Genes under selection: Mhc and others

Jacob Höglund

DOI: 10.1093/acprof:oso/9780199214211.003.0005

Abstract and Keywords

This chapter focuses on genes under selection. Much of what is known about ‘ecologically relevant’ genetic variation at the level of DNA sequences comes from studies of genes of the major histocompatibility complex (Mhc genes). This gene family codes for cell-surface proteins involved in immunoresistance in vertebrates. The chapter briefly reviews evidence for selection on Mhc loci, links to parasite resistance, and consequences of lost genetic variation at Mhc loci. It also considers other candidate genes. Examples of such that may be relevant in conservation are genes coding for animal pigmentation (such as mc1r) and clock genes (involved in photoperiodism). It is shown that selection may both maintain genetic variation, through balancing selection, and erode it, through purifying and directional selection.

Keywords: major histocompatibility complex, gene family, genetic variation, selection, conservation, pigmentation genes, photoperiodism
Genes under selection: Mhc and others

For good reasons workers in the field of molecular population genetics have by tradition used neutral genetic markers to study evolutionary processes. By being able to ignore selection, such markers have allowed estimation of the strength and importance of mutation and recombination, genetic drift, and migration in shaping genetic diversity among and within populations. However, variation at neutral loci cannot provide direct information on selective processes involved in the interaction between individuals and their environment, nor on the capacity for future adaptive changes (Meyers and Bull 2002, van Tienderen et al. 2002, Sommer 2005). The mechanisms that maintain and promote adaptive genetic diversity in natural populations is a central issue in evolutionary ecology and conservation (Orr and Coyne 1992, Hedrick 2001, Boake 2002, Sommer 2005). How adaptive genetic diversity is apportioned across both space and time provides insight into how adaptation may progress under novel or changing environmental conditions, and the extent to which populations may be prone to stochastic extinction through the erosion of genetic diversity. Such an endeavour would need the study of coding genes and the regulatory mechanisms that underlie adaptive phenotypes in natural populations.

That genetic diversity has been estimated by using neutral genetic markers was also partly driven by the fact that coding loci were hard to access in non-model species. In the past, it was often assumed that neutral and adaptive variation are correlated (Hedrick 2002). Although the relationship may sometimes hold, the correlation between neutral and adaptive genetic diversity is usually rather weak (Hedrick 2001) and sometimes even absent (e.g. Madsen et al. 2000).

This chapter will focus on genes under selection. Much of what is known about ‘ecologically relevant’ genetic variation at the level of DNA sequences comes from studies of genes of the major histocompatibility complex (Mhc genes). This gene family codes for cell-surface proteins involved in immunoresistance in vertebrates. I will briefly, since there are a number of excellent reviews on this topic, review evidence for selection on Mhc loci, links to parasite resistance, and consequences of lost genetic variation at Mhc loci. Not all immune genes belong to the Mhc family and there is a growing concern that immunoeological studies should address other immunity genes (Acevedo-Whitehouse and Cunningham 2006).
At the end of the chapter other candidate genes will also be covered. Examples of such that may be relevant in conservation are genes coding for animal pigmentation (such as \textit{mc1r}) and \textit{clock} genes (involved in photoperiodism).

5.1 \textit{Mhc} genes
In 2006 Piertney and Oliver stated that ‘our understanding of how selection can act to maintain adaptive polymorphism in natural populations remains based on a small number of key gene regions, such as the major histocompatibility complex (Mhc)’. This cluster of genes has been extensively studied in both model and non-model species during the last decades (see reviews by Brown and Eklund 1994, Apanius et al. 1997, Edwards and Hedrick 1998, Jordan and Bruford 1998, Penn and Potts 1998, 1999, Tregenza and Wedell 2000, Zelano and Edwards 2002, Bernatchez and Landry 2003, Garrigan and Hedrick 2003, Mays and Hill 2004, Ziegler et al. 2005, Piertney and Oliver 2006, Sommer 2005).

Mhc genes are among the best candidates for the study of adaptive genetic diversity as they are extraordinarily variable and of obvious ecological relevance. The cell-surface proteins encoded by Mhc class I are found on all cells and bind to epitopes from antigens derived from intracellular pathogens, such as viruses, and present these on the cell surface (Fig. 5.1). Class II molecules are only found on specialized immune cells, for example macrophages, that engulf extracellular parasites and bind epitopes derived from such extracellular pathogens. These can then be recognized by the helper cells that trigger the production of specific antibodies by B cells. In this process MHC class II molecules are involved in the signalling between B and T cells. Both molecules are therefore important in the triggering of the adaptive immune response. Hence there is a direct link between Mhc genes and individual fitness. Furthermore, vertebrate Mhc genes are among the most variable loci known in humans, with over 500 alleles found at a single locus (Robinson et al. 2003).

There is considerable variation in the organization and size of the Mhc among vertebrates. In humans, the Mhc complex contains 421 loci (Horton et al. 2004). In domestic chicken the classical region (BF/BL) is much smaller—about 20 genes—and is therefore sometimes referred to as the minimal essential Mhc (Kaufman et al. 1999, Kaufman 2000). Because the ability of MHC to bind to broad arrays of pathogens is related to a high allelic sequence variation in the region coding for the antigen-binding sites (Doherty and Zinkernagel 1975), this high level of polymorphism is likely to be
Genes under selection: Mhc and others

maintained by balancing selection resulting from heterozygote or rare-allele advantage (Takahata and Nei 1990). In (p.83)

addition, MHC disassortative mating preferences (Landry and Bernatchez 2001, Penn 2002, Zelano and Edwards 2002, Milinski 2006), as well as prenatal foetal incompatibilities in mammals (Ober 1999), can contribute to the maintenance of extreme levels of polymorphism.

