Immunology & Cell Biology 2022; **100**: 497–506
www.wileyonlinelibrary.com/journal/icb

PERSPECTIVE

OPEN

Immunoglobulin genes, reproductive isolation and vertebrate speciation

Andrew M Collins[1], Corey T Watson[2] & Felix Breden[3]

1 School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia
2 Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY, USA
3 Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada

Keywords
Dobzhansky–Muller incompatibility, immunoglobulin, reproductive isolation, self/non-self-discrimination, speciation, tolerance

Correspondence
Andrew M Collins, School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia.
E-mail: a.collins@unsw.edu.au

Received 24 March 2022;
Revised 19 June 2022;
Accepted 21 June 2022

doi: 10.1111/imcb.12567

Immunology & Cell Biology 2022; **100**: 497–506

Abstract

Reproductive isolation drives the formation of new species, and many genes contribute to this through Dobzhansky–Muller incompatibilities (DMIs). These incompatibilities occur when gene divergence affects loci encoding interacting products such as receptors and their ligands. We suggest here that the nature of vertebrate immunoglobulin (IG) genes must make them prone to DMIs. The genes of these complex loci form functional genes through the process of recombination, giving rise to a repertoire of heterodimeric receptors of incredible diversity. This repertoire, within individuals and within species, must defend against pathogens but must also avoid pathogenic self-reactivity. We suggest that this avoidance of autoimmunity is only achieved through a coordination of evolution between heavy- and light-chain genes, and between these genes and the rest of the genome. Without coordinated evolution, the hybrid offspring of two diverging populations will carry a heavy burden of DMIs, resulting in a loss of fitness. Critical incompatibilities could manifest as incompatibilities between a mother and her divergent offspring. During fetal development, biochemical differences between the parents of hybrid offspring could make their offspring a target of the maternal immune system. This hypothesis was conceived in the light of recent insights into the population genetics of IG genes. This has suggested that antibody genes are probably as susceptible to evolutionary forces as other parts of the genome. Further repertoire studies in human and nonhuman species should now help determine whether antibody genes have been part of the evolutionary forces that drive the development of species.

INTRODUCTION

The principal function of immunoglobulin (IG) genes of the vertebrate adaptive immune system is clear. The antibodies that they encode are central to the immune response to invading pathogens. However, we propose here that IG genes have an additional role that is far removed from the microbial world. We propose that they help maintain the reproductive isolation that separates one species from another.

The potential role of IG genes as drivers of speciation can be explained by the Dobzhansky–Muller incompatibility (DMI) model.[1,2] This model suggests that genes of loci encoding interacting products may contribute to the reproductive isolation of different populations if gene divergence leads to poor interactions between the products of the genes in hybrid offspring. Genes that contribute to reproductive isolation are referred to as speciation genes if genetic divergence at the loci occurred before speciation was complete.[3] We believe that IG genes may be particularly susceptible to DMIs, and that they may be able to directly contribute to the development of reproductive isolation. They should therefore be considered candidate speciation genes.

IG genes encode the heavy- and light-chain polypeptides that combine to function as heterodimeric

IG molecules. Many genes work together to encode each IG molecule, and the hybrid offspring of diverging populations may carry heavy- and light-chain gene sets that interact poorly. We refer here to incompatibilities between paired heavy- and light-chain genes as *specific* DMIs. As discussed below, in the hybrid offspring of divergent populations, DMIs of this kind could lead to a disruption of processes by which the immune system is normally able to control its reactivity to "self" molecules, with serious implications for the viability of the individual, and for its reproductive fitness.

We believe there could also be "higher-order" incompatibilities, in which sets of IG genes essentially become incompatible with much of the surrounding genome. We refer to these incompatibilities as *generalized* DMIs, and they could particularly manifest *via* a mechanism that has evolved to provide immunological protection to the newborn. All mammalian offspring are initially protected from infection by maternal antibodies that are acquired either *via* placental transfer or *via* the ingestion of colostrum. Generalized incompatibilities between the maternal antibody repertoire and the biochemistry of hybrid offspring could have profound consequences during fetal development or in the perinatal period. A detailed outline of this kind of incompatibility is also provided below.

IG genes encode the highly variable receptor molecules that cover the surface of all B cells, with each cell carrying receptors that are essentially unique to that cell. It is *via* these receptors that lymphocytes are able to target microbial antigens. There are millions of variant genes that encode the IG repertoire, and these variant genes are generated by recombination processes involving large sets of variable (V), diversity (D) and joining (J) genes. These genes are unique in that they cannot function in their germline configuration. Rather, functional gene products are generated by processes that join multiple genes of different types into a functional unit.[4–6] In humans, there are approximately 50 unique but similar heavy-chain V genes, as well as 23 unique heavy-chain D gene sequences and 6 unique heavy-chain J genes.[7] Functional heavy-chain VDJ rearrangements are produced by the recombination of one gene from each of the three gene sets, and the combinatorial diversity that is made possible by the size of the gene sets is massively expanded by additional processes that add to and remove nucleotides from the joining gene ends.[4–6] Functional human light-chain VJ genes are similarly produced from large sets of V and J genes from the kappa locus on chromosome 2, or from genes of the lambda V and J loci on chromosome 22.[8,9] The recombination process is illustrated in Figure 1.

