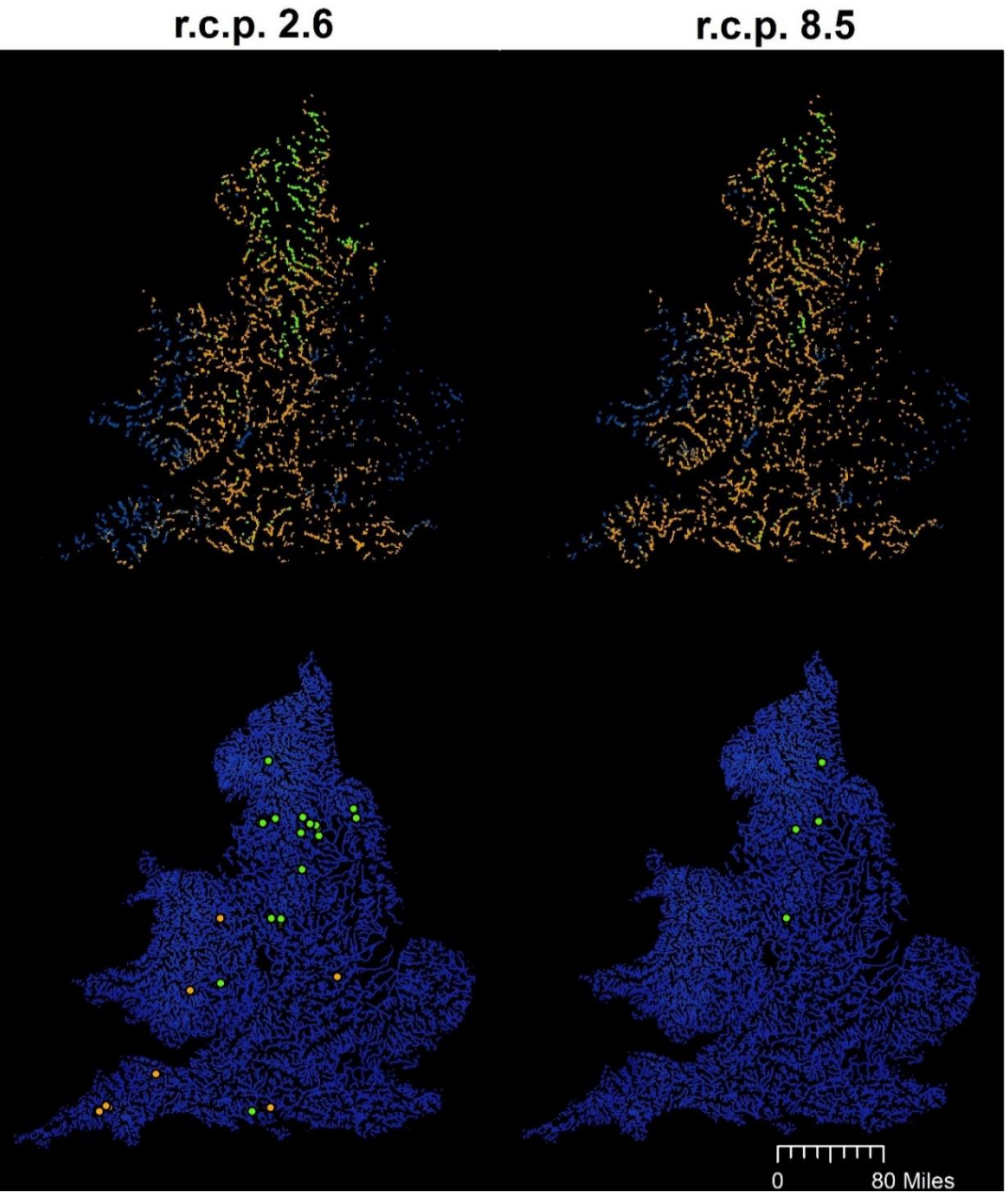


Figure 6.4 Classified predictions of habitat suitability estimates for 2050 for r.c.p. 2.6 and 8.5 when locally either orthophoshate, nitrite or copper concentrations were reduced: *top panels show classified suitabilities using the Minimum Training Presence Threshold (MTP) and Equal Training Sensitivity and Specificity (ESS) threshold; bottom panels show populations that benefit by habitat improvement to either high or low/medium suitablitiy*

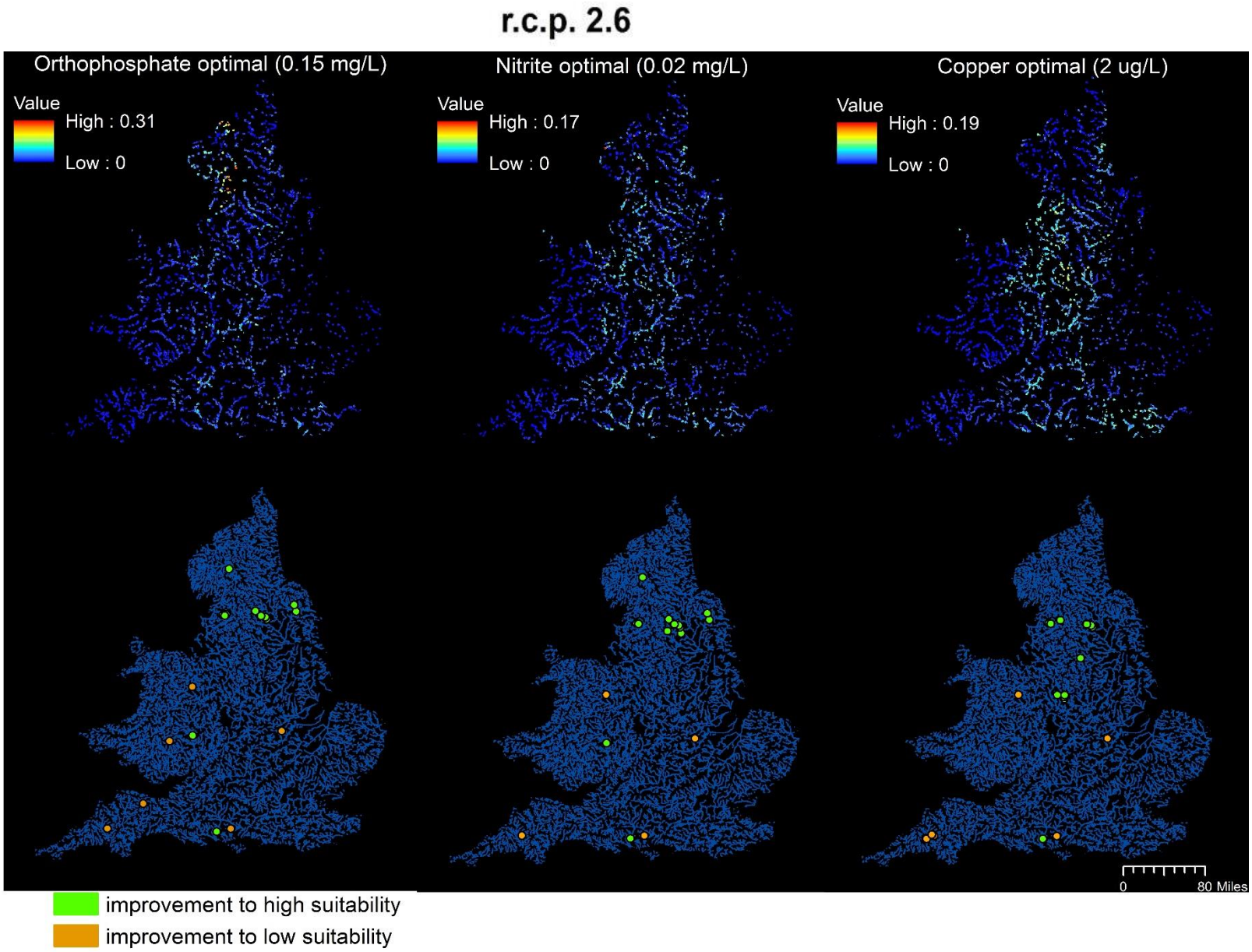


Figure 6.5 Projections of change in habitat suitability estimates for 2050 for r.c.p. 2.6 when reducing Orthophosphate, Nitrite or Copper concentrations to maximum values estimated not to affect habitat suitability for grayling: *top panels: increase in suitability observed across the study area; bottom panels: sites with current occurrences of grayling that are positively affected by reducing either Orthophosphate, Nitrite or Copper concentrations*

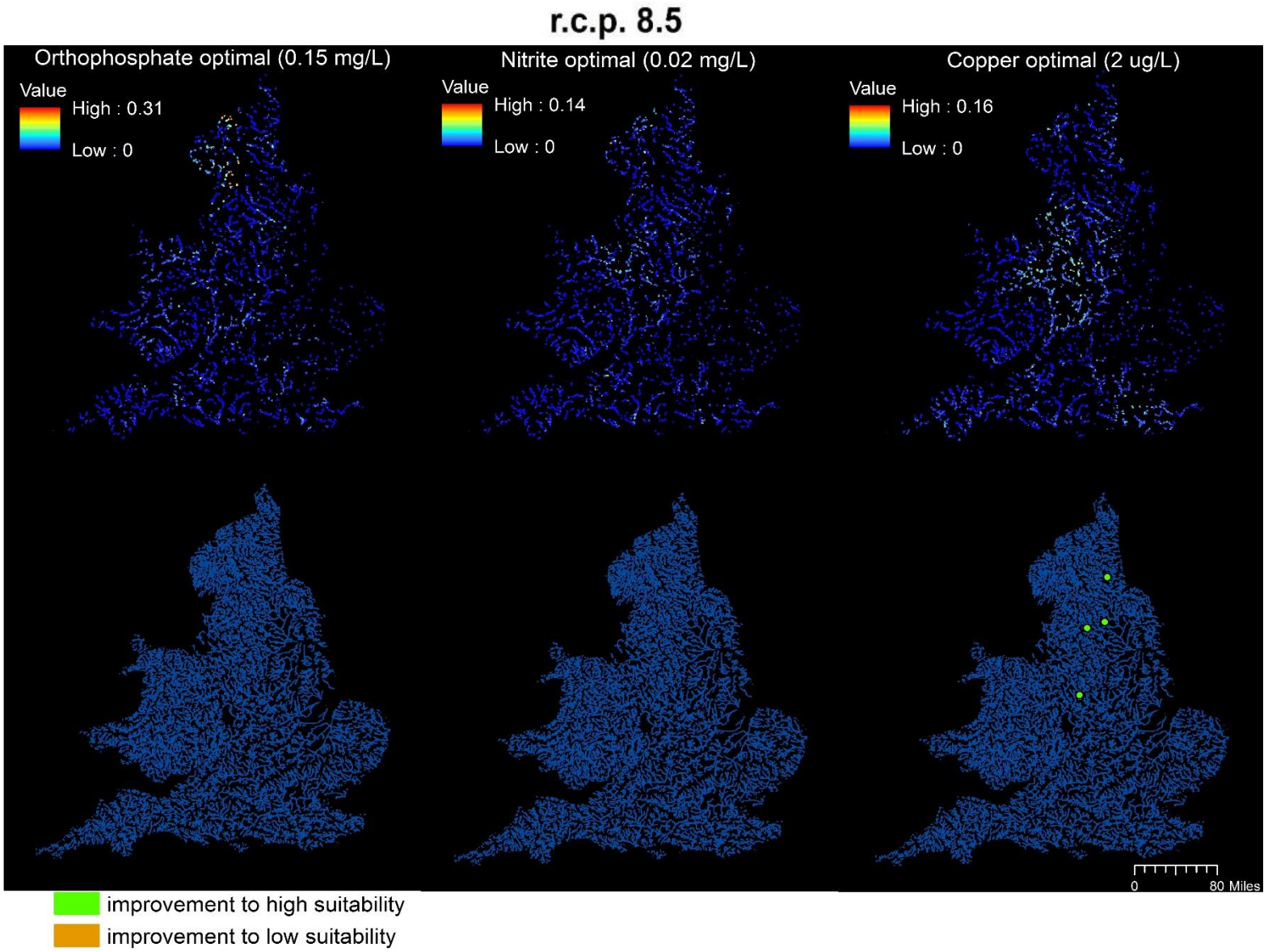


Figure 6.6 Projections of change in habitat suitability estimates for 2050 for r.c.p. 8.5 when reducing Orthophosphate, Nitrite or Copper concentrations to maximum values estimated not to affect habitat suitability for grayling: *top panels: increase in suitability observed across the study area; bottom panels: sites with current occurrences of grayling that are positively affected by reducing either Orthophosphate, Nitrite or Copper concentrations*

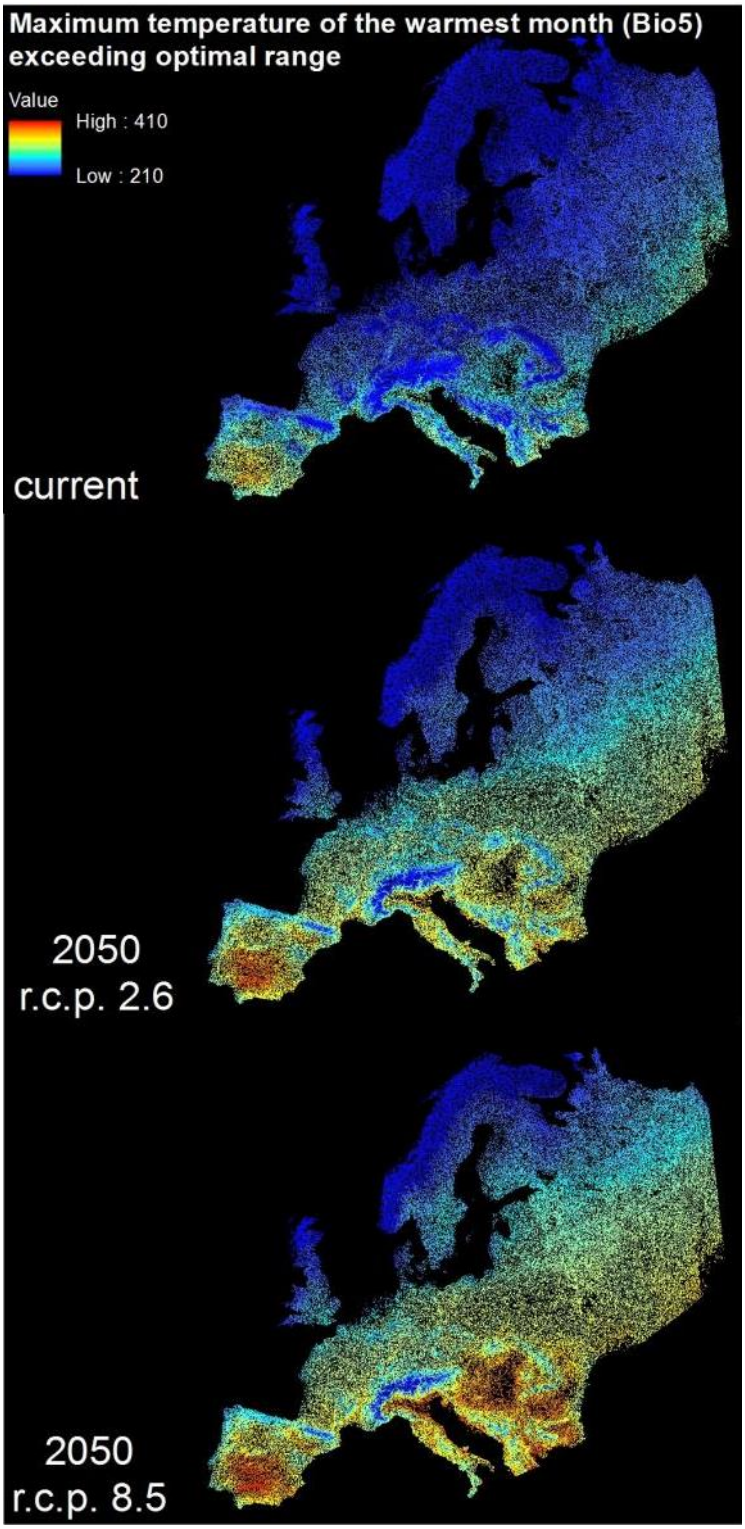


Figure 6.7 Maximum temperature of the warmest month (Bio5) (°C*10) across the whole area of distribution of European grayling under current conditions (top) and for 2050: *considering the r.c.p. 2.6 scenario of change (middle) and the r.c.p. 8.5 scenario of change (bottom), in relation to grayling sensitivity (habitat suitability is estimated to be impacted above 21°C, chapter 5)*

Discussion

This study showed significant reductions of suitable range for grayling under future climate predictions. However, the models presented demonstrate that adjustments in environmental parameters maintain habitat suitability under climate change. It is clear that modelling of future distributions should take both climate and non-climate related parameters into account.

Particular environmental parameters of importance (for mitigating range changes) included orthophosphates, nitrites and metal pollution. Organic pollution and eutrophication, associated with high concentrations of phosphate and nitrate are one of the major anthropogenic impacts on freshwater systems (Birk et al. 2012; Blabolil et al. 2016). Harmful effects for fish are caused by oxygen depletion, resulting from increased phytoplankton growth (Elshout et al. 2013). Because of temperature dependencies, the impact of eutrophication is expected to further increase under climate change conditions (Ficke et al. 2007; Moss et al. 2011). Metals have toxic effects on fish, mainly caused by a disruption of mechanisms important for ion regulation (Alsop and Wood 2011). Metal pollution is considered a serious threat to freshwater ecosystems (Förstner and Wittmann 2012) and is associated with historic mining activities within the UK (Macklin et al. 1997).

As climate change occurs, species will need to respond either adaptively or by tracking range changes (Hoffmann and Sgrò 2011). That there is a scope for adaptive responses to climate change in salmonids has been shown by several studies. Eliason et al. (2011) showed that populations of sockeye salmon (*Onchorhynchus nerka*) exhibit different tolerance limits to temperatures, reflecting historic temperature ranges of their habitat. An adaptive population divergence in response to temperature has been shown by Kavanagh et al. (2010) arising quickly and under constraining conditions of continued gene flow and previous bottlenecks in grayling. Jensen et al. (2008) showed significant differences in the amount of heritable variation in phenotypic plasticity between populations in response to temperature regimes in brown trout (*Salmo trutta*). Increasing flexibility is thought to be particularly important under conditions of climate change, which are predicted to be more fluctuating.

The high importance of isothermality in the habitat suitability model for grayling within the UK, might indicate that there are selection pressures on increased flexibility present.

Responses to climate change in grayling life-history traits have also been described. Data from a 62 year long survey on grayling revealed a shift to earlier spawning in congruence with increasing water temperatures (Wedekind and Küng 2010). An adaptive advantage of this shift in spawning timing is likely with increasing temperatures in spring, as temperatures above 16°C are lethal to grayling eggs (Jungwirth and Winkler 1984). Wedekind and Küng (2010) notice however unequal sex-ratios towards males, which resulted to be a persistent pattern, best explained by the mean temperature juveniles experienced in their first summer (Wedekind et al. 2013), which is thought to be the result of sex-specific differences in growth rate and survival under higher temperatures (Hutchings 2006; Martins et al. 2012). Unequal sex-ratios reduce the effective size of a population and therefore increases the risk of inbreeding (Crow and Kimura 1970) and reduce the adaptive capacity, as the efficiency of selection is directly related to the effective population size (Wright 1931).

Maximum temperatures in summer (Maximum air temperatures in the warmest period (Bio5)), also showed to be the parameter subject to most change under future projections (Figure 6.2). Whilst air temperatures generally show a linear relationship to water temperatures, the latter are also affected by flow, water volume, shading and wind shelters and deviations from linearity have been particularly shown, when maximum air temperatures exceed 25°C (Erickson and Stefan 2000; Webb et al. 2003). This was also evident from the river surface water temperature dataset used here (Orr et al. 2010). Differences in the effect of maximum air temperatures on mean water temperatures may reflect local differences in buffering capacity of the water body. Local thermal refugia have been shown to be actively used by juvenile salmonids and are potentially important for their survival in a warming climate (Sutton and Soto 2012). Further investigations on local thermal refugia within grayling rivers that are expected to suffer from temperature increases, like those in the South East, and management actions, like the creation of riparian woodland zones would be recommended (Malcolm et al. 2008). Also, the effect of groundwater abstraction should be evaluated in this context and in regard to increasing the risk of low summer flows, which can be critical for grayling (Solomon and Sambrook 2004; Riley et al. 2009).

Reducing non-climate related environmental stress has been highlighted to be among the most important management action in the face of climate change (Heller and Zavaleta 2009). Whilst this study does not try to give accurate predictions of the future distribution of grayling, its main goal was to show the potential of different habitat improvement strategies to increase habitat suitability for grayling under conditions of climate change and to give specific suggestions on local priority actions. Effective conservation of grayling within the UK and other parts in the northern area of the species distribution is particularly warranted to safe-guard the species persistence, as the effects of climate change are expected to be less drastic than for the southern parts. The importance of evaluating priorities and invoking management actions on habitat improvement within the continental distribution of grayling is emphasised, given an expected stronger effect of climate change.

Chapter 7: Initial description of the microbiome associated with European grayling

Abstract

The study of the microbiome, the entirety of microbial organisms associated with an individual, offers multiple possibilities to study the evolutionary dynamics between hosts and pathogenic and mutualistic micro-organisms, including the role of host-genetics and environmental factors. Here a preliminary study on the microbiome associated with the gills and mouth of grayling is presented, with the aim to inform the experimental design for future large-scale studies. In particular the difference in targeting the V1-2 region or the V3 region of the bacterial 16S rRNA gene is highlighted, showing a higher phylogenetic resolution for the V1-2 region in most, but not all cases. The communities associated with the gills are compared to those of the mouth and the results confirm the initial hypothesis that the gill communities reflect the more stable environmental community to a higher degree than the mouth community, which shows shifts in relative abundances. As the first study to investigate the microbiome of grayling, results towards the definition of a core microbiome associated with grayling are presented, which consist of genera found in samples of two individuals from different environments in both the gill and mouth community.

Introduction

The development of new sequencing technologies, referred to as next-generation sequencing (NGS) has revolutionised most fields of biological research (Koboldt et al. 2013). Particularly, microbiological research has entered a new era, characterised by culture-independent assessment of microbial diversity using genomic tools (Xu 2006; Hugenholtz and Tyson 2008). The discovery of an underappreciated, vast diversity of microorganisms and their high degree of complex relationships with eukaryotic organisms, including those of mutualistic, commensal and pathogenic nature has significantly changed established views of microbial communities and their role in the evolution and functioning of higher organisms (Zilber-Rosenberg and Rosenberg 2008; Fraune and Bosch 2010; Fumagalli et al. 2011). An organism and the entirety of microbial organisms associated with it (its microbiome), can act as one unit for selection, adaptation and evolution (Rosenberg and Zilber-Rosenberg 2016). This is illustrated by examples where differences in microbiome composition affect mate choice and offspring survival leading to reproductive isolation, driving the process of speciation (Brucker and Bordenstein 2013; Hurst and Jiggins 2013; Arbuthnott et al. 2016). This is highly relevant for conservation biology, as alterations of microbiome composition in captive breeding programmes for example, which are often characterised by artificial environmental sterility to avoid disease outbreaks, could affect reintroduction success (Redford et al. 2012) and contribute to outbreeding depression (Brucker and Bordenstein 2013).

The important role of host-associated microbial communities in disease susceptibility has also been recently recognized (Koskella 2014). With the prospect of infectious disease becoming a major threat to species within a warming climate (Altizer et al. 2013), the understanding of evolutionary resilience to pathogens as part of a complex microbial-host co-evolution is warranted (Llewellyn et al. 2014). This includes extending our knowledge about dynamics of microbial communities in natural systems to understand the risk of disease-mediated extinctions in relation to other pressures like habitat loss or pollution (Smith et al. 2009). Habitat degradation and environmental stress can cause shifts in the microbiome composition, generally reducing diversity, which increases susceptibility to opportunistic pathogens (Thurber et al. 2008; Verbrugghe et al. 2012; Amato et al. 2013;

Boutin et al. 2013). Monitoring of the microbiome composition across populations could therefore be a good indication of the general physiological state and be used in a conservation context (Allendorf et al. 2010).

European grayling (*Thymallus thymallus*) constitute a good system to study the interactions of host immune genetics, the microbiome and the environment for a number of reasons. Grayling is a species of conservation concern and has a broad distribution in freshwater systems across Europe (Gum et al. 2009). Microbiomes of natural fish populations have hardly been studied (Sevellec et al. 2014) and grayling potentially encounter diverse microbial selection pressures across the range. Grayling show the highest sensitivity to high temperatures among salmonids (Ibbotson et al. 2001; Jonsson and Jonsson 2009) and high vulnerability to human impacts such as water quality alterations (Oberdorff et al. 2002) and pollution (Buhl and Hamilton 1990, 1991; Vuorinen et al. 1998; Uiblein et al. 2001). They may therefore provide a good indicator species for habitat quality for salmonids, which are a family of high economic value (Youngson et al. 2003). Captive breeding and stocking programmes are moreover a common management strategy in salmonids and grayling in particular (Fraser 2008; Persat et al. 2016). Research on potential effects of an artificially sterile rearing environment on an individuals' microbiome and the consequences of this are warranted and may partly explain generally lower survival of hatchery reared fish in comparison to wild reared individuals Harbicht et al. (2014).

Here, a preliminary study on the microbiome associated with different body parts of grayling is presented to inform future large scale studies. Whilst studies of microbial diversity associations mostly target internal organs, like the kidney, to be able to infer infection through the identification of pathogenic strains (Dionne et al. 2009; Sevellec et al. 2014), conservation oriented studies, are often limited to non-invasive sampling techniques. The suitability of swab samples, taken from the mouth or gills of fish, is therefore evaluated to inform about an individuals' microbiome. Evidence of pathogens, infecting different tissues, from swabs of the mouth cavity of the infected individual has been given for various species and distinct pathogens (Cohen et al. 1989; Icenhour et al. 2002; Komar et al. 2002; Prickett et al. 2008). The microflora of gills are thought to mainly reflect that of their environment (Cahill 1990; Austin 2006). Moreover the phylogenetic resolution when targeting different regions of the bacterial 16S gene is evaluated. In particular, results for

the V3 region (Muyzer et al. 1993) and the V 1-2 region (Klindworth et al. 2012) of the bacterial 16S gene are compared, in addition to two different purification protocols (gel extraction and Ampure). Four samples were sequenced on an Ion Torrent PGM bench top sequencer to evaluate the experimental design for future studies. Through the identification of genera present in both gill and mouth samples and for individuals from different rivers first results towards the definition of a core microbiome associated with grayling are furthermore presented.

Materials and Methods

Samples of two fish (fish 1, from the river Ure and fish 2, from the river Wharfe) were used in this preliminary study designed to inform a future large-scale study. Sampling was done in January 2014 by taking swabs from the mouth and gills of each fish, which were stored in the same buffer (0.4M NaCl,10mM Tris HCL, 2nM EDTA, 2%SDS, 0.001% Proteinase K) used for DNA extraction following Dawnay et al. (2011). The mouth sample of fish 1 was used to compare two different primer pairs, one described by Muyzer et al. (1993) targeting the V3 region of the bacterial 16S rRNA gene and one described by Klindworth et al. (2012), targeting the V1-2 region of the of the bacterial 16S gene. The other three samples were only amplified for the V1-2 region. PCRs were done in 20 µl volumes using Taq polymerase (Bioline). PCRs were performed in Prime (Bibby) PCR cyclers or in a ABI 1 PCR cycler with following thermal profile for the V3 region: 94°C for 5 min, followed by 20 cycles at 95°C for 30 s, 65°C for 45 s, decreasing the temperature by 0.5°C each cycle, and 30 s at 72°C, followed by 10 cycles at 94°C for 30 s, 55°C for 45 s and 72°C for 30s and a final extension of 3 min. at 72°C. The thermal profile for the V1-2 region was: 7 min. at 94°C, followed by 10 cycles at 94°C for 30 s, 62°C for 30 s, decreasing the temperature by 1°C per cycle and 72°C for 30 s, followed by 20 cycles at 94°C for 30 s, 53°C for 30 s and 72°C for 30 s and a final extension for 7 min. at 72°C.

The gill sample of fish 2 was amplified twice (sample 1.2a and 1.2b) in order to assess the effect of different sample treatment and specifically purification on the results. In particular,

the effect of higher sample loss during preparation on estimates of relative abundances within the sample was tested, predicting a loss of the least abundant taxa. Sample 1.2a was therefore purified by slicing the PCR band out of the gel and pulse spinning it through sterile glass wool to remove the agarose, prior to purification with AmpureXP (Beckman and Coulter) at a ratio 1: 1.3, which was the only purification applied to the other three samples. For fish 2 only the mouth sample was analysed. PCR products were quantified using a Qubit® 2.0 Fluorometer (Invitrogen) and 26 pM of each sample were loaded to Ion Sphere Particles (ISPs) using Ion PGM™ Template OT2 200 kit (LifeTechnologies™, USA) according to the manufacturer's instructions for high-throughput sequencing. For sample one, 13 pM of both PCRs, using different primers, were pooled. Sample 1.2a exhibited a lower concentration and only 20 pM were loaded. Multiplexed sequencing was conducted using a 316™ chip (LifeTechnologies™) on an Ion Torrent Personal Genome Machine (LifeTechnologies™). Sequences were binned by sample and filtered within the PGM software to remove low quality reads. Data were then exported as FastQ files.

The mouth sample of fish one, which was multiplexed for two different 16S regions, was demultiplexed by region specific primers in Mothur (Schloss et al. 2009). Then, for all samples a quality filter was applied based on a minimum Phred score of 20 within 80 % of the sequence, using FASTX-Toolkit (Hannon Lab, USA). The sequences were concatenated and sorted by similarity in a single fasta file for further processing and analysis in QIIME 1.8 (Caporaso et al. 2010b). Sequences with less than ten reads were removed. Putative chimeric sequences were identified using USEARH (Edgar 2010) and removed from the analysis. Operational taxonomic units (OTUs) were picked with a minimum pairwise identity of 97%. The most abundant sequence in each OTU was selected as a representative to assign a taxonomic classification based on the Greengenes database (DeSantis et al. 2006) using the RDP classifier (Wang et al. 2007), clustering the sequences at 97% similarity with a 0.80 confidence threshold. Multiple alignments of the representative sequences for each OTU with a minimum length of 150bp were created using PyNast (Caporaso et al. 2010a). Alpha diversity metrics were calculated on rarefied OTU tables to asses sampling depth coverage using observed species, Chao1 (Chao 1984), Shannon's diversity index (Shannon and Weaver 1948) and Good's coverage (Good 1953). Beta diversity metrics among samples were calculated

using weighted and unweighted Unifrac distances (Lozupone et al. 2007) and Bray-Curtis similarity (Bray and Curtis 1957). Venn diagrams were created to visualize sample similarity using the Venny 2.1 web server (Oliveros 2007).