One approach to studying selection at Mhc loci has been to identify balancing selection in the current generation. The tools used have been observed deviations from Hardy-Weinberg equilibria, Mendelian expectations, or expectations about random associations (Garrigan and Hedrick 2003). Furthermore, associations have been looked for between specific genotypes and fitness on exposure to certain environments. Associations between specific Mhc alleles and disease resistance or susceptibility have been found in a number of species including humans (Sommer 2005). When looking at Mhc evolution over evolutionary times the most common approach has been to examine the ratio of non-synonymous to synonymous substitutions (dN/dS) in the sequences coding for the molecule. There are two hypotheses, both involving balancing selection, to explain the variation in

Figure 5.1 Schematic picture of MHC class I and II molecules. To the left, the molecules are seen from the side with the cell surfaces at the bottom. Antigen-binding sites are shown by the black areas and the approximate positions of α and β chains are indicated. To the right the molecules are shown from the top with the antigen-binding sites in black.
the Mhc genes. These are (1) heterozygote advantage and/or (2) frequency-dependent selection in response to parasites and pathogens (reviewed in Penn and Potts 1998, Hedrick 2002). There is not yet any consensus on which of these hypotheses is more important, although present evidence seems to lean towards some form of frequency-dependent selection (Sommer 2005, Hedrick 2006). Both mechanisms can explain why Mhc diversity is often high, even in species or populations were neutral markers indicate a loss of genetic variation due to random genetic drift (e.g. Aguilar et al. 2004, van Oosterhout et al. 2006).

It should be clear from the above that exactly how pathogens maintain a high level of Mhc diversity is still debated and the issue needs further investigation (Penn 2002, Zelano and Edwards 2002, Milinski 2006, Piertney and Oliver 2006). To clarify these issues, isolation of Mhc markers in non-model species is needed. This has until recently been hampered by interspecific variation in Mhc architecture. Since species vary considerably in the number of functional and non-functional Mhc genes, an important prerequisite to studying MHC diversity is to know how many duplications of Mhc genes are present in the species of interest, and whether or not these loci are expressed (Strand et al. 2007). Incomplete knowledge may lead to misleading conclusions, for instance if variation in pseudo-genes is associated with ecological factors. There is thus a need to understand how Mhc diversity is selected for and maintained in natural populations. Studies of associations between Mhc diversity, or MHC profile, and condition parameters (and mate choice) are frequent in the recent Mhc literature and include mammals, birds, and fish (Piertney and Oliver 2006). To better understand these interactions Mhc genes other than those coding for the classically studied MHC class II need to be targeted (Acevedo-Whitehouse and Cunningham 2006).
5.1.1 Mhc and conservation in mammals

The MHC was first discovered in humans in the 1950s in studies on skin graft rejection. The link to immunology was soon detected and at present more than 420 genes are known in this gene complex, of which 252 are expressed; about 70 of these are potentially associated with immunity (Beck and Trowsdale 1999). In humans the Mhc genes reside on chromosome 6 but may be regulated by genes located on other chromosomes (Reith and Mach 2001). It appears that Mhc structure and organization is quite similar in our close relatives, the great apes, with humans and chimpanzees sometimes sharing the same alleles. This trans-species polymorphism is a common observation in many mammalian studies (Klein et al. 1998, Garrigan and Hedrick 2003) and is explained by balancing selection maintaining variation for long periods. As a consequence, often the most similar Mhc sequence is not in the same species but in a related one (Hedrick 2006). In artiodactyls, balancing selection appears to have maintained allelic lineages for over 20 million years (Gutierrez-Espeleta et al. 2001).

Within-species genetic variation at Mhc loci can either be similar to that at neutral loci or, because of past balancing selection, exceed the neutral variation (the third possibility that neutral variation exceeds Mhc variation is to my knowledge never observed). Historical demographic events have been implicated to explain why Swedish beaver, Castor fiber, and moose, Alces alces, possess a low number of Mhc alleles (Ellegren et al. 1993, 1996, Mikko and Andersson 1995; see also Sommer 2005 for a list of similar examples, and Babik et al. 2005 for more on low Mhc diversity in Eurasian beavers). Bottlenecks and founder effects have according to this explanation been stronger than the power of selection in shaping current levels of Mhc variation. In these cases the reduced Mhc polymorphism is thus correlated with low genome-wide genetic variation (Hedrick 2002). African cheetahs, Aonyx jubatus, have been cited as the prime example in which low Mhc diversity correlates with a genome-wide loss of diversity, presumably due to a genetic bottleneck about 10000 years ago (O’Brien et al. 1985). However, the details of this case have been debated. Another famous example is the Northern elephant seal, Mirounga angustirostris, which became almost extinct due to hunting about 100 years ago. This species is low in presumably neutral...
Genes under selection: Mhc and others

allozymes, mitochondrial DNA, mini- and microsatellite loci, as well as adaptive Mhc class II genetic variation (Hoelzel et al. 1999, Weber et al. 2004).

Other cases (reviewed by Sommer 2005) show that Mhc diversity may be maintained, at least for some time, despite the species being subject to loss of overall genetic variation. Although a direct link between pathogen-mediated population decline and low Mhc variation has been difficult to demonstrate in natural populations (Guiterrez-Espelata et al. 2001), recent studies have indicated that although Mhc allele numbers are low in many bottlenecked species, a high degree of divergence between alleles can still be observed. Moreover, genetic diversity at antigen-binding sites exceeds that at other Mhc codons in a range of threatened and fragmented species (Sommer 2005).