Together, the IG gene sets encode a repertoire of tens of millions of different antibodies with distinct specificities. It was long thought that the powers of the recombination process would ensure that everyone had all necessary specificities within their IG repertoires. It has recently become clear, however, that the presence or absence of particular IG genes and allelic variants can influence the ability of individuals to mount antigen-specific responses, and this has been reported from human studies[10] and mouse studies.[11] Critically, the IG gene loci found within a given species are not necessarily similar to those of even closely related sister taxa. Genetic data from various species, including the house mouse (discussed below), demonstrate that these loci are not highly conserved across the tree of life, but are likely subject to rapid evolution. Divergence of the IG loci in separating populations must result from the changing nature of the microbial challenges that are the targets of the IG response, and this likely evolutionary drive addresses a deficiency that has previously been noted for the DMI model of speciation.[12] The DMI model focuses on gene incompatibilities without addressing possible drivers of gene divergence that give rise to the incompatibilities.[12]

Speciation genes can be present in regions of high genomic differentiation,[13,14] and the search for such regions has therefore become a pathway to the identification of candidate speciation genes. This approach, however, has not previously pointed to any involvement of the IG loci in speciation. We believe that this is a consequence of the fact that the genomic differentiation of the IG loci has been hidden by the complexity of these regions. Complex loci such as these are unsuited to investigation using short-read sequencing, and are poorly represented in all but the highest quality genome assemblies.[15–17] Almost no data have been available until recently on the population genetics of IG genes in any species.[18] This has allowed these complex loci of the vertebrate adaptive immune system to escape scrutiny by evolutionary biologists and even by immunologists.

Our understanding of the population genetics of IG genes is now progressing, with the development of inferential approaches to gene identification,[19–22] and more recently from genomic studies using long-read sequencing.[16,17,23] The magnitude of IG gene variation that these approaches have identified within the human population and in other species has transformed our understanding of antibody gene evolution. This in turn led to our belief that IG genes are outstanding candidate speciation genes. Many genes may participate in the speciation process, varying in their relative contributions or "effect sizes," but we believe that the combination of general and specific DMIs outlined here could give IG genes a high epistatic "effect size."

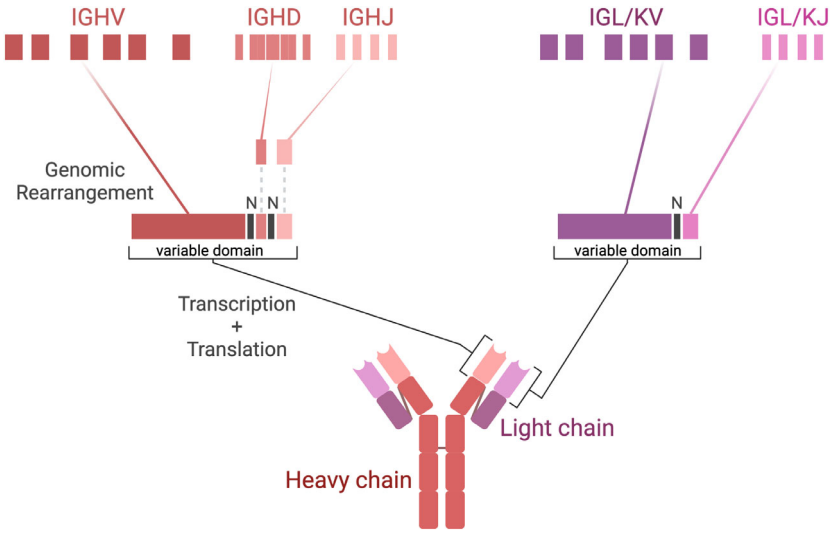

Figure 1. An overview of key elements contributing to the diversity of the B-cell and immunoglobulin (IG) repertoires. A basic schematic of the germline IGH locus is shown, consisting of clusters of tandemly arrayed heavy-chain V, D and J and light-chain V and J genes. Only a handful of the genes of each cluster are shown and the figure is not to scale. Constant region genes are located downstream from the J genes of each locus, and are not shown. During the initial formation of the naïve repertoire, single IGH V, D and J gene segments on one of two chromosomes in a given B-cell are somatically recombined; at each of these steps, varying numbers of nucleotides are removed from the joining gene ends, and random nucleotides are added at the D–J and V–D junctions. Light-chain V-J rearrangements then takes place in a similar manner. Random additions are labeled "N." The variable regions of the IG molecules that are encoded by the V(D)J gene rearrangements are indicated, and are expressed in association with products of constant region genes. Two identical heavy chains and two identical light chains are ultimately paired through disulfide bonds to form a functional antibody.

COEVOLUTION OF GENES IS REQUIRED FOR SELF-TOLERANCE

For over 70 years, our understanding of the adaptive immune systems of vertebrate species has been guided by the metaphor of "self and nonself."[24] Although the metaphor has sometimes been challenged,[25,26] and although the concept of self is not a precise scientific term,[27] the immune system generally responds to molecules of the external world (nonself), and *vigorous* responses to molecules of the immune system's host (self) are usually associated with autoimmune disease or the response to neoantigens of cancerous tissues (altered self). Critical to our belief that IG genes have a role as speciation genes is our belief that a failure of an individual to mount a vigorous response to "self" is, in effect, a failure of the individual to make vigorous responses to "self" molecules that might characterize its species. The biochemical self is not easily distinguishable from the chemistry of a species, because the biochemical make-up of any individual is mostly shared with other individuals of their species. The immune system's "awareness" of self can therefore be looked at as an awareness of the species.

In any individual, each lymphocyte of the adaptive immune system carries antigen receptors on its surface that are more or less unique to that lymphocyte. These receptors are encoded by genes that are utilized through a unique process known as V(D)J gene recombination, which is influenced by both heritable and stochastic components (Figure 1). A consequence of the generation of millions of recombined IG genes is that the resulting B-cell repertoire will include many self-reactive cells.[28,29] Despite the generation of these autoreactive B and T cells, a state of *self-tolerance* is usually maintained. Central, peripheral and systemic processes act to delete, silence or suppress any self-reactive B cells,[30,31] and to ensure that clonal activation is focused upon non-self-pathogens.[32,33]

One process that limits the generation of self-reactive cells is a central selection process that is critical to the present hypothesis (Figure 2). V(D)J recombination is usually said to involve a single heavy-chain VDJ rearrangement and a single light-chain VJ rearrangement. In fact, if this initial process leads to self-reactivity, developing B cells can be selected for further rounds of rearrangement of their light-chain genes. This process is called receptor editing,[34,35] and it leads to new heavy-chain/light-chain pairs that may be less self-reactive (Figure 2). The nature and organization of the light-chain genes is such that multiple rounds of light-chain rearrangement are possible,[36] while the genes of the heavy-chain locus can only rearrange once. Some heavy-

and light-chain pairings are not autoreactive and are therefore "permitted." Others are self-reactive, and such pairings may trigger receptor editing. Genes that give rise to "forbidden" pairs may be tolerated within the genome as long as many suitable pairings exist for each gene, and as long as receptor editing offers a means by which suitable pairings can usually be found.

The danger of self-reactive pairs is revealed if the capacity of an individual to edit light-chain rearrangements is compromised. Reduced receptor editing is associated with several autoimmune conditions in the human and the mouse.[37,38] When the system of gene recombination and editing is fully functional, occurring independently in billions of developing lymphocytes, an individual can generate a diverse antibody repertoire that targets the external world without overloading the systems that normally control any residual self-reactivity.

SPECIFIC DMIS BECAUSE OF HEAVY- AND LIGHT-CHAIN GENE INTERACTIONS

IG genes must evolve in response to the changing microbial world. In fact, as IG genes encode

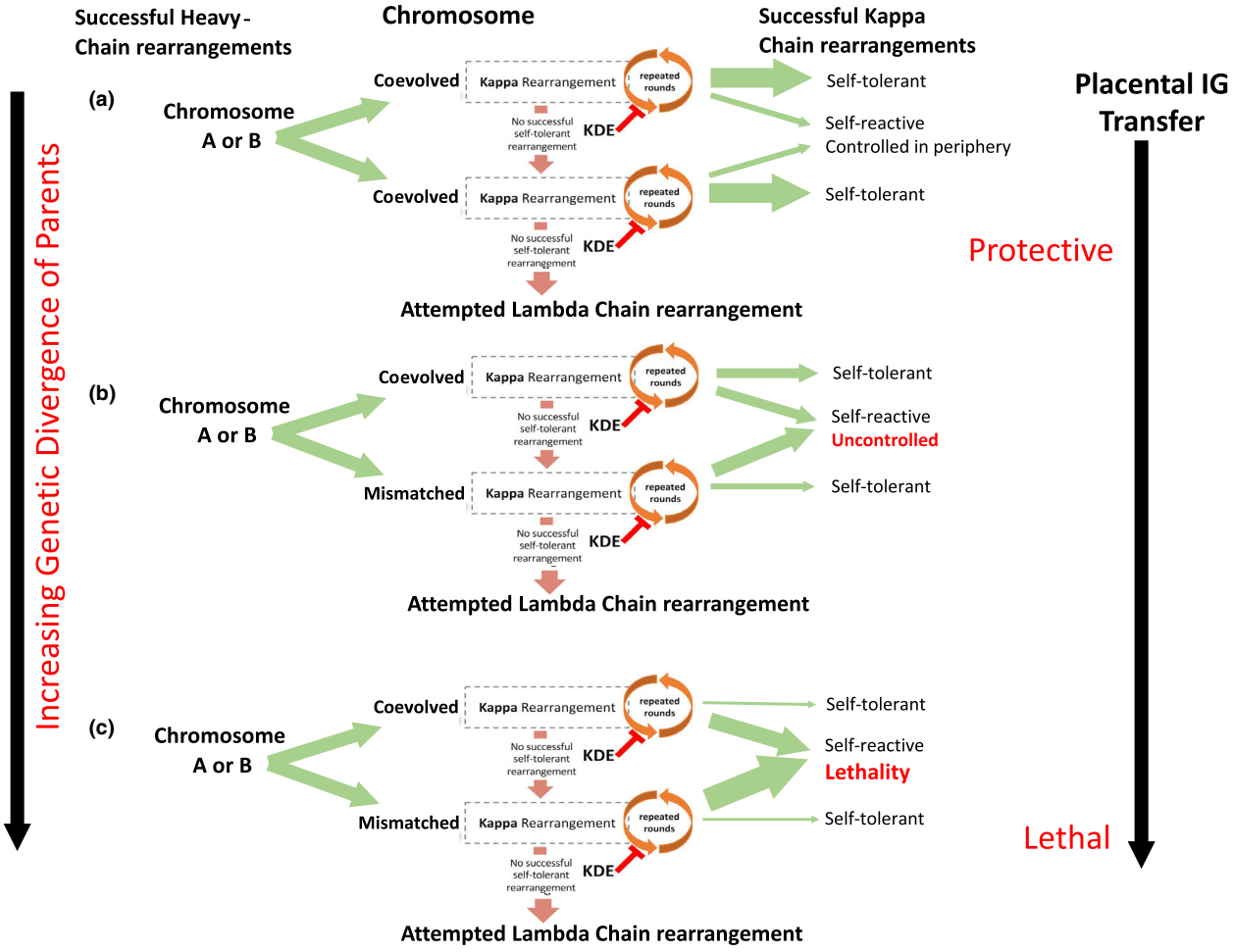

Figure 2. The outcome of kappa chain gene rearrangement after successful heavy-chain gene rearrangement in the offspring of parents with varying amounts of genetic divergence. Recombination of a rearranging locus may be terminated by activity of the kappa deleting element (KDE). This will drive recombination to the κ locus on the alternate chromosome, or from the second κ locus to the λ gene loci. Differing outcomes are shown where heavy- and light-chain genes have coevolved, and where they are mismatched. Even in the offspring of highly divergent parents, some antibodies will form from heavy- and light-chain genes that have coevolved, and some that are mismatched. Where there is little divergence between parents **(a)** multiple rounds of kappa chain rearrangement lead to a repertoire with little self-reactivity that is easily controlled. As genetic divergence between the parents increases **(b)**, there is increasing uncontrollable self-reactivity in the expressed repertoire. Genetic divergence reaches a point **(c)** where self-reactivity is potentially lethal. The lethality of the expressed repertoire, in an individual, is paralleled by the increasing lethality of the placental transfer of antibodies from mother to fetal offspring. Only kappa light-chain rearrangements are shown. If kappa chain rearrangements are unproductive, VJ recombination begins at the lambda loci. IG, immunoglobulin.

heterodimeric molecules, the heavy- and light-chain gene loci must coevolve to ensure that a repertoire of antibodies suited to the pathogens of the moment can be generated, while still maintaining antibody structure and effector functions. Importantly, they must also coevolve to limit the extent of self-reactive heavy- and light-chain pairings. When divergent populations reconnect, and sets of IG genes that have not coevolved in this way are brought together, specific DMIs could result. The consequences of this may be illustrated by the laboratory mouse. Inbred mouse strains are composites of different mouse subspecies, and they appear to be burdened by incompatibilities in their IG loci. We believe this partly explains the susceptibility of particular strains to autoimmune disease.[36]

To understand the IG genes of laboratory mice and the divergence of their gene sets, we must consider the loci of the three *Mus musculus* subspecies from which laboratory mice are largely derived. The subspecies are believed to have diverged only 350 000 years ago,[14] yet the divergence of the heavy-chain gene loci of the three subspecies is extreme. In a study of wild-derived inbred strains of mice, we identified 87 IG heavy-chain V gene sequences in the CAST/EiJ strain that is derived from the *Mus musculus castaneus* subspecies.[39] Exact matches to the CAST/EiJ sequences were only seen in 5 of 92 V gene sequences from the *M. m. musculus*–derived PWD/PhJ strain, and in 2 of 78 V gene sequences from the *M. m. domesticus*–derived LEWES/EiJ strain.[39] Some divergence was also seen in the heavy-chain D and J gene sets, while single-nucleotide polymorphism analysis and our unpublished data suggest that the sets of light-chain genes of the three subspecies are also highly divergent.[36,40]

The loci of common inbred mouse strains are also highly divergent both from one another and from the loci of the wild-derived inbred strains, for as a consequence of their early breeding histories, the heavy and light chain loci of classical inbred mouse strains seem to be mosaics of genes derived from each of the different mouse sub-species.[39,41] For example, just 4 of the 109 expressed IG heavy-chain V genes of the C57BL/6 mouse are found among the 162 expressed V gene sequences identified in the BALB/c mouse.[41,42] The mixing of heavy- and light-chain genes through human management has been so extreme that many of the self-tolerant heavy- and light-chain pairings of wild mice may have been lost.

In the course of just a century, the breeding of these animals, in the absence of purifying selection, has created numerous strains that are so susceptible to self-reactivity that they have become widely used models for the study of human autoimmune disease. Importantly, these strains were not bred for this purpose; their autoimmunity is accidental. We have noted that most mouse strains that suffer from a susceptibility to autoimmune disease have glaring mismatches in the likely origins of their heavy- and light-chain loci.[36] That is, single-nucleotide polymorphism analysis suggests that their heavy-chain genes are predominantly derived from one subspecies, while their light-chain genes are predominantly derived from another (Figure 3). Incompatible sets of heavy- and light-chain genes will increase the likelihood that any lymphocyte will produce self-reactive antibodies, and beyond a certain level of incompatibility, measures for the control of self-reactive antibodies may well be overwhelmed (Figure 2).

As an illustrative example, heavy- and light-chain loci appear to be mismatched in the non-obese diabetic/ShiLtJ mouse that is an important model of type 1 diabetes.[43] Single-nucleotide polymorphism analysis (Figure 3) suggests that non-obese diabetic/ShiLtJ mice carry a heavy-chain locus that is *M. m. musculus*–like, but they seem to carry a kappa light-chain locus that is generally *M. m. castaneus*–derived.[36,39] Similar mispairings are seen in the (NZB × NZW)F1, MRL (129 × C57BL/6)F1 and BXSB strains that provide models of systemic lupus erythematosis[44–46] (Figure 3).

Not all inbred mice that seem prone to autoimmunity harbor obviously mismatched loci, and we do not suggest that IG genes provide the only pathway to autoimmune disease. However, it is possible that mismatched loci in some of these susceptible strains are not yet detectable. We lack a detailed knowledge of the IG loci in any inbred mouse strains other than a few wild-derived strains and the BALB/c and C57BL/6 strains.[39,42,47,48] If the identification of the subspecies origins of IG genes is based on single-nucleotide polymorphism data (Figure 3), these data may presently fail to capture some important variation within the loci. It is also true that not all strains that carry apparently mismatched loci have been reported to be susceptible to autoimmunity. It is nevertheless striking that most of the key strains used in research of autoimmunity carry these kinds of mismatches, and as a consequence, they would have little reproductive fitness outside the confines of an animal breeding facility.

COEVOLUTION IS REQUIRED TO AVOID GENERALIZED DMIS BETWEEN IG GENES

The divergence between the IG loci of the different subspecies of the house mouse and the reduced fitness of inbred mouse strains that carry apparently mismatched heavy- and light-chain loci suggests to us that the genes of these loci are candidate speciation genes. Unfortunately, the limited available data make it difficult

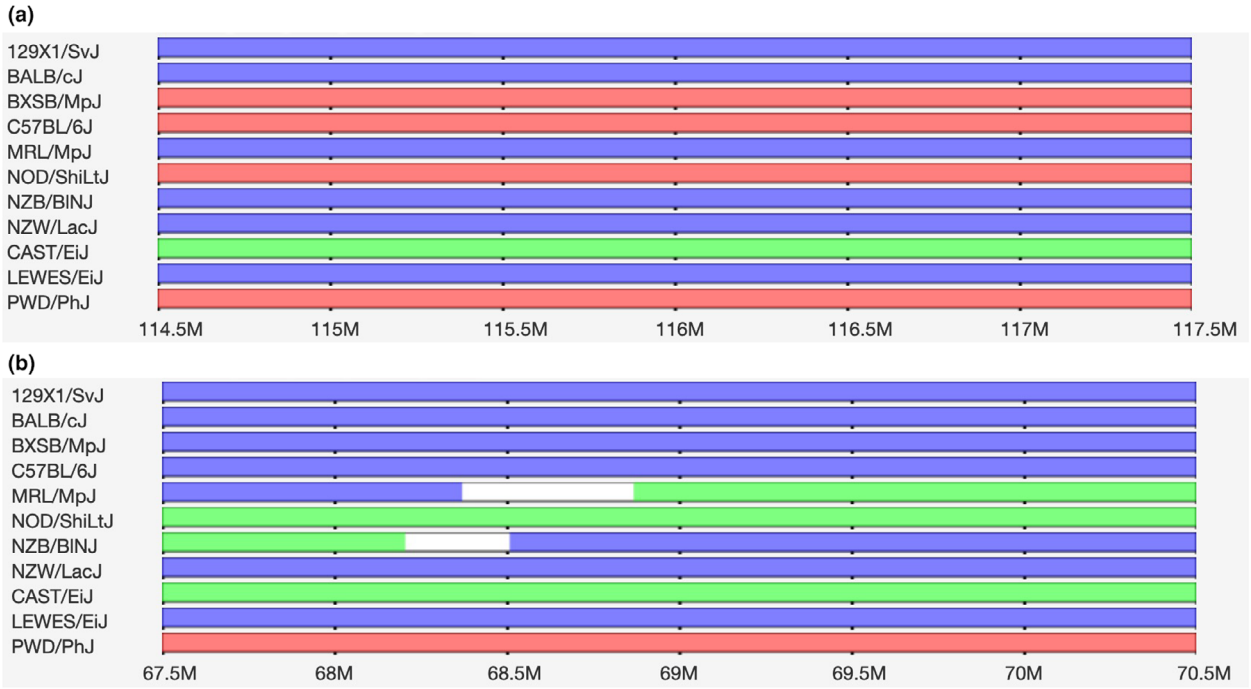

Figure 3. Single-nucleotide polymorphism–based predictions of the *Mus musculus* subspecies origins of germline haplotypes found in the immunoglobulin (IG) **(a)** IGHV and **(b)** IGKV loci among selected classical inbred and wild-derived mouse strains, including strains that are particularly susceptible to autoimmunity. The MRL/MpJ strain is the parental strain of the MRL/lpr strain, and although MRL/MpJ mice carry a normal Fas gene, they are also prone to autoimmunity. The lowest three strains shown are wild derived, with genomes that are representative of the three major subspecies of *Mus musculus*. *M. m. domesticus*–derived sequences are shown in blue, *M. m. musculus*–derived sequences are shown in red and *M. m. castaneus*–derived sequences are shown in green. Data from Yang and colleagues.[74] Graphics by Mouse Phylogeny Viewer.[75]

at present to investigate our general hypothesis in other closely related species or subspecies. It is possible, however, to explore the likely consequences of gene flow between reconnecting vicariant populations by reference to a well-described human condition, as well as to recent studies of hominid evolution.

If a population of a species migrates into new territories, with a new microbiota, the IG loci will be under novel selection pressures. Both the heavy- and light-chain loci will evolve to directly engage with new microbial challenges, and to minimize the self-reactivity of new heavy- and light-chain pairings. Later reconnection of vicariant populations would be difficult if incompatibilities had emerged between the heavy-chain genes of one population, and the light-chain genes of another. In hybrid offspring of the two populations, heavy chains encoded by genes from one parent, through association with light chains encoded by genes from the other parent, could increase the burden of autoreactivity for the offspring. If many incompatibilities had developed between the heavy- and light-chain sets of the two populations, they could be on an immunological pathway to reproductive isolation and speciation (Figure 4).

Additional incompatibilities would strengthen this drive to isolation, for evolution of the genome of the migrating population would not be confined to the IG genes. Other changes in the genome would require coevolution of the IG genes to maintain self-tolerance (i.e. the ability of products of the IG loci to avoid recognition of self-proteins across the genome). Gradual genome-wide changes would ultimately lead to a new biochemical "self" in the migrating population.

Reconnection of these two populations would now be compromised if differences between the "selves" of the two populations had become too great. The immune system of hybrid offspring of the two populations would face specific DMIs arising from the lack of coevolution of the heavy-chain genes of one population with the light-chain genes from the other. They would also suffer from incompatibilities between each parentally derived antibody repertoire and "changed self." This would occur because the chromosomal organization of the IG genes would ensure that essentially 50% of the antibody repertoire of hybrid offspring would match antibodies of the maternal or paternal repertoire. These specific and general incompatibilities would reduce fitness *via*

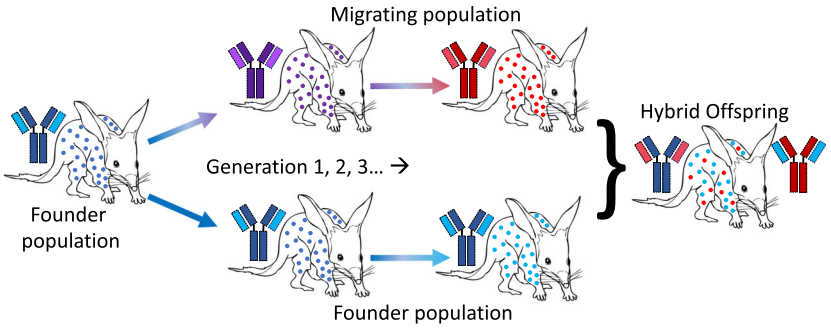

Figure 4. Genomic changes and antibody repertoire changes in diverging populations. In response to their changing environments and changing pathogens, migrating populations are subject to major changes in the genome, including major changes in the expressed antibody repertoire. Over time, a founder population is subject to minor changes in the genome and minor changes in the expressed antibody repertoire. Compatibility between heavy and light chains, and between the immunoglobulin (IG) repertoire and the biochemical "self" are indicated by color coding. If the two populations reconnect, the hybrid offspring (at right) show IG Dobzhansky–Muller incompatibilities of four different kinds: founder IG heavy chains with migrating light chains; migrating heavy chains with founder light chains; founder heavy- and light-chain pairs with migrating neoantigens and migrating heavy- and light-chain pairs with founder antigens.

pathological processes akin to those seen in chronic antibody-mediated autoimmune diseases.

An additional antibody-mediated process resulting from *generalized* DMIs could also present a potentially lethal acute autoimmune-like challenge to a hybrid fetus. Maternally derived antibodies of all eutherian species function to protect the fetus and/or newborn.[49] In humans and many other species, maternal antibodies are delivered across the placenta during early fetal development.[50] In others, colostrum-derived maternal antibodies are taken up from the gastrointestinal tract in the first days after birth.[51] However they are acquired, these maternal antibodies act to protect the newborn, and they persist while the immune defenses of the newborn develop. But these antibodies are not always protective, as is shown by the human condition hemolytic disease of the fetus and newborn (HDFN).[52]

HDFN arises from differences between the blood group antigens of a mother, and the blood group antigens of her offspring.[53] Although the maternal antibody repertoire should not be reactive with cells of the mother, the maternal immune system may come to target the "foreign" cells of the developing fetus, with potentially fatal consequences. HDFN, as well as rare pathology associated with intravenous IG therapy,[54,55] reminds us that despite the biochemical similarities of all members of a species, some biologically critical differences between individuals remain. Although the usual pathway to the development of HDFN is different to the processes we suggest are involved in speciation, HDFN highlights the potentially fatal consequences of the transplacental passage of maternal antibodies. During the evolution of eutherian species, biochemical differences within a breeding population will grow. The likelihood and the severity of deleterious consequences of maternally derived antibodies could also grow. It is therefore reasonable to propose that matings of individuals from distinct groups within a breeding population could result in the placental transfer of a deadly suite of antibodies.

The immunological consequences of the reconnection of populations would not all be negative. A radiating population that had responded genetically to the pathogens of new territories could be a beneficial source of genes if somehow individuals from a "founder" population reconnected with these adapted individuals in those new territories. If the benefits of the introgression of pathogen-specific genes outweighed the reproductive consequences of autoreactivity, such genes could become fixed in the evolving founder population. Eventually, further genetic modification could help resolve any resulting autoreactivity while maintaining the beneficial specificities in the antibody repertoire. Introgressions that seem to have benefited the immune system are certainly known.

Many recent studies have presented evidence that admixtures with Neanderthals and Denisovans have made valuable contributions to the modern human genome, including important contributions to the way modern humans respond to pathogens.[56–58] Such studies have generally focused on receptors of the innate immune system[59,60] and on some molecules (e.g. complement proteins) that straddle the innate/adaptive divide.[61,62] The introgression of a Denisovan HLA-B allele has also been identified,[63] and the human leukocyte antigen system is central to the operation of T-cell–mediated adaptive defenses. We believe that the lack of reports of introgressions involving IG genes may simply be a consequence of our lack of understanding of the population genetics of human IG genes, and of our inability, at present, to determine the IG genes of our hominin ancestors. As our understanding improves, it

will be important to consider both the positive and the negative outcomes of any introgressions that may be identified.

CHALLENGES

Our hypothesis that IG genes are important speciation genes is not the first proposal that genes of the immune system could serve this role. For example, the genes of the vertebrate major histocompatibility complex have been a focus of speciation studies in teleost fish.[64] Neither is it the first time that it has been proposed that pathogens could be significant drivers of reproductive isolation between diverging populations. Karvonen and Seehausen[65] proposed that parasites, rather than habitat, trophic specialization or resource competition, have been important drivers of divergence and reproductive isolation within and between populations. Our hypothesis extends these proposals, and provides an explanation for why IG genes could play a special role in speciation.

The nature of the IG loci will make the investigation of the hypothesis challenging. It will not be a simple matter to explore genetic variation in the IG loci of the vertebrate species that are so often used for evolutionary studies, for we lack the necessary baseline studies. A handful of IG genes have been reported from many teleost fish species, but the loci have only been well documented in catfish,[66] zebrafish[67] and salmonid species.[68,69] Despite decades of research, we are only now appreciating the extent of individual genetic variation in the human IG loci,[17,18,20] while the study of intraspecies genetic variation in teleost species is confined to a single study of the rainbow trout.[69]

Mouse models may offer a better way forward. Hybridization zones where mouse subspecies interact have already been targets for the investigation of speciation.[70] A better understanding of the population genetics of the IG loci in mouse subspecies should be pursued. The IG loci of available inbred mouse strains also need to be properly documented, and will then offer special opportunities for the study of such things as the relationships between maternal antibody repertoires and hybrid fetal viability. The development of congenic mouse strains with altered IG loci will also allow the more direct testing of the hypothesis. This will also allow a more general question to be addressed: can evolution act upon germline IG genes to reduce the likelihood of self-reactivity? Antibody repertoire studies, in the mouse, human and other species, should also offer important insights—particularly studies that seek to identify whether or not processes that maintain central tolerance create "holes" in the antibody repertoire.[71] The existence of such holes is critical to our hypothesis, for it is these holes that we suggest would be "plugged" in hybrid offspring, to their detriment.

Our discussions here have been confined to the possible role of genes of the immune system in vertebrate speciation, but it is possible that genes of the immune system have played important roles in speciation in other groups of organisms. Interestingly, the immune system of plants has been implicated in speciation, and hybrid necrosis in plants has been described as an autoimmune-like response resulting from DMIs.[72,73] This raises the intriguing possibility that throughout evolutionary history, defense systems that allow the discrimination of self from nonself may have been driving the formation of reproductive isolates within the living world.

ACKNOWLEDGMENTS

The development of the hypothesis outlined here was triggered by discussions initiated by Chris Dewhurst at the Rock and Roll Café, Mullumbimby, NSW, Australia. We thank Michael Wade for his review of an earlier version of the manuscript. Open access publishing facilitated by University of New South Wales, as part of the Wiley - University of New South Wales agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Dobzhansky TG. Genetics and the Origin of Species. New York, NY: Columbia University Press; 1937.
2. Muller HJ. Bearing of the *Drosophila* work on systematics. In: Huxley JS, ed. The New Systematics. Oxford: Clarendon Press; 1940: 185–268.
3. Nosil P, Schluter D. The genes underlying the process of speciation. *Trends Ecol Evol* 2011; **26**: 160–167.
4. Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983; **302**: 575–581.
5. Jackson KJ, Kidd MJ, Wang Y, Collins AM. The shape of the lymphocyte receptor repertoire: lessons from the B cell receptor. *Front Immunol* 2013; **4**: 263.
6. Elhanati Y, Sethna Z, Marcou Q, Callan CG Jr, Mora T, Walczak AM. Inferring processes underlying B-cell repertoire diversity. *Philos Trans R Soc Lond B Biol Sci* 2015; **370**: 20140243.
7. Matsuda F, Ishii K, Bourvagnet P, *et al.* The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med* 1998; **188**: 2151–2162.
8. Kawasaki K, Minoshima S, Nakato E, *et al.* Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes. *Eur J Immunol* 2001; **31**: 1017–1028.
9. Kawasaki K, Minoshima S, Nakato E, *et al.* One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res* 1997; **7**: 250–261.

10. Avnir Y, Watson CT, Glanville J, *et al.* IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. *Sci Rep* 2016; **6**: 20842.
11. Raposo B, Dobritzsch D, Ge C, *et al.* Epitope-specific antibody response is controlled by immunoglobulin V_H polymorphisms. *J Exp Med* 2014; **211**: 405–411.
12. Wu CI, Ting CT. Genes and speciation. *Nat Rev Genet* 2004; **5**: 114–122.
13. Wolf JB, Ellegren H. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet* 2017; **18**: 87–100.
14. Phifer-Rixey M, Bomhoff M, Nachman MW. Genome-wide patterns of differentiation among house mouse subspecies. *Genetics* 2014; **198**: 283–297.
15. Rhie A, McCarthy SA, Fedrigo O, *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 2021; **592**: 737–746.
16. Ford M, Haghshenas E, Watson CT, Sahinalp SC. Genotyping and copy number analysis of immunoglobulin heavy chain variable genes using long reads. *iScience* 2020; **23**: 101508.
17. Rodriguez OL, Gibson WS, Parks T, *et al.* A novel framework for characterizing genomic haplotype diversity in the human immunoglobulin heavy chain locus. *Front Immunol* 2020; **11**: 2136.
18. Watson CT, Glanville J, Marasco WA. The individual and population genetics of antibody immunity. *Trends Immunol* 2017; **38**: 459–470.
19. Boyd SD, Gaeta BA, Jackson KJ, *et al.* Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol* 2010; **184**: 6986–6992.
20. Gadala-Maria D, Gidoni M, Marquez S, *et al.* Identification of subject-specific immunoglobulin alleles from expressed repertoire sequencing data. *Front Immunol* 2019; **10**: 129.
21. Kidd MJ, Chen Z, Wang Y, *et al.* The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements. *J Immunol* 2012; **188**: 1333–1340.
22. Kirik U, Greiff L, Levander F, Ohlin M. Parallel antibody germline gene and haplotype analyses support the validity of immunoglobulin germline gene inference and discovery. *Mol Immunol* 2017; **87**: 12–22.
23. Zhang JY, Roberts H, Flores DSC, *et al.* Using *de novo* assembly to identify structural variation of eight complex immune system gene regions. *PLoS Comput Biol* 2021; **17**: e1009254.
24. Burnet FM, Fenner F. The Production of Antibodies. 2nd ed. Melbourne: Macmillan; 1949.
25. Fuchs E. Reply from Ephraim Fuchs. *Immunol Today* 1993; **14**: 236–237.
26. Matzinger P. Tolerance, danger, and the extended family. *Annu Rev Immunol* 1994; **12**: 991–1045.
27. Tauber AI. The immune self: theory or metaphor? *Immunol Today* 1994; **15**: 134–136.
28. Sng J, Ayoglu B, Chen JW, *et al.* AIRE expression controls the peripheral selection of autoreactive B cells. *Sci Immunol* 2019; **4**: eaav6778.
29. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC. Predominant autoantibody production by early human B cell precursors. *Science* 2003; **301**: 1374–1377.
30. Cashman KS, Jenks SA, Woodruff MC, *et al.* Understanding and measuring human B-cell tolerance and its breakdown in autoimmune disease. *Immunol Rev* 2019; **292**: 76–89.
31. Theofilopoulos AN, Kono DH, Baccala R. The multiple pathways to autoimmunity. *Nat Immunol* 2017; **18**: 716–724.
32. Cohen IR, Young DB. Autoimmunity, microbial immunity and the immunological homunculus. *Immunol Today* 1991; **12**: 105–110.
33. Cohen IR. Activation of benign autoimmunity as both tumor and autoimmune disease immunotherapy: a comprehensive review. *J Autoimmun* 2014; **54**: 112–117.
34. Halverson R, Torres RM, Pelanda R. Receptor editing is the main mechanism of B cell tolerance toward membrane antigens. *Nat Immunol* 2004; **5**: 645–650.
35. Prak EL, Weigert M. Light chain replacement: a new model for antibody gene rearrangement. *J Exp Med* 1995; **182**: 541–548.
36. Collins AM, Watson CT. Immunoglobulin light chain gene rearrangements, receptor editing and the development of a self-tolerant antibody repertoire. *Front Immunol* 2018; **9**: 2249.
37. Panigrahi AK, Goodman NG, Eisenberg RA, Rickels MR, Naji A, Luning Prak ET. RS rearrangement frequency as a marker of receptor editing in lupus and type 1 diabetes. *J Exp Med* 2008; **205**: 2985–2994.
38. Vander Heiden JA, Stathopoulos P, Zhou JQ, *et al.* Dysregulation of B cell repertoire formation in myasthenia gravis patients revealed through deep sequencing. *J Immunol* 2017; **198**: 1460–1473.
39. Watson CT, Kos JT, Gibson WS, *et al.* A comparison of immunoglobulin IGHV, IGHD and IGHJ genes in wild-derived and classical inbred mouse strains. *Immunol Cell Biol* 2019; **97**: 888–901.
40. Kos JT, Safanova Y, Shields KM, *et al.* Characterization of extensive diversity In immunoglobulin light chain Variable germline genes across biomedically important mouse strains. *bioRxiv* