Abstract

The major histocompatibility complex (MHC) plays a key role in adaptive immunity. It encodes proteins that are responsible for presenting pathogenic antigens to T lymphocytes, triggering the adaptive immune response. The MHC genes are divided into two functional classes. MHC-I recognises intracellular antigens, such as proteins encoded by viral genes, while MHC-II recognises extracellular pathogens, such as bacteria or parasites. Both classes are highly variable and represent the most polymorphic genes in jawed vertebrates. The higher the MHC polymorphism, the wider the spectrum of pathogens recognised. To date, more than 36,000 unique allele sequences have been identified in the best-studied species - humans. Extreme MHC polymorphism results from two contributing factors: extensive gene duplication and the maintenance of high allelic polymorphism of individual MHC genes. The former causes variation in the number of MHC genes between species and between haplotypes within species, and contributes to the common sharing of the same allelic sequences between duplicated genes. It creates a complex picture of the MHC region with multiple genes of both MHC classes, where an individual gene often shows high allelic polymorphism overlapping with the allelic polymorphism of other MHC genes.

The long-term maintenance of MHC polymorphism is made possible by balancing selection acting on MHC genes. Although the specific mechanisms are still the subject of debate among researchers, there is little doubt that the overall selection is driven by the arms race between hosts and pathogens. In response to pathogen adaptations, the target of selection undergoes recurrent changes that prevent the fixation of a single MHC allele. As a result, multiple MHC alleles are maintained in the population, usually at very low frequencies, providing a reservoir of MHC alleles that can be used for future adaptations from standing genetic variation. However, for an allele to be maintained or selected, it must be present in the population. There are three possible sources of new MHC alleles: de novo mutations, recombination and gene conversion involving the existing allele, and introgression. Adaptive introgression is thought to be a widespread phenomenon and may be the major source of new functional MHC variants. Its role in shaping MHC variation remains, however, poorly understood.

Introgression is a process of genetic exchange between different but closely related species. It occurs in areas known as hybrid zones, where species usually come into secondary contact after a period of isolation and produce viable and often fertile offspring. The hybrids produced, usually with fitness reduced compared to the parental species, may nevertheless transfer genes of foreign origin into the genetic background of the recipient species through the process of

backcrossing. The transfer of the neutral variant depends largely on the extent of isolation accumulated prior to secondary contact, with lower chances of neutral introgression between species that diverged earlier.

Adaptive introgression is a special case of introgression where a transferred gene variant is under positive selection. The theory predicts that, regardless of the strength of the accumulated barrier, gene transfer would be only slightly delayed compared to unrestricted gene flow. This ease of gene transfer, the prevalence of hybrid zones, and the clear selective advantage of novel MHC variants suggest that adaptive introgression is a major contributor to the observed levels of MHC polymorphism. Although several studies have postulated adaptive MHC introgression in individual hybrid zones, a large-scale comparative analysis of the prevalence of the process has been lacking.

This work aims to test the prevalence of adaptive MHC introgression in vertebrates. It takes advantage of multiple, densely sampled hybrid zones to address this question in a comparative framework. Such an approach allowed us to look for repeatable patterns of MHC introgression and to compare their rate directly with the rate of genome-wide (neutral) introgression. We used four independent tests adapted to the specificities of the MHC region: randomisation test, comparison of shared MHC polymorphism between isolated species and those with adjacent ranges, and geographic and genomic clines. The randomisation test was used to test for increased MHC similarity between species close to contact zones but beyond the reach of genome-wide introgression. Comparison of isolated and adjacent species tested for the prevalence of MHC introgression. Geographic clines tested whether the width of MHC clines exceeded that of genome-wide clines. Finally, genomic clines tested whether the rate of transition in MHC ancestry was lower relative to genome-wide ancestry.

Here I present three chapters that address the issue of adaptive MHC introgression. The first examines six hybrid zones between species of *Triturus* newts, the second focuses on six hybrid zones between *Podarcis* lizards, and the last provides a comparative analysis of MHC introgression in 29 hybrid zones from all major vertebrate groups. For all hybrid zones studied, we amplified MHC genes, analysed MHC variability and collected previously published estimates of genome-wide ancestry. To test whether the rate of MHC introgression is higher than the rate of neutral introgression, indicating its adaptiveness, we directly compared the rate of introgression between MHC and genome-wide markers.

The six hybrid zones presented in the first chapter are formed by six species of *Triturus* newts belonging to two separate clades: crested and marbled newts. Four of them inhabit the Balkan Peninsula and Anatolia (crested), while the rest occupy Western Europe (marbled). Our analyses showed much lower MHC-II variation with only a single functional MHC-II locus. In addition, the MHC variation in marbled newts was generally lower than in crested newts. In terms of introgression, we found a widespread signal of MHC introgression and strong support for its adaptiveness. The MHC introgression was detected even in the hybrid zone, where virtually no genome-wide admixture was detected. We also showed increased MHC similarity between species with adjacent ranges, compared to geographically isolated species, when controlling for the time of divergence between species. Interestingly, the comparison between MHC classes also suggested an increased rate of MHC-I introgression.

The seven species of Iberian *Podarcis* lizards analysed in the second chapter present a more uniform picture. They are all closely related and occupy similar ecological niches. The MHC variability of both MHC classes was comparable. Similar to the first chapter, we found evidence for widespread MHC introgression and strong evidence that it is adaptive. Again, similar to *Triturus,* we found elevated MHC allele sharing between species with abutting ranges. In contrast to previous results, we did not find differences in introgression rates between MHC classes. Instead, we found some patterns of asymmetric MHC introgression, which we hypothesised to be caused by differences in the level of MHC variation between hybridising species.

The final chapter combines the previous results with nine additional genera from all major vertebrate groups. We analysed MHC introgression independently in each system and then combined the results using meta-analytic methods. We observed a substantial variation in MHC diversity between genera, but similar MHC diversity for species within a genus. We found strong evidence for adaptive MHC introgression, confirmed by each of the tests applied. We found no differences in the rate of MHC introgression between MHC classes. Finally, we detected evidence for an asymmetry of MHC introgression towards species with lower MHC diversity.

In conclusion, this work provides the first comparative evidence for the prevalence of adaptive MHC introgression. We have shown that the process of MHC introgression could be a major source of novel MHC variants, confirming a long-standing hypothesis. In addition, our work raises new questions about the fitness effects of individual MHC variants and the processes behind asymmetric MHC introgression. The tools presented here can also be applied to the

studies of adaptive introgression in other genetic systems under balancing selection, such as incompatibility genes in plants. I believe this is an important contribution to the field with implications for the fields of evolutionary immunology, conservation biology and speciation research.