In summary, a few studies of mammals hint at the importance of Mhc variability in conservation (reviewed by Sommer 2005) although others indicate that species can persist, at least in the short term, despite being devoid of Mhc variability (Ellegren et al. 1993, 1996, Mikko and Andersson 1995). The importance of Mhc variability with respect to the severity of human impact is even less well studied. Theoretically one would expect a genotype-by-environment interaction whereby low variability might not lead to extinction when environmental conditions are benign, whereas adverse effects of low variability would become apparent under adverse environmental conditions due to human-induced changes such as pollution and habitat fragmentation.

5.1.2 Mhc and conservation in birds

Most of what is known of the genomic organization of Mhc in birds mostly comes from studies of the domestic chicken (Zoorob et al. 1993, Kaufman et al. 1999). Although more genomic information from other bird species is on the way, there is (p.86) a need for specifically targeted studies of the comparative genomics of bird Mhc. The chicken Mhc gene family differs from mammalian Mhc by consisting of two independently assorting clusters of genes, the B and Y (formerly Rfp-Y) regions (Miller et al. 1996). Both these regions map to microchromosome number 16 in the chicken, and both contain Mhc class I and II genes (Miller et al. 2004). The B genes are polymorphic and expressed (Goto et al. 2002) and
have been found to be correlated with resistance to several diseases in chickens (Kaufman 2000).

In chicken and some other birds there appear to be two expressed separate class II B genes (Freeman Gallant et al. 2002) but the number of both class I and II B genes may be manifold in other species (Westerdahl et al. 2000). Less is known about the Y genes. At least one Mhc class I Y locus (YF) is expressed and may be active in the immune function of the chicken (Hunt et al. 2006). However, to date it is not clear whether the Mhc class II Y (YLB) genes are functional in the chicken, as all the YLB loci mapped to date are apparently pseudo-genes (Shiina et al. 2006). The Mhc class II B (BLB) and YLB genes have only been characterized in chicken (e.g. Miller et al. 1996) and black grouse (Strand et al. 2007), but the ring-necked pheasant (Wittzell et al. 1995) and other birds also seem to have this division of Mhc class IIB genes. Studies of possible YLB genes will add to the understanding of the selection and evolution of Mhc genes in general. So far the Mhc class II studies of non-model bird species have, with the exception of our study on black grouse (Strand et al. 2007), focused on BLB or BLB-like genes (Table 5.1).

Table 5.1 Examples of MHC studies in non-model bird species.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene(s)</th>
<th>Finding</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bobwhite quail, <em>Colinus virginianus</em></td>
<td>B haplotypes</td>
<td>Polymorphism detected</td>
<td>Drake <em>et al.</em> 1999</td>
</tr>
<tr>
<td>Great snipe, <em>Gallinago media</em></td>
<td><em>Mhc class II B</em></td>
<td>Polymorphisms</td>
<td>Ekblom <em>et al.</em> 2003</td>
</tr>
<tr>
<td>Organism</td>
<td>Gene(s)</td>
<td>Finding</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------------------</td>
<td>--------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Savannah sparrow, <em>Passerculus sandwichensis</em></td>
<td><em>Mhc</em> class II B</td>
<td>Polymorphisms</td>
<td>Freeman-Gallant <em>et al.</em> 2002</td>
</tr>
<tr>
<td>Hawaiian honeycreepers (Drepanidinae)</td>
<td><em>Mhc</em> class II B and pseudogenes</td>
<td>Polymorphisms in functional genes</td>
<td>Jarvi <em>et al.</em> 2002</td>
</tr>
<tr>
<td>New Zealand robins (Petroicidae)</td>
<td><em>Mhc</em> class II B</td>
<td>Genes transcribed</td>
<td>Miller and Lambert 2004</td>
</tr>
<tr>
<td><em>Acrocephalus</em>, warblers</td>
<td><em>Mhc</em> class I</td>
<td>Polymorphisms in both inbred and outbred species</td>
<td>Westerdahl <em>et al.</em> 1999, Richardson and Westerdahl 2003</td>
</tr>
<tr>
<td>Bengalese finch, <em>Lonchura striata</em></td>
<td><em>Mhc</em> class II B</td>
<td>Presence of locus verified</td>
<td>Vincek <em>et al.</em> 1995</td>
</tr>
</tbody>
</table>

Several studies suggest that BLB genes are important in conservation. The Chatham Island black robin, *Petroica traversi*, found only, as the name indicates, on the Chatham Islands off New Zealand, is a highly inbred, endangered passerine with extremely low levels of genetic variation. Miller and Lambert (2004) investigated *Mhc* class II variation in both the black robin and its non-endangered relative, the South Island robin, *Petroica australis australis*. To test whether *Mhc* genes were under balancing selection they compared *Mhc* variation in the black robin with artificially bottlenecked populations of the South Island robin, and with their respective source populations. The black robin was monomorphic at the studied class II B loci, while both source and bottlenecked populations of South Island robin were found to have retained moderate levels of variation. Thus it was concluded that genetic drift must have outweighed balancing selection in the case of the black robin and consequently this species is extremely vulnerable to the introduction of new pathogens to the population.
The adaptive radiation and speciation of Hawaiian honeycreepers is a textbook example, but they currently face one of the highest extinction rates in the world. The introduction of avian malaria to the Hawaiian islands is thought to be a major threat to extant honeycreepers. Jarvi et al. (2004) studied class II Mhc variation in four species of honeycreeper. Phylogenetic analyses revealed two clusters of genes and the authors found that variation in one cluster was high, with \( d_N > d_S \) and levels of diversity similar to other studies of Mhc B genes in birds. The second cluster was nearly invariant, as in the studies of the Y genes in chicken and black grouse mentioned above. The presence of balancing selection was supported by transpecies polymorphisms and high \( d_N/d_S \) ratios at putative antigen-binding site codons. When comparing two species, mitochondrial DNA control region sequences were invariant in one species, but were highly variable in another. However, Mhc class II B variation appeared comparable. Thus, even though honeycreepers have been subjected to strong bottlenecks, it was concluded that balancing selection had been strong enough to maintain MHC variation.

The Galápagos Islands harbour the endemic Galápagos penguin, *Spheniscus mendiculus*, which is the only penguin that occurs on the equator. This species relies on food brought about by the nutrient-rich upwellings from the Humboldt stream and experiences severe population declines when ocean temperatures rise during so-called El Niño events, which occur irregularly. The reduced genetic diversity in this species are likely caused by the bottlenecks brought about by El Niño. Bollmer et al. (2007) characterized the amount of genetic variation at the Mhc in Galápagos penguins, and compared it with published data from other penguin species. They found that the Galápagos penguin had the lowest Mhc diversity of the eight penguin species studied. The authors explained an excess of non-synonymous mutations and a pattern of trans-specific evolution by (p.88) suggesting that balancing selection may have been acting on the penguin Mhc. Thus this case mirrors the honeycreepers in that Mhc variation seemed to be upheld despite lost variation at neutral loci.

In studies from my own research group of a threatened lek-breeding wader, the great snipe *Gallinago media*, we found a high number of Mhc alleles (50 from 175 individuals; Ekblom et al. 2007). This, together with a higher rate of non-
synonymous than synonymous substitutions in the peptide-binding sites, and high Tajima's $D$ value in certain regions of the gene, suggests a history of balancing selection (Ekblom et al. 2008). Furthermore, genetic differentiation in the $Mhc$ between two ecologically distinct distributional regions (Scandinavian mountain populations and Eastern European lowland populations) was present after statistically controlling for the effect of selectively neutral microsatellite variation (Fig. 5.2). This suggests that spatially varying selection is generating this structure and that this mechanism contributes to the balancing selection. $Mhc$ variation in great snipe can thus be seen as a form of local adaptation to different environments. If this pattern is common, the implications for conservation are important. It suggests that local populations may be adapted to the local parasite fauna and that translocations of birds between populations may not do any good to their conservation status under certain circumstances.

Again, as is the case with mammals, studies on birds are equivocal on the conservation effects of lost $Mhc$ variation. As in the case with the black robins reviewed above, species can persist despite having very little $Mhc$ variation. In other case studies in which the study species perhaps have not been bottlenecked as severely, balancing selection seems to uphold $MHC$ variation despite past and present population size reductions. Future studies are needed to resolve whether species can persist despite being reduced in $Mhc$ variation. In theory it is just a matter of time before these $Mhc$-reduced species are hit by a new pathogen to which they cannot respond. If so, such taxa are doomed to extinction.

5.1.3 $Mhc$ and conservation in reptiles and amphibians

$Mhc$ structure and variation is poorly known in reptiles, but broad taxonomic studies have involved crocodiles and tuataras (see Edwards et al. 1995, Miller et al. 2005). There is to date no complete genomic mapping for any reptilian species. Similarly, very little information on $Mhc$ variation and patterns of evolution are available for amphibians, a group known to be declining rapidly worldwide. Fungal diseases are most likely involved in these declines (Daszak et al. 1999, Pounds et al. 2006) and therefore information on $Mhc$ could contribute to devising appropriate conservation strategies. $Mhc$ class I and II has been characterized in two species of Urodela, in the
Genes under selection: Mhc and others

axolotl *Ambystoma mexicana* (Sammut et al. 1997, 1999, Laurens et al. 2001, Richman et al. 2007) and class II in the tiger salamander *Ambystoma tigrinum* (Bos and DeWoody 2005), and in a single (p.89)

... 

... 

Figure 5.2 Point estimates of pairwise genetic distance ($F_{ST}$) between populations, within the same region (W) and in different regions (B), for Mhc class II genes and microsatellites. Different estimators of $F_{ST}$ are used for the different markers ($\Phi_{ST}$ for Mhc; $R_{ST}$ and Weir–Cockerham $F_{ST}$ for microsatellites). For Mhc, estimates are larger for pairs of populations located in different regions than for estimates within region, and more so than expected from the corresponding patterns of (b, c) microsatellite pairwise estimates between regions (filled symbols) and within regions (open symbols) (from Ekblom et al. 2007, reprinted with permission from the publisher).

...
genes under selection: Mhc and others

Genetic variation for each of the marker systems. It was found that this was only borne out in the case of Mhc (Fig. 5.3). Thus, it was argued that mini- and microsatellite techniques may provide ambiguous information concerning the relationship between population size and genetic variability. This is somewhat surprising since, due to balancing selection as argued above, a marker such as the Mhc may be the one expected to be deviant from a relationship between population size and genetic variability.

Some of the authors from the study cited above have also studied Mhc class I variation in isolated Hungarian populations of meadow vipers, Vipera ursinii, and compared them with larger Ukranian populations (Ujvari et al. 2002). Genetic variability at the class I loci was lower for Hungarian snakes than for Ukranian populations. In Hungary, birth deformities, chromosomal abnormalities, and low juvenile survival was found, which strongly suggests that the Hungarian vipers are experiencing inbreeding depression. This study thus supports the notion that low Mhc variability is somehow tied to inbreeding depression. Unfortunately there is no information on overall genetic variability in these populations and therefore it is not possible to rule out the possibility that the observed inbreeding depression may be tied to an overall loss of genetic variation rather than being specifically connected with loss of MHC variability.

Figure 5.3 Mean Mhc class I band sharing and relative population size in six sand lizard populations (open circles) and five adder populations (filled circles) (from Madsen et al. 2000, reprinted with permission from the publisher).
The variability of the peptide-binding region of Mhc class II in the fire-bellied toad Bombina bombina, which is of conservation concern in at least parts of its range, has been investigated and eight distinct alleles in 20 individuals were identified (Hauswaldt et al. 2007). All substitutions but one were non-synonymous, (p.91) and many of the highly polymorphic sites corresponded with amino acid positions known to be involved in antigen binding. The level of Mhc variation found in fire-bellied toads was thus comparable with what has been found in other amphibians. Future studies are needed to resolve whether Mhc variation correlates with population size and is related to vulnerability to pathogens and extinction risk.

Similarly Babik et al. (2008) examined Mhc variation in Alpine newts Mesotriton alpestris from three allopatric population groups in Poland at the north-eastern margin of the distribution of this species. They found two putative expressed Mhc II loci with contrasting levels of variation. One locus exhibited low polymorphism. The other locus was highly polymorphic (37 alleles in 149 individuals), and showed evidence of balancing selection with populations varying substantially in allelic richness. The Mhc variation at this locus correlated with variation in microsatellites. The authors argued the observed regional differences could be explained by increased levels genetic drift with increasing distance from glacial refugia. This implies that selection and drift interplayed to produce the pattern of Mhc variation observed in marginal populations of the alpine newt and that marginal populations are more prone to extinction.

The number of studies on reptile and amphibian Mhc genes in a conservation context are too few to allow any firm conclusions. Technical advances in primer design (Hauswaldt et al. 2007) provide great promise for future studies of anuran non-model species. Such are highly relevant under the current decline in amphibians and especially relevant because the decline is most probably related to the spread of a pathogen (Berger et al. 1998, Daszak et al. 1999, 2003).
### 5.1.4 Mhc and conservation in fish

*Mhc* variation is relatively well studied in fish. In teleosts, MHC class I and II are not found on the same chromosome (Stet *et al.* 2003). By far the largest number of expressed class I and class II alleles are described for salmonids in which there is only one expressed class II locus (Grimholt *et al.* 2000). The fish model, the zebrafish *Danio rerio* and the three-spined stickleback *Gasterosteus aculeatus*, have been fully sequenced and in the three-spined stickleback *Mhc* genetics and ecology have been studied extensively (Milinski 2006). In sticklebacks there is more than one class II locus.

Genetic variation at eight microsatellite loci and sequence variation at exon 2 of the *Mhc* class II B genes in two wild populations of the Trinidadian guppy, *Poecilia reticulata*, were studied by van Oosterhout and coworkers (2006). They compared genetic variation in a small and isolated population upstream a system of rapids separating this population from a larger downstream population. Microsatellite diversity in the small population upstream was lower and the populations were genetically differentiated when considering microsatellites. (p.92) However, the two populations were not differentiated by *Mhc* and showed similar levels of allelic richness. The authors used computer simulations to suggest that the observed level of genetic variation in the two populations can be maintained with overdominant selection acting at three *Mhc* loci. This explanation requires that selection intensities varies among the populations and this is indeed what was found. Estimated selection intensities and parasite abundances suggested that large differences in selection intensity may exist between populations. Thus it is possible that high levels of *Mhc* diversity could be maintained in the small upstream population despite strong genetic drift.

*Mhc* studies on salmonids have been plentiful. There may be two reasons for this. First, there is a link between olfaction and *Mhc* and salmonids rely heavily on olfactory communication (Höglund 1961). MHC molecules are volatile and believed to be recognizable by olfaction (Wedekind and Füri 1987). Second, salmonids are of considerable economic importance in many fisheries and thus both pure and applied research resources have been avialbale for *Mhc* studies. As a result of this research, *Mhc* loci are used in the identification of

Patterns of population differentiation at neutral markers and \textit{Mhc} genes have been studied in wild Atlantic salmon, \textit{Salmo salar} (Landry and Bernatchez 2001). Variation at a \textit{Mhc} class II B locus and microsatellites were compared among 14 samples from seven different rivers and seven subpopulations within a single river system covering a variety of habitats and different geographical scales. It was shown that balancing selection was acting on the sites involved in antigen presentation and thus could explain a high level of polymorphism within populations. The comparison of population structure at \textit{Mhc} and microsatellites on large geographical scales revealed a correlation between patterns of differentiation despite important differences in habitat type among populations. This indicated that genetic drift and migration have been more important than selection in shaping population differentiation at the \textit{Mhc} locus. On the other hand there were strong discrepancies between patterns of population differentiation among the two types of marker within rivers, which suggested a role of selection in shaping population structure at this scale. Taken together these results suggest that both selection and drift are influencing \textit{Mhc} gene diversity in wild Atlantic salmon. This is a very similar result to the one found in the great snipe study reviewed above and confirms that translocations as a conservation measure should be considered with caution.

Studies of \textit{Mhc} genetic variation among populations of chinook salmon, \textit{Oncorhynchus tshawytscha}, at three class I loci and one class II locus showed that populations from different river drainages were differentiated (Miller et al. 1997). As in Atlantic salmon it appears as though \textit{Mhc} variation has been shaped by a combination of selection and genetic drift. The force of genetic drift has \textbf{(p.93)} been influenced by repeated bottlenecks and isolation by distance in separate glacial refugia. Again, these populations seem at least partly adapted to local parasitic faunas and any conservation measures should be taken with this in mind.

The conservation status of another salmonid fish, the brown trout, \textit{Salmo trutta}, varies across its distribution. Many populations are threatened by various types of human activity, like environmental degradation, harvesting, and development
of hydroelectric dams (Laikre and Ryman 1996). Campos et al. (2006) studied levels and distribution of genetic variation in nine isolated populations of Brown trout in northern Spain and tested the importance of preservation of genetic variability for the survival of a set of isolated populations from the same river drainage system. They screened genetic variation at three different markers: mitochondrial DNA, microsatellites, and the Mhc class II locus. Genetic variation was similar at Mhc loci and microsatellites: populations polymorphic for microsatellite loci were also polymorphic at the Mhc loci (Fig. 5.4). They also observed high levels of differentiation among populations. Thus, in this case genetic drift seemed to have eroded the effect of balancing selection and was seen as the predominant evolutionary force shaping genetic variation in the smaller populations.

(94) It has been noted that local brown trout populations may be highly differentiated and adapted to their local stream and thus it may be important to take genetic variation between populations into account in conservation programmes (Laikre 1999). As MHC molecules have the potential to coevolve in response to selection pressures imposed by local parasite faunas, Mhc variability may be of a special concern in maintaining local adaptations. Thus it may be argued that the basic unit for management and conservation of brown trout (or any other organism) are the
local populations (Laikre 1999). If genetic variation and population structure at the Mhc and any other locus is mainly driven by neutral processes then local adaptation is prohibited, but spatially varying selection may lead to local adaptation, especially if migration is limited. If this is the case, admixture of local populations may lead to outbreeding depression as has recently been shown in a Swedish brown trout population (Grahn and Forsberg 2008). In the River Dalälven in Sweden a hydroelectric power plant has hindered migration to previously used spawning grounds since 1915 and artificial breeding and stocking have been provided as a substitute. It is likely that local populations within the river have become admixed during this process and experimental data show that in this admixed population Mhc homozygous males have an advantage in spawning competition and in production of young. The explanation for this surprising and counterintuitive result is that the Mhc diversity in the artificially admixed population is too high and above the optimal level. Reasons for why high variation may lead to fitness loss include autoimmune responses (Nowak et al. 1992, Reusch, et al. 2001) and loss of local adaptation to prevailing conditions.

As in the case with other animals the evidence for Mhc polymorphism being maintained by balancing selection is somewhat ambiguous in fish but it seems clear that Mhc variation in many populations is indeed maintained by balancing selection imposed by local parasitic faunas. This emphasizes what was hinted by studies on other vertebrates: local adaptation in Mhc is prevalent and admixture of previously locally adapted populations may have adverse effects, as in the case of Swedish brown trout.
5.1.5 Summary: Mhc and immunogenetics in conservation

Because of the role of the MHC in the immune defence of vertebrates, Mhc variability is arguably important for the viability of natural populations. As reviewed above, many studies have shown that populations exhibiting low levels of variability at the Mhc or with certain haplotypes are susceptible to diseases and therefore prone to extinction (see also O’Brien et al. 1985, Paterson et al. 1998, Langefors et al. 2001, Arkush et al. 2002). However, other studies have presented evidence that populations with no or low variability at Mhc loci are still persisting (Slade and McCallum 1992, Ellegren et al. 1993, Seddon and Baverstock 1999, (p.95) Hedrick et al. 2000, Miller and Lambert 2004, Weber et al. 2004). The role and importance of Mhc genes in conservation are thus debated. There has been a strong focus on Mhc class II genes in many immunogenetic studies with a conservation focus. While these studies have provided much insight into disease resistance in wild and threatened populations, it is clear that conservation immunogenetic studies will benefit by including more immune genes and loci in the future (Acevedo-Whitehouse and Cunningham 2006).

An example of such a study is one on Danish brown trout populations (Jensen et al. 2008). These authors used eight neutral microsatellite loci and two microsatellite loci embedded in the sequence encoding the protein TAP (which stands for transporter associated with antigen processing) to study temporal and geographic differentiation. Tap genes encode molecules that associate with MHC class I molecules when foreign peptides are transported across the membrane of the endoplasmatic reticulum and thus are important in launching an immune response to intracellular parasites. Thus the genetic variation at these loci could be influenced by parasite- and pathogen-driven selection. The observed neutral genetic variation suggested that population structure was temporally unstable within regions, although stable over time among regions. Statistical tests designed to detect selective sweeps found evidence of selection at the two Tap markers, indicating both a regional and microgeographical effect. Moreover, signals of divergent selection among temporal samples within localities suggest that selection also might fluctuate at a temporal scale. These results suggest that immune genes other than the classical Mhc classes I and II
might be subject to selection and warrant further studies of functional polymorphism of such genes in natural populations.

5.2 Other candidate genes relevant for conservation

*Mhc* genes have been by far the most commonly studied candidate genes in the context of conservation. Other genes have been less studied, partly because relevant genomic information has been scarce in non-model species. As technical advances proceed there is no reason for not including other ecologically important genes when studying threatened and endangered species. Below I briefly review a few genes which have been studied in non-model species in an evolutionary ecology framework. Such studies obviously have a bearing on conservation issues (Segelbacher and Höglund 2008).

5.2.1 Pigmentation genes: *mc1r*

The study of animal pigmentation has a long history in ecological genetics (Hoekstra 2006). The classical studies of banding patterns in *Cepaea* snails and *Biston betularia* serve as only two examples of how the study of evolutionary genetics of coloration have played an important role in understanding how populations may adapt to local differences in selective regimes (in these two cases ultimately driven by visual predators). It is clear that pigmentation has a strong genetic component and that populations quickly can adapt to local conditions (Majerus 1998). Pigmentation genes should therefore be very relevant in a conservation context.

Although there are several types of animal pigments the most studied and well-known system is that of melanin-based pigmentation. Melanin is produced by specialized cells, so-called melanocytes. Melanin production, or melanogenesis, in vertebrates is a complex process that includes the inception, migration, and regulation of melanocytes (Jackson 1994). Melanocytes can synthesize either eumelanin or phaeomelanin, or produce no pigment at all. Increased eumelanin synthesis leads to darker skin, hair, or feathers, increased production of phaeomelanin produces red or brown phenotypes, and no melanin synthesis results in albinism (Fig. 5.5).
The physiological pathways and the genes involved in melanin-based pigmentation have recently become quite well established. In mammals the best-known pathway is the one mediated by the cell-surface protein MC1R (melanocortin 1 receptor or α melanocyte-stimulating hormone receptor). Here circulating levels of the agonist α melanocyte-stimulating hormone (αMSH) activate MC1R which triggers the production of a messenger molecule cAMP which activates a complex pathway involving tyrosinase (Tyr) and tyrosinase-related protein 1 (TyRp1), ultimately leading to synthesis of eumelanin. If Agouti, which is the inverse antagonist of αMSH, binds to MC1R, the outcome is no synthesis of melanin or phaeomelanin. Melanin synthesis is believed to follow a similar pathway in other animals but the details are less well known (Mundy 2006, Hoekstra 2006).

The mc1r gene is a short gene (the single exon extends approximately 1000bp) expressed in melanocytes in skin and developing feather buds or analogue tissues in vertebrates. In humans it is known that mutations on mc1r are often correlated with phenotypic variation such as red hair and light skin (Makova and Norton 2005). Studies linking phenotypic variation with sequence polymorphisms have also been published in both domesticated (e.g. Kerje et al. 2003, Våge et al. 2005) and wild (reviews by Mundy 2006, Hoekstra 2006) animals. Recently, mc1r evolution have been shown to evolve faster in lineages of galliforms that show more plumage dimorphism. This is probably due to varying intensities of sexual selection (Nadeau et al. 2007a).

As in the case with immune genes there is a strong focus on a single candidate gene in the study of pigmentation genes. In a recent review of pigmentation mutations segregating in wild vertebrate populations Hoekstra (2006) listed 14 studies of which 12 were on mc1r. Clearly there is a need for studies of more (p.97)
loci involved in melanin synthesis and of other pigments. For example, studies in Japanese quail, *Coturnix japonica*, have shown that there is an association between a single-nucleotide substitution in the gene encoding Tyrp1 and plumage colour (Nadeau *et al.* 2007b). To date I am aware of no studies directly relating variation in any pigmentation gene to conservation issues. However, studies of pigmentation in a conservation context would probably challenge the view that preserving genetic diversity per se is all that matters in conservation genetics. Since pigmentation is often strongly related to the ecological background of the organism there is often a match between the environment and the most optimal phenotype (Hoekstra *et al.* 2003, 2005). As in the case of Mhc, transplantation to boost numbers or genetic variation could introduce alleles that have negative consequences.
5.2.2  

**Figure 5.5** Schematic representation of the pathways regulating mammalian melanogenesis and phenotypic effects on individual hair pigment and pattern. (a) The binding of circulating α melanocyte-stimulating hormone (αMSH) to MC1R activates the synthesis of the enzyme tyrosinase (Tyr) via cAMP. Within the cell, tyrosine is oxidized to dopaquinone, a reaction catalysed by Tyr. cAMP affects the enzymatic activity of Tyr as well as the eumelanin-specific enzymes, tyrosinase-related protein 1 (Tyrp1) and dopachrome tautomerase (Dct). When all three of these enzymes are working, eumelanin (brown to black pigment) is deposited in melanosomes. However, when Agouti, the inverse agonist of MC1R, binds to MC1R with the aid of the extracellular protein Atrn, intracellular cAMP levels are repressed and this leads to production of phaeomelanin (yellow to red pigment) which is also dependent on the incorporation of cystine, whose uptake is at least partially regulated by xCT (a protein regulating cystine uptake in melanocytes; in mice it is a gene product of the Slc7a11 locus). (b) Illustration of how overall coat colour in mammals is determined by the density and distribution of melanin on individual hairs. Pigmentation on individual hairs ranges from fully pigmented with dark eumelanin to complete absence of pigment resulting in albino hairs (from Hoekstra 2006, reprinted with permission from the publisher).

**Photoperiodism:** *Clock and other genes*

Several recent studies have pointed out the circadian clock (i.e. synchronization of an organism to daily rhythms; Bell-Pedersen *et al.* 2005) as one of the aspects of animal behaviour best characterized at the molecular level (Fidler and Gwinner...
For example, in vertebrates the gene *Clock* encodes a protein that heterodimerizes with a second protein, BMAL1, to produce a transcription-activating complex which is important in the molecular control of vertebrate circadian rhythms (Panda *et al.* 2002). In humans, a single nucleotide polymorphism in *Clock* correlates with variation in sleeping-time preferences (Mishima *et al.* 2005). Not only circadian rhythms appear influenced by *Clock*; there is evidence that *Clock* polymorphisms are associated with differences in spawning times in rainbow trout, *Oncorhynchus mykiss* (Leder *et al.* 2006).

Johnsen *et al.* (2007) studied allelic variation in a region of the avian *Clock* which encodes a polyglutamine repeat (*Clk*polyQcds), in two species of passerine birds, the bluethroat, *Luscinia svecica*, which is a migrant, and the non-migratory blue tit, *Cyanistes caeruleus*. Multiple *Clk*polyQcds alleles were found within populations of both species. When testing for population differentiation they found that observed allele frequency variation among populations at the *Clk*polyQcds and at neutral microsatellite loci could not be explained by the same underlying demographic processes in blue tits. In this species allelic variation in the *Clk*polyQcds showed evidence of being maintained by selection for microevolutionary adaptation to differences in photoperiod. This could not be detected among bluethroats, possible because of low statistical power due to small sample sizes.

The allelic variation in this case was found in polyglutamine repeats in the coding region of the gene. As pointed out by Johnsen and coworkers, it has been hypothesized that the relatively high mutation rates of repeat sequences may account for rapid morphological evolution among mammals (Fondon and Garner 2004) and they suggest further that the potential of coding region repeats for rapid evolution might be selected for, as it provides plasticity in the face of fluctuating selective pressures (see also Wren *et al.* 2000). They also proposed that investigating changing frequencies of allelic variants of genes encoding circadian clock components may warrant attention in the context of adaptation to rapid climate change. When climate changes, many parameters related to biorhythms are predicted to change accordingly. So if, for example, passerine bird populations are adapted to respond to...
changes in photoperiodism to time their maximum reproductive output with a phenological peak in food abundance, such populations either have to respond genetically or face extinction (Dias and Blondel 1997).

In plants there has been a quest to find the genes involved in the regulation of ecologically important traits such as flowering time, seed set, bud set, and annual differences in growth. Like circadian and phenological rhythms in animals, flowering time is an important life-history trait that coordinates the life cycle with local environmental conditions (Roux et al. 2006). The genetic basis of flowering time in plants have been studied extensively in the model species Arabidopsis thaliana. Such work has revealed a complex network of genes involved in flowering-time regulation (Fig. 5.6). There appears to be four major pathways and many potential candidate genes (Bernier and Périlleux 2005).

(p.100) It is predicted that flowering time should correlate with latitudinal differences among populations. However, no clear patterns have been found (Stinchcombe et al. 2004, Shindo et al. 2005). It was suggested that geographical trends may be masked by other selective regimes than photoperiod that may vary locally (Roux et al. 2006). Despite the inconsistencies among studies it seems clear that photoreceptor genes are major agents of natural variation in Arabidopsis flowering and growth response as shown by genome-wide scans of association of 65 loci with latitude (Balasubramanian et al. 2006). In this study the most associated locus across 163

Figure 5.6  Simplified overall network of flowering-time regulation (from Roux et al. 2006, reprinted with permission from the publisher).
strains was *Phyc* (phytochrome C photoreceptor), suggesting that *Phyc* is under diversifying selection.

A major challenge is to transfer the results in *Arabidopsis* to other plants. In rice, genus *Oryza*, two independent floral pathways have been detected: one is mediated by *Hdf1* (heading date 1), which is an orthologue of the *Arabidopsis* gene *CO* (Constans), and the other is *Ehd1* (early heading date 1), an orthologue of *FT* (flowering time locus). Both *CO* and *FT* play important roles in the timing of flowering in *Arabidopsis*. Similarly, indel variation in a *CO* orthologue in *Brassica nigra* has been shown to be associated with variation in flowering time in this species (Österberg *et al.* 2002), and an *FT* homolog has been implicated in controlling growth rhythm in conifers (Gyllenstrand *et al.* 2007). Likewise, nucleotide polymorphisms in the *phytochrome B2* locus in aspen *Populus tremula* have been found to be associated with the timing of bud set (Ingvarsson *et al.* 2008).

However, there are also important differences between *Arabidopsis* and wild species. Slotte and coworkers (2007) used gene-expression differences between pairs of early- and late-flowering *Capsella bursa-pastoris* ecotypes and compared their responses to changes in temperature (vernalization). Using *Arabidopsis* microarrays they found differences among the ecotypes. In contrast, in *Arabidopsis FLC* (flowering time locus C) was not differentially expressed prior to vernalization and the gibberellin and photoperiodic pathways appeared similar.

The picture that seems to emerge is that photoperiodism is evolutionarily rather conserved among plants. Thus there is scope for using a candidate gene approach to studying and preserving genetic diversity in threatened plant populations. However, as shown by the complexity among biological pathways such an approach is not without complications.

5.3 Conclusions
It is a conservation genetic paradigm that genetic variation is a prerequisite for any population's ability to adapt to a changing environment. Since small and (p.101) fragmented populations are signified by low levels of genetic variation it follows that they are thus less able to adapt when conditions change. Population fragmentation and isolation thus have extremely detrimental effects on both the fitness and viability
of extant populations, and also the evolutionary potential of species (see papers in Ferrière et al. 2004). This line of reasoning may lead to the conclusion that all that matters in conservation genetics is to preserve genetic variation and the more the better. However, as has been argued in this chapter, studies on ecologically relevant candidate genes to some extent challenge this view. Of course conservation genetics should still focus on the preservation of genetic variation and on detecting the processes that are important in preserving natural populations of threatened species, but preservation of genetic diversity must be done with knowledge and caution. Thus it is important to understand local adaptation, which is the topic of the next chapter.

Most studies that have attempted to monitor genetic diversity within and among threatened populations have used so-called neutral genetic markers to quantify variation (McKay and Latta 2002, Sommer 2005). As argued previously in this volume these markers are excellent for estimating effective population size, migration rates, and other population genetic processes since, on the whole, they are not affected by selection and hence genetic variation is mainly determined by genetic drift. However, it is now questioned whether neutral genetic variation is a suitable proxy for the ecologically meaningful genetic variation required to maintain populations as viable entities capable of adapting to habitat and environmental change (e.g. Madsen et al. 2000, Hedrick 2001). A recent review used the following citation from Frankham (1999) to illustrate the present situation on how to study genetic variation in natural populations using neutral markers, quantitative trait loci, and ecological traits: ‘A major unresolved issue [in conservation] is the relationship between molecular measures of genetic diversity and quantitative genetic variation’ (McKay and Latta 2002). Studies of candidate genes are bridges to understanding local adaptation. As has been discussed in this chapter, selection may both maintain genetic variation, through balancing selection, and erode it, through purifying and directional selection.
Genes under selection: Mhc and others