Results

Summary statistics for the four 16S samples run on an Ion Torrent PGM are shown in Table 1. Sample 1.2a, which had been purified using gel-extraction prior to AmpureXP purification and was loaded in lower concentration, showed the lowest number of reads of all samples (Table 7.1). Good's coverage estimations >99% have been obtained and rarefaction curves (Figure 7.1) show that a plateau is approached for all samples. In total, 2508 operational taxonomic units (OTUs) were observed, covering 31 phyla (Figure 7.2) and 413 genera. Highest α- diversity was found for the gill sample of fish 1 (sample 1.2b), sequenced for the V1-2 region, whilst lowest α- diversity was found for the mouth sample of fish 2 (sample 2) (Table 7.2).

Table 7.1 Summary statistics of Ion Torrent PGM data for four 16S samples

Sample	Sample description	Purification	16 S region analysed	Individual	Sampling location	No. of reads	No. of reads used in analysis
1.1 a	Mouth swab	AmpureXP	V3	1	Ure	232,555	73,939
1.1 b	Mouth swab	AmpureXP	V1-2	1	Ure		58,706
1.2 a	Gill swab	Gel-extraction+ AmpureXP	V1-2	1	Ure	60,178	39,282
1.2 b	Gill swab	AmpureXP	V1-2	1	Ure	249,816	170,892
2	Mouth swab	AmpureXP	V1-2	2	Wharfe	203,901	133,696

Information given for each sample: name, description of the sampling and processing method, individual the sample was derived from, location where the individual was sampled, number of reads it generated on an Ion Torrent PGM, number of sequences finally used after filtering for chimeric or low quality sequences, having a minimum length of 200 bp for the V 1-2 region and a minimum length of 120 for the V 3 region that aligned with >= 97 % identity to the Greengenes databas and percentage of initial reads used in the presented analysis

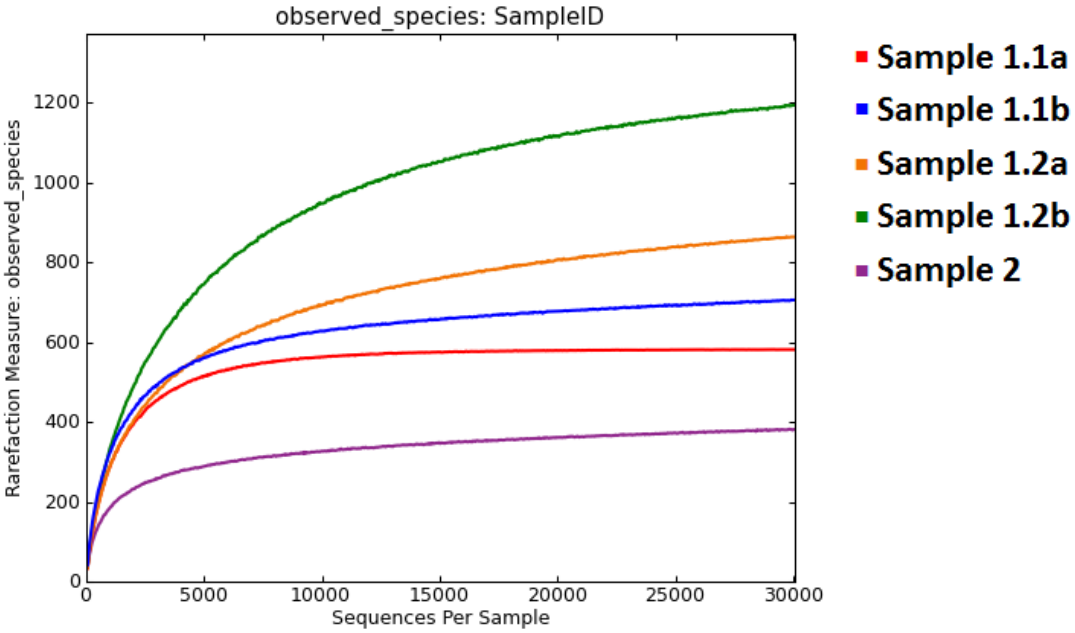


Figure 7.1 Rarefaction curves*: generated by averaging 100 repititions of random sampling of observed species within each sample as a function of sequences per sample*

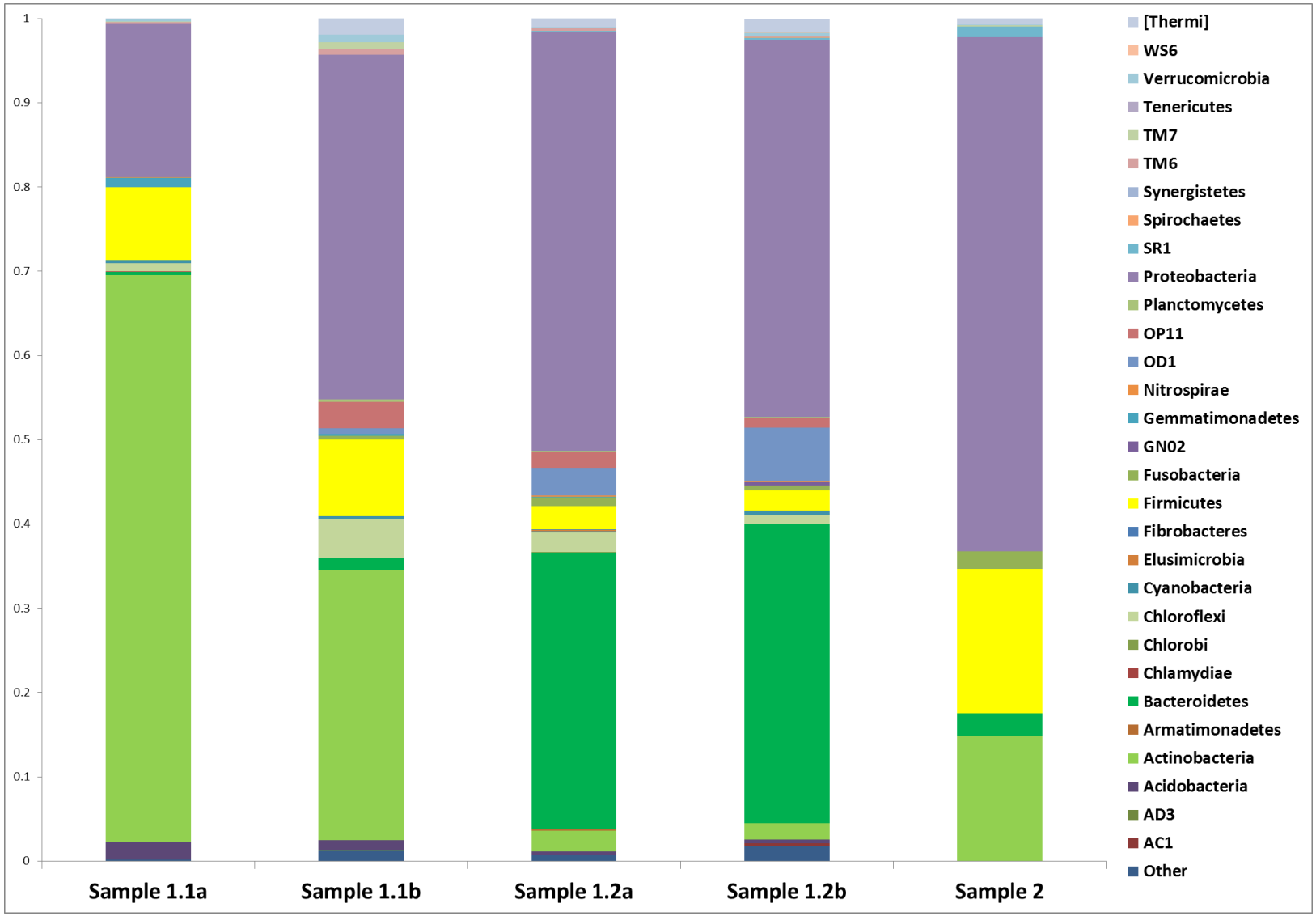


Figure 7.2 Phylogenetic composition of bacterial diversity found in each sample at the phylum level: *contributions are given as percentages of the whole sample*

Table 7.2 α- diversity estimates for all samples

measurement	Sample 1.1a	Sample 1.1b	Sample 1.2a	Sample 1.2b	Sample 2
Observed species total	581	705	863	1193	381
Chao1 index	582	788	968	1295	438
Shannon index	6.7	7.8	7.1	7.4	6.7

Comparison of the V3 and V1-2 region

Proteobacteria had the highest contribution for all samples sequenced for the V1-2 region, whilst Actinobacteria had the highest contribution when the V3 region was sequenced (Figure 7.2). Comparing the results from the V3 and V1-2 region amplified from the same mouth sample (sample 1.1a versus sample 1.1b), higher phylogenetic resolution was achieved for the V1-2 region (Table 7.3). Of all genera identified, 28.3% were found in both samples, 21.3% were only identified for the V3 region and 50% were only identified for the V1-2 region (Figure 7.3). The mean frequency of OTUs exclusively identified in sample 1.1a was 0.28 ±0.5% and the maximum was 2.5%, consisting of genus *Ralstonia*. The mean frequency of OTUs exclusively identified in sample 1.1b was 0.31±0.7% and the maximum was 4.6%, consisting of the genus *Methylobacterium*.

Table 7.3 Comparison of resolution between the V 1-2 region and the V 3 region: *Percentage of sequences that can be assigned at the phylogenetic level of order, family, genus and species respectively for the V 1-2 and V 3 region;*

	V3	V 1-2
Order	94	93
Family	42	80
genus	26	50
Species	4	8

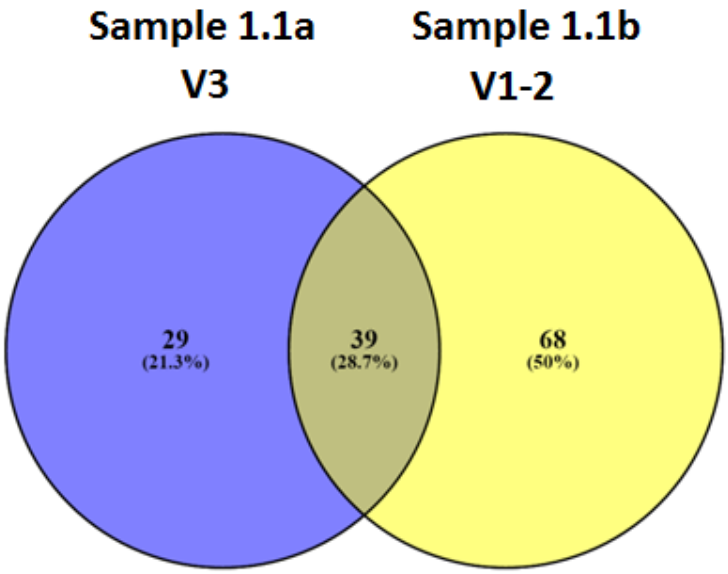


Figure 7.3 Venny diagram comparing number of OTUs with phylogenetic resolution at the genus level for V3 and the V1-2 region of the 16S rRNA gene: *comparison of sample 1.1 and 1.2, which have been amplified for the V3 and the V1-2 region of the 16S gene respectively*

Effect of sample preparation

The more than four-fold lower coverage of sample 1.2a in comparison to the same sample with a higher coverage (sample 1.2b), shows a loss of 230 OTUs (Table 7.2). For all genera identified, 35 (23.8%) were only found in sample 1.2b and 9 (6.1%) were only found in sample 1.2a (Figure 7.4). All genera exclusively identified in only one of the gill samples of fish 1 showed very low frequency in the sample they were present. The mean frequency of OTUs exclusively identified in sample 1.2a was 0.21 ±0.07% and the maximum was 0.9%. The mean contribution of OTUs exclusively identified in sample 1.2b was 0.03 ±0.04% and the maximum was 0.13%.

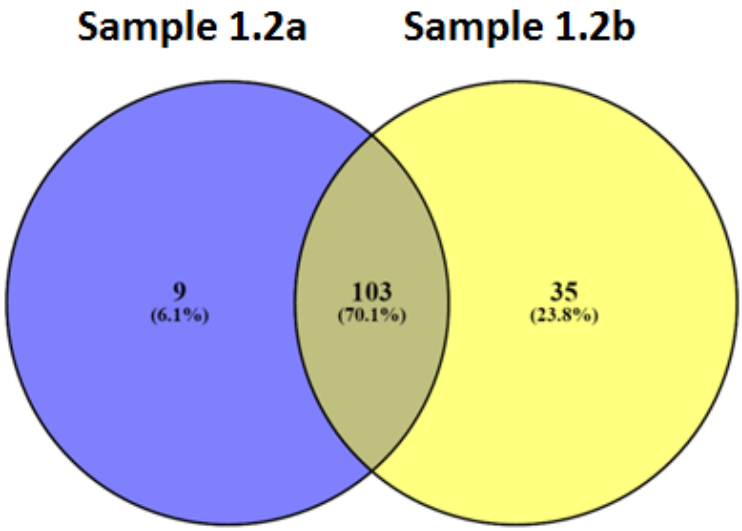


Figure 7.4 Venny diagramm comparing number of OTUs with phylogenetic resolution at genus level for different sample treatment: *comparison of sample 2.1 and 2.2, representing different treatments of the same sample amplified for the V1-2 region of the 16S gene; Sample 2.1 was purified using both gel-extraction and AmpureXP and had a coverage 4-fold lower than sample 2.2, which was purified only using AmpureXP*

Comparison of mouth and gill samples

Comparing the phylogenetic composition between mouth and gill samples of the same fish, shows that Actinobacteria have a much higher contribution in mouth samples, whilst Bacteroidetes have a much higher contribution in gill samples (Figure 7.2). Figure 7.5 shows the comparison of the mouth and gill sample of fish 1 (sample 1.1b and 1.2b respectively) and the mouth sample of fish 2 (sample 2), all amplified for the V1-2 region and purified using only AmpureXP, for all OTUs with phylogenetic resolution at the genus level. A core microbiome, which is found in both gill and mouth samples and for different fish from different environments, consist of 48 genera in total (Figure 7.5, Appendix 4). The gill sample shows a higher number of distinct genera than either of the mouth samples (Figure 7.5).

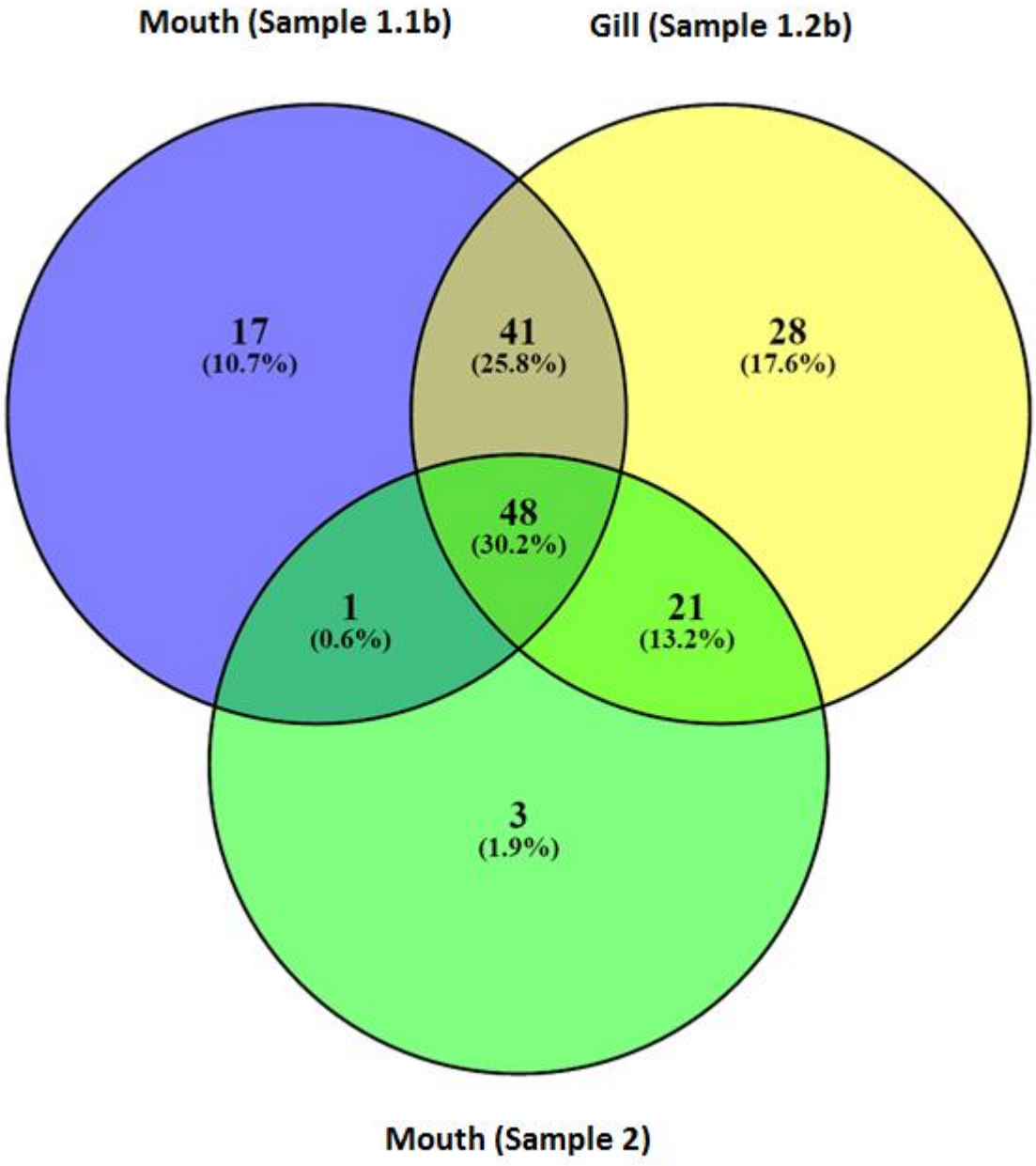


Figure 7.5 Venny diagramm comparing number of OTUs with phylogenetic resolution at genus level of samples from gill and mouth and mouth samples of different individuals: *and shared between sample 1.2, 2.2 and 3, representing the mouth and gill samples of fish 1 and the mouth sample of fish 2 respectively*

β-diversity between sample location and individuals

A comparison of β-diversity between samples shows that the mouth and gill samples of an individual relate more closely than mouth samples of two distinct individuals, which come from different locations, when using unweighted clustering based on presence only (Figure 7.6). When using the weighted clustering method, that takes abundance and phylogenetic distance into account, the distance between mouth and gill is greater than the distance between individuals (Figure 7.6). Sample 1.1a, which was amplified for the V3 region of the 16S gene clustered as an outgroup for both methods.

Comparison of community composition between individuals

A comparison of the relative abundance of different genera, with species of known opportunistic salmonid pathogens (Austin et al. 1983 ; Austin and Stobie 1992; Ostland et al. 1994; Gil et al. 2000; Hisar et al. 2002; Nematollahi et al. 2003; Austin and Austin 2007*;* Kayis et al. 2009; Navarrete et al. 2009; Merrifield et al. 2010), probiotics *(*Merrifield et al. 2010*,* Mukai et al. 2002) or that have been associated with either stressed or healthy fish in *Salvelinus fontinalis* is shown in Figure 7.7 (Boutin et al. 2013). In total nine genera were identified, which contain known pathogenic species and one pathogen was identified at the species level (*Micrococcus luteus* (Austin and Stobie 1992)) (Figure 7.7). Of these six were found in sample 2 at the highest frequency. The genera *Exiguobacterium* and *Bacteroides*, which have been found positively associated with stress by Boutin et al. (2013b) and *Lactobacillus reuteri,* which is known for probiotic effects (Mukai et al. 2002) were exclusively found in sample 2 (Figure 7.7). A total of five genera with known pathogenic species have been also found in samples from fish 1 (Figure 7.7). An additional four genera positively associated with stress and two negatively associated with stress have been only found in samples from fish 1 (Figure 7.7). Out of twelve genera analysed and assessed in samples of fish 1, two were present in both mouth and gill samples, three were only present in gill samples and seven only in mouth samples, with two respectively being only resolved by one of two the primers (Figure 7.7).

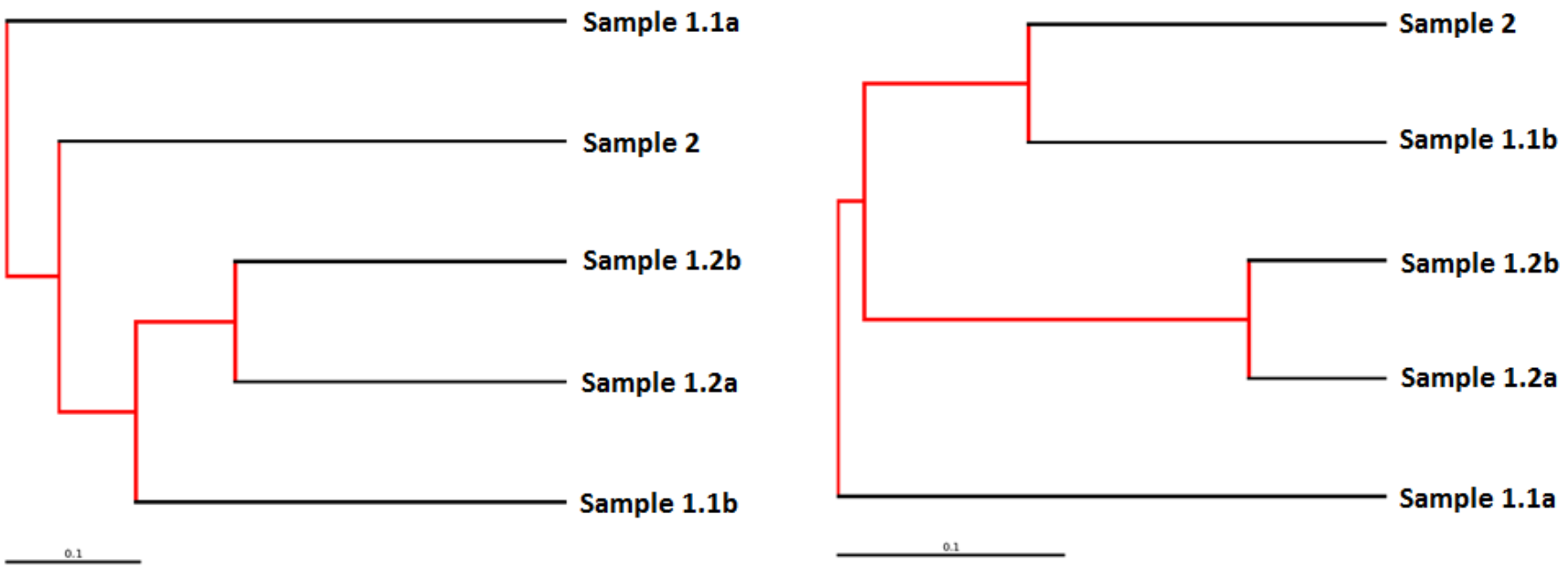


Figure 7.6 β-diversity as the average distance between samples: *the unweighted pair-group method (left) and the weighted pair-group method (right) with arithmetic mean (UPGMA) was used for clustering; jackknifing was done by resampling 100 times; Bray Curtis distances are shown*

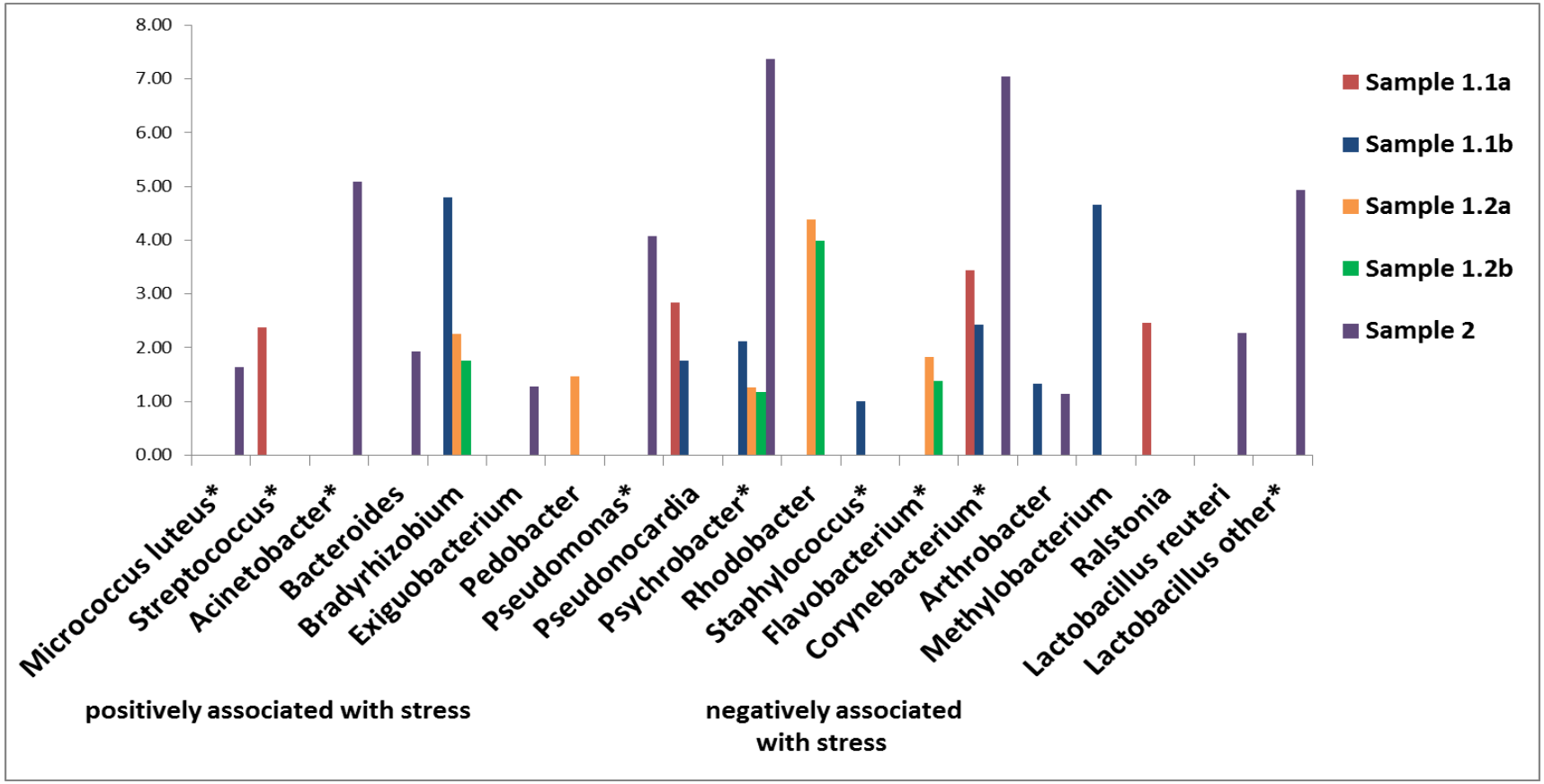


Figure 7.7 relative abundance of genera across samples: *genera with known opportunistic salmonid pathogens are indicated with (*)(Micrococcus luteus (Austin and Stobie 1992); Streptococcus iniae (Merrifield et al. 2010); Acinetobacter spp (Austin and Austin 2007), junii (Navarrete et al. 2009); Pseudomonas anguilliseptica, chlororaphis, fluorescens, pseudoalcalignes, putida and luteola (Austin and Austin 2007 ; Kayis et al. 2009); Psychrobacter immobilis (Hisar et al. 2002); Staphylococcus epidermidis, warneri (Gil et al. 2000); Flabobacterium psychrophilum, branchiophilum (Ostland et al. 1994; Nematollahi et al. 2003); Corynebacterium xerosis (Austin et al. 1983); Lactobacillus spp. (Austin and Austin 2007)), that are positively (Acinetobacter, Bacteroides, Bradyrhizobium, Exiguobacterium, Pedobacter, Pseudomonas,Pseudonocardia, Psychrobacter, Rhodobacter, Staphylococcus) or negatively (Flavobacterium,Corynebacterium, Arthrobacter, Methylobacterium, Ralstonia) associated with stress (Boutin et al. 2013b), or are potential probiotics (Lactobacillus plantarum, sakei, fluorescens, reuteri (Merrifield et al. 2010, Mukai et al. 2002)) are shown for each sample; only contributions with a minimum of 1% are shown ;*

Discussion

This study highlights important technical considerations to study the microbiome in natural fish populations and gives valuable recommendations for future studies. Also, the first results on the microbiome associated with grayling are presented. These results confirm general findings of the core microbiome of teleosts, with Proteobacteria, Actinobacteria, Firmicutes and Bacteroidetes as phylogenetic groups with high contribution (Llewellyn et al. 2014). The presented results suggest a clear difference in bacterial communities of the gills and mouth of grayling as well as within individuals, which is successfully resolved by the swab sampling approach. The gill samples showed a higher number of total OTUs and a higher Chao diversity estimate, which does not take abundance into account (Colwell and Coddington 1994). However, considering both richness and evenness, the Shannon diversity estimate is higher for mouth than gill samples of the same fish. This confirms our expectation that the microflora of gills, which show less evenly distributed abundances of OTUs, reflect the community of the environment to a higher degree than mouth samples, which may reflect species specific microbiome composition (Cahill 1990; Austin 2006). Also, when abundance is considered, the weighted measure of β-diversity shows more pronounced differences between gills and mouth, whilst when only presence is considered, the difference between individuals is stronger than the difference within one individual. The availability of both sample types allows comparison of relative abundances within gill samples, having a higher representation of the environment, to the mouth samples, which may reflect the characteristics of an individuals' microbiome more closely. This is supported by the higher resolution of genera associated with the physiological state in mouth than in gill samples (Figure 7.7).

A higher phylogenetic resolution could be clearly shown for the primer set for the V1-2 region of the 16S gene, which may partly be explained by covering a longer sequence in comparison to primer set for the V3 region, which increases the possibilities for comparison. Therefore, if one primer set has to be chosen our results support the choice for the V1-2 region, which is in agreement to the recommendations of Klindworth et al. (2012). However,

for some genera the V3 region was shown to be superior to resolve phylogeny. As it was demonstrated that both primer sets can be successfully used in a single barcoded sample, thus reducing library preparation cost, this approach would be recommended to take advantage of the complementary properties of both regions.

The gel-extraction prior to AmpureXP purification resulted in a lower sample concentration, so that less than the recommended amount was loaded for sequencing, which resulted in the low coverage observed for sample 1.2a (Table 7.1). Whilst Good's coverage estimates indicated that an increase in coverage would not result in significantly higher diversity estimates, almost a quarter of all genera identified in the same sample, but purified using only AmpureXP (sample 1.2b), were not present in sample 1.2a. All of these genera were present at very low frequency within sample 1.2b (the maximum frequency was 0.13%), confirms our hypothesis that more extensive purification would result in a predominant loss of the least abundant taxa. This results highlights that firstly the loss of sample should be minimized through processing and secondly be highly standardized for all samples to allow meaningful comparison. In sample 1.2a we also identified 6.7% of genera at low frequency, which were not present in sample 1.2b (Figure 7.4). Some stochasticity is expected within low frequency taxa to achieve sufficiently high coverage to be considered, which could explain this pattern. Similarly, Porazinska et al. (2010) found that just over 90% of species were reproducibly identified across replicate samples, using a metagenomics approach to study nematode diversity. Ramond et al. (2015) suggest a DNA normalisation procedure, involving a nuclease treatment of samples before preparation, to overcome the predominant loss of underrepresented taxa, which can be of particular interest in metagenomic studies.

There are clear differences in the relative abundance of genera with known relevant pathogenic species between samples of the two individuals (Figure 7.7). In accordance with the observation that lower diversity of commensal bacteria increases the risk for opportunistic pathogens (Boutin et al. 2013b), the sample form fish 2 (sample 2) showed both lowest diversity measures (Table 7.2) and generally higher contribution of genera with known pathogenic species, of which one was identified at the species level (*Micrococcus luteus* (Austin and Stobie 1992))(Figure 7.7). However, *Lactobacillus reuteri,* a species with

known probiotic effects (Mukai et al. 2002) was also identified in fish 2. Probiotic species can provide beneficial effects to the host in a number of ways, which include competitive exclusion of pathogenic species, the release of nutrients or enzymes that can be used by the host and immunostimulation (Pérez-Sánchez et al. 2014). A larger study would be required to elucidate the role of probiotics in grayling.Boutin et al. (2013b) identified family genetic background as the second most influential factor in explaining variation of skin mucus composition in brook charr (*Salvelinus fontinalis*) after exposure to stress and Bolnick et al. (2014) showed that certain MHC alleles increased or decreased the abundance of certain bacterial taxa in the gut microbiome of threespine stickleback (*Gasterosteus aculeatus*). A future study assessing both the microbiome of individual grayling and the MH genotype across multiple populations and environmental conditions would allow to assess how immune genetic variation affects the microbiome composition, including well known probiotic taxa, which could include trade-offs between maintaining beneficial probiotics and potentially harmful pathogens (Bolnick et al. 2014). Moreover, it could identify the degree of variation of selection pressures grayling are exposed to across different environments and give valuable insights on evolutionary dynamics between the host genetics and pathogens and the importance of environmental conditions (Spor et al. 2011). Assessing the potential adaptive value of certain MH variants, by assessing their influence on the microbiome composition, would further elucidate mechanisms of selection, like the degree of heterogeneity across space and the relative importance of a rare allele advantage. Altogether the study of the microbiome in comparison to immune genetic variation offers numerous possibilities to enhance our understanding on the complex interactions involved in evolutionary processes.

Chapter 8: Final discussion

Global biodiversity is being lost at alarming rates (Barnosky et al. 2012) with potential harmful effects to humanity (Cardinale et al. 2012). Because patterns of species decline are typically not linear, which can also be described as a vortex ultimately leading to extinction (Fagan and Holmes 2006), the loss of biodiversity is likely to be further accelerated (Ceballos et al. 2015), indicated by high rates of population loss (Ceballos and Ehrlich 2002). Loss of genetic diversity is a main factor that can contribute to a disproportionate increase in extinction risk over short time-scales (Spielman et al. 2004; Frankham 2005a; O'Grady et al. 2006; Blomqvist et al. 2010). International agreements aiming to mitigate loss of biodiversity recognise the importance of genetic factors, e.g. the Convention on Biological Diversity (CBD)(www.cbd.int) lists the preservation of genetic diversity as a major priority. This requires the capacity to accurately assess genetic variation in non-model species and evaluate it in an ecologically meaningful context. This includes understanding the limitations of methods used for genetic assessment and the possible impacts they have on conservation decisions (Bonin et al. 2007). Furthermore, to promote preservation of genetic diversity in natural populations, possible management actions need to be evaluated for their ability to meet this goal. In order to be able to pro-actively prevent the loss of genetic diversity, a better understanding of evolutionary forces acting upon natural populations is warranted.

In this PhD study, genetic variation at functionally important immune genes (major histocompatibility (MH) genes) was assessed in a non-model species (European grayling), based on data from closely related taxa (chapter 2). The experimental design and analytical pipeline are optimised for accuracy, high through-put and economy and can be easily applied to other study systems. The availability of data obtained from neutral markers for the same individuals and populations (Dawnay et al. 2011) genotyped at the MH in this study, allowed a direct comparison of results from both markers types. This made it possible to evaluate the impact of marker choice on management decisions and to investigate how

selection affects functional genetic variation by comparison with variation at markers, predominantly affected by demographic processes and drift.

No measurement of genetic diversity was correlated between ten microsatellite loci and the MH class II DAA and DAB loci across twelve populations in our study, confirming the suggestion that neutral markers may be inappropriate to predict functional genetic diversity (Reed and Frankham 2001; Kirk and Freeland 2011). This is important to be aware of where neutral markers are solely used to assess genetic diversity and distinctiveness, particularly if the aim is to set conservation priorities (Bozzi et al. 2012; Al-Qamashoui et al. 2014). It has been shown that adaptive genetic diversity (defined as functional genetic variation that translates to differences in individual fitness) can be affected differently by demographic processes than neutral markers, due to the simultaneous effects of drift and selection (Ejsmond and Radwan 2011; Sutton et al. 2011). For example Coughlan et al. (2006) found significant losses of genetic diversity measured for the MH I UBA in sea trout (*Salmo trutta)*, but not for neutral markers, during a period of recorded disease outbreaks in nearby aquacultures. The authors suggest periods of increased pressures of disease mediated, directional selection as the cause for the observed pattern. Periods of directional selection are in accordance with a frequency-dependent mechanism of selection, or heterogeneous selection in space and time as has been suggested to operate on the MHC (Piertney and Oliver 2005; Consuegra et al. 2011). This does not exclude the possibility of heterozygote advantage, for which empirical evidence has also been given (Evans and Neff 2009; Kekäläinen et al. 2009), but rather highlights that the mode of selection is context dependent. An advantage of heterozygosity is mainly expected in situations where exposure to multiple pathogenic strains is common (Penn et al. 2002; Oliver et al. 2009), whilst episodes of directional selection are expected when there is a dominant pathogen and specific alleles that confer resistance (De Boer et al. 2004).

Directional selection for particular MH variants has been found by Grimholt et al. (2003), who identified certain class II alleles to be significantly associated with increased resistance toward salmon anaemia virus (ISAV) and furuncolosis (*Aeromonas salmonicida),* which was higher for individuals carrying a resistance conferring variant in a homozygous state. Similarly, Eizaguirre et al. (2009) showed in three-spined sticklebacks (*Gasterosteus*

aculeatus) that individuals with a particular MH II haplotype had significantly lower burden of the most common parasite (*Gyrodactylus sp.*). In such situations, when there are specific 'good genes' referring highest resistance to the most prevalent pathogens mate choice is expected to be assortative to combine currently advantageous genes (Milinski 2006). Indeed, Eizaguirre et al. (2009) found that fish of both sexes carrying the haplotype that conferred highest resistance towards *Gyrodactylus sp.* had significantly higher likelihood of reproducing, leading to a pattern of assortative mating. Support for a frequency dependent mechanism in this case was given by a study that investigated the same population in the following generation, where the frequency of the previously selected haplotype had strongly increased and the infection intensities of *Gyrodactylus sp.* was strongly reduced (Lenz et al. 2009). At the same time, a positive association of resistance with the previously beneficial haplotype towards *Gyrodactylus sp*. was no longer present and mate choice for intermediate MH diversity was observed.

High inbreeding coefficients for the MH class II, but not for neutral markers were found for some populations in this study (chapter 4). One explanation for this pattern would be that periods of directional selection and potentially assortative mating under such conditions, affect immune genetic variation. However, evidence for balancing selection acting as a dominant evolutionary force on the grayling MH was given both historically (chapter 2), as well as more recently (chapter 4). This suggests, that directional selection is episodic and in a context of either frequency-dependent selection or heterogeneous selection in time. Highest inbreeding coefficients (F_{IS}) at the MH II were found for a population, for which evidence exists that the availability of spawning grounds is reduced (Lewis 2006). This possibly increases competition for mating opportunities, which is expected to increase the strength of sexual selection (Milinski 2006; Eizaguirre et al. 2009). This highlights the importance of the maintenance and restoration of suitable spawning grounds to preserve productivity and adaptive genetic variation.

Another important component of genetic diversity is distinctiveness across populations (Olden et al. 2004). In this study, measurements of population differentiation were correlated between the MH II and microsatellites and population clustering based on genetic distance followed mainly the same pattern for both marker types (chapter 4). Pfrender et al. (2000) found, similarly to this study, no correlations for diversity measurements, but consistent divergence between quantitative trait loci (QTLs) and neutral markers and the authors conclude that neutral markers might be generally suitable to assess population distinctiveness. However, a number of studies find that adaptive divergence is not reflected by neutral markers (McKay and Latta 2002; Gomez-Mestre and Tejedo 2004; Ballentine and Greenberg 2010). The MHC is, however, not necessarily an ideal marker to resolve local adaptation, because balancing selection is expected to maintain higher variation than for neutral markers, when selection pressures are similar (Fraser and Neff 2010; O'Farrell et al. 2012).

Evidence for local adaption at small scales has been given in grayling for a young meta-population in Norway, where population divergence in response to pronounced differences in temperature has been observed within one lake system (Kavanagh et al. 2010; Junge et al. 2011). The study of Junge et al. (2011) suggests adaptive phenotypic divergence between populations inhabiting different temperature regimes, which is not reflected by neutral markers. Kavanagh et al. (2010) give evidence for a genetic basis for the observed phenotypic divergence by comparing differences in developmental traits of offspring from adults inhabiting either warmer or colder habitats under the same conditions. This divergence has arisen within 9-12 generations and on a spatial scale of about 10 km distance, under constraining conditions of continued gene flow, previous bottlenecks due to founder effects and small population sizes. These findings on local adaptations arising quickly and on small spatial scales has been documented for other salmonid species as well (Fraser et al. 2011). This highlights the importance of integrating assessment of adaptive differentiation for the definition of evolutionary significant units for management (Crandall et al. 2000; Garcia de Leaniz et al. 2007; Gebremedhin et al. 2009). New genotyping-by-sequencing (GBS) techniques, such as restriction-site-associated DNA sequencing (RADseq) or transcriptome sequencing are promising tools to identify genetic regions potentially involved in local adaptation, e.g. by the identification of F_{st} outliers, in non-model species (Davey et al. 2011;

Hoban et al. 2016). However, to decipher the genetic basis for evolutionary change, that is driven by natural selection, different approaches, like quantitative trait loci (QTL) mapping, controlled crosses, or genome-wide-association studies (GWAS) need to be combined to link genetic variation to phenotypic traits and fitness (Stinchcombe and Hoekstra 2007; Bernatchez et al. 2010; Narum et al. 2013; Vitti et al. 2013). Together these approaches have the potential to advance our understanding of the genetics of local adaptation in the future, which will facilitate the assignation of evolutionary significant units for management (Funk et al. 2012).

Another objective of this PhD study was to evaluate the effect of the supplementation of populations with hatchery-reared fish (stocking) on immune genetic variation in grayling. Whilst neutral genetic marker variation did not show significant differences between purely native and stocked populations (Dawnay et al. 2011), there are a number of factors that support the hypothesis that adaptive genetic variation can be adversely affected (Fraser 2008; Araki and Schmid 2010). The early life stages experience highest rates of mortality in salmonids in the wild, which makes selection most likely to have a strong effect (Einum and Fleming 2000; Evans et al. 2010; de Eyto et al. 2011). Therefore, the relaxation of selection during these critical early life stages could significantly impact on adaptive evolutionary processes (Araki et al. 2007, 2009). Also, natural mate choice, which is considered an important mechanism of selection, is not allowed for within hatchery breeding programmes (Quader 2005; Consuegra and Leaniz 2008). Additionally, individuals used to produce the hatchery offspring have an increased reproductive success compared to wild reproducing individuals, where offspring mortality at young life stages is high. Overrepresentation of hatchery produced offspring in the supplemented population can lead to increased variation in family sizes, which reduces the effective population size N_e (Ryman and Laikre 1991). Because the effective population size directly relates to the rate of inbreeding (Crow and Kimura 1970), a supplementation programme can increase the loss of genetic diversity, though overall productivity is increased (Ryman and Laikre 1991). In addition, at functional genes, the efficiency of selection to prevent the loss of beneficial mutations and eliminate deleterious mutations is positively related to the effective population size (Wright 1931), and is thus negatively affected by a decrease in N_e. Empirically it has been shown that supplementation programmes do not always decrease N_e and increase the rate of

inbreeding however (Hedrick et al. 2000; Duchesne and Bernatchez 2002). Increasing the number of individuals used to produce the hatchery offspring and equalizing contributions to the offspring generation among them has been shown to reduce these effects (Hedrick et al. 2000; Duchesne and Bernatchez 2002).

Here, higher inbreeding among supplemented (stocked) populations was not found, indicating no strong effect of the limited number of individuals used to create hatchery stocks on genetic variation within the supplemented population or a pronounced reduction of the effective population size (chapter 4). However, expected heterozygosity at the MH II was significantly lower in stocked than purely native populations (chapter 4). Moreover, whilst purely native populations showed significantly lower differentiation at the MH II than observed at neutral markers, indicating that balancing selection is acting to retain variation at the MH, this was not the case for stocked populations (chapter 4). These results suggest that the efficiency of selection is impaired in stocked populations compared to purely native populations. Because no data were available on genetic variation within the stocked populations prior to starting the supplementation programme, one might speculate that those populations selected for management, showed recent population declines and a predominant role of drift leading to the observed pattern of reduced efficiency of selection. But, grayling populations selected for stocking did not show evidence for population bottlenecks based on microsatellites more frequently than native populations, or exhibit lower effective population size (Dawnay et al. 2011), which does not support the hypothesis that management was primarily in response of recent population decline. There was no bias within the selection of populations included in this study either, which incorporated two native and one stocked population that showed evidence for a past bottleneck (Dawnay et al. 2011, chapter 4). Considering a direct effect of the stocking process on the efficiency of selection to act within the supplemented population seems therefore a valid conclusion. This highlights the importance of assessing genetic variation at genes under selection to effectively evaluate the suitability of stocking in sustaining long-term population viability (Fraser 2008).

In order to inform management, the objective of hatchery releases needs to be clearly formulated. If the only purpose of supportive breeding is to conserve populations at risk, it

would be recommended only in cases of populations with critical low population sizes and high likelihood of extinction due to environmental stochasticity, given the considerable risk of harmful effects (Fleming and Petersson 2001; Araki and Schmid 2010). Particularly in the light of climate change, the findings on the impairment of the efficiency of natural selection through the supplementation of populations with hatchery reared fish in grayling (chapter 4) indicate that careful considerations on the necessity of stocking to prevent population extinction are demanded. This is because under the conditions of climate change population persistence may depend on their evolutionary potential (Hoffmann and Sgrò 2011). Climate change is expected to pressure species in various ways. On one hand there are direct challenges like thermal stress and increased climatic fluctuations. Evidence exists that salmonids may have the scope for adaptive genetic responses to cope with this challenge, at least to some degree (Jensen et al. 2008; Eliason et al. 2011). Another threat imposed by climate change is, that existing environmental stressors might be exacerbated by cumulative or synergistic effects (Brook et al. 2008), although a large proportion of combined effects have been suggested to be of an antagonistic nature in freshwater ecosystems (Jackson et al. 2016). However, eutrophication is an example of an environmental stressor, which is expected to be reinforced by climate change (Moss et al. 2011) and this study highlighted that under conditions of climate change the reduction of eutrophication enhances habitat suitability for grayling by up to 31% and is thus suggested to be an important target for climate change mitigation (chapter 6). Another important risk of climate change is that transmission rates, prevalence and severity of infectious disease are expected to increase under conditions of rising temperatures (Marcogliese 2008; Altizer et al. 2013). Hari et al. (2006) for example document, that for brown trout (*Salmo trutta*) populations that experience higher water temperatures in the context of climate change, declines are strongly driven by increased incidences of Proliferative Kidney Disease (PKD). This highlights the importance of standing immune genetic variation to increase the chances of a species to cope with more severe pathogen pressures under climate change conditions (Dionne et al. 2009). As this study showed that stocking (chapter 4) was associated with reduced immune genetic variation, the clear recommendation to best invest management resources for the conservation of grayling is to improve habitat quality, as this could be shown to improve the probability of population persistence under conditions of climate change (chapter 6).

Explicit suggestions of local priorities under conditions of climate change are given in chapter 6.

However, where the supplementation of populations is required, either because of the high risk of extinction of the population to be supplemented or because social demands on management priorities are to compensate losses for fisheries, various improvements to current breeding programmes could potentially reduce negative effects and merit further investigation in grayling. For example, allowing for natural mate choice within semi-natural spawning channels could significantly enhance the possibilities for sexual selection and offspring fitness (Wedekind et al. 2001; Pitcher and Neff 2007; Consuegra and Leaniz 2008). Further, maintaining rearing conditions more closely to that of the natural environment, for example by using untreated river water may reduce the disruption of natural selection as it increases exposure to natural microbes, leading to differences in survival based on immune competence (Evans et al. 2010). However, the risk of epidemic outbreaks is increased, particularly if the hatchery stock is kept at high density (Amos and Thomas 2002). Careful consideration of the stocking efforts within one river in relation to the carrying capacity of the environment and the natural production of the existing grayling population are warranted as well, to reduce competition and density-dependent effects on the population (Einum and Fleming 2001; Keeley 2001).

Altogether, this study makes important contributions to inform management of European grayling to support long-term viable populations. Conservation research approaches used here may be applied to other endangered species. The importance of including adaptive markers in conservation genetic studies has clearly been shown. The MHC is a useful candidate to study evolutionary processes in comparison to demographic forces, because the effect of selection can be strong with quick evolutionary dynamics between hosts and pathogens. Future research, that includes the assessment of the microbiome in conjunction with the MHC could be fruitful in providing a more predictive framework, based on the investigation of immune genetic variation in relation to microbial selection pressures and environmental conditions (Amato 2013). Predicting the potential for future adaption is of utmost concern for conservation genetic research (Allendorf 2017). This requires a holistic approach that integrates demographic forces, environmental conditions and selection

pressures. In this respect it would be important to examine immune genetic variation and its implication on evolutionary potential in the broader context of the whole genome (Pearse 2016). In summary, genetic research and environmental modelling methods are clearly able to make important contributions to the conservation of biodiversity, but successful implementations ultimately depend on the priorities of the society and decision making parties.

Appendices

Appendix 1 Analysis scripts

Here, programming scripts used to perform the analysis of Illumina NGS data in Python and R are provided. The main purpose is to document the main scripts, whilst for the future the provision of more flexible tools available online for other users, would be suggested.

Appendix 1.1 Main Python-scripts for Sommer pipeline: pre-processing scripts are not provided here; comments are in red;

```
from __future__ import division
from Bio import SeqIO
from Bio.SeqRecord import SeqRecord
from Bio.Seq import Seq
from Bio import motifs
import pickle
import xlwt
import xlrd
from xlrd import open_workbook
from collections import Counter
import operator
import matplotlib.pyplot as plt
```

```
def compare_seqs(alleles,selection):
#input: all sequences within amplicon (alleles (IDs)), selection (indexed sequences)
#generates for each pair of sequences all possible chimer combinations and searches if they are a sequence within the amplicon
#returns in dictionary 'found' all potential chimers as keys with the parents as another dictionary containing entries of the number of possibilities to form chimer
    found = {}
    reps_dic ={}
    id_list = []
    seqlist = []
    for q in range(len(alleles)):                                   #for every variant in amplicon
        if str(alleles[q]) in selection.keys():
            reps_dic[str(alleles[q])]= selection[str(alleles[q])].seq   #get sequence as entry in reps_dic which has IDs as key
        else:
            print alleles[q], ' not found'
    for v in reps_dic.keys():                                       #create two lists
       seqlist.append(str(reps_dic[v]))                             #one with sequence
       id_list.append(v)                                            #one with sequence IDs in the same order
    index1 = 0
    for g in range(len(seqlist)):                                   #every sequence in list (index1) is combined with every other sequence (index)
        index2 = 0
        for y in range(len(seqlist)-index2):
            #print id_list[index1],id_list[index2]                  #to see all combinations searched
            if len(seqlist[index1])>= len(seqlist[index2]):         #determine length of longer sequence
                limit = len(seqlist[index2])
            else:
                limit = len(seqlist[index1])
            for i in range(limit-1):
                if seqlist[index1][i] != seqlist[index2][i]:        #starting from the first mismatch
                    COUNT = 0
                    for a in range(i+1,len(seqlist[0])):
                       pot_chim = str(seqlist[index1][:a])+str(seqlist[index2][a:]) #potential chim is defined as the combination of the two seqs at every position a
                       if str(pot_chim) in seqlist and str(pot_chim)!= seqlist[index2][i] and str(pot_chim)!= seqlist[index1][i]:
                          COUNT += 1
                          for x in reps_dic.keys():
                             if str(pot_chim) == str(reps_dic[x]):
                                Q = x
                                parents = str(id_list[index1])+' ' +str(id_list[index2])
                                if Q not in found.keys():
                                    found[Q] = {}
                                    found[Q][str(parents)] = 0
                                else:
                                    if str(parents) in found[Q].keys():
                                        found[Q][str(parents)]+= 1
                                    else:
                                        found[Q][str(parents)] = 0
                          break
                index2 += 1
            index1 += 1
        return found
```

```
def filter_chims(found_chims,alleles):
    chimeric_ids = []
    allele_count = Counter(alleles)
    g = allele_count.most_common(1)
    for a in found_chims.keys():
        for i in found_chims[a].keys():
            X = i.split(' ')
            if allele_count[a]>= allele_count[X[0]] or allele_count[a]>= allele_count[X[1]] or a==X[0] or a==X[1]:   #filter chimers of itself, and those with higher
                del found_chims[a][i]                                                                                #frequency than parents
            else:
                chimeric_ids.append(a)
    chim_ids = set(chimeric_ids)
    return found_chims,list(chim_ids)
```

```
def find_difference(alleles,most_common,infile2,T):
#input: all varints of amplicon (alleles), indexed sequence records (infile2),
#Threshold (T) for maximum number of differences for low difference category(here max_two as T =2)
#compares all sequence variants within amplicon to the three most common sequences
#returns two lists, max_two and more_two, each variant is asssigned to based on number of differences to most common sequences
        max_two_dif = []
        more_dif = []
        for q in alleles:           #compare each variant in amplicon
            if q in infile2.keys():
                i = infile2[q]      #get sequence record of variant
            else:
                pass
            for e in most_common:       #to three most common
                if q == e:              #no comparison to self
                    pass
                else:
                    z = 0               #count sequence differences as z
                    if e in infile2.keys():
                        e = infile2[e]
                        if len(i) != len(e):        #when length difference
                                limit = max([len(i),len(e)])    #longer defines length limit for comparison
                                z = abs(len(i)-len(e))          #difference in length counts as sequence difference counted in z
                        else:
                            limit = len(i)
                        for x in range(0,limit-1):              #across whole sequence length per base sequence comparison
                            if str(i[x])!= str(e[x]):           #if different count in z
                                z +=1
                        if z <=T:                               #variants with a maximum T differnces assigned max_two
                            max_two_dif.append(q)
                        else:
                            more_dif.append(q)                  #variants with a more than T differnces assigned more_two
                    else:
                        'not found in trim_sel keys: ',e        #if a sequence record is not found show error message
        return max_two_dif,more_dif


def define_cluster(dic_per_indiv,infile2,T):
#input sequences per individual/amplicon (dic_per_indiv), indexed sequences (infile2),
#threshold (T) for maximum number of differences for low difference category (here max_two as T =2)
#sorts all variants per amplicon in the following bins: 'chims' (chimeric sequence of two more common variants), 'putative alleles' (most common sequence),
#'more_two' (maximum of two differences between one of three most common sequences, 'more two' (more than two bp different to three most common)
#returns dictionary of indivs containing dictionary of bins
        variants = {}
        putative_allele_2 = []
        putative_allele_3 = []
        most_common = []
        all_variants = []
        all_put_als = []
        chim_set_joined = []
        chim_dic = {}
        for i in dic_per_indiv.keys():
                all_variants2 = []
                #print i
                #alleles = set(dic_per_indiv[i]).difference(set(private_alleles))
                variants[i] = {}
                alleles = dic_per_indiv[i]
                found_chims = compare_seqs(list(set(alleles)),infile2)
                chims = filter_chims(found_chims,alleles)
                chim_set = list(set(chims[1]))
                chim_d = chims[0]
                chim_dic[i] = chim_d
                chim_set_joined = chim_set_joined + chim_set
                put_al =[]
                for x in dic_per_indiv[i]:
                    if x not in chim_set:
                        put_al.append(x)
                    all_variants2.append(x)
                all_variants = all_variants + list(set(all_variants2))      #needs to be set to just get it one of each indiv
                count = Counter(put_al)
                sorted_count = sorted(count.items(),key=operator.itemgetter(1),reverse=True)
                c = 0
                for a in sorted_count:
                   if c == 0:
                       variants[i]['putative_alleles']=[]
                       variants[i]['putative_alleles'].append(a[0])
                       all_put_als.append(a[0])
                       C = put_al.count(a[0])
                       for n in range(C):
                           put_al.remove(a[0])
                       most_common.append(a[0])                     #compare all seqs to three most common
                   if c <= 3:
                       most_common.append(a[0])
                   c +=1
                clusters = find_difference(list(set(put_al)),most_common,infile2,T)
                variants[i]['chims']= chim_set
                variants[i]['max_two']= list(set(clusters[0]))
                variants[i]['more_two']= list(set(clusters[1]))
        return variants,all_variants,list(set(all_put_als)),list(set(chim_set_joined)),chim_dic
```

```
def sommer_stepII(variants_dic,variant_dic2,all_alleles,dic_per_indiv1,dic_per_indiv2,Dic_per_indiv1,Dic_per_indiv2,all_put_als,all_chims,T):
#input: variant dictionaries, list of all alleles with one count per individual where it was observed,dictionaries with all amplicon reads for each individual,
#with and without singletons; list of all variants classified putative allele so far, list of all chimers with one count per individual where it was observed,
#read depth threshold
#compares amplicon reads between replicates or against all other amplicons if no rep. available (Sommer step II)
#output to categories ti continue with Sommer step III
    C = 0
    C2 =0
    for i in variant_dic2.keys():
        f = 'x'                                                                     #flag to indicate not replicated sample
        i = str(i)
        if len(dic_per_indiv1[i]) < T and len(dic_per_indiv2[i]) < T:               #if both samples below threshold coverage T
            pass                                                                    #don't consider
        elif len(dic_per_indiv1[i]) < T and len(dic_per_indiv2[i]) >= T:            #if only one over T compare to all other variants
            C2 +=1
            not_in_repl = []
            ch = []
            for variant in Dic_per_indiv2[i]:
                if all_alleles.count(variant)==1:                                   #keep only if present in other indiv(count all_alleles min =2)
                    not_in_repl.append(variant)
                if all_chims.count(variant)>1:                                      #keep only if present in other indiv(count all_alleles min =2)
                    ch.append(variant)
        elif len(dic_per_indiv2[i]) < T and len(dic_per_indiv1[i]) >= T:
            C2 +=1
            not_in_repl = []
            ch = []
            for variant in Dic_per_indiv[i]:
                if all_alleles.count(variant)==1:                                   #keep only if present in other indiv(count all_alleles min =2)
                    not_in_repl.append(variant)
                if all_chims.count(variant)>1:                                      #keep only if present in other indiv(count all_alleles min =2)
                    ch.append(variant)
        else:                                                                       #else replicated sample
            f = 'X'                                                                 #change flag to indicate replicated sample
            C +=1
            not_in_repl = list(set(Dic_per_indiv[i]).symmetric_difference(set(Dic_per_indiv2[i])))  #symmetric difference between samples represents reads
            ch = []                                                                                 #found in only one rep
        put_not_in_rep = list(set(all_put_als).intersection(not_in_repl))           #distinguish between not replicated reads that have been scored as putative
        put_not = list(set(not_in_repl).difference(set(put_not_in_rep)))            #allele in other indiv. and those that havn't
        variants_dic[i]['unclassified']= list(set(variants_dic[i]['max_two']).intersection(put_not_in_rep)) #modification: if variant category max two not in rep.,
        variant_dic2[i]['unclassified']= list(set(variant_dic2[i]['max_two']).intersection(put_not_in_rep)) #but put.allele in other indiv. then unclassified
        variants_dic[i]['artefacts']= list(set(variants_dic[i]['max_two']).intersection(put_not))          #otherwise artefact
        variant_dic2[i]['artefacts'] = list(set(variant_dic2[i]['max_two']).intersection(put_not))
        m2 = list(set(variant_dic2[i]['max_two']).difference(set(not_in_repl)))     #if max two in rep. keep for frequency evaluation in next step
        m1 = list(set(variants_dic[i]['max_two']).difference(set(not_in_repl)))
        variants_dic[i]['max_two'] = m1
        variant_dic2[i]['max_two'] = m2
        maybe1 = list(set(variants_dic[i]['more_two']).intersection(not_in_repl))   #those that are not in replicate, further evaluated in maybe
        maybe2 = list(set(variant_dic2[i]['more_two']).intersection(not_in_repl))
        maybe = list(set(maybe1 +maybe2))
        keep2 = []
        for y in maybe:
            if all_alleles.count(y)>1:                                              #keep only if present in other indiv(count all_alleles min =2)
                keep2.append(y)
        s1 = list(set(variants_dic[i]['more_two']).intersection(set(keep2)))
        s2 = list(set(variant_dic2[i]['more_two']).intersection(set(keep2)))
        variant_dic2[i]['more_twoB']= s2                                            #more two not present in rep, but in other indiv, further evluated in next step
        variants_dic[i]['more_twoB']= s1

        n2 = list(set(variant_dic2[i]['more_two']).difference(set(not_in_repl)))
        n1 = list(set(variants_dic[i]['more_two']).difference(set(not_in_repl)))
        sum_more_two_not_kept2 =list(set(variant_dic2[i]['more_two']).difference(set(variant_dic2[i]['more_twoB']+n2)))
        sum_more_two_not_kept1 =list(set(variants_dic[i]['more_two']).difference(variants_dic[i]['more_twoB']+n1))
        variant_dic2[i]['more_two']= n2
        variants_dic[i]['more_two']= n1
        variant_dic2[i]['artefacts'].append(sum_more_two_not_kept2)
        variants_dic[i]['artefacts'].append(sum_more_two_not_kept1)
        rem = []
        rem2 = []
        for a in variants_dic[i]['chims']:
            if a in not_in_repl  or a in variant_dic2[i]['chims']:  #chimers not in replicate or called chim in replicate artefact
                if a not in all_put_als:                            #if not called as putative allele elsewhere so far (highest read number)
                    variants_dic[i]['artefacts'].append(a)          #artefact
                    rem.append(a)
                else:
                    variants_dic[i]['more_two'].append(a)
                    rem.append(a)
            elif f != 'X' and a in ch:                              #if not replicated sample comparison across all other amplicons
                if a not in all_put_als:
                    variants_dic[i]['artefacts'].append(a)
                    rem.append(a)
                else:
                    variants_dic[i]['more_two'].append(a)           #else appended to 'more_two' for frequency comparison
                    rem.append(a)
            else:
                    variants_dic[i]['more_two'].append(a)
                    rem.append(a)
        for a in variant_dic2[i]['chims']:
            if a in not_in_repl  or a in variants_dic[i]['chims']:
                if a not in all_put_als:
                    variant_dic2[i]['artefacts'].append(a)
                    rem2.append(a)
                else:
                    variant_dic2[i]['more_two'].append(a)           #else appended to 'more_two' for frequency comparison
```

```
            if a in not_in_repl  or a in variants_dic[i]['chims']:
                if a not in all_put_als:
                    variant_dic2[i]['artefacts'].append(a)
                    rem2.append(a)
                else:
                    variant_dic2[i]['more_two'].append(a)           #else appended to 'more_two' for frequency comparison
                    rem2.append(a)
            elif f != 'X' and a in ch:
                if a not in all_put_als:
                    variant_dic2[i]['artefacts'].append(a)
                    rem2.append(a)
                else:
                    variant_dic2[i]['more_two'].append(a)
                    rem2.append(a)
            else:
                    variant_dic2[i]['more_two'].append(a)
                    rem2.append(a)
        if 'chimsB' not in variants_dic[i].keys():                  #all sequences called as chimers recorded in ChimsB
            variants_dic[i]['chimsB']=[]
        for el in rem:
            variants_dic[i]['chimsB'].append(el)
            variants_dic[i]['chims'].remove(el)
        if 'chimsB' not in variant_dic2[i].keys():
            variant_dic2[i]['chimsB']=[]
        for el2 in rem2:
            variant_dic2[i]['chimsB'].append(el2)
            variant_dic2[i]['chims'].remove(el2)
    print 'comparisons made between replicates: ',C
    print 'single samples above T: ',C2
    return variants_dic,variant_dic2




def artefact_freqs(variants_sorted,dic_per_indiv,T1):
    #input: classifications per indiv. from sommer stepII, all reads per indiv, threshold T1
    #for variants in more_twoB (not present in rep., but other indiv.): if putative allele or unclassified in other indiv. then unclassified otherwise artefact
    #calculates frequency of all artefacts within amplicons, holds for each amplicon highest artefact freq. in 'art_freq' category
    #returns new classified dic, updated for more_twoB classificatiions and maximum artefact frequecy across all amplicons
    for q in variants_sorted.keys():
        if 'more_twoB' in variants_sorted[q].keys():
            if len(variants_sorted[q]['more_twoB'])==0:
                z = 1
            for x in variants_sorted[q]['more_twoB']:
                z =0
                g = 0
                for p in dic_per_indiv.keys():
                    if p != q  and len(dic_per_indiv[p])>=20:
                        if x in dic_per_indiv[p]:
                            z = 1
                            #print x,' found in other'
                            if x in variants_sorted[p]['more_twoB'] or x in variants_sorted[p]['max_two'] or x in variants_sorted[p]['putative_alleles']:
                                g = 1
                            else:
                                pass
                if g ==0:
                                if 'artefacts' in variants_sorted[q].keys():
                                    variants_sorted[q]['artefacts'].append(x)
                                else:
                                    variants_sorted[q]['artefacts'] = []
                                    variants_sorted[q]['artefacts'].append(x)
                else:
                                if 'unclassified' in variants_sorted[q].keys():
                                    variants_sorted[q]['unclassified'].append(x)
                                else:
                                    variants_sorted[q]['unclassified'] = []
                                    variants_sorted[q]['unclassified'].append(x)
                                        _
                if z != 1:
                    if 'artefacts' in variants_sorted[q].keys():
                                        variants_sorted[q]['artefacts'].append(x)
                    else:
                                        variants_sorted[q]['artefacts'] = []
                                        variants_sorted[q]['artefacts'].append(x)
    for i in variants_sorted.keys():
        variants_sorted[i]['art_freq']=0
        if len(dic_per_indiv[i]) > T1:
            if 'artefacts' in variants_sorted[i].keys():
                for item in variants_sorted[i]['artefacts']:                        #dissolve list items in list
                        if type(item) == list:
                            for element in item:
                                variants_sorted[i]['artefacts'].append(element)
                            variants_sorted[i]['artefacts'].remove(item)
                if len(variants_sorted[i]['artefacts']) != 0:
                    most_common =0
                    for art in variants_sorted[i]['artefacts']:
                        if dic_per_indiv[i].count(art)>most_common:
                            most_common = dic_per_indiv[i].count(art)
                    freq = most_common/len(dic_per_indiv[i])
                    variants_sorted[i]['art_freq'] = freq
    frequencies = []
    for f in variants_sorted.keys():
        frequencies.append(variants_sorted[f]['art_freq'])
        Max = max(frequencies)
    return variants_sorted,Max
```

```
def compare_freqs(variants_art_freq,all_alleles,all_put_als,dic_per_indiv,Max):
    #input: variant dic from previous step, list of all variants(count equivalent to number of indivs in which present, list of all putative alleles,highest art.freq.
    #compares frequencies against highest artefact frequency, Sommer step III
    all_put_als2 = all_put_als[:]
    for i in variants_art_freq.keys():
        if 'putative_alleles' not in variants_art_freq[i].keys():
            variants_art_freq[i]['putative_alleles'] = []
        if 'unclassified' not in variants_art_freq[i].keys():
            variants_art_freq[i]['unclassified'] = []
        length = len(dic_per_indiv[i])
        if length >= 20:
            count1 = Counter(variants_art_freq[i]['max_two'])
            for e in count1.keys():
                if (dic_per_indiv[i].count(e)/length)> Max:
                    variants_art_freq[i]['putative_alleles'].append(e)
                    all_put_als2.append(e)
                else:
                    variants_art_freq[i]['unclassified'].append(e)
                variants_art_freq[i]['max_two'].remove(e)
            count2 = Counter(variants_art_freq[i]['more_two'])
            for o in count2.keys():
                if (dic_per_indiv[i].count(o)/length)> Max:
                    variants_art_freq[i]['putative_alleles'].append(o)
                    all_put_als2.append(o)
                else:
                    variants_art_freq[i]['unclassified'].append(o)
                variants_art_freq[i]['more_two'].remove(o)
            rem = []
            flag = 0
            for u in variants_art_freq[i]['chims']:
                if all_alleles.count(u)>1:                                  #if chim is present in other individual
                    if u in all_put_als:                                        #as a puative allele
                        #print 'put ',u
                        variants_art_freq[i]['unclassified'].append(u)              #here unclassified as well
                        flag = 1
                        if 'chimsB' in variants_art_freq[i].keys():
                            variants_art_freq[i]['chimsB'].append(u)
                        else:
                            variants_art_freq[i]['chimsB'] = []
                            flag = 1
                            variants_art_freq[i]['chimsB'].append(u)
                        rem.append(u)
                        #all_put_als.append(u)
                    else:
                        variants_art_freq[i]['artefacts'].append(u)
                        flag = 1
                        if 'chimsB' in variants_art_freq[i].keys():
                            variants_art_freq[i]['chimsB'].append(u)
                        else:
                            variants_art_freq[i]['chimsB'] = []
                            flag = 1
                            variants_art_freq[i]['chimsB'].append(u)
                        rem.append(u)
                else:
                    variants_art_freq[i]['artefacts'].append(u)
                    flag = 1
                    variants_art_freq[i]['chimsB'].append(u)
                    rem.append(u)
                if flag == 0:
                    print i,u
            for el in rem:
                    variants_art_freq[i]['chims'].remove(el)
    return variants_art_freq,list(set(all_put_als2))
```

```
def write_variants(rep_dic,fastq2):
    #write fasta file of putative alleles
    alleles = []
    for i in rep_dic:
        if i in fastq2.keys():
            alleles.append(fastq2[i])
    name = path + locus + '_putative_alleles.fasta'
    SeqIO.write(alleles,name,"fasta")
```

```
def compare_freqs_eff_cor(variants_art_freq,all_alleles,all_put_als,dic_per_indiv,Max,all_eff):
    #compares variant frequencies to highest art. freq., but takes low allele efficiencies into account
    all_put_als2 = all_put_als[:]
    for i in variants_art_freq.keys():
        if 'putative_alleles' not in variants_art_freq[i].keys():
            variants_art_freq[i]['putative_alleles'] = []
        if 'unclassified' not in variants_art_freq[i].keys():
            variants_art_freq[i]['unclassified'] = []
        length = len(dic_per_indiv[i])
        if length > 0:
            count1 = Counter(variants_art_freq[i]['max_two'])
            for e in count1.keys():
                if e in all_eff.keys():
                    cor = dic_per_indiv[i].count(e)/all_eff[e]
                    dif = cor-dic_per_indiv[i].count(e)
                    if cor/(length +dif)> Max:
                        variants_art_freq[i]['putative_alleles'].append(e)
                        all_put_als2.append(e)
                    else:
                        variants_art_freq[i]['unclassified'].append(e)
                        variants_art_freq[i]['max_two'].remove(e)
                elif dic_per_indiv[i].count(e)/len(dic_per_indiv[i])>Max:
                    variants_art_freq[i]['putative_alleles'].append(e)
                    all_put_als2.append(e)
                    variants_art_freq[i]['max_two'].remove(e)
                else:
                        variants_art_freq[i]['unclassified'].append(e)
                        variants_art_freq[i]['max_two'].remove(e)
            count2 = Counter(variants_art_freq[i]['more_two'])


          for o in count2.keys():
              if o in all_eff.keys():
                  cor = dic_per_indiv[i].count(o)/all_eff[o]
                  dif = cor-dic_per_indiv[i].count(o)
                  if cor/(length+dif) > Max:
                      variants_art_freq[i]['putative_alleles'].append(o)
                      all_put_als2.append(o)
                  else:
                      variants_art_freq[i]['unclassified'].append(o)
                      variants_art_freq[i]['more_two'].remove(o)
              elif dic_per_indiv[i].count(o)/len(dic_per_indiv[i])>Max:
                  variants_art_freq[i]['putative_alleles'].append(o)
                  all_put_als2.append(o)
              else:
                      variants_art_freq[i]['unclassified'].append(o)
                      variants_art_freq[i]['more_two'].remove(o)
          rem = []
          flag = 0
          for u in variants_art_freq[i]['chims']:
              if all_alleles.count(u)>1:                                  #if chim is present in other individual
                  #print 'chim in other indiv',u
                  if u in all_put_als:                                        #as a puative allele
                      #print 'put ',u
                      variants_art_freq[i]['unclassified'].append(u)              #here unclassified as well
                      flag = 1
                      if 'chimsB' in variants_art_freq[i].keys():
                          variants_art_freq[i]['chimsB'].append(u)
                      else:
                          variants_art_freq[i]['chimsB'] = []
                          flag = 1
                          variants_art_freq[i]['chimsB'].append(u)
                      rem.append(u)
                      #all_put_als.append(u)
                  else:
                      variants_art_freq[i]['artefacts'].append(u)
                      flag = 1
                      if 'chimsB' in variants_art_freq[i].keys():

                         variants_art_freq[i]['chimsB'].append(u)
                    else:
                         variants_art_freq[i]['chimsB'] = []
                         flag = 1
                         variants_art_freq[i]['chimsB'].append(u)
                    rem.append(u)
                    #all_put_als.append(u)
              else:
                  variants_art_freq[i]['artefacts'].append(u)
                  flag = 1
                  if 'chimsB' in variants_art_freq[i].keys():
                      variants_art_freq[i]['chimsB'].append(u)
                  else:
                      variants_art_freq[i]['chimsB'] = []
                      flag = 1
                      variants_art_freq[i]['chimsB'].append(u)
                  rem.append(u)
           else:
              variants_art_freq[i]['artefacts'].append(u)
              flag = 1
              variants_art_freq[i]['chimsB'].append(u)
              rem.append(u)
           if flag == 0:
              print (i,u)
        for el in rem:
              variants_art_freq[i]['chims'].remove(el)
return variants_art_freq,list(set(all_put_als2))
```

Appendix 1.2 R-scripts

1.2.1 R-script to calculate the position of the maximum degree of change (DOC) across all variants in each amplicon in the dataset (following Lighten et al. 2014b); comments are in green;

```
DA.indivs <-read.table(PATH,row.names=1)
a<-c(1,2,3,4,5,6,7,8,9,10)
colnames(DA.indivs)<-a


ROCs_per_indiv<-function(x){
  #calculates the difference in cumulative read depth between all variants for each Amplicon
  ROCs<-c()
  for(i in 2:10){
    ROC<-as.numeric(x[i]-x[i-1])
  ROCs<-c(ROCs,ROC)}
  return(ROCs)}


DOCs_per_indiv<-function(x){
  #calculates degree of change (DOC) for every variant to the next variant for each amplicon
  ROCs<-c()
  for(i in 1:9){
    ROC<-as.numeric(x[i])/as.numeric(x[i+1])
    ROCs<-c(ROCs,ROC)}
  return(ROCs)}

DOCs_max_indiv<-function(x){
  #returns variant position with highest degree of change
  return(which.max(x))}

ROCs<-apply(DA.indivs,1,function(x) ROCs_per_indiv(x))
ROC<-t(ROCs) #transpose columns and rows
R2<-as.data.frame(ROC)
DOCs<-apply(R2,1,function(x) DOCs_per_indiv(x))
DOC<-t(DOCs)
R<-as.data.frame(DOC)
a<-c(1,2,3,4,5,6,7,8,9)
colnames(R)<-a
M<-apply(R,1,function(x) DOCs_max_indiv(x))
A<-cbind(row.names(DA.indivs),M)
A<-as.data.frame(A)
A2 <- data.frame(lapply(A, as.character), stringsAsFactors=FALSE)
write.csv(A2,outPATH,col.names =TRUE)
```

1.2.2 R-script to estimate copy number scenario with highest fit (following Lighten et al. 2014b), example for one locus (equivalent for more loci)

```
DA.indivs <-read.table(PATH,row.names=1,)  #infile (txt) specify path

DA.matrix<-DA.indivs[,c(-1,-1)] #without row names

G<-nrow(DA.indivs)  #number of samples
#part 1 1locus: make table of sum of squares for each scenario and sample

modes <- c(2:10)
exp_freq <- function(x,D,A) (D/A)*x
matrix.allele_dist <- as.matrix(restrictedparts(2,10,include.zero=TRUE,decreasing=TRUE))#make all possible copy number distributions
DA.matrix<-DA.indivs[,c(-1,-1)]                                                          #for 2 loci across 10 reads



square_values<-function(x,y) {
  (x-y)^2}

per_indiv_Exp<-function(observed,matrix.allele_dist){
  allele_exp<-apply(matrix.allele_dist,(1:2),function(x) exp_freq(x,sum(observed),2))} # calculate expected frequencies for
                                                                                       # each copy number sceanrio
exp_values<-apply(DA.matrix,1,function(x) per_indiv_Exp(x,matrix.allele_dist))


vectors<-function(x,y){
  x<-as.integer(rep(x,2))
  y<-as.integer(y)
  result<-mapply(square_values,x,y,SIMPLIFY = 'array')
  return(result)}


squares<-function(x,y,z){
  data_result<-data.frame(matrix(ncol = 2, nrow = 0))     #create new dataframe for sum of square values
  for (i in 1:G){
    A<-z[i,1]   #sample name
    t<-vectors(x[i,],y[,i])     #DA.matrix in rows, exp_values in columns
    a<-sum(t[1:10])             #sum of squares scenario 1
    b<-sum(t[11:20])            #sum of squares scenario 2
    row<-c(a,b)
    data_result[toString(A),]<-row}
  return(data_result)}


SQ<-as.data.frame(squares(DA.matrix,exp_values,DA.indivs))
DA_data_result<-SQ
c<-c('1_0','1_1')
colnames(DA_data_result)<-c

#part2
#1lofus   # make not summed squares as vectors for each scenario and individual for testing difference in variance


modes <- c(2:10)
exp_freq <- function(x,D,A) (D/A)*x
matrix.allele_dist <- as.matrix(restrictedparts(2,10,include.zero=TRUE,decreasing=TRUE))
DA.matrix<-DA.indivs[,c(-1,-1)]


square_values<-function(x,y) {
  (x-y)^2}

per_indiv_Exp<-function(observed,matrix.allele_dist){
  allele_exp<-apply(matrix.allele_dist,(1:2),function(x) exp_freq(x,sum(observed),2))}

exp_values<-apply(DA.matrix,1,function(x) per_indiv_Exp(x,matrix.allele_dist))


vectors<-function(x,y){
  x<-as.integer(rep(x,2))
  y<-as.integer(y)
  result<-mapply(square_values,x,y,SIMPLIFY = 'array')
  return(result)}


squares<-function(x,y,z){
  data_result<-data.frame(matrix(ncol = 20, nrow = 0))
  for (i in 1:G){
    A<-z[i,1]
    t<-vectors(x[i,],y[,i])
    data_result[toString(A),]<-t}
  return(data_result)}
```

```
SQ<-as.data.frame(squares(DA.matrix,exp_values,DA.indivs))

min_val<-function(x1,y1){       #x contains sum of squares, y vectors of squares
  min<- x1[1]
  col<- 1
  l<-length(x1)
  for (i in 1:l){
    if(x1[i]<min){              #if the sum of squares of more complex scenario is lower than simpler model apply F-ratio test
      j<-i*10                   #for every sum of sqares vector of ten not summed squares
      J<-col*10                 #col saves current lowest sum, i is new lowest tested
      a<- c(as.integer(y1[(j-9):j]))
      b<-c(as.integer(y1[(J-9):J]))
      res= var.test(a,b,alternative='less')
      if(res$p.value<0.05){     #if test significant, i is new col, minimum is tested sum of squares
        min<-x1[i]
        col<- i}}}
  res =c(as.numeric(min),colnames(x1)[col])
  return(res) }

sum_indiv<-function(x,y){
  res<-data.frame(matrix(ncol = 2, nrow = 0))
  for (i in 1:G){
    r<-rownames(x)[i]
    res[toString(r),]<-min_val(x[i,],y[i,])}      #find for every individual the lowest sum of squares, that significantly reduces variance
  return(res)}
```

```
DA_res_cor<-as.data.frame(sum_indiv(DA_data_result,SQ))
write.csv(DA_res_cor,PATH)  #outfile specify path
```

Appendix 2 Ewen Watterson test results

Table A.2 *Ewen Watterson test results*

		Marker Microsatellites								MH	
population		**Ogo2**	**BFRO018**	**BFRO011**	**One2**	**BFRO010**	**BFRO006**	**BFRO012**	**BFRO016**	**DAA**	**DAB**
AIR	**N**	39	39	39	39	39	39	39	39	31	29
	(obs. F/ exp. F)-1	-0.34	-0.36	-0.37	-0.39	-0.14	-0.09	-0.24	-0.14	-0.28	-0.04
	p Slatkin	0.11	0.1	0.07	0.07	0.4	0.35	0.25	0.3	0.2	0.53
	p Watterson	0.11	0.07	0.07	0.09	0.28	0.35	0.28	0.3	0.15	0.51
CLD	**N**	64	64	64	64	64	64	64	64	40	37
	(obs. F/ exp. F)-1	-0.27	-0.1	-0.2	-0.18	-0.46	-0.27	-0.34	-0.39	-0.22	-0.08
	p Slatkin	0.19	0.47	0.23	0.33	0.02	0.19	0.11	0.01	0.24	0.44
	p Watterson	0.52	0.25	0.23	0.2	0.02	0.19	0.11	0.01	0.24	0.36
DBD	**N**	39	39	39	39	39	39	39	39	35	36
	(obs. F/ exp. F)-1	-0.35	0.1	-0.25	-0.38	-0.05	-0.28	-0.32	-0.36	-0.15	-0.12
	p Slatkin	0.08	0.71	0.2	0.07	0.45	0.15	0.1	0.06	0.38	0.4
	p Watterson	0.08	0.76	0.2	0.14	0.28	0.15	0.1	0.06	0.61	0.63
DEE	**N**	52	52	52	52	52	52	52	52	27	26
	(obs. F/ exp. F)-1	-0.34	-0.43	-0.36	-0.45	0.29	-0.45	-0.47	-0.45	-0.4	-0.31
	p Slatkin	0.11	0.05	0.1	0.03	0.8	0.02	0.02	0.02	0.03	0.09
	p Watterson	0.09	0.09	0.08	0.06	0.84	0.02	0.02	0.02	0.05	0.14
DOV	**N**	50	50	50	50	50	50	50	50	26	35
	(obs. F/ exp. F)-1	-0.09	0	-0.37	-0.39	0.04	-0.04	-0.1	-0.06	0	0.06
	p Slatkin	0.43	0.59	0.09	0.06	0.56	0.4	0.42	0.35	0.56	0.68
	p Watterson	0.46	0.62	0.27	0.08	0.33	0.4	0.43	0.35	0.62	0.42
EDN	**N**	45	45	45	45	45	45	45	45	33	36
	(obs. F/ exp. F)-1	-0.04	0	-0.34	-0.21	0.12	-0.33	-0.36	-0.37	-0.08	-0.04
	p Slatkin	0.46	0.51	0.1	0.3	0.62	0.13	0.08	0.04	0.45	0.51
	p Watterson	0.25	0.66	0.07	0.37	0.62	0.33	0.06	0.04	0.23	0.28
HAV	**N**	58	58	58	58	58	58	58	58	37	33

	(obs. F/ exp. F)-1	-0.18	-0.39	-0.27	-0.52	0.08	-0.37	-0.4	-0.27	-0.43	-0.3
	p Slatkin	0.32	0.07	0.19	0.01	0.52	0.05	0.06	0.17	0.04	0.14
	p Watterson	0.33	0.03	0.5	0	0.52	0.05	0.08	0.17	0.04	0.18
ITH	**N**	50	50	50	50	50	50	50	50	34	20
	(obs. F/ exp. F)-1	-0.24	-0.22	0.17	-0.28	-0.39	-0.11	-0.06	0.07	-0.29	-0.26
	p Slatkin	0.21	0.22	0.68	0.19	0.03	0.32	0.5	0.51	0.16	0.14
	p Watterson	0.21	0.22	0.65	0.1	0.03	0.32	0.17	0.51	0.16	0.14
SEV	**N**	39	39	39	39	39	39	39	39	31	30
	(obs. F/ exp. F)-1	-0.28	-0.15	-0.04	-0.12	NA	-0.07	-0.35	-0.37	-0.32	-0.28
	p Slatkin	0.18	0.39	0.54	0.43	N.A.	0.45	0.1	0.07	0.1	0.15
	p Watterson	0.18	0.21	0.79	0.29	N.A.	0.27	0.3	0.1	0.09	0.18
URE	**N**	58	58	58	58	58	58	58	58	31	30
	(obs. F/ exp. F)-1	-0.1	-0.43	0.02	-0.1	-0.32	0.42	0.03	0.33	-0.48	-0.13
	p Slatkin	0.42	0.05	0.58	0.48	0.13	0.97	0.61	0.86	0.01	0.43
	p Watterson	0.25	0.06	0.34	0.51	0.13	0.97	0.58	0.88	0.01	0.46
WLA	**N**	48	48	48	48	48	48	48	48	34	25
	(obs. F/ exp. F)-1	-0.26	-0.54	-0.37	0.21	-0.05	-0.35	-0.46	0.05	-0.49	-0.45
	p Slatkin	0.19	0	0.04	0.71	0.37	0.1	0.01	0.48	0	0.41
	p Watterson	0.19	0	0.04	0.55	0.37	0.1	0.02	0.48	0.01	0.53
WLB	**N**	51	51	51	51	51	51	51	51	N.A.	N.A.
	(obs. F/ exp. F)-1	-0.34	-0.38	-0.35	-0.22	NA	-0.37	-0.49	-0.3	N.A.	N.A.
	p Slatkin	0.11	0.09	0.09	0.28	N.A.	0.06	0.01	0.15	N.A.	N.A.
	p Watterson	0.11	0.06	0.09	0.37	N.A.	0.06	0.01	0.15	N.A.	N.A.
WYE	**N**	55	55	55	55	55	55	55	55	30	22
	(obs. F/ exp. F)-1	-0.11	-0.26	-0.15	0.31	NA	-0.21	-0.19	0.12	-0.33	-0.37
	p Slatkin	0.41	0.25	0.4	0.8	N.A.	0.28	0.32	0.57	0.11	0.05
	p Watterson	0.43	0.17	0.25	0.75	N.A.	0.15	0.37	0.57	0.27	0.05

Appendix 3 Supporting documents of chapter 5

Table A3.1 *Pearson correlation coefficients of Bioclimatic parameters across the European study area*

Layer	Bio 1	Bio 2	Bio 3	Bio 4	Bio 5	Bio 6	Bio 7	Bio 8	Bio 9	Bio 10	Bio 11	Bio 12	Bio 13	Bio 14	Bio 15	Bio 16	Bio 17	Bio 18	Bio 19
Bio 1	1.00	0.46	0.77	-0.59	0.76	0.92	-0.53	0.04	0.83	0.82	0.93	0.07	0.05	-0.03	0.00	0.06	0.04	-0.44	0.31
Bio 2	0.46	1.00	0.45	0.00	0.71	0.23	0.20	0.17	0.38	0.58	0.31	-0.29	-0.25	-0.32	0.16	-0.26	-0.29	-0.42	-0.14
Bio 3	0.77	0.45	1.00	-0.87	0.33	0.89	-0.77	-0.35	0.81	0.34	0.90	0.34	0.27	0.25	-0.08	0.29	0.30	-0.22	0.51
Bio 4	-0.59	0.00	-0.87	1.00	0.05	-0.86	0.97	0.52	-0.70	-0.02	-0.84	-0.52	-0.44	-0.43	0.13	-0.47	-0.48	0.01	-0.61
Bio 5	0.76	0.71	0.33	0.05	1.00	0.44	0.15	0.38	0.52	0.98	0.49	-0.35	-0.31	-0.43	0.15	-0.32	-0.38	-0.61	-0.10
Bio 6	0.92	0.23	0.89	-0.86	0.44	1.00	-0.82	-0.22	0.86	0.53	1.00	0.31	0.25	0.21	-0.09	0.27	0.27	-0.28	0.51
Bio 7	-0.53	0.20	-0.77	0.97	0.15	-0.82	1.00	0.49	-0.62	0.04	-0.78	-0.56	-0.47	-0.50	0.19	-0.50	-0.54	-0.08	-0.62
Bio 8	0.04	0.17	-0.35	0.52	0.38	-0.22	0.49	1.00	-0.36	0.40	-0.21	-0.52	-0.43	-0.40	0.11	-0.45	-0.44	0.01	-0.61
Bio 9	0.83	0.38	0.81	-0.70	0.52	0.86	-0.62	-0.36	1.00	0.55	0.88	0.25	0.20	0.09	-0.02	0.21	0.17	-0.48	0.55
Bio 10	0.82	0.58	0.34	-0.02	0.98	0.53	0.04	0.40	0.55	1.00	0.57	-0.29	-0.26	-0.36	0.11	-0.27	-0.32	-0.57	-0.05
Bio 11	0.93	0.31	0.90	-0.84	0.49	1.00	-0.78	-0.21	0.88	0.57	1.00	0.27	0.23	0.16	-0.05	0.24	0.23	-0.32	0.48
Bio 12	0.07	-0.29	0.34	-0.52	-0.35	0.31	-0.56	-0.52	0.25	-0.29	0.27	1.00	0.93	0.83	-0.20	0.94	0.87	0.58	0.88
Bio 13	0.05	-0.25	0.27	-0.44	-0.31	0.25	-0.47	-0.43	0.20	-0.26	0.23	0.93	1.00	0.64	0.11	0.99	0.68	0.55	0.82
Bio 14	-0.03	-0.32	0.25	-0.43	-0.43	0.21	-0.50	-0.40	0.09	-0.36	0.16	0.83	0.64	1.00	-0.61	0.66	0.98	0.68	0.62
Bio 15	0.00	0.16	-0.08	0.13	0.15	-0.09	0.19	0.11	-0.02	0.11	-0.05	-0.20	0.11	-0.61	1.00	0.10	-0.59	-0.20	-0.10
Bio 16	0.06	-0.26	0.29	-0.47	-0.32	0.27	-0.50	-0.45	0.21	-0.27	0.24	0.94	0.99	0.66	0.10	1.00	0.70	0.56	0.84
Bio 17	0.04	-0.29	0.30	-0.48	-0.38	0.27	-0.54	-0.44	0.17	-0.32	0.23	0.87	0.68	0.98	-0.59	0.70	1.00	0.63	0.69
Bio 18	-0.44	-0.42	-0.22	0.01	-0.61	-0.28	-0.08	0.01	-0.48	-0.57	-0.32	0.58	0.55	0.68	-0.20	0.56	0.63	1.00	0.15
Bio 19	0.31	-0.14	0.51	-0.61	-0.10	0.51	-0.62	-0.61	0.55	-0.05	0.48	0.88	0.82	0.62	-0.10	0.84	0.69	0.15	1.00

3.2 Pearson correlation coefficients of parameters across the UK study area

Legend for Table A3.2.1 *number code for each parameters used in table of Pearson correlation coefficients*

Parameter description	Parameter code	Parameter description	Parameter code	Parameter description	Parameter code
Autumn mean water temperature	1	Bio 16	11	Orthophosphate concentration	21
Bio 3	2	Bio 18	12	Ph	22
Bio 4	3	BOD	13	Phalacrocorax carbo density	23
Bio 5	4	Calcium concentration	14	Q10 of annual flow	24
Bio 6	5	Chloride concentration	15	Q95 of annual flow	25
Bio 8	6	Copper concentration	16	Spring mean water temperature	26
Bio 9	7	Annual average flow	17	Summer mean water temperature	27
Bio 11	8	Land Cover	18	Suspended solid concentration	28
Bio 14	9	Nitrate concentration	19	Total ammonium concentration	29
Bio 15	10	Nitrite concentration	20	Winter mean water temperature	30

Table 3.2.1 left part Pearson correlation coefficients

Layer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00	-0.12	0.29	0.50	0.27	0.11	0.03	0.39	-0.30	-0.05	-0.22	-0.37	0.05	0.26	0.25
2	-0.12	1.00	-0.15	0.10	-0.09	-0.03	-0.11	-0.04	-0.20	-0.23	-0.28	-0.16	0.11	0.07	0.09
3	-0.29	-0.15	1.00	0.74	-0.41	-0.18	-0.20	-0.23	-0.42	-0.48	-0.46	-0.38	0.05	0.50	0.30
4	-0.50	-0.10	-0.74	1.00	-0.17	-0.20	-0.09	-0.40	-0.61	-0.41	-0.60	-0.70	-0.04	0.57	0.38
5	-0.27	-0.09	-0.41	-0.17	1.00	-0.02	-0.29	-0.89	-0.10	-0.30	-0.14	-0.10	-0.05	-0.15	-0.00
6	-0.11	-0.03	-0.18	-0.20	-0.02	1.00	-0.33	-0.05	-0.46	-0.26	-0.40	-0.30	-0.01	-0.33	-0.18
7	-0.03	-0.11	-0.20	-0.09	-0.29	-0.33	1.00	-0.20	-0.53	-0.71	-0.64	-0.28	-0.03	-0.43	-0.29
8	-0.39	-0.04	-0.23	-0.40	-0.89	-0.05	-0.20	1.00	-0.20	-0.16	-0.08	-0.37	-0.03	-0.08	-0.12
9	-0.30	-0.20	-0.42	-0.61	-0.10	-0.46	-0.53	-0.20	1.00	-0.61	-0.92	-0.90	-0.01	-0.73	-0.45
10	-0.05	-0.23	-0.48	-0.41	-0.30	-0.26	-0.71	-0.16	-0.61	1.00	-0.83	-0.50	-0.01	-0.56	-0.38
11	-0.22	-0.28	-0.46	-0.60	-0.14	-0.40	-0.64	-0.08	-0.92	-0.83	1.00	-0.84	-0.01	-0.70	-0.46
12	-0.37	-0.16	-0.38	-0.70	-0.10	-0.30	-0.28	-0.37	-0.90	-0.50	-0.84	1.00	-0.04	-0.63	-0.41
13	-0.05	-0.11	-0.05	-0.04	-0.05	-0.01	-0.03	-0.03	-0.01	-0.01	-0.01	-0.04	1.00	-0.01	-0.20
14	-0.26	-0.07	-0.50	-0.57	-0.15	-0.33	-0.43	-0.08	-0.73	-0.56	-0.70	-0.63	-0.01	1.00	-0.40
15	-0.25	-0.09	-0.30	-0.38	-0.00	-0.18	-0.29	-0.12	-0.45	-0.38	-0.46	-0.41	-0.20	-0.40	1.00
16	-0.25	-0.14	-0.21	-0.33	-0.09	-0.08	-0.30	-0.18	-0.29	-0.36	-0.36	-0.25	-0.08	-0.26	-0.38
17	-0.01	-0.12	-0.06	-0.03	-0.01	-0.01	-0.17	-0.01	-0.05	-0.13	-0.09	-0.02	-0.01	-0.08	-0.01
18	-0.12	-0.11	-0.19	-0.28	-0.05	-0.12	-0.20	-0.12	-0.24	-0.29	-0.29	-0.21	-0.01	-0.23	-0.16

19	0.16	0.10	0.26	0.33	-0.09	0.15	-0.28	0.05	-0.44	-0.35	-0.43	-0.35	0.18	0.46	0.32
20	0.18	0.09	0.24	0.27	-0.04	0.17	-0.30	0.05	-0.34	-0.36	-0.38	-0.24	0.09	0.36	0.36
21	0.27	0.14	0.29	0.40	0.00	0.21	-0.33	0.13	-0.47	-0.42	-0.48	-0.40	0.05	0.44	0.36
22	0.02	0.07	0.26	0.22	-0.14	0.10	-0.27	-0.06	-0.28	-0.35	-0.34	-0.21	-0.17	0.37	0.11
23	0.33	-0.01	0.32	0.39	0.02	0.10	-0.11	0.13	-0.25	-0.14	-0.21	-0.23	0.06	0.22	0.14
24	-0.03	0.09	-0.09	-0.08	0.02	-0.03	-0.16	-0.02	-0.01	-0.09	-0.04	0.02	-0.02	-0.11	-0.03
25	0.01	0.16	-0.01	0.04	0.02	0.01	-0.17	0.01	-0.08	-0.17	-0.13	-0.05	-0.01	-0.03	0.01
26	0.67	-0.06	0.36	0.58	0.21	0.22	-0.11	0.38	-0.49	-0.19	-0.40	-0.50	0.05	0.47	0.33
27	0.59	-0.02	0.35	0.49	0.11	0.33	-0.21	0.25	-0.48	-0.26	-0.42	-0.44	0.06	0.35	0.31
28	0.09	0.07	-0.04	0.06	0.11	0.00	-0.05	0.12	-0.07	-0.02	-0.06	-0.07	0.22	0.04	0.18
29	0.07	0.09	0.08	0.09	-0.01	0.04	-0.16	0.02	-0.08	-0.16	-0.13	-0.02	0.24	0.08	0.22
30	0.51	-0.12	0.05	0.32	0.34	-0.14	0.23	0.42	-0.09	0.19	0.02	-0.24	-0.04	0.18	0.04

Table 3.2.1 right part

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Layer
0.25	-0.01	-0.12	0.16	0.18	0.27	0.02	0.33	-0.03	0.01	0.67	0.59	0.09	0.07	0.51	1
0.14	0.12	-0.11	0.10	0.09	0.14	0.07	-0.01	0.09	0.16	-0.06	-0.02	0.07	0.09	-0.12	2
0.21	-0.06	-0.19	0.26	0.24	0.29	0.26	0.32	-0.09	-0.01	0.36	0.35	-0.04	0.08	0.05	3
0.33	-0.03	-0.28	0.33	0.27	0.40	0.22	0.39	-0.08	0.04	0.58	0.49	0.06	0.09	0.32	4
0.09	0.01	-0.05	-0.09	-0.04	0.00	-0.14	0.02	0.02	0.02	0.21	0.11	0.11	-0.01	0.34	5
0.08	-0.01	-0.12	0.15	0.17	0.21	0.10	0.10	-0.03	0.01	0.22	0.33	0.00	0.04	-0.14	6
-0.30	-0.17	0.20	-0.28	-0.30	-0.33	-0.27	-0.11	-0.16	-0.17	-0.11	-0.21	-0.05	-0.16	0.23	7
0.18	-0.01	-0.12	0.05	0.05	0.13	-0.06	0.13	-0.02	0.01	0.38	0.25	0.12	0.02	0.42	8
-0.29	-0.05	0.24	-0.44	-0.34	-0.47	-0.28	-0.25	-0.01	-0.08	-0.49	-0.48	-0.07	-0.08	-0.09	9
-0.36	-0.13	0.29	-0.35	-0.36	-0.42	-0.35	-0.14	-0.09	-0.17	-0.19	-0.26	-0.02	-0.16	0.19	10
-0.36	-0.09	0.29	-0.43	-0.38	-0.48	-0.34	-0.21	-0.04	-0.13	-0.40	-0.42	-0.06	-0.13	0.02	11
-0.25	-0.02	0.21	-0.35	-0.24	-0.40	-0.21	-0.23	-0.02	-0.05	-0.50	-0.44	-0.07	-0.02	-0.24	12
-0.08	-0.01	0.01	-0.18	-0.09	-0.05	-0.17	-0.06	-0.02	-0.01	-0.05	-0.06	-0.22	-0.24	-0.04	13
-0.26	-0.08	-0.23	-0.46	-0.36	-0.44	-0.37	-0.22	-0.11	-0.03	-0.47	-0.35	-0.04	-0.08	-0.18	14
-0.38	-0.01	-0.16	-0.32	-0.36	-0.36	-0.11	-0.14	-0.03	-0.01	-0.33	-0.31	-0.18	-0.22	-0.04	15
1.00	0.03	-0.21	-0.31	-0.47	-0.52	-0.11	-0.11	-0.02	-0.05	-0.26	-0.29	-0.12	-0.22	-0.00	16
-0.03	1.00	-0.02	-0.05	-0.04	-0.07	-0.07	-0.13	-0.96	-0.94	-0.03	-0.10	-0.07	-0.04	-0.06	17
-0.21	-0.02	1.00	-0.13	-0.15	-0.17	-0.12	-0.07	-0.01	-0.06	-0.17	-0.14	-0.04	-0.11	-0.05	18

0.31	-0.05	-0.13	1.00	0.45	0.35	0.15	0.09	-0.06	-0.02	0.28	0.20	0.09	0.14	0.12	**19**
0.47	-0.04	-0.15	0.45	1.00	0.46	0.27	0.08	-0.05	-0.02	0.25	0.28	0.09	0.23	-0.02	**20**
0.52	0.07	-0.17	0.35	0.46	1.00	0.20	0.15	0.05	0.10	0.31	0.36	0.10	0.13	-0.01	**21**
0.11	0.07	-0.12	0.15	0.27	0.20	1.00	0.03	0.05	0.09	0.15	0.18	-0.02	-0.06	-0.06	**22**
0.11	0.13	-0.07	0.09	0.08	0.15	0.03	1.00	0.09	0.16	0.35	0.33	0.02	0.09	0.17	**23**
0.02	0.96	-0.01	-0.06	-0.05	0.05	0.05	0.09	1.00	0.96	0.00	0.08	0.07	0.04	-0.07	**24**
0.05	0.94	-0.06	-0.02	-0.02	0.10	0.09	0.16	0.96	1.00	0.06	0.11	0.07	0.05	-0.04	**25**
0.26	0.03	-0.17	0.28	0.25	0.31	0.15	0.35	0.00	0.06	1.00	0.69	0.11	0.08	0.47	**26**
0.29	0.10	-0.14	0.20	0.28	0.36	0.18	0.33	0.08	0.11	0.69	1.00	0.09	0.08	0.04	**27**
0.12	0.07	-0.04	0.09	0.09	0.10	-0.02	0.02	0.07	0.07	0.11	0.09	1.00	0.08	0.03	**28**
0.22	0.04	-0.11	0.14	0.23	0.13	-0.06	0.09	0.04	0.05	0.08	0.08	0.08	1.00	0.00	**29**
0.00	-0.06	-0.05	0.12	-0.02	-0.01	-0.06	0.17	-0.07	-0.04	0.47	0.04	0.03	0.00	1.00	**30**

Appendix 3.3 UK climate only model outputs

Table A3.3.1 Parameter contributions for the climate only model fine-tuned for the UK as subset area

Variable	Percent contribution	Permutation importance
Bio 11	3.1	5.8
Bio 16	15.9	27.5
Bio 18	16.4	17.6
Bio 3	21.1	13.0
Bio 4	8.8	2.2
Bio 5	27.6	23.5
Bio 8	5.5	8.1
Bio 9	1.5	2.4

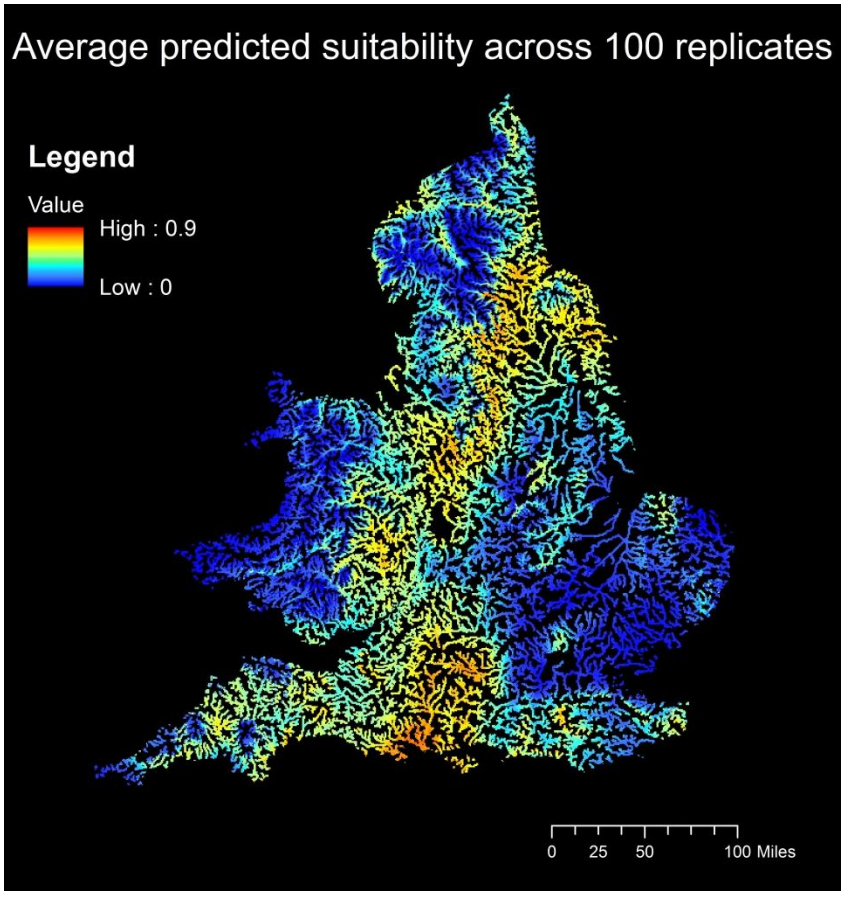


Figure A3.3.1 Suitability estimates for the climate only model fine-tuned for the UK as a subset area

Appendix 4 Grayling core microbiome

Table A.4 Grayling core microbiome*: identified as the genera present in both mouth and gill samples of the same individual and across mouth samples of different individuals*

family	genus	Mouth fish 1 (sample 1.2) % contribution	Gill fish 1 (sample 2.2) % contribution	Mouth fish 2 (sample 3) % contribution
Mycoplasmataceae	Mycoplasma	0.009	0.005	0.006
Xanthomonadaceae	Stenotrophomonas	0.354	0.002	0.637
Pseudomonadaceae	Pseudomonas	0.160	0.658	4.081
Moraxellaceae	Psychrobacter	2.122	1.183	7.362
Moraxellaceae	Acinetobacter	0.104	0.187	5.094
Pasteurellaceae	Haemophilus	0.019	0.001	0.504
Halomonadaceae	Halomonas	1.114	0.006	4.202
Enterobacteriaceae	Trabulsiella	0.083	0.003	1.002
Comamonadaceae	Variovorax	0.068	0.438	0.001
Comamonadaceae	Delftia	0.250	0.011	0.192
Sphingomonadaceae	Sphingomonas	1.523	1.250	0.001
Rhodobacteraceae	Rhodobacter	0.640	3.987	0.256
Rhodobacteraceae	Paracoccus	0.119	0.009	0.417
Rhodobacteraceae	Amaricoccus	0.496	0.009	0.144
Rhizobiaceae	Agrobacterium	0.002	0.003	3.967
Phyllobacteriaceae	Mesorhizobium	0.317	0.311	0.045
Methylobacteriaceae	Methylobacterium	4.654	0.101	0.232
Hyphomicrobiaceae	Rhodoplanes	0.838	0.362	0.004
Hyphomicrobiaceae	Devosia	0.761	0.032	0.006
Brucellaceae	Pseudochrobactrum	2.037	0.036	3.214
Bradyrhizobiaceae	Bradyrhizobium	4.787	1.767	0.396
Bradyrhizobiaceae	Afipia	0.007	0.013	0.028
Caulobacteraceae	Mycoplana	0.003	0.153	11.272
Fusobacteriaceae	Fusobacterium	0.429	0.626	2.073
[Tissierellaceae]	Anaerococcus	0.055	0.244	0.361
Streptococcaceae	Streptococcus	0.453	0.060	0.442
Lactobacillaceae	Lactobacillus	0.031	0.018	7.211
Staphylococcaceae	Staphylococcus	1.000	0.020	0.003
Planococcaceae	Planomicrobium	0.123	0.004	0.035
Sphingobacteriaceae	Sphingobacterium	0.012	0.038	0.334
Flavobacteriaceae	Flavobacterium	0.049	1.387	0.231
[Weeksellaceae]	Chryseobacterium	0.036	3.225	0.108
Prevotellaceae	Prevotella	0.252	0.647	0.001

Nocardiaceae	Rhodococcus	0.044	0.022	0.223
Micrococcaceae	Renibacterium	0.058	0.004	0.455
Micrococcaceae	Micrococcus	0.247	0.032	2.488
Micrococcaceae	Kocuria	0.187	0.028	0.005
Micrococcaceae	Citricoccus	0.002	0.005	0.001
Micrococcaceae	Arthrobacter	1.335	0.020	1.145
Microbacteriaceae	Microbacterium	0.278	0.022	0.956
Intrasporangiaceae	Kytococcus	0.031	0.013	0.001
Intrasporangiaceae	Knoellia	0.027	0.011	0.550
Intrasporangiaceae	Janibacter	0.671	0.007	0.133
Dietziaceae	Dietzia	0.032	0.044	0.021
Dermabacteraceae	Brachybacterium	0.055	0.011	0.277
Corynebacteriaceae	Corynebacterium	2.432	0.068	7.049
Brevibacteriaceae	Brevibacterium	1.073	0.043	0.810
Deinococcaceae	Deinococcus	1.903	1.547	0.714

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinforma. Oxf. Engl. 21, 2104–2105.

Acevedo-Whitehouse, K., Cunningham, A.A., 2006. Is MHC enough for understanding wildlife immunogenetics? Trends Ecol. Evol. 21, 433–438.

Adams, E.J., Parham, P., 2001. Species-specific evolution of MHC class I genes in the higher primates. Immunol. Rev. 183, 41–64.

Agudo, R., Alcaide, M., Rico, C., Lemus, J.A., Blanco, G., Hiraldo, F., Donázar, J.A., 2011. Major histocompatibility complex variation in insular populations of the Egyptian vulture: inferences about the roles of genetic drift and selection. Mol. Ecol. 20, 2329–2340.

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D., Wayne, R.K., 2004. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc. Natl. Acad. Sci. U. S. A. 101, 3490–3494.

Aissa-Grouz, N., Garnier, J., Billen, G., Mercier, B., Martinez, A., 2015. The response of river nitrification to changes in wastewater treatment (The case of the lower Seine River downstream from Paris). Ann. Limnol. - Int. J. Limnol. 51, 351–364.

Alabaster, J.S., Lloyd, R.S., 2013. Water Quality Criteria for Freshwater Fish. Elsevier.

Alcaide, M., 2010. On the relative roles of selection and genetic drift in shaping MHC variation. Mol. Ecol. 19, 3842–3844.

Allendorf, F.W., 2017. Genetics and the conservation of natural populations: allozymes to genomes. Mol. Ecol. 26, 420–430.

Allendorf, F.W., Hohenlohe, P.A., Luikart, G., 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697–709.

Allendorf, F.W., Thorgaard, G.H., 1984. Tetraploidy and the Evolution of Salmonid Fishes, in: Turner, B.J. (Ed.), Evolutionary Genetics of Fishes, Monographs in Evolutionary Biology. Springer US, pp. 1–53.

Allendorf, F.W., Leary, R.F., Spruell, P., Wenburg, J.K., 2001. The problems with hybrids: setting conservation guidelines. Trends Ecol. Evol. 16, 613–622.

Al-Qamashoui, B., Simianer, H., Kadim, I., Weigend, S., 2014. Assessment of genetic diversity and conservation priority of Omani local chickens using microsatellite markers. Trop. Anim. Health Prod. 46, 747–752.

Alsop, D., Wood, C.M., 2011. Metal uptake and acute toxicity in zebrafish: Common mechanisms across multiple metals. Aquat. Toxicol. 105, 385–393.

Altizer, S., Harvell, D., Friedle, E., 2003. Rapid evolutionary dynamics and disease threats to biodiversity. Trends Ecol. Evol. 18, 589–596.

Altizer, S., Ostfeld, R.S., Johnson, P.T., Kutz, S., Harvell, C.D., 2013. Climate change and infectious diseases: from evidence to a predictive framework. science 341, 514–519.

Amato, K.R., 2013. Co-evolution in context: The importance of studying gut microbiomes in wild animals. Microbiome Sci. Med. 1.

Amato, K.R., Yeoman, C.J., Kent, A., Righini, N., Carbonero, F., Estrada, A., Rex Gaskins, H., Stumpf, R.M., Yildirim, S., Torralba, M., Gillis, M., Wilson, B.A., Nelson, K.E., White, B.A., Leigh, S.R., 2013. Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes. ISME J. 7, 1344–1353.

Amos, K., Thomas, J., 2002. Disease interactions between wild and cultured fish: observations and lessons learned in the Pacific Northwest. Bull.-Eur. Assoc. FISH Pathol. 22, 94–101.

Anderson, R.P., Gonzalez Jr., I., 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. Ecol. Model. 222, 2796–2811.

Aoyagi, K., Dijkstra, J.M., Xia, C., Denda, I., Ototake, M., Hashimoto, K., Nakanishi, T., 2002. Classical MHC Class I Genes Composed of Highly Divergent Sequence Lineages Share a Single Locus in Rainbow Trout (*Oncorhynchus mykiss*). J. Immunol. 168, 260–273.

Araki, H., Cooper, B., Blouin, M.S., 2009. Carry-over effect of captive breeding reduces reproductive fitness of wild-born descendants in the wild. Biol. Lett. 5, 621–624.

Araki, H., Cooper, B., Blouin, M.S., 2007. Genetic Effects of Captive Breeding Cause a Rapid, Cumulative Fitness Decline in the Wild. Science 318, 100–103.

Araki, H. and Schmid, C., 2010. Is hatchery stocking a help or harm?: evidence, limitations and future directions in ecological and genetic surveys. Aquaculture, 308, pp.S2-S11.

Araújo, M.B., Cabeza, M., Thuiller, W., Hannah, L., Williams, P.H., 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. Glob. Change Biol. 10, 1618–1626.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33, 1677–1688.

Araújo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. Ecology 93, 1527–1539.

Araújo, M.B., Thuiller, W., Pearson, R.G., 2006. Climate warming and the decline of amphibians and reptiles in Europe. J. Biogeogr. 33, 1712–1728.

Arbuthnott, D., Levin, T.C., Promislow, D.E.L., 2016. The impacts of *Wolbachia* and the microbiome on mate choice in *Drosophila melanogaster*. J. Evol. Biol. 29, 461–468.

Austin, B., 2006. The bacterial microflora of fish, Revised. Sci. World J. 6, 931–945.

Austin, B., Austin, D.A., 2007. Bacterial fish pathogens: disease of farmed and wild fish. Springer Science & Business Media.

Austin, B., Embley, T. m., Goodfellow, M., 1983. Selective isolation of *Renibacterium salmoninarum*. FEMS Microbiol. Lett. 17, 111–114.

Austin, B., Stobie, M., 1992. Recovery of *Micrococcus luteus* and presumptive *Planococcus sp*. from moribund fish during an outbreak of rainbow trout, *Oncorhynchus mykiss* (Walbaum), fry syndrome in England. J. Fish Dis. 15, 203–206.

Austin, M.P., Van Niel, K.P., 2011. Improving species distribution models for climate change studies: variable selection and scale. J. Biogeogr. 38, 1–8.

Avise, J.C., 2009. Perspective: conservation genetics enters the genomics era. Conserv. Genet. 11, 665–669.

Ayllon, F., Martinez, J.L., Garcia-Vazquez, E., 2006. Loss of regional population structure in Atlantic salmon, *Salmo salar L*., following stocking. ICES J. Mar. Sci. J. Cons. 63, 1269–1273.

Babik, W., 2010. Methods for MHC genotyping in non-model vertebrates. Mol. Ecol. Resour. 10, 237–251.

Babik, W., Taberlet, P., Ejsmond, M.J., Radwan, J., 2009. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. Mol. Ecol. Resour. 9, 713–719.

Babik, W., Durka, W., Radwan, J., 2005. Sequence diversity of the MHC DRB gene in the Eurasian beaver (*Castor fiber*). Mol. Ecol. 14, 4249–4257.

Ballentine, B. and Greenberg, R., 2010. Common garden experiment reveals genetic control of phenotypic divergence between swamp sparrow subspecies that lack divergence in neutral genotypes. PloS one, 5(4), p.e10229.

Bardonnet, A., Gaudin, P., Thorpe, J.E., 1993. Diel rhythm of emergence and of first displacement downstream in trout (*Salmo trutta*), Atlantic salmon (*S. salar*) and grayling (*Thymallus thymallus*). J. Fish Biol. 43, 755–762.

Barnosky, A.D., Hadly, E.A., Bascompte, J., Berlow, E.L., Brown, J.H., Fortelius, M., Getz, W.M., Harte, J., Hastings, A., Marquet, P.A., Martinez, N.D., Mooers, A., Roopnarine, P., Vermeij, G., Williams, J.W., Gillespie, R., Kitzes, J., Marshall, C., Matzke, N., Mindell, D.P., Revilla, E., Smith, A.B., 2012. Approaching a state shift in Earth's biosphere. Nature 486, 52–58.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B., Ferrer, E.A., 2011. Has the Earth's sixth mass extinction already arrived? Nature 471, 51–57.

Bean, W.T., Stafford, R., Brashares, J.S., 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. Ecography 35, 250–258.

Beauchamp, D.A., 1990. Movements, habitat use, and spawning strategies of Arctic grayling in a

subalpine lake tributary. Northwest Sci. 64.

Beck, J., Böller, M., Erhardt, A., Schwanghart, W., 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecol. Inform. 19, 10–15.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., Courchamp, F., 2012. Impacts of climate change on the future of biodiversity. Ecol. Lett. 15, 365–377.

Bergström, U., Sundblad, G., Downie, A.-L., Snickars, M., Boström, C., Lindegarth, M., 2013. Evaluating eutrophication management scenarios in the Baltic Sea using species distribution modelling. J. Appl. Ecol. 50, 680–690.

Bernatchez, L., Landry, C., 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J. Evol. Biol. 16, 363–377.

Bernatchez, L., Renaut, S., Whiteley, A.R., Derome, N., Jeukens, J., Landry, L., Lu, G., Nolte, A.W., Østbye, K., Rogers, S.M., St-Cyr, J., 2010. On the origin of species: insights from the ecological genomics of lake whitefish. Philos. Trans. R. Soc. B Biol. Sci. 365, 1783–1800.

Bingulac-Popovic, J., Figueroa, F., Sato, A., Talbot, W.S., Johnson, S.L., Gates, M., Postlethwait, J.H., Klein, J., 1997. Mapping of Mhc class I and class II regions to different linkage groups in the zebrafish, Danio rerio. Immunogenetics 46, 129–134.

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecol. Indic. 18, 31–41.

Blabolil, P., Logez, M., Ricard, D., Prchalová, M., Říha, M., Sagouis, A., Peterka, J., Kubečka, J., Argillier, C., 2016. An assessment of the ecological potential of Central and Western European reservoirs based on fish communities. Fish. Res., Ecology of Fish in Lakes and Reservoirs 173, Part 1, 80–87.

Bletz, M.C., Loudon, A.H., Becker, M.H., Bell, S.C., Woodhams, D.C., Minbiole, K.P.C., Harris, R.N., 2013. Mitigating amphibian chytridiomycosis with bioaugmentation: characteristics of effective probiotics and strategies for their selection and use. Ecol. Lett. 16, 807–820.

Blomqvist, D., Pauliny, A., Larsson, M., Flodin, L.-Å., 2010. Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. BMC Evol. Biol. 10, 33.

Bolnick, D.I., Snowberg, L.K., Caporaso, J.G., Lauber, C., Knight, R., Stutz, W.E., 2014. Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. Mol. Ecol. 23, 4831–4845.

Bonin, A., Nicole, F., Pompanon, F., Miaud, C., Taberlet, P., 2007. Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. Conserv. Biol. 21, 697–708.

Bonneaud, C., Sorci, G., Morin, V., Westerdahl, H., Zoorob, R., Wittzell, H., 2004. Diversity of Mhc class I and IIB genes in house sparrows (*Passer domesticus*). Immunogenetics 55, 855–865.

Bontrop, R.E., Otting, N., de Groot, N.G., Doxiadis, G.G.M., 1999. Major histocompatibility complex class II polymorphisms in primates. Immunol. Rev. 167, 339–350.

Bosch, T.C.G., McFall-Ngai, M.J., 2011. Metaorganisms as the new frontier. Zoology 114, 185–190.

Bottrill, M.C., Joseph, L.N., Carwardine, J., Bode, M., Cook, C., Game, E.T., Grantham, H., Kark, S., Linke, S., McDonald-Madden, E., Pressey, R.L., Walker, S., Wilson, K.A., Possingham, H.P., 2008. Is conservation triage just smart decision making? Trends Ecol. Evol. 23, 649–654.

Boutin, S., Bernatchez, L., Audet, C., Derôme, N., 2013. Network analysis highlights complex interactions between pathogen, host and commensal microbiota. PLoS ONE 8, e84772.

Boyson, J.E., Shufflebotham, C., Cadavid, L.F., Urvater, J.A., Knapp, L.A., Hughes, A.L., Watkins, D.I., 1996. The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. J. Immunol. 156, 4656–4665.

Bozzi, R., Alvarez, I., Crovetti, A., Fernandez, I., De Petris, D., Goyache, F., 2012. Assessing priorities for conservation in Tuscan cattle breeds using microsatellites. animal 6, 203–211.

Brambilla, M., Ficetola, G.F., 2012. Species distribution models as a tool to estimate reproductive parameters: a case study with a passerine bird species. J. Anim. Ecol. 81, 781–787.

Braud, V.M., Allan, D.S., McMichael, A.J., 1999. Functions of nonclassical MHC and non-MHC-encoded class I molecules. Curr. Opin. Immunol. 11, 100–108.

Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of Southern Wisconsin. Ecol. Monogr. 27, 326–349.

Britton, J.R., Harvey, J.P., Cowx, I.G., Holden, T., Feltham, M.J., Wilson, B.R., Davies, J.M., 2002. Compensatory responses of fish populations in a shallow eutrophic lake to heavy depredation pressure by cormorants and the implications for management. Manag. Ecol. Lake Reserv. Fish. 171–183.

Brodersen, J., Seehausen, O., 2014. Why evolutionary biologists should get seriously involved in ecological monitoring and applied biodiversity assessment programs. Evol. Appl. 7, 968–983.

Brook, B.W., Sodhi, N.S., Bradshaw, C.J.A., 2008. Synergies among extinction drivers under global change. Trends Ecol. Evol. 23, 453–460.

Brown, C.J., Saunders, M.I., Possingham, H.P., Richardson, A.J., 2013. Managing for interactions between local and global stressors of ecosystems. PLOS ONE 8, e65765.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., Wiley, D.C., 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364, 33–39.

Brown, J.L., 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. Methods Ecol. Evol. 5, 694–700.

Brucker, R.M., Bordenstein, S.R., 2013. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. Science 341, 667–669.

Buckley, J.T., Roch, M., McCarter, J.A., Rendell, C.A., Matheson, A.T., 1982. Chronic exposure of coho salmon to sublethal concentrations of copper—I. Effect on growth, on accumulation and distribution of copper, and on copper tolerance. Comp. Biochem. Physiol. Part C Comp. Pharmacol. 72, 15–19.

Bucklin, D.N., Basille, M., Benscoter, A.M., Brandt, L.A., Mazzotti, F.J., Romañach, S.S., Speroterra, C., Watling, J.I., 2015. Comparing species distribution models constructed with different subsets of environmental predictors. Divers. Distrib. 21, 23–35.

Buhl, K.J., Hamilton, S.J., 1991. Relative sensitivity of early life stages of arctic grayling, coho salmon, and rainbow trout to nine inorganics. Ecotoxicol. Environ. Saf. 22, 184–197.

Buhl, K.J., Hamilton, S.J., 1990. Comparative toxicity of inorganic contaminants released by placer mining to early life stages of salmonids. Ecotoxicol. Environ. Saf. 20, 325–342.

Butchart, S.H.M., Walpole, M., Collen, B., Strien, A. van, Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.C., Watson, R., 2010. Global Biodiversity: Indicators of recent declines. Science 328, 1164–1168.

Cabeza, M., Araújo, M.B., Wilson, R.J., Thomas, C.D., Cowley, M.J.R., Moilanen, A., 2004. Combining probabilities of occurrence with spatial reserve design. J. Appl. Ecol. 41, 252–262.

Cahill, A.E., Aiello-Lammens, M.E., Fisher-Reid, M.C., Hua, X., Karanewsky, C.J., Ryu, H.Y., Sbeglia, G.C., Spagnolo, F., Waldron, J.B., Warsi, O., Wiens, J.J., 2013. How does climate change cause extinction? Proc R Soc B 280, 20121890.

Cahill, M.M., 1990. Bacterial flora of fishes: A Review. Microb. Ecol. 19, 21–41.

Callaghan, D.A., Kirby, J.S., Bell, M.C., Spray, C.J., 1998. Cormorant *Phalacrocorax carbo* occupancy and impact at stillwater game fisheries in England and Wales. Bird Study 45, 1–17.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R., 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26, 266–267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010b. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012. Biodiversity loss and its impact on humanity. Nature 486, 59–67.

Carilli, J.E., Norris, R.D., Black, B.A., Walsh, S.M. and McField, M., 2009. Local stressors reduce coral resilience to bleaching. Plos one, 4(7), p.e6324.

Castric, V., Bernatchez, L., Belkhir, K., Bonhomme, F., 2002. Heterozygote deficiencies in small lacustrine populations of brook charr Salvelinus Fontinalis Mitchill (Pisces, Salmonidae): a test of alternative hypotheses. Heredity 89, 27–35.

Ceballos, G., Ehrlich, P.R., 2002. Mammal Population Losses and the Extinction Crisis. Science 296, 904–907.

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., Palmer, T.M., 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Sci. Adv. 1.

Chao, A., 1984. Nonparametric Estimation of the Number of Classes in a Population. Scand. J. Stat. 11, 265–270.

Chaplin, D.D., 2010. Overview of the Immune Response. J. Allergy Clin. Immunol. 125, S3-23.

Charlesworth, D., Willis, J.H., 2009. The genetics of inbreeding depression. Nat. Rev. Genet. 10, 783–796.

Charmantier, A., Garant, D., 2005. Environmental quality and evolutionary potential: lessons from wild populations. Proc. R. Soc. B Biol. Sci. 272, 1415–1425.

Cheng, Y., Stuart, A., Morris, K., Taylor, R., Siddle, H., Deakin, J., Jones, M., Amemiya, C.T., Belov, K., 2012. Antigen-presenting genes and genomic copy number variations in the Tasmanian devil MHC. BMC Genomics 13, 87.

Chevin, L.M., Lande, R. and Mace, G.M., 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biol, 8(4), p.e1000357.

Chowdhury, M.J., Girgis, M., Wood, C.M., 2016. Revisiting the mechanisms of copper toxicity to rainbow trout: Time course, influence of calcium, unidirectional Na+ fluxes, and branchial Na+, K+ ATPase and V-type H+ ATPase activities. Aquat. Toxicol. 177, 51–62.

Cohen, J.I., Feinstone, S., Purcell, R.H., 1989. Hepatitis A virus infection in a chimpanzee: duration of viremia and detection of virus in saliva and throat swabs. J. Infect. Dis. 160, 887–890.

Colwell, R.K., Coddington, J.A., 1994. Estimating terrestrial biodiversity through extrapolation. Philos. Trans. Biol. Sci. 345, 101–118.

Colwell, R.K., Rangel, T.F., 2009. Hutchinson's duality: The once and future niche. Proc. Natl. Acad. Sci. 106, 19651–19658.

Comber, S.D.W., Merrington, G., Sturdy, L., Delbeke, K., van Assche, F., 2008. Copper and zinc water quality standards under the EU Water Framework Directive: The use of a tiered approach to

estimate the levels of failure. Sci. Total Environ. 403, 12–22.

Consuegra, S., de Eyto, E., McGinnity, P., Stet, R.J.M., Jordan, W.C., 2011. Contrasting responses to selection in class I and class IIα major histocompatibility-linked markers in salmon. Heredity 107, 143–154.

Consuegra, S., Leaniz, C.G. de, 2008. MHC-mediated mate choice increases parasite resistance in salmon. Proc. R. Soc. Lond. B Biol. Sci. 275, 1397–1403.

Consuegra, S., Megens, H.J., Leon, K., Stet, R.J.M., Jordan, W.C., 2005a. Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. Immunogenetics 57, 16–24.

Consuegra, S., Megens, H.-J., Schaschl, H., Leon, K., Stet, R.J.M., Jordan, W.C., 2005b. Rapid evolution of the MH class I locus results in different allelic compositions in recently diverged populations of Atlantic salmon. Mol. Biol. Evol. 22, 1095–1106.

Correll, D.L., 1998. The role of phosphorus in the eutrophication of receiving waters: A Review. J. Environ. Qual. 27, 261.

Crain, C.M., Kroeker, K., Halpern, B.S., 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecol. Lett. 11, 1304–1315.

Crandall, K.A., Bininda-Emonds, O.R., Mace, G.M., Wayne, R.K., 2000. Considering evolutionary processes in conservation biology. Trends Ecol. Evol. 15, 290–295.

Crow, J.F. and Kimura, M., 1970. An introduction to population genetics theory. An introduction to population genetics theory.

Cruickshank, T.E., Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23, 3133–3157.

Cullen, M., Perfetto, S.P., Klitz, W., Nelson, G., Carrington, M., 2002. High-resolution patterns of meiotic recombination across the human Major Histocompatibility Complex. Am. J. Hum. Genet. 71, 759–776.

Danchin, E.G.J., Pontarotti, P., 2004. Towards the reconstruction of the bilaterian ancestral pre-MHC region. Trends Genet. 20, 587–591.

Daszak, P., Cunningham, A.A., Hyatt, A.D., 2000. Emerging infectious diseases of wildlife-- Threats to Biodiversity and Human Health. Science 287, 443–449.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510.

Dawnay, N., Dawnay, L., Hughes, R.N., Cove, R., Taylor, M.I., 2011. Substantial genetic structure among stocked and native populations of the European grayling (*Thymallus thymallus*, Salmonidae) in the United Kingdom. Conserv. Genet. 12, 731–744.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C., Mace, G.M., 2011. Beyond predictions: Biodiversity conservation in a changing climate. Science 332, 53–58.

De Boer, R.J., Borghans, J.A., Van Boven, M., Keşmir, C., Weissing, F.J., 2004. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics 55, 725–731.

de Eyto, E., McGinnity, P., Huisman, J., Coughlan, J., Consuegra, S., Farrell, K., O'Toole, C., Tufto, J., Megens, H.J., Jordan, W., Cross, T., Stet, R.J.M., 2011. Varying disease-mediated selection at different life-history stages of Atlantic salmon in fresh water. Evol. Appl. 4, 749–762.

Defoirdt, T., Halet, D., Vervaeren, H., Boon, N., Van de Wiele, T., Sorgeloos, P., Bossier, P., Verstraete, W., 2007. The bacterial storage compound poly-β-hydroxybutyrate protects *Artemia franciscana* from pathogenic *Vibrio campbellii*. Environ. Microbiol. 9, 445–452.

Delport, W., Poon, A.F.Y., Frost, S.D.W., Pond, S.L.K., 2010a. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26, 2455–2457.

Delport, W., Scheffler, K., Botha, G., Gravenor, M.B., Muse, S.V. and Pond, S.L.K., 2010. CodonTest: modeling amino acid substitution preferences in coding sequences. PLoS Comput Biol, 6(8), p.e1000885.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072.

DeWoody, Y.D., DeWoody, J.A., 2005. On the estimation of genome-wide heterozygosity using molecular markers. J. Hered. 96, 85–88.

Diaz, R.J., Breitburg, D.L., 2009. Chapter 1 The hypoxic environment, in: Physiology, B.F. (Ed.), Hypoxia. Academic Press, pp. 1–23.

Ding, A., Goudet, J., 2005. Heterozygote advantage and the maintenance of polymorphism for multilocus traits. Theor. Popul. Biol. 68, 157–166.

Dionne, M., Miller, K.M., Dodson, J.J., Bernatchez, L., 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. Philos. Trans. Biol. Sci. 364, 1555–1565.

Dionne, M., Miller, K.M., Dodson, J.J., Caron, F., Bernatchez, L., 2007. Clinal variation in MHC diversity with temperature: evidence for the role of host–pathogen interaction on local adaptation in Atlantic salmon. Evolution 61, 2154–2164.

Dirscherl, H., Yoder, J.A., 2014. Characterization of the Z lineage Major histocompatability complex class I genes in zebrafish. Immunogenetics 66, 185–198.

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J., Collen, B., 2014. Defaunation in the Anthropocene. Science 345, 401–406.

Dormann, C.F., 2007. Promising the future? Global change projections of species distributions. Basic Appl. Ecol. 8, 387–397.

Duchesne, P., Bernatchez, L., 2002. An analytical investigation of the dynamics of inbreeding in multi-generation supportive breeding. Conserv. Genet. 3, 45–58.

Duftner, N., Koblmüller, S., Weiss, S., Medgyesy, N., Sturmbauer, C., 2005. The impact of stocking on the genetic structure of European grayling (*Thymallus thymallus, Salmonidae*) in two alpine rivers. Hydrobiologia 542, 121–129.

Duplouy, A., Ikonen, S., Hanski, I., 2013. Life history of the *Glanville fritillary* butterfly in fragmented versus continuous landscapes. Ecol. Evol. 3, 5141–5156.

Ekblom, R., Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107, 1–15.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Edmands, S., 2007. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. Mol. Ecol. 16, 463–475.

Ehrlich, P.R., Ehrlich, A.H., 2013. Can a collapse of global civilization be avoided? Proc. R. Soc. B Biol. Sci. 280.

Einum, S., Fleming, I.A., 2000. Highly fecund mothers sacrifice offspring survival to maximize fitness. Nature 405, 565–567.

Einum, S., Fleming, I.A., 2001. Implications of stocking: ecological interactions between wild and released salmonids. Nord. J. Freshw. Res. 56–70.

Eizaguirre, C., Lenz, T.L., Kalbe, M., Milinski, M., 2012. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat. Commun. 3, 621.

Eizaguirre, C., Yeates, S.E., Lenz, T.L., Kalbe, M., Milinski, M., 2009. MHC-based mate choice combines good genes and maintenance of MHC polymorphism. Mol. Ecol. 18, 3316–3329.

Ejsmond, M.J., Radwan, J., 2011. MHC diversity in bottlenecked populations: a simulation model. Conserv. Genet. 12, 129–137.

Elbers, J.P., Taylor, S.S., 2016. Major Histocompatibility Complex polymorphism in reptile conservation. Herpetol. Conserv. Biol. 11, 1–12.

Eliason, E.J., Clark, T.D., Hague, M.J., Hanson, L.M., Gallagher, Z.S., Jeffries, K.M., Gale, M.K., Patterson, D.A., Hinch, S.G., Farrell, A.P., 2011. Differences in thermal tolerance among sockeye salmon populations. Science 332, 109–112.

Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. Methods Ecol. Evol. 1, 330–342.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.

Elshout, P.M.F., Dionisio Pires, L.M., Leuven, R.S.E.W., Wendelaar Bonga, S.E., Hendriks, A.J., 2013. Low oxygen tolerance of different life stages of temperate freshwater fish species. J. Fish Biol. 83, 190–206.

Elwood, J.W., Waters, T.F., 1969. Effects of Floods on Food Consumption and Production Rates of a Stream Brook Trout Population. Trans. Am. Fish. Soc. 98, 253–262.

Epps, C.W., Wehausen, J.D., Bleich, V.C., Torres, S.G., Brashares, J.S., 2007. Optimizing dispersal and corridor models using landscape genetics. J. Appl. Ecol. 44, 714–724.

Erickson, T.R., Stefan, H.G., 2000. Linear air/water temperature correlations for streams during open water periods. J. Hydrol. Eng. 5, 317–321.

Erp, S.H.M. van, Dixon, B., Figueroa, F., Egberts, E., Stet, R.J.M., 1996. Identification and characterization of a new major histocompatibility complex class I gene in carp (*Cyprinus carpio L.*). Immunogenetics 44, 49–61.

Esselman, P.C., Allan, J.D., 2011. Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of northeastern Mesoamerica. Freshw. Biol. 56, 71–88.

Estes, L.D., Bradley, B.A., Beukes, H., Hole, D.G., Lau, M., Oppenheimer, M.G., Schulze, R., Tadross, M.A., Turner, W.R., 2013. Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. Glob. Ecol. Biogeogr. 22, 1007–1018.

Evans, M.L., Neff, B.D., 2009. Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). Mol. Ecol. 18, 4716–4729.

Evans, M.L., Neff, B.D., Heath, D.D., 2010. MHC-mediated local adaptation in reciprocally translocated Chinook salmon. Conserv. Genet. 11, 2333–2342.

Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3, 87–112.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567.

Eyto, E. de, McGinnity, P., Consuegra, S., Coughlan, J., Tufto, J., Farrell, K., Megens, H.J., Jordan, W., Cross, T., Stet, R.J.M., 2007. Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. Proc. R. Soc. B Biol. Sci. 274, 861–869.

Fabricius, E., Gustafson, K.-J., 1955. Observations on the spawning behaviour of the grayling, *Thymallus thymallus* (L.). National Swedish Board of Fisheries.

Fagan, W.F., Holmes, E.E., 2006. Quantifying the extinction vortex. Ecol. Lett. 9, 51–60.

Falconer, D.S., Mackay, T.F., Frankham, R., 1996. Introduction to quantitative genetics (4th edn).

Trends Genet. 12, 280.

Felsenstein, J., 1993. PHYLIP: phylogenetic inference package, version 3.5c. Department of Genetics, University of Washington.

Ficke, A.D., Myrick, C.A., Hansen, L.J., 2007. Potential impacts of global climate change on freshwater fisheries. Rev. Fish Biol. Fish. 17, 581–613.

Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L., Gurr, S.J., 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484, 186–194.

Fisher, R.A., 1930. The Theory of Natural Selection. Oxford University Press, London.Flajnik, M., Ohta, Y., Namikawa-Yomada, C., Nonaka, M., 1999. Insight into the primordial MHC from studies in ectothermic vertebrates. Immunol. Rev. 167, 59–67.

Flajnik, M.F., Kasahara, M., 2001. Comparative Genomics of the MHC: Glimpses into the evolution of the adaptive immune system. Immunity 15, 351–362.

Fleming, I.A., Petersson, E., 2001. The ability of released, hatchery salmonids to breed and contribute to the natural productivity of wild populations. Nord. J. Freshw. Res. 71–98.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global consequences of land use. Science 309, 570–574.

Foll, M., Gaggiotti, O., 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics 180, 977–993.

Forest, F., Grenyer, R., Rouget, M., Davies, T.J., Cowling, R.M., Faith, D.P., Balmford, A., Manning, J.C., Procheş, Ş., van der Bank, M., Reeves, G., Hedderson, T.A.J., Savolainen, V., 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445, 757–760.

Förstner, U., Wittmann, G.T., 2012. Metal pollution in the aquatic environment. Springer Science & Business Media.

Frankham, R., 2015. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. Mol. Ecol. 24, 2610–2618.

Frankham, R., 2010. Where are we in conservation genetics and where do we need to go? Conserv. Genet. 11, 661–663.

Frankham, R., 2008. Genetic adaptation to captivity in species conservation programs. Mol. Ecol. 17, 325–333.

Frankham, R., 2005a. Genetics and extinction. Biol. Conserv. 126, 131–140.

Frankham, R., 2005b. Stress and adaptation in conservation genetics. J. Evol. Biol. 18, 750–755.

Frankham, R., Manning, H., Margan, S.H., Briscoe, D.A., 2000. Does equalization of family sizes reduce genetic adaptation to captivity? Anim. Conserv. Forum 3, 357–363.

Fraser, B.A., Neff, B.D., 2010. Parasite mediated homogenizing selection at the MHC in guppies. Genetica 138, 273.

Fraser, D.J., 2008. How well can captive breeding programs conserve biodiversity? A review of salmonids. Evol. Appl. 1, 535–586.

Fraser, D.J., Weir, L.K., Bernatchez, L., Hansen, M.M., Taylor, E.B., 2011. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity 106, 404–420.

Fraser, N.H.C., Metcalfe, N.B., Thorpe, J.E., 1993. Temperature-Dependent Switch between Diurnal and Nocturnal Foraging in Salmon. Proc. R. Soc. Lond. B Biol. Sci. 252, 135–139.

Fraune, S., Bosch, T.C.G., 2010. Why bacteria matter in animal development and evolution. BioEssays 32, 571–580.

Frenz, C., Klinger, H., Schumacher, H., 1997. Zwischenbericht zur Situation von Äsche (*Thymallus thymallus L.*) und Bachforelle (*Salmo trutta L.*) in der Lenne (NRW)–Lebensraum, Kormorane und Angelfischerei. Nat. Landsch. 72, 401–407.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L. and Nielsen, R., 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet, 7(11), p.e1002355.

Funk, W.C., McKay, J.K., Hohenlohe, P.A., Allendorf, F.W., 2012. Harnessing genomics for delineating conservation units. Trends Ecol. Evol. 27, 489–496.

Galan, M., Guivier, E., Caraux, G., Charbonnel, N., Cosson, J.F., 2010. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. BMC Genomics 11, 296.

Garcia de Leaniz, C., Fleming, I.A., Einum, S., Verspoor, E., Jordan, W.C., Consuegra, S., Aubin-Horth, N., Lajus, D., Letcher, B.H., Youngson, A.F., Webb, J.H., Vøllestad, L.A., Villanueva, B., Ferguson, A., Quinn, T.P., 2007. A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. Biol. Rev. 82, 173–211.

Gardiner, R., 2000. The origins and present distribution of grayling. Complete Book Grayl. Robert Hale Publ. Lond. 15–22.

Gardiner, R., 1991. Scottish grayling: history and biology of the populations. Proc Inst Fish Manag Annu Study Cour 22, 171–178.

Gebremedhin, B., Ficetola, G.F., Naderi, S., Rezaei, H.-R., Maudet, C., Rioux, D., Luikart, G., Flagstad, Ø., Thuiller, W., Taberlet, P., 2009. Frontiers in identifying conservation units: from neutral markers to adaptive genetic variation. Anim. Conserv. 12, 107–109.

Gil, P., Vivas, J., Gallardo, C.S., Rodrıguez, L.A., 2000. First isolation of *Staphylococcus warneri*, from diseased rainbow trout, *Oncorhynchus mykiss* (Walbaum), in Northwest Spain. J. Fish Dis. 23, 295–

298.

Gilbert, S.F., Sapp, J., Tauber, A.I., Thomson, H.E.J.D., Stearns, A.E.S.C., 2012. A symbiotic view of life: we have never been individuals. Q. Rev. Biol. 87, 325–341.

Giles, N., Westgarth, J., Hewlett,N., 2004. Management advice for trout, grayling and Arctic char fisheries. Environment Agency Fisheries Technical Manual No. 7.

Giljohann, K.M., Hauser, C.E., Williams, N.S.G., Moore, J.L., 2011. Optimizing invasive species control across space: willow invasion management in the Australian Alps. J. Appl. Ecol. 48, 1286–1294.

Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T., Martin, J.-F., 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics 12, 245.

Goecks, J., Nekrutenko, A., Taylor, J., 2010. Galaxy: a
comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 11, R86.

Gómez, D., Conejeros, P., Marshall, S.H., Consuegra, S., 2010. MHC evolution in three salmonid species: a comparison between class II alpha and beta genes. Immunogenetics 62, 531–542.

Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J. 3, 1314–1317.

Gomez-Mestre, I., Tejedo, M., 2004. Contrasting patterns of quantitative and neutral genetic variation in locally adapted populations of the natterjack toad, *Bufo calamita*. Evol. Int. J. Org. Evol. 58, 2343–2352.

Good, I.J., 1953. The population frequencies of species and the estimation of population parameters. Biometrika 40, 237–264.

Goudet, J., 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9. 3).

Griffiths, R.A., Pavajeau, L., 2008. Captive breeding, reintroduction, and the conservation of amphibians. Conserv. Biol. 22, 852–861.

Grimholt, U., 2016. MHC and Evolution in Teleosts. Biology 5, 6.

Grimholt, U., Hordvik, I., Fosse, V.M., Olsaker, I., Endresen, C., Lie, Ø., 1993. Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). Immunogenetics 37, 469–473.

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S., Stet, R.J.M., 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55, 210–219.

Grimholt, U., Olsaker, I., Lingaas, F., Lie, Ø., 1997. Sequence analysis of MHC class I α2 domain exon variants in one diploid and two haploid Atlantic salmon pedigrees. Anim. Genet. 28, 385–390.

Grimholt, U., Tsukamoto, K., Azuma, T., Leong, J., Koop, B.F., Dijkstra, J.M., 2015. A comprehensive analysis of teleost MHC class I sequences. BMC Evol. Biol. 15, 32.

Gross, R., Kühn, R., Baars, M., Schröder, W., Stein, H., Rottmann, O., 2001. Genetic differentiation of European grayling populations across the Main, Danube and Elbe drainages in Bavaria. J. Fish Biol. 58, 264–280.

Grueber, C.E., Waters, J.M., Jamieson, I.G., 2011. The imprecision of heterozygosity-fitness correlations hinders the detection of inbreeding and inbreeding depression in a threatened species. Mol. Ecol. 20, 67–79.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R., Wintle, B.A., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Glob. Ecol. Biogeogr. 24, 276–292.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. Ecol. Lett. 16, 1424–1435.

Gum, B., 2007. Genetic characterisation of European grayling (*Thymallus thymallus*) populations: implications for conservation and management. PhD thesis.

Gum, B., Gross, R., Geist, J., 2009. Conservation genetics and management implications for European grayling, *Thymallus thymallus*: synthesis of phylogeography and population genetics. Fish. Manag. Ecol. 16, 37–51.

Gum, B., Gross, R., Kuehn, R., 2005. Mitochondrial and nuclear DNA phylogeography of European grayling (*Thymallus thymallus*): evidence for secondary contact zones in central Europe. Mol. Ecol. 14, 1707–1725.

Gum, B., Gross, R., Rottmann, O., Schröder, W., Kühn, R., 2003. Microsatellite variation in Bavarian populations of European grayling (*Thymallus thymallus*): Implications for conservation. Conserv. Genet. 4, 659–672.

Haddeland, P.J., Junge, C., Serbezov, D. and Vøllestad, L.A., 2015. Genetic parentage analysis confirms a polygynandrous breeding system in the European grayling (*Thymallus thymallus*). PloS one, 10(3), p.e0122032.

Halet, D., Defoirdt, T., van Damme, P., Vervaeren, H., Forrez, I., van de Wiele, T., Boon, N., Sorgeloos, P., Bossier, P., Verstraete, W., 2007. Poly-β-hydroxybutyrate-accumulating bacteria protect

gnotobiotic Artemia franciscana from pathogenic *Vibrio campbellii*. FEMS Microbiol. Ecol. 60, 363–369.

Hansen, T.H. , Huang, S., Arnold, P.L., Fremont, D.H., 2007. Patterns of nonclassical MHC antigen presentation. Nat. Immunol. 8, 563–568.

Hansen, M.M., Olivieri, I., Waller, D.M., Nielsen, E.E., The GeM Working Group, 2012. Monitoring adaptive genetic responses to environmental change. Mol. Ecol. 21, 1311–1329.

Hansson, B., Westerberg, L., 2008. Heterozygosity-fitness correlations within inbreeding classes: local or genome-wide effects? Conserv. Genet. 9, 73–83.

Harbicht, A., Wilson, C.C., Fraser, D.J., 2014. Does human-induced hybridization have long-term genetic effects? Empirical testing with domesticated, wild and hybridized fish populations. Evol. Appl. 7, 1180–1191.

Hari, R.E., Livingstone, D.M., Siber, R., Burkhardt-Holm, P., Guettinger, H., 2006. Consequences of climatic change for water temperature and brown trout populations in Alpine rivers and streams. Glob. Change Biol. 12, 10–26.

Harvell, C.D., 2002. Climate warming and disease risks for terrestrial and marine biota. Science 296, 2158–2162.

Hashimoto, K., Okamura, K., Yamaguchi, H., Ototake, M., Nakanishi, T., Kurosaiva, Y., 1999. Conservation and diversification of MHC class I and its related molecules in vertebrates. Immunol. Rev. 167, 81–100.

Hedrick, 1999. Balancing selection and MHC. Genetica 7–214.

Hedrick, P.W., 2001. Conservation genetics: where are we now? Trends Ecol. Evol. 16, 629–636.

Hedrick, P.W., Garcia-Dorado, A., 2016. Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends Ecol. Evol. 31, 940–952.

Hedrick, P.W., 1999. Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation. Evolution 53, 313–318

Hedrick, P.W., Rashbrook, V.K., Hedgecock, D., 2000. Effective population size of winter-run chinook salmon based on microsatellite analysis of returning spawners. Can. J. Fish. Aquat. Sci. 57, 2368–2373.

Heggenes, J., Qvenild, T., Stamford, M.D., Taylor, E.B., 2006. Genetic structure in relation to movements in wild European grayling (*Thymallus thymallus*) in three Norwegian rivers. Can. J. Fish. Aquat. Sci. 63, 1309–1319.

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G., Körber, J.-H., 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. Glob. Ecol. Biogeogr. 16, 754–763.

Heller, N.E., Zavaleta, E.S., 2009. Biodiversity management in the face of climate change: a review of 22 years of recommendations. Biol. Conserv. 142, 14–32.

Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. 89, 10915–10919.

Hess, C.M., Edwards, S.V., 2002. The Evolution of the Major Histocompatibility Complex in Birds Scaling up and taking a genomic approach to the major histo compatibilty complex (MHC) of birds reveals surprising departures from generalities found in mammals in both large-scale structure and the mechanisms shaping the evolution of the MHC. BioScience 52, 423–431.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

Hilscherova, K., Dusek, L., Kubik, V., Cupr, P., Hofman, J., Klanova, J., Holoubek, I., 2007. Redistribution of organic pollutants in river sediments and alluvial soils related to major floods. J. Soils Sediments 7, 167–177.

Hirzel, A.H., Le Lay, G., 2008. Habitat suitability modelling and niche theory. J. Appl. Ecol. 45, 1372–1381.

Hisar, O., Yanik, T., Hisar, S.A., 2002. Clinical and pathological investigation of Psychrobacter immobilis infection in Rainbow trout (Oncorhynchus Mykiss, Walbaum).

Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., Poss, M.L., Reed, L.K., Storfer, A., Whitlock, M.C., 2016. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am. Nat. 188, 379–397.

Hoffmann, A.A., Hercus, M.J., 2000. Environmental stress as an evolutionary force. BioScience 50, 217–226.

Hoffmann, A.A., Parsons, P.A. (Peter A., 1933-, 1991. Evolutionary genetics and environmental stress. Oxford University Press.

Hoffmann, A.A., Sgrò, C.M., 2011. Climate change and evolutionary adaptation. Nature 470, 479–485.

Hoffmann, A.A., Willi, Y., 2008. Detecting genetic responses to environmental change. Nat. Rev. Genet. 9, 421–432.

Hofstetter, A.R., Sullivan, L.C., Lukacher, A.E., Brooks, A.G., 2011. Diverse roles of non-diverse molecules: MHC class Ib molecules in host defence and control of autoimmunity. Curr. Opin. Immunol. 23, 104–110.

Hollert, H., Dürr, M., Erdinger, L., Braunbeck, T., 2000. Cytotoxicity of settling particulate matter and sediments of the Neckar River (Germany) during a winter flood. Environ. Toxicol. Chem. 19, 528–534.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6, e1000862.

Huchard, E., Albrecht, C., Schliehe-Diecks, S., Baniel, A., Roos, C., Peter, P.M.K., Brameier, M., 2012. Large-scale MHC class II genotyping of a wild lemur population by next generation sequencing. Immunogenetics 64, 895–913.

Hugenholtz, P., Tyson, G.W., 2008. Microbiology: Metagenomics. Nature 455, 481–483.

Hulme, P.E., 2005. Adapting to climate change: is there scope for ecological management in the face of a global threat? J. Appl. Ecol. 42, 784–794.

Hulstere, D. d', Philippart, J.C., 1982. Observations on hatching and post-hatching behaviour in the European grayling [*Thymallus thymallus* (L.) pisciculture in Belgium]. Cah. Ethologie Appl. Belg.

Hunn, J.B., 1985. Role of calcium in gill function in freshwater fishes. Comp. Biochem. Physiol. A Physiol. 82, 543–547.

Hurst, G.D.D., Jiggins, C.D., 2013. Evolutionary biology: A gut feeling for isolation. Nature 500, 412–413.

Hutchings, J.A., 2006. Survival consequences of sex-biased growth and the absence of a growth–mortality trade-off. Funct. Ecol. 20, 347–353.

Hutchinson, G.E., 1957. Concluding remarks.: Cold Sprig Harbor Symposia on Quantitative Biology. Yale University New Haven.

Ibbotson, A.T., Cove, R.J., Ingraham, A., Gallagher, M., Hornby, D.D., Furse, M., Williams, C., 2001. A review of grayling ecology, status and management practice: recommendations for future management in England and Wales. Environment Agency.

Icenhour, C.R., Rebholz, S.L., Collins, M.S., Cushion, M.T., 2002. Early Acquisition of Pneumocystis carinii in Neonatal Rats as Evidenced by PCR and Oral Swabs. Eukaryot. Cell 1, 414–419.

Imholt, C., Malcolm, I.A., Bacon, P.J., Gibbins, C.N., Soulsby, C., Miles, M., Fryer, R.J., 2011. Does diurnal temperature variability affect growth in juvenile Atlantic salmon Salmo salar? J. Fish Biol. 78, 436–448.

Jackson, M.C., Loewen, C.J.G., Vinebrooke, R.D., Chimimba, C.T., 2016. Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. Glob. Change Biol. 22, 180–189.

Jeffery, K.J.., Bangham, C.R.., 2000. Do infectious diseases drive MHC diversity? Microbes Infect. 2, 1335–1341.

Jensen, A.J., Johnsen, B.O., 1999. The functional relationship between peak spring floods and survival and growth of juvenile Atlantic salmon (*Salmo salar*) and Brown trout (*Salmo trutta*). Funct. Ecol. 13, 778–785.

Jensen, L.F., Hansen, M.M., Pertoldi, C., Holdensgaard, G., Mensberg, K.-L.D., Loeschcke, V., 2008. Local adaptation in brown trout early life-history traits: implications for climate change adaptability. Proc. R. Soc. Lond. B Biol. Sci. 275, 2859–2868.

Jiménez-Valverde, A., 2012. Insights into the area under the receiver operating characteristic curve

(AUC) as a discrimination measure in species distribution modelling. Glob. Ecol. Biogeogr. 21, 498–507.
Jiménez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to either–or presence–absence. Acta Oecologica 31, 361–369.

Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P., Lobo, J.M., 2011. Use of niche models in invasive species risk assessments. Biol. Invasions 13, 2785–2797.

Johnson, M.S., 2000. Measuring and interpreting genetic structure to minimize the genetic risks of translocations. Aquac. Res. 31, 133–143.

Johnson, M.T.J., Stinchcombe, J.R., 2007. An emerging synthesis between community ecology and evolutionary biology. Trends Ecol. Evol. 22, 250–257.

Johnson, N.A., Vallejo, R.L., Silverstein, J.T., Welch, T.J., Wiens, G.D., Hallerman, E.M., Palti, Y., 2008. Suggestive association of Major histocompatibility IB genetic markers with resistance to bacterial cold water disease in Rainbow trout (Oncorhynchus mykiss). Mar. Biotechnol. 10, 429–437.

Jonsson, B., Jonsson, N., 2009. A review of the likely effects of climate change on anadromous Atlantic salmon *Salmo salar* and brown trout *Salmo trutta*, with particular reference to water temperature and flow. J. Fish Biol. 75, 2381–2447.

Jorgensen, T.C., Weatherley, L.R., 2003. Ammonia removal from wastewater by ion exchange in the presence of organic contaminants. Water Res. 37, 1723–1728.

Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. Mamm. Protein Metab. 3, 132.

Junge, C., Vøllestad, L.A., Barson, N.J., Haugen, T.O., Otero, J., Sætre, G.-P., Leder, E.H., Primmer, C.R., 2011. Strong gene flow and lack of stable population structure in the face of rapid adaptation to local temperature in a spring-spawning salmonid, the European grayling (*Thymallus thymallus*). Heredity 106, 460–471.

Jungwirth, M., Winkler, H., 1984. The temperature dependence of embryonic development of grayling (*Thymallus thymallus*), Danube salmon (Hucho hucho), Arctic char (*Salvelinus alpinus*) and Brown trout (*Salmo trutta fario*). Aquaculture 38, 315–327.

Kaufman, J., Salomonsen, J., Flajnik, M., 1994. Evolutionary conservation of MHC class I and class II molecules—different yet the same. Semin. Immunol. 6, 411–424.

Kavanagh, K.D., Haugen, T.O., Gregersen, F., Jernvall, J., Vøllestad, L.A., 2010. Contemporary temperature-driven divergence in a Nordic freshwater fish under conditions commonly thought to hinder adaptation. BMC Evol. Biol. 10, 350.

Kawaguchi, Y., Nakano, S., 2001. Contribution of terrestrial invertebrates to the annual resource budget for salmonids in forest and grassland reaches of a headwater stream. Freshw. Biol. 46, 303–316.

Kayis, S., Capkin, E., Altinok, Ii., 2009. Bacteria in Rainbow trout (*Oncorhynchus mykiss*) in the Southern Black Sea Region of Turkey - A Survey.

Kearney, M., 2006. Habitat, environment and niche: what are we modelling? Oikos 115, 186–191.

Keeley, E.R., 2001. Demographic responses to food and space competition by juvenile steelhead trout. Ecology 82, 1247–1259.

Keesing, F., Belden, L.K., Daszak, P., Dobson, A., Harvell, C.D., Holt, R.D., Hudson, P., Jolles, A., Jones, K.E., Mitchell, C.E., Myers, S.S., Bogich, T., Ostfeld, R.S., 2010. Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature 468, 647–652.

Kekäläinen, J., Vallunen, J.A., Primmer, C.R., Rättyä, J., Taskinen, J., 2009. Signals of Major Histocompatibility complex overdominance in a wild salmonid population. Proc. Biol. Sci. 276, 3133–3140.

Kircher, M., Kelso, J., 2010. High-throughput DNA sequencing – concepts and limitations. BioEssays 32, 524–536.

Kirk, H., Freeland, J.R., 2011. Applications and implications of neutral versus non-neutral markers in molecular ecology. Int. J. Mol. Sci. 12, 3966–3988.

Kiryu, I., Dijkstra, J.M., Sarder, R.I., Fujiwara, A., Yoshiura, Y., Ototake, M., 2005. New MHC class Ia domain lineages in rainbow trout (*Oncorhynchus mykiss*) which are shared with other fish species. Fish Shellfish Immunol. 18, 243–254.

Klein, D., Ono, H., O'huigin, C., Vincek, V., Goldschmidt, T., Klein, J., 1993a. Extensive MHC variability in cichlid fishes of Lake Malawi. Nature 364, 330–334.

Klein, J., Satta, Y., O'huigin, C., Takahata, N., 1993b. The molecular descent of the major histocompatibility complex. Annu. Rev. Immunol. 11, 269–295.

Klein, J., Sato, A., O'huigin, C., 1998. Evolution by gene duplication in the major histocompatibility complex. Cytogenet Cell Genet 80.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. gks808.

Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., Mardis, E.R., 2013. The Next-Generation Sequencing Revolution and Its Impact on Genomics. Cell 155, 27–38.

Kohn, M.H., Murphy, W.J., Ostrander, E.A., Wayne, R.K., 2006. Genomics and conservation genetics. Trends Ecol. Evol. 21, 629–637.

Komar, N., Lanciotti, R., Bowen, R., Langevin, S., Bunning, M., 2002. Detection of West Nile virus in oral and cloacal swabs collected from bird carcasses. Emerg. Infect. Dis. 8, 741–742.

Koskella, B., 2014. Research highlights for issue 5: the role of the microbiome in shaping evolution. Evol. Appl. 7, 519–520.

Koskinen, M.T., Nilsson, J., Veselov, A.J., Potutkin, A.G., Ranta, E., Primmer, C.R., 2002. Microsatellite data resolve phylogeographic patterns in European grayling, *Thymallus thymallus*, Salmonidae. Heredity 88, 391–401.

Koskinen, M.T., Ranta, E., Piironen, J., Veselov, A., Titov, S., Haugen, T.O., Nilsson, J., Carlstein, M., Primmer, C.R., 2000. Genetic lineages and postglacial colonization of grayling (*Thymallus thymallus*, Salmonidae) in Europe, as revealed by mitochondrial DNA analyses. Mol. Ecol. 9, 1609–1624.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Divers. Distrib. 19, 1366–1379.

Kremen, C., Cameron, A., Moilanen, A., Phillips, S.J., Thomas, C.D., Beentje, H., Dransfield, J., Fisher, B.L., Glaw, F., Good, T.C., others, 2008. Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. Science 320, 222–226.

Kristiansen, H., Dølving, K.B., 1996. The migration of spawning stocks of grayling *Thymallus thymallus*, in Lake Mjøsa, Norway. Environ. Biol. Fishes 47, 43–50.

Kroglund, F., Rosseland, B.O., Teien, H.-C., Salbu, B., Kristensen, T., Finstad, B., 2007. Water quality limits for Atlantic salmon (*Salmo salar L.*) exposed to short term reductions in pH and increased aluminum simulating episodes. Hydrol. Earth Syst. Sci. Discuss. 4, 3317–3355.

Kroupova, H., Machova, J., Piackova, V., Blahova, J., Dobsikova, R., Novotny, L., Svobodova, Z., 2008. Effects of subchronic nitrite exposure on rainbow trout (*Oncorhynchus mykiss*). Ecotoxicol. Environ. Saf. 71, 813–820.

Kroupova, H., Machova, J., Svobodova, Z., 2005. Nitrite influence on fish: a review. Vet. Med.-PRAHA- 50, 461.

Kruiswijk, C.P., Hermsen, T.T., Westphal, A.H., Savelkoul, H.F.J., Stet, R.J.M., 2002. A novel functional class I lineage in Zebrafish (*Danio rerio*), Carp (*Cyprinus carpio*), and Large Barbus (*Barbus intermedius*) Showing an Unusual Conservation of the Peptide Binding Domains. J. Immunol. 169, 1936–1947.

Kulski, J.K., Shiina, T., Anzai, T., Kohara, S., Inoko, H., 2002. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol. Rev. 190, 95–122.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Lacy, R.C., 1997. Importance of genetic variation to the viability of mammalian populations. J. Mammal. 78, 320.

Lange, V., Böhme, I., Hofmann, J., Lang, K., Sauter, J., Schöne, B., Paul, P., Albrecht, V., Andreas, J.M., Baier, D.M., Nething, J., Ehninger, U., Schwarzelt, C., Pingel, J., Ehninger, G., Schmidt, A.H., 2014. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics 15, 63.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinforma. Oxf. Engl. 23, 2947–2948.

Larson, W.A., Seeb, J.E., Dann, T.H., Schindler, D.E., Seeb, L.W., 2014. Signals of heterogeneous selection at an MHC locus in geographically proximate ecotypes of sockeye salmon. Mol. Ecol. 23, 5448–5461.

Laurén, D.J., McDonald, D.G., 1985. Effects of copper on branchial ionoregulation in the rainbow trout, *Salmo gairdneri* Richardson. J. Comp. Physiol. B 155, 635–644.

Lauzon, H.L., Dimitroglou, A., Merrifield, D.L., Ringø, E., Davies, S.J., 2014. Probiotics and prebiotics-concepts, definitions and history. Aquac. Nutr. Gut Health Probiotics Prebiotics 169–184.

Lenz, T.L., Eizaguirre, C., Scharsack, J.P., Kalbe, M., Milinski, M., 2009. Disentangling the role of MHC-dependent “good genes” and “compatible genes” in mate-choice decisions of three-spined sticklebacks Gasterosteus aculeatus under semi-natural conditions. J. Fish Biol. 75, 2122–2142.

Lewin, J., Macklin, M.G., 1987. Metal mining and floodplain sedimentation in Britain. Int. Geomorphol. 1986, 1009–1027.

Lewis, V., 2006, Habitat advisory visit to the river Severn, Caersws, Wales; Caersws angling association, The Wild Trout Trust Ltd. Registered Charity No 1162478; Sheffield, UK.

Li, M., Stoneking, M., 2012. A new approach for detecting low-level mutations in next-generation sequence data. Genome Biol. 13, R34.

Lighten, J., Oosterhout, C., Bentzen, P., 2014a. Critical review of NGS analyses for de novo genotyping multigene families. Mol. Ecol. 23, 3957–3972.

Lighten, J., van Oosterhout, C., Paterson, I.G., McMullan, M., Bentzen, P., 2014b. Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). Mol. Ecol. Resour. 14, 753–767.

Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28, 385–393.

Llewellyn, M.S., Boutin, S., Hoseinifar, S.H., Derome, N., 2014. Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. Front. Microbiol. 5.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151.

Lorenzen, K., Beveridge, M.C.M., Mangel, M., 2012. Cultured fish: integrative biology and management of domestication and interactions with wild fish. Biol. Rev. 87, 639–660.

Lozupone, C.A., Hamady, M., Kelley, S.T., Knight, R., 2007. Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. Appl. Environ. Microbiol. 73, 1576–1585.

Lukacs, M., Harstad, H., Bakke, H., Beetz-Sargent, M., McKinnel, L., Lubieniecki, K., Koop, B., Grimholt, U., 2010. Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. BMC Genomics 11, 154.

Lukacs, M.F., Harstad, H., Grimholt, U., Beetz-Sargent, M., Cooper, G.A., Reid, L., Bakke, H.G., Phillips, R.B., Miller, K.M., Davidson, W.S., Koop, B.F., 2007. Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (*Salmo salar*). BMC Genomics 8, 251.

Luoto, M., Virkkala, R., Heikkinen, R.K., 2007. The role of land cover in bioclimatic models depends on spatial resolution. Glob. Ecol. Biogeogr. 16, 34–42.

Lynch, M., O'Hely, M., 2001. Captive breeding and the genetic fitness of natural populations. Conserv. Genet. 2, 363–378.

Mace, G.M., Norris, K., Fitter, A.H., 2012. Biodiversity and ecosystem services: a multilayered relationship. Trends Ecol. Evol. 27, 19–26.

Macklin, M.G., Hudson-Edwards, K.A., Dawson, E.J., 1997. The significance of pollution from historic metal mining in the Pennine orefields on river sediment contaminant fluxes to the North Sea. Sci. Total Environ., U.K. Fluxes to the North Sea, Land Ocean Interaction Study (LOIS) Rivers Basins Research, the First Two Years 194, 391–397.

MacRae, R.K., Smith, D.E., Swoboda-Colberg, N., Meyer, J.S., Bergman, H.L., 1999. Copper binding affinity of rainbow trout (*Oncorhynchus mykiss*) and brook trout (*Salvelinus fontinalis*) gills: Implications for assessing bioavailable metal. Environ. Toxicol. Chem. 18, 1180–1189.

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., Ravikesavan, R., 2013. Gene duplication as a major force in evolution. J. Genet. 92, 155–161.

Malcolm, I.A., Bacon, P.J., Middlemas, S.J., Fryer, R.J., Shilland, E.M., Collen, P., 2014. Relationships between hydrochemistry and the presence of juvenile brown trout (*Salmo trutta*) in headwater streams recovering from acidification. Ecol. Indic., Threats to upland waters 37, Part B, 351–364.

Malcolm, I.A., Soulsby, C., Hannah, D.M., Bacon, P.J., Youngson, A.F., Tetzlaff, D., 2008. The influence of riparian woodland on stream temperatures: implications for the performance of juvenile salmonids. Hydrol. Process. 22, 968–979.

Mallet, J.P., Charles, S., Persat, H., Auger, P., 1999. Growth modelling in accordance with daily water temperature in European grayling (*Thymallus thymallus L.*). Can. J. Fish. Aquat. Sci. 56, 994–1000.

Manel, S., Holderegger, R., 2013. Ten years of landscape genetics. Trends Ecol. Evol. 28, 614–621.

Marcogliese, D.J., 2008. The impact of climate change on the parasites and infectious diseases of aquatic animals. Rev. Sci. Tech. Int. Off. Epizoot. 27, 467–484.

Mardis, E.R., 2013. Next-Generation Sequencing Platforms. Annu. Rev. Anal. Chem. 6, 287–303.

Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. 24, 133–141.

Marsden, C.D., Verberkmoes, H., Thomas, R., Wayne, R.K., Mable, B.K., 2013. Pedigrees, MHC and microsatellites: an integrated approach for genetic management of captive African wild dogs (*Lycaon pictus*). Conserv. Genet. 14, 171–183.

Martins, E.G., Hinch, S.G., Patterson, D.A., Hague, M.J., Cooke, S.J., Miller, K.M., Robichaud, D., English, K.K., Farrell, A.P., 2012. High river temperature reduces survival of sockeye salmon (*Oncorhynchus nerka*) approaching spawning grounds and exacerbates female mortality. Can. J. Fish. Aquat. Sci. 69, 330–342.

Mayr, E., 1954. Change of genetic environment and evolution.

McClellan, D.A., Ellison, D.D., 2010. Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. Int. J. Bioinforma. Res. Appl. 6, 120–133.

McClelland, E.K., Ming, T.J., Tabata, A., Kaukinen, K.H., Beacham, T.D., Withler, R.E., Miller, K.M., 2013. Patterns of selection and allele diversity of class I and class II major histocompatibility loci across the species range of sockeye salmon (*Oncorhynchus nerka*). Mol. Ecol. 22, 4783–4800.

McClelland, E.K., Ming, T.J., Tabata, A., Miller, K.M., 2011. Sequence analysis of MHC class I α2 from sockeye salmon (*Oncorhynchus nerka*). Fish Shellfish Immunol. 31, 507–510.

McIntyre, J.K., Baldwin, D.H., Beauchamp, D.A., Scholz, N.L., 2012. Low-level copper exposures increase visibility and vulnerability of juvenile coho salmon to cutthroat trout predators. Ecol. Appl. 22, 1460–1471.

McKay, J.K., Latta, R.G., 2002. Adaptive population divergence: markers, QTL and traits. Trends Ecol. Evol. 17, 285–291.

McKim, J.M., Benoit, D.A., 1971. Effects of long-term exposures to copper on survival, growth, and reproduction of Brook trout (*Salvelinus fontinalis*). J. Fish. Res. Board Can. 28, 655–662.

Meraner, A., Cornetti, L., Gandolfi, A., 2014. Defining conservation units in a stocking-induced genetic melting pot: unraveling native and multiple exotic genetic imprints of recent and historical secondary contact in Adriatic grayling. Ecol. Evol. 4, 1313–1327.

Meraner, A., Gandolfi, A., 2012. Phylogeography of European grayling, *Thymallus thymallus* (*Actinopterygii, Salmonidae*), within the Northern Adriatic basin: evidence for native and exotic mitochondrial DNA lineages. Hydrobiologia 693, 205–221.

Meraner, A., Unfer, G., Gandolfi, A., 2013. Good news for conservation: mitochondrial and microsatellite DNA data detect limited genetic signatures of inter-basin fish transfer in *Thymallus thymallus* (*Salmonidae*) from the Upper Drava River. Knowl. Manag. Aquat. Ecosyst. 1.

Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36, 1058–1069.

Merrifield, D.L., Dimitroglou, A., Foey, A., Davies, S.J., Baker, R.T.M., Bøgwald, J., Castex, M., Ringø, E., 2010. The current status and future focus of probiotic and prebiotic applications for salmonids. Aquaculture 302, 1–18.

Metzker, M.L., 2010. Sequencing technologies — the next generation. Nat. Rev. Genet. 11, 31–46.

Meyer, D., Single, R.M., Mack, S.J., Erlich, H.A., Thomson, G., 2006. Signatures of Demographic History and Natural Selection in the Human Major Histocompatibility Complex Loci. Genetics 173, 2121–2142.

Meyer, L., 2001. Spawning migration of grayling *Thymallus thymallus* (L., 1758) in a Northern German lowland river. Arch. Für Hydrobiol. 152, 99–117.

Meyer-Lucht, Y., Sommer, S., 2005. MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). Mol. Ecol. 14, 2233–2243.

Milinski, M., 2006. The major histocompatibility complex, sexual selection, and mate choice. Annu Rev Ecol Evol Syst 37, 159–186.

Miller, H.C., Allendorf, F., Daugherty, C.H., 2010. Genetic diversity and differentiation at MHC genes in island populations of tuatara (*Sphenodon spp.*). Mol. Ecol. 19, 3894–3908.

Miller, H.C., Lambert, D.M., 2004. Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (*Petroicidae*). Mol. Ecol. 13, 3709–3721.

Miller, J., 2010. Species distribution modeling. Geogr. Compass 4, 490–509.

Miller, K.M., Kaukinen, K.H., Schulze, A.D., 2002. Expansion and contraction of major histocompatibility complex genes: a teleostean example. Immunogenetics 53, 941–963.

Miller, K.M., Li, S., Ming, T.J., Kaukinen, K.H., Schulze, A.D., 2006. The salmonid MHC class I: more ancient loci uncovered. Immunogenetics 58, 571–589.

Miller, K.M., Winton, J.R., Schulze, A.D., Purcell, M.K., Ming, T.J., 2004. Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. Environ. Biol. Fishes 69, 307–316.

Moonsamy, P.V., Williams, T., Bonella, P., Holcomb, C.L., Höglund, B.N., Hillman, G., Goodridge, D., Turenchalk, G.S., Blake, L.A., Daigle, D.A., Simen, B.B., Hamilton, A., May, A.P., Erlich, H.A., 2013. High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array™ system for simplified amplicon library preparation. Tissue Antigens 81, 141–149. 1

Moss, B., Kosten, S., Meerhof, M., Battarbee, R., Jeppesen, E., Mazzeo, N., Havens, K., Lacerot, G., Liu, Z., De Meester, L., others, 2011. Allied attack: climate change and eutrophication. Inland Waters 1, 101–105.

Mukai, T., Asasaka, T., Sato, E., Mori, K., Matsumoto, M., Ohori, H., 2002. Inhibition of binding of *Helicobacter pylori* to the glycolipid receptors by probiotic *Lactobacillus reuteri*. FEMS Immunol. Med. Microbiol. 32, 105–110.

Murray, K.A., Retallick, R.W.R., Puschendorf, R., Skerratt, L.F., Rosauer, D., McCallum, H.I., Berger, L., Speare, R., VanDerWal, J., 2011. Assessing spatial patterns of disease risk to biodiversity: implications for the management of the amphibian pathogen, *Batrachochytrium dendrobatidis*. J. Appl. Ecol. 48, 163–173.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S.L.K., Scheffler, K., 2013. FUBAR: A fast, unconstrained bayesian approximation for inferring selection. Mol. Biol. Evol. 30, 1196–1205.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K. and Pond, S.L.K., 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet, 8(7), p.e1002764.

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P., 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods Ecol. Evol. 5, 1198–1205.

Muyzer, G., Waal, E.C. de, Uitterlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59, 695–700.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., Hohenlohe, P.A., 2013. Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. 22, 2841–2847.

Navarrete, P., Espejo, R.T., Romero, J., 2009. Molecular analysis of microbiota along the digestive tract of juvenile Atlantic salmon (*Salmo salar L.*). Microb. Ecol. 57, 550.

Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283–292.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Nei, M., Kumar, S., 2000. Molecular evolution and phylogenetics. Oxford University Press.

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121–152.

Nematollahi, A., Decostere, A., Pasmans, F., Haesebrouck, F., 2003. *Flavobacterium psychrophilum* infections in salmonid fish. J. Fish Dis. 26, 563–574.

Newaj-Fyzul, A., Al-Harbi, A.H., Austin, B., 2014. Review: Developments in the use of probiotics for disease control in aquaculture. Aquaculture, Special Issue on Aquaculture conference: To the next 40 years of sustainable global aquaculture- Part A 431, 1–11.

Norden, C.R., 1961. Comparative osteology of representative salmonid fishes, with particular reference to the grayling (*Thymallus arcticus*) and its phylogeny. J. Fish. Res. Board Can. 18, 679–791.

Northcote, T.G., 1995. Comparative biology and management of Arctic and European grayling (*Salmonidae, Thymallus*). Rev. Fish Biol. Fish. 5, 141–194.

Nykänen, M., Huusko, A., 2002. Suitability criteria for spawning habitat of riverine European grayling. J. Fish Biol. 60, 1351–1354.

Nykänen, M., Huusko, A., Lahti, M., 2004. Movements and habitat preferences of adult grayling (*Thymallus thymallus L*.) from late winter to summer in a boreal river. Arch. Für Hydrobiol. 161, 417–432.

Oberdorff, T., Pont, D., Hugueny, B., Porcher, J.-P., 2002. Development and validation of a fish-based index for the assessment of "river health" in France. Freshw. Biol. 47, 1720–1734.

O'Connor, E.A., Strandh, M., Hasselquist, D., Nilsson, J.-Å., Westerdahl, H., 2016. The evolution of highly variable immunity genes across a passerine bird radiation. Mol. Ecol. 25, 977–989.

O'Farrell, B., Dennis, C., Benzie, J.A., McGinnity, P., Carlsson, J., de Eyto, E., Coughlan, J.P., Igoe, F., Meehan, R., Cross, T.F., 2012. Balancing selection on MHC class I in wild brown trout *Salmo trutta*. J. Fish Biol. 81, 1357–1374.

O'Grady, J.J., Brook, B.W., Reed, D.H., Ballou, J.D., Tonkyn, D.W., Frankham, R., 2006. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. Biol. Conserv. 133, 42–51.

Ohta, Y., Okamura, K., McKinney, E.C., Bartl, S., Hashimoto, K., Flajnik, M.F., 2000. Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proc. Natl. Acad. Sci. 97, 4712–4717.

Okamura, K., Nakanishi, T., Kurosawa, Y., Hashimoto, K., 1993. Expansion of genes that encode MHC class I molecules in cyprinid fishes. J. Immunol. 151, 188–200.

Okita, T.W., Hwang, Y.S., Hnilo, J., Kim, W.T., Aryan, A.P., Larson, R., Krishnan, H.B., 1989. Structure and expression of the rice glutelin multigene family. J. Biol. Chem. 264, 12573–12581.

Olden, J.D., LeRoy Poff, N., Douglas, M.R., Douglas, M.E., Fausch, K.D., 2004. Ecological and evolutionary consequences of biotic homogenization. Trends Ecol. Evol. 19, 18–24.

Oliver, M.K., Piertney, S.B., 2012. Selection maintains MHC diversity through a natural population bottleneck. Mol. Biol. Evol. 29, 1713–1720.

Oliver, M.K., Telfer, S., Piertney, S.B., 2009. Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). Proc. R. Soc. Lond. B Biol. Sci. 276, 1119–1128.

Oliveros, J.C., 2007. VENNY. An interactive tool for comparing lists with Venn Diagrams.

Ormerod, S.J., Jones, M.E., Jones, M.C., Phillips, D.R., 2004. The effects of riparian forestry on invertebrate drift and brown trout in upland streams of contrasting acidity. Hydrol. Earth Syst. Sci. Discuss. 8, 578–588.

Orr, H.G., des Clers, S., Simpson, G.L., Hughes, M., Battarbee, R.W., Cooper, L., Dunbar, M.J., Evans, R., Hannaford, J., Hannah, D.M., Laize, C.L.R., Richards, K.S., Watts, G., Wilby, R.L., 2010. Changing water temperatures: a surface water archive for England and Wales, in: Kirby, C. (Ed.), Role of

hydrology in managing consequences of a changing global environment. British Hydrological Society, Newcastle University, Newcastle upon Tyne, United Kingdom.

Ostland, V.E., Lumsden, J.S., MacPhee, D.D., Ferguson, H.W., 1994. Characteristics of *Flavobacterium branchiophilum*, the cause of salmonid bacterial gill disease in Ontario. J. Aquat. Anim. Health 6, 13–26.

Ottewell, K.M., Bickerton, D.C., Byrne, M., Lowe, A.J., 2016. Bridging the gap: a genetic assessment framework for population-level threatened plant conservation prioritization and decision-making. Divers. Distrib. 22, 174–188.

Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R. (Kuke), Hedrick, P.W., 2010. Conservation genetics in transition to conservation genomics. Trends Genet. 26, 177–187.

Ovidio, M., Parkinson, D., Sonny, D., Philippart, J.-C., 2004. Spawning movements of European grayling *Thymallus thymallus* in the River Aisne (Belgium). Folia Zool. 53, 87.

Ozaki, A., Khoo, S.-K., Yoshiura, Y., Ototake, M., Sakamoto, T., Dijkstra, J.M., Okamoto, N., 2007. Identification of Additional Quantitative Trait Loci (QTL) Responsible for Susceptibility to Infectious Pancreatic Necrosis Virus in Rainbow Trout. 魚病研究 42, 131–140.

Palachek, R.M., Tomasso, J.R., 1984. Toxicity of Nitrite to Channel Catfish (*Ictalurus punctatus*), Tilapia (*Tilapia aurea*), and Largemouth Bass (*Micropterus salmoides*): Evidence for a Nitrite Exclusion Mechanism. Can. J. Fish. Aquat. Sci. 41, 1739–1744.

Parham, P., Ohta, T., 1996. Population biology of antigen presentation by MHC class I molecules. Science 272, 67–74.

Pavey, S.A., Sevellec, M., Adam, W., Normandeau, E., Lamaze, F.C., Gagnaire, P.-A., Filteau, M., Hebert, F.O., Maaroufi, H., Bernatchez, L., 2013. Nonparallelism in MHCIIβ diversity accompanies nonparallelism in pathogen infection of lake whitefish (*Coregonus clupeaformis*) species pairs as revealed by next-generation sequencing. Mol. Ecol. 22, 3833–3849.

Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28, 2537–2539.

Pearse, D.E., 2016. Saving the spandrels? Adaptive genomic variation in conservation and fisheries management. J. Fish Biol. 89, 2697–2716.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Glob. Ecol. Biogeogr. 12, 361–371.

Pearson, R.G., Dawson, T.P., Liu, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27, 285–298.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34, 102–117.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L.,

Segurado, P., Dawson, T.P., Lees, D.C., 2006. Model-based uncertainty in species range prediction. J. Biogeogr. 33, 1704–1711.

Penn, D.J., Damjanovich, K., Potts, W.K., 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc. Natl. Acad. Sci. 99, 11260–11264.

Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarrés, J.F., Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L., Cooper, H.D., Gilman, E.L., Guénette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R., Walpole, M., 2010. Scenarios for global biodiversity in the 21st century. Science 330, 1496–1501.

Pérez-Sánchez, T., Ruiz-Zarzuela, I., de Blas, I., Balcázar, J.L., 2014. Probiotics in aquaculture: a current assessment. Rev. Aquac. 6, 133–146.

Persat, H., 1996. Threatened populations and conservation of the European grayling, *Thymallus thymallus* (L., 1758), in: Conservation of Endangered Freshwater Fish in Europe. Springer, pp. 233–247.

Persat, H., Mattersdorfer, K., Charlat, S., Schenekar, T., Weiss, S., 2016. Genetic integrity of the European grayling (*Thymallus thymallus*) populations within the Vienne River drainage basin after five decades of stockings. CYBIUM 40, 7–20.

Persat, H., Pattee, E., 1981. growth rate of young grayling in some French rivers. Proc.Int. Assoc. Theor. Appl. Limnol.

Pertoldi, C., Bijlsma, R., Loeschcke, V., 2007. Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. Biodivers. Conserv. 16, 4147–4163.

Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddemeier, R.H., Stockwell, D.R.B., 2002. Future projections for Mexican faunas under global climate change scenarios. Nature 416, 626–629.

Pfrender, M.E., Spitze, K., Hicks, J., Morgan, K., Latta, L., Lynch, M., 2000. Lack of concordance between genetic diversity estimates at the molecular and quantitative-trait levels. Conserv. Genet. 1, 263–269.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197.

Phillips, S.J., Dudík, M., Schapire, R.E., 2004. A Maximum Entropy Approach to Species Distribution Modeling, in: Proceedings of the Twenty-First International Conference on Machine Learning, ICML '04. ACM, New York, NY, USA, p. 83–.

Pienkowski, M.W., Watkinson, A.R., Kerby, G., Suter, W., 1998. The effect of predation by wintering

cormorants *Phalacrocorax carbo* on grayling *Thymallus thymallus* and trout (Salmonidae) populations: two case studies from Swiss rivers. Reply. J. Appl. Ecol. 35, 611–616.

Piertney, S.B., Oliver, M.K., 2005. The evolutionary ecology of the major histocompatibility complex. Heredity 96, 7–21.
Piertney, S.B., Webster, L.M.I., 2010. Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. Genetica 138, 419–432.

Piertney, S.B., Webster, L.M.I., 2008. Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. Genetica 138, 419–432.

Piontkivska, H., Nei, M., 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol. Biol. Evol. 20, 601–609.

Pitcher, T.E., Neff, B.D., 2007. Genetic quality and offspring performance in Chinook salmon: implications for supportive breeding. Conserv. Genet. 8, 607–616.

Playle, R.C., 1998. Modelling metal interactions at fish gills. Sci. Total Environ. 219, 147–163.

Polce, C., Termansen, M., Aguirre-Gutiérrez, J., Boatman, N.D., Budge, G.E., Crowe, A., Garratt, M.P., Pietravalle, S., Potts, S.G., Ramirez, J.A., Somerwill, K.E., Biesmeijer, J.C., 2013. Species distribution models for crop pollination: A modelling framework applied to Great Britain. PLOS ONE 8, e76308.

Poncin, P., 1996. A field observation on the influence of aggressive behaviour on mating success in the European grayling. J. Fish Biol. 48, 802–804.

Pond, S.L.K., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23, 1891–1901.

Pont, D., Hugueny, B., Rogers, C., 2007. Development of a fish-based index for the assessment of river health in Europe: the European Fish Index. Fish. Manag. Ecol. 14, 427–439.

Porazinska, D.L., Sung, W., Giblin-Davis, R.M., Thomas, W.K., 2010. Reproducibility of read numbers in high-throughput sequencing analysis of nematode community composition and structure. Mol. Ecol. Resour. 10, 666–676.

Prickett, J., Simer, R., Christopher-Hennings, J., Yoon, K.-J., Evans, R.B., Zimmerman, J.J., 2008. Detection of Porcine reproductive and respiratory syndrome virus infection in porcine oral fluid samples: a longitudinal study under experimental conditions. J. Vet. Diagn. Invest. 20, 156–163.

Prudhomme, C., Davies, H., 2009a. Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate. Clim. Change 93, 177–195.

Prudhomme, C., Davies, H., 2009b. Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: future climate. Clim. Change 93, 197–222.

Puschendorf, R., Carnaval, A.C., VanDerWal, J., Zumbado-Ulate, H., Chaves, G., Bolaños, F., Alford, R.A., 2009. Distribution models for the amphibian chytrid *Batrachochytrium dendrobatidis* in Costa Rica: proposing climatic refuges as a conservation tool. Divers. Distrib. 15, 401–408.

Quader, S., 2005. Mate choice and its implications for conservation and management. Curr. Sci. 89, 1220–1229.

Rached, L.A., McDermott, M.F., Pontarotti, P., 1999. The MHC Big Bang. Immunol. Rev. 167, 33–45.

Radosavljevic, A., Anderson, R.P., 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. J. Biogeogr. 41, 629–643.

Raes, J., Van de Peer, Y., 2003. Gene duplication, the evolution of novel gene functions, and detecting functional divergence of duplicates in silico. Appl. Bioinformatics 2, 91–101.

Ramond, J.-B., Makhalanyane, T. p., Tuffin, M. i., Cowan, D. a., 2015. Normalization of environmental metagenomic DNA enhances the discovery of under-represented microbial community members. Lett. Appl. Microbiol. 60, 359–366.

Redford, K.H., Segre, J.A., Salafsky, N., del Rio, C.M., McAloose, D., 2012. Conservation and the Microbiome. Conserv. Biol. 26, 195–197.

Reed, D.H., Frankham, R., 2001. How Closely Correlated Are Molecular and Quantitative Measures of Genetic Variation? a Meta-Analysis. Evolution 55, 1095–1103.

Reiley, M.C., 2007. Science, policy, and trends of metals risk assessment at EPA: How understanding metals bioavailability has changed metals risk assessment at US EPA. Aquat. Toxicol., A Tribute to Rick Playle: The interface of toxicology, physiology, and modelling in improving water quality regulations for metals; Proceedings of a Symposium: "SETAC 27th Annual Meeting. A Tribute to Rick Playle: The interface of toxicology, physiology, and modelling in improving water quality regulations for metals" 84, 292–298.

Renner, I.W., Warton, D.I., 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. Biometrics 69, 274–281.

Reusch, T.B.H., Langefors, Å., 2005. Inter- and intralocus recombination drive MHC class IIB gene diversification in a Teleost, the Three-spined stickleback *Gasterosteus aculeatus*. J. Mol. Evol. 61, 531–541.

Reusch, T.B.H., Schaschl, H., Wegner, K.M., 2004. Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. Immunogenetics 56, 427–437.

Rhodes, J.R., Wiegand, T., Mcalpine, C.A., Callaghan, J., Lunney, D., Bowen, M., Possingham, H.P., 2006. Modeling species' distributions to improve conservation in Semiurban landscapes: Koala case study. Conserv. Biol. 20, 449–459.

Riley, W.D., Maxwell, D.L., Pawson, M.G., Ives, M.J., 2009. The effects of low summer flow on wild salmon (*Salmo salar*), trout (*Salmo trutta*) and grayling (*Thymallus thymallus*) in a small stream. Freshw. Biol. 54, 2581–2599.

Robasky, K., Lewis, N.E., Church, G.M., 2014. The role of replicates for error mitigation in next-generation sequencing. Nat. Rev. Genet. 15, 56–62.

Robert, A., 2009. Captive breeding genetics and reintroduction success. Biol. Conserv. 142, 2915–2922.

Roberts, S.C., Gosling, L.M., 2003. Genetic similarity and quality interact in mate choice decisions by female mice. Nat. Genet. 35, 103–106.

Rodgers, J.R., Cook, R.G., 2005. MHC class Ib molecules bridge innate and acquired immunity. Nat. Rev. Immunol. 5, 459–471.

Rodríguez, J.P., Brotons, L., Bustamante, J., Seoane, J., 2007. The application of predictive modelling of species distribution to biodiversity conservation. Divers. Distrib. 13, 243–251.

Rosenberg, E. and Zilber-Rosenberg, I., 2016. Microbes drive evolution of animals and plants: the hologenome concept. MBio, 7(2), pp.e01395-15.

Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.

Ryman, N., Laikre, L., 1991. Effects of supportive breeding on the genetically effective population size. Conserv. Biol. 5, 325–329.

Saper, M.A., Bjorkman, P.J., Wiley, D.C., 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. J. Mol. Biol. 219, 277–319.

Satake, A., Araki, H., 2012. Stocking of captive-bred fish can cause long-term population decline and gene pool replacement: predictions from a population dynamics model incorporating density-dependent mortality. Theor. Ecol. 5, 283–296.

Sato, A., Figueroa, F., Murray, B.W., Málaga-Trillo, E., Zaleska-Rutczynska, Z., Sültmann, H., Toyosawa, S., Wedekind, C., Steck, N., Klein, J., 2000. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics 51, 108–116.

Savage, A.E., Zamudio, K.R., 2011. MHC genotypes associate with resistance to a frog-killing fungus. Proc. Natl. Acad. Sci. 108, 16705–16710.

Schad, J., Sommer, S., Ganzhorn, J.U., 2004. MHC variability of a small lemur in the littoral forest fragments of southeastern Madagascar. Conserv. Genet. 5, 299–309.

Schierup, M.H., Vekemans, X., Charlesworth, D., 2000. The effect of subdivision on variation at multi-allelic loci under balancing selection. Genet. Res. 76, 51–62.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Horn, D.J.V., Weber, C.F., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.

Schneider, S.C., 2011. Impact of calcium and TOC on biological acidification assessment in Norwegian rivers. Sci. Total Environ. 409, 1164–1171.

Schöffmann, J., 2000. The grayling species (*Thymallinae*) of three different catchment areas of Mongolia. Grayl. Soc. J. Spring.

Schwartz, M.K., Luikart, G., Waples, R.S., 2007. Genetic monitoring as a promising tool for conservation and management. Trends Ecol. Evol. 22, 25–33.

Scott, A., 1985. Distribution, growth, and feeding of postemergent grayling *Thymallus thymallus* in an English river. Trans. Am. Fish. Soc. 114, 525–531.

Sempeski, P., Gaudin, P., 1996. Size-related shift in feeding strategy and prey-size selection in young grayling (*Thymallus thymallus*). Can. J. Zool. 74, 1597–1603.

Sempeski, P., Gaudin, P., 1995. Habitat selection by grayling—I. Spawning habitats. J. Fish Biol. 47, 256–265.

Sempeski, P., Gaudin, P., Persat, H., Grolet, O., 1995. Diet selection in early-life stages of grayling (*Thymallus thymallus*).

Sepil, I., Moghadam, H.K., Huchard, E., Sheldon, B.C., 2012. Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. BMC Evol. Biol. 12, 68.

Setchell, J.M., Charpentier, M.J.E., Abbott, K.M., Wickings, E.J., Knapp, L.A., 2010. Opposites attract: MHC-associated mate choice in a polygynous primate. J. Evol. Biol. 23, 136–148.

Sevellec, M., Pavey, S.A., Boutin, S., Filteau, M., Derome, N., Bernatchez, L., 2014. Microbiome investigation in the ecological speciation context of lake whitefish (*Coregonus clupeaformis*) using next-generation sequencing. J. Evol. Biol. 27, 1029–1046.

Sgrò, C.M., Lowe, A.J., Hoffmann, A.A., 2011. Building evolutionary resilience for conserving biodiversity under climate change. Evol. Appl. 4, 326–337.

Shafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K.D., Fior, S., Hajibabaei, M., Hill, J.A., Hoezel, A.R., Höglund, J., Jensen, E.L., Krause, J., Kristensen, T.N., Krützen, M., McKay, J.K., Norman, A.J., Ogden, R., Österling, E.M., Ouborg, N.J., Piccolo, J., Popović, D., Primmer, C.R., Reed, F.A., Roumet, M., Salmona, J., Schenekar, T., Schwartz, M.K., Segelbacher, G., Senn, H., Thaulow, J., Valtonen, M., Veale, A., Vergeer, P., Vijay, N., Vilà, C., Weissensteiner, M., Wennerström, L., Wheat, C.W., Zieliński, P., 2015. Genomics and the challenging translation into conservation practice. Trends Ecol. Evol. 30, 78–87.

Shannon, C.-W., Weaver, W., 1948. W (1949) The mathematical Theory of Communication. Press UoI Ed.

Shiina, T., Dijkstra, J.M., Shimizu, S., Watanabe, A., Yanagiya, K., Kiryu, I., Fujiwara, A., Nishida-Umehara, C., Kaba, Y., Hirono, I., Yoshiura, Y., Aoki, T., Inoko, H., Kulski, J.K., Ototake, M., 2005. Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (*Oncorhynchus mykiss*), a species with a presumably recent tetraploid ancestry. Immunogenetics 56, 878–893.

Shiina, T., Shimizu, S., Hosomichi, K., Kohara, S., Watanabe, S., Hanzawa, K., Beck, S., Kulski, J.K., Inoko, H., 2004. Comparative Genomic Analysis of Two Avian (Quail and Chicken) MHC Regions. J. Immunol. 172, 6751–6763.

Shum, B.P., Guethlein, L., Flodin, L.R., Adkison, M.A., Hedrick, R.P., Nehring, R.B., Stet, R.J.M., Secombes, C., Parham, P., 2001. Modes of salmonid MHC class I and II evolution differ from the primate paradigm. J. Immunol. 166, 3297–3308.

Slatkin, M., 1994. An exact test for neutrality based on the Ewens sampling distribution. Genet. Res. 64, 71–74.

Smith, K.F., Acevedo-Whitehouse, K., Pedersen, A.B., 2009. The role of infectious diseases in biological conservation. Anim. Conserv. 12, 1–12.

Smith, K.F., Sax, D.F., Lafferty, K.D., 2006. Evidence for the role of infectious disease in species extinction and endangerment. Conserv. Biol. 20, 1349–1357.

Solomon, D.J., Sambrook, H.T., 2004. Effects of hot dry summers on the loss of Atlantic salmon, *Salmo salar*, from estuaries in South West England. Fish. Manag. Ecol. 11, 353–363.

Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2, 16.

Sommer, S., Courtiol, A., Mazzoni, C.J., 2013. MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics 14, 542.

Sovová, T., Boyle, D., Sloman, K.A., Vanegas Pérez, C., Handy, R.D., 2014. Impaired behavioural response to alarm substance in rainbow trout exposed to copper nanoparticles. Aquat. Toxicol. 152, 195–204.

Spielman, D., Brook, B.W., Frankham, R., 2004. Most species are not driven to extinction before genetic factors impact them. Proc. Natl. Acad. Sci. U. S. A. 101, 15261–15264.

Spor, A., Koren, O., Ley, R., 2011. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat. Rev. Microbiol. 9, 279–290.

Spurgin, L.G., Richardson, D.S., 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. R. Soc. B Biol. Sci. 277, 979–988.

Stanton, J.C., Pearson, R.G., Horning, N., Ersts, P., Reşit Akçakaya, H., 2012. Combining static and dynamic variables in species distribution models under climate change. Methods Ecol. Evol. 3, 349–357.

Star, B., Nederbragt, A.J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T.F., Rounge, T.B., Paulsen, J., Solbakken, M.H., Sharma, A., Wetten, O.F., Lanzén, A., Winer, R., Knight, J., Vogel, J.-H., Aken, B., Andersen, Ø., Lagesen, K., Tooming-Klunderud, A., Edvardsen, R.B., Tina, K.G., Espelund, M., Nepal, C., Previti, C., Karlsen, B.O., Moum, T., Skage, M., Berg, P.R., Gjøen, T., Kuhl, H., Thorsen, J., Malde, K., Reinhardt, R., Du, L., Johansen, S.D., Searle, S., Lien, S., Nilsen, F., Jonassen, I., Omholt, S.W., Stenseth, N.C., Jakobsen, K.S., 2011. The genome sequence of Atlantic cod reveals a unique

immune system. Nature 477, 207–210.

Stet, R. J. M., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B.P., Grimholt, U, 2002. Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. Immunogenetics 54, 320–331.
Stet, R.J.M., Kruiswijk, C.P., Dixon, B., 2003. Major histocompatibility lineages and immune gene function in Teleost fishes: The road not taken. Crit. Rev. Immunol. 23, 473–488.

Stinchcombe, J.R., Hoekstra, H.E., 2007. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100, 158–170.

Strandh, M., Westerdahl, H., Pontarp, M., Canbäck, B., Dubois, M.-P., Miquel, C., Taberlet, P., Bonadonna, F., 2012. Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. Proc. R. Soc. Lond. B Biol. Sci. 279, 4457–4463.

Sullivan, L.C., Hoare, H.L., McCluskey, J., Rossjohn, J., Brooks, A.G., 2006. A structural perspective on MHC class Ib molecules in adaptive immunity. Trends Immunol. 27, 413–420.

Sunnucks, P., 2000. Efficient genetic markers for population biology. Trends Ecol. Evol. 15, 199–203.

Sutton, J.T., Nakagawa, S., Robertson, B.C., Jamieson, I.G., 2011. Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. Mol. Ecol. 20, 4408–4420.

Sutton, J.T., Robertson, B.C., Jamieson, I.G., 2015. MHC variation reflects the bottleneck histories of New Zealand passerines. Mol. Ecol. 24, 362–373.

Sutton, R., Soto, T., 2012. Juvenile coho salmon behavioural characteristics in Klamath river summer thermal refugia. River Res. Appl. 28, 338–346.

Sutton, R.J., Deas, M.L., Tanaka, S.K., Soto, T., Corum, R.A., 2007. Salmonid observations at a Klamath River thermal refuge under various hydrological and meteorological conditions. River Res. Appl. 23, 775–785.

Swatdipong, A., Primmer, C.R., Vasemägi, A., 2010. Historical and recent genetic bottlenecks in European grayling, *Thymallus thymallus*. Conserv. Genet. 11, 279–292.

Synes, N.W., Osborne, P.E., 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. Glob. Ecol. Biogeogr. 20, 904–914.

Szulkin, M., Bierne, N., David, P., 2010. Heterozygosity-fitness correlations: a time for reappraisal. Evolution 64, 1202–1217.

Takahata, N., 1995. MHC diversity and selection. Immunol. Rev. 143, 225–247.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. mst197.

Taylor, J.S., Raes, J., 2004. duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38, 615–643.

Tellier, A., Moreno-Gámez, S., Stephan, W., 2014. Speed of adaptation and genomic footprints of host–parasite coevolution under arms race and trench warfare dynamics. Evolution 68, 2211–2224.

Thalwitzer, S., Wachter, B., Robert, N., Wibbelt, G., Müller, T., Lonzer, J., Meli, M.L., Bay, G., Hofer, H., Lutz, H., 2010. Seroprevalences to viral pathogens in free-ranging and captive Cheetahs (*Acinonyx jubatus*) on Namibian farmland. Clin. Vaccine Immunol. 17, 232–238.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Townsend Peterson, A., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145–148.

Thuiller, W., Albert, C., Araújo, M.B., Berry, P.M., Cabeza, M., Guisan, A., Hickler, T., Midgley, G.F., Paterson, J., Schurr, F.M., Sykes, M.T., Zimmermann, N.E., 2008. Predicting global change impacts on plant species' distributions: Future challenges. Perspect. Plant Ecol. Evol. Syst., Space matters - Novel developments in plant ecology through spatial modelling 9, 137–152.

Thurber, R.L.V., Barott, K.L., Hall, D., Liu, H., Rodriguez-Mueller, B., Desnues, C., Edwards, R.A., Haynes, M., Angly, F.E., Wegley, L., Rohwer, F.L., 2008. Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral *Porites compressa*. Proc. Natl. Acad. Sci. 105, 18413–18418.

Tilman, D., 1980. Resources: A Graphical-Mechanistic Approach to Competition and Predation. Am. Nat. 116, 362–393.

Tomasso, J.R., 1986. Comparative toxicity of nitrite to freshwater fishes. Aquat. Toxicol. 8, 129–137.

Tosar, J.P., Rovira, C., Naya, H., Cayota, A., 2014. Mining of public sequencing databases supports a non-dietary origin for putative foreign miRNAs: underestimated effects of contamination in NGS. RNA 20, 754–757.

Traherne, J.A., 2008. Human MHC architecture and evolution: implications for disease association studies. Int. J. Immunogenet. 35, 179–192.

Tyler, S.J., Ormerod, S.J., 1991. The influence of stream acidification and riparian land-use on the breeding biology of Grey Wagtails *Motacilla cinerea* in Wales. Ibis 133, 286–292.

Uiblein, F., Jagsch, A., Honsig-Erlenburg, W., Weiss, S., 2001. Status, habitat use, and vulnerability of the European ggrayling in Austrian waters. J. Fish Biol. 59, 223–247.

Ujvari, B., Belov, K., 2011. Major Histocompatibility Complex (MHC) Markers in Conservation Biology. Int. J. Mol. Sci. 12, 5168–5186.

van Dijk, G.M., van Liere, L., Admiraal, W., Bannink, B.A., Cappon, J.J., 1994. Present state of the water quality of European rivers and implications for management. Sci. Total Environ. 145, 187–195.

Vásquez-Carrillo, C., Friesen, V., Hall, L., Peery, M.Z., 2014. Variation in MHC class II B genes in marbled murrelets: implications for delineating conservation units. Anim. Conserv. 17, 244–255.

Verbrugghe, E., Boyen, F., Gaastra, W., Bekhuis, L., Leyman, B., Van Parys, A., Haesebrouck, F., Pasmans, F., 2012. The complex interplay between stress and bacterial infections in animals. Vet. Microbiol. 155, 115–127.

Verschuere, L., Rombaut, G., Sorgeloos, P., Verstraete, W., 2000. Probiotic bacteria as biological control agents in aquaculture. Microbiol. Mol. Biol. Rev. 64, 655–671.

Vetemaa, M., Eschbaum, R., Albert, A., Saks, L., Verliin, A., Jürgens, K., Kesler, M., Hubel, K., Hannesson, R., Saat, T., 2010. Changes in fish stocks in an Estonian estuary: overfishing by cormorants? ICES J. Mar. Sci. J. Cons. 67, 1972–1979.

Vilà, C., Sundqvist, A.-K., Flagstad, Ø., Seddon, J., Rnerfeldt, S.B., Kojola, I., Casulli, A., Sand, H., Wabakken, P., Ellegren, H., 2003. Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. Proc. R. Soc. Lond. B Biol. Sci. 270, 91–97.

Vitti, J.J., Grossman, S.R., Sabeti, P.C., 2013. Detecting natural selection in genomic data. Annu. Rev. Genet. 47, 97–120.

von Salomé, J., Kukkonen, J.P., 2008. Sequence features of HLA-DRB1 locus define putative basis for gene conversion and point mutations. BMC Genomics 9, 1.

Vuorinen PJ, Keinaenen M., Peuranen M., Tigerstedt C, 1998. Effects of iron, aluminium, dissolved humic material and acidity on grayling (*Thymallus thymallus*) in laboratory exposures, and a comparison of sensitivity with brown trout (*Salmo trutta*). Boreal Environment Research pp.405-419.

Wade, M.J., 2007. The co-evolutionary genetics of ecological communities. Nat. Rev. Genet. 8, 185–195.

Wahlund, S., 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas 11, 65–106.

Waller, D.M., 2015. Genetic rescue: a safe or risky bet? Mol. Ecol. 24, 2595–2597.

Wan, J., Wang, C., Han, S., Yu, J., 2014. Planning the priority protected areas of endangered orchid species in northeastern China. Biodivers. Conserv. 23, 1395–1409.

Wang, L., Espinoza, H.M., Gallagher, E.P., 2013. Brief exposure to copper induces apoptosis and alters mediators of olfactory signal transduction in coho salmon. Chemosphere 93, 2639–2643.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267.

Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21, 335–342.

Warren, D.L., Seifert, S.N., 2010. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21, 335–342.

Warren, D.L., Wright, A.N., Seifert, S.N., Shaffer, H.B., 2014. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. Divers. Distrib. 20, 334–343.

Watterson, G.A., 1986. The homozygosity test after a change in population size. Genetics 112, 899–907.

Watterson, G.A., 1978. The homozygosity test of neutrality. Genetics 88, 405–417.

Webb, B.W., Clack, P.D., Walling, D.E., 2003. Water–air temperature relationships in a Devon river system and the role of flow. Hydrol. Process. 17, 3069–3084.

Wedekind, C., Evanno, G., Székely, T., Pompini, M., Darbellay, O., Guthruf, J., 2013. Persistent unequal sex ratio in a population of Grayling (*Salmonidae*) and possible role of temperature increase. Conserv. Biol. 27, 229–234.

Wedekind, C., Küng, C., 2010. Shift of spawning season and effects of climate warming on developmental stages of a Grayling (*Salmonidae*). Cambio En Época Desove Ef. Calentamiento Climático Sobre Etapas Desarro. Un Tímalo Salmonidae 24, 1418–1423.

Wedekind, C., Müller, R., Spicher, H., 2001. Potential genetic benefits of mate selection in whitefish. J. Evol. Biol. 14, 980–986.

Wehrly, K.E., Wang, L., Mitro, M., 2007. Field-based estimates of thermal tolerance limits for trout: incorporating exposure time and temperature fluctuation. Trans. Am. Fish. Soc. 136, 365–374.

Weijters, M.J., Janse, J.H., Alkemade, R., Verhoeven, J.T.A., 2009. Quantifying the effect of catchment land use and water nutrient concentrations on freshwater river and stream biodiversity. Aquat. Conserv. Mar. Freshw. Ecosyst. 19, 104–112.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the analysis of population structure. Evolution 38, 1358–1370.

Weiss, S., Persat, H., Eppe, R., Schlötterer, C., Uiblein, F., 2002. Complex patterns of colonization and refugia revealed for European grayling *Thymallus thymallus*, based on complete sequencing of the mitochondrial DNA control region. Mol. Ecol. 11, 1393–1407.

Weiss, S.J., Kopun, T., Sušnik Bajec, S., 2013. Assessing natural and disturbed population structure in European grayling, *Thymallus thymallus*: melding phylogeographic, population genetic and jurisdictional perspectives for conservation planning. J. Fish Biol. 82, 505–521.

Wenger, S.J., Isaak, D.J., Luce, C.H., Neville, H.M., Fausch, K.D., Dunham, J.B., Dauwalter, D.C., Young, M.K., Elsner, M.M., Rieman, B.E., Hamlet, A.F., Williams, J.E., 2011. Flow regime, temperature, and biotic interactions drive differential declines of trout species under climate change. Proc. Natl. Acad. Sci. 108, 14175–14180.

Whiteley, A.R., Fitzpatrick, S.W., Funk, W.C., Tallmon, D.A., 2015. Genetic rescue to the rescue. Trends Ecol. Evol. 30, 42–49.

Whitham, T.G., Bailey, J.K., Schweitzer, J.A., Shuster, S.M., Bangert, R.K., LeRoy, C.J., Lonsdorf, E.V., Allan, G.J., DiFazio, S.P., Potts, B.M., Fischer, D.G., Gehring, C.A., Lindroth, R.L., Marks, J.C., Hart, S.C., Wimp, G.M., Wooley, S.C., 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nat. Rev. Genet. 7, 510–523.

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: Assessing the assumptions and uncertainties. Proc. Natl. Acad. Sci. 106, 19729–19736.

Willi, Y., Van Buskirk, J., Hoffmann, A.A., 2006. Limits to the adaptive potential of small populations, in: Annual Review of Ecology Evolution and Systematics. Annual Reviews, Palo Alto, pp. 433–458.

Williams, S.E., Hoffman, E.A., 2009. Minimizing genetic adaptation in captive breeding programs: A review. Biol. Conserv. 142, 2388–2400.

Williams, S.E., Shoo, L.P., Isaac, J.L., Hoffmann, A.A., Langham, G., 2008. Towards an integrated framework for assessing the vulnerability of species to climate change. PLOS Biol 6, e325.

Williamson, K.S., Murdoch, A.R., Pearsons, T.N., Ward, E.J., Ford, M.J., 2010. Factors influencing the relative fitness of hatchery and wild spring Chinook salmon (*Oncorhynchus tshawytscha*) in the Wenatchee River, Washington, USA. Can. J. Fish. Aquat. Sci. 67, 1840–1851.

Wimp, G.M., Young, W.P., Woolbright, S.A., Martinsen, G.D., Keim, P., Whitham, T.G., 2004. Conserving plant genetic diversity for dependent animal communities. Ecol. Lett. 7, 776–780.

Winfield, I.J., Crawshaw, D.H., Durie, N.C., 2003. Management of the cormorant, Phalacrocorax carbo, and endangered whitefish, Coregonus lavaretus, populations of Haweswater, UK. Interact. Fish Birds Implic. Manag. 335–344.

Witkowski, A., Kowalewski, M., 1988. Migration and structure of spawning population of European grayling *Thymallus thymallus* (L.) in the Dunajec basin. Arch. Für Hydrobiol. 112, 279–297.

Woolland, J.V., 1986. Grayling in the Welsh Dee. Part 2. Movements. J. Grayl. Soc. 35–37.

Wright, L.I., Tregenza, T., Hosken, D.J., 2008. Inbreeding, inbreeding depression and extinction. Conserv. Genet. 9, 833.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.

Xu, J., 2006. INVITED REVIEW: Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. Mol. Ecol. 15, 1713–1731.

Yang, D., Liu, Q., Yang, M., Wu, H., Wang, Q., Xiao, J., Zhang, Y., 2012. RNA-seq liver transcriptome analysis reveals an activated MHC-I pathway and an inhibited MHC-II pathway at the early stage of vaccine immunization in zebrafish. BMC Genomics 13, 319.

Youngson, A.F., Jordan, W.C., Verspoor, E., McGinnity, P., Cross, T., Ferguson, A., 2003. Management of salmonid fisheries in the British Isles: towards a practical approach based on population genetics. Fish. Res., The Scientific Basis for Management of Salmonid stocks in the British Isles 62, 193–209.

Zagalska-Neubauer, M., Babik, W., Stuglik, M., Gustafsson, L., Cichoń, M., Radwan, J., 2010. 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. BMC Evol. Biol. 10, 395.

Zangenberg, G., Huang, M.-M., Arnheim, N., Erlich, H., 1995. New HLA–DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nat. Genet. 10, 407–414.

Zhang, J., Chiodini, R., Badr, A., Zhang, G., 2011. The impact of next-generation sequencing on genomics. J. Genet. Genomics Yi Chuan Xue Bao 38, 95–109.

Zhang, Y.X., Chen, S.L., Liu, Y.G., Sha, Z.X., Liu, Z.J., 2006. Major Histocompatibility Complex Class IIB Allele Polymorphism and Its Association with Resistance/Susceptibility to *Vibrio anguillarum* in Japanese Flounder (*Paralichthys olivaceus*). Mar. Biotechnol. 8, 600–610.

Zilber-Rosenberg, I., Rosenberg, E., 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol. Rev. 32, 723–735.

Zinkernagel, R.M., Doherty, P.C., 1974. Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. Nature 248, 701–702.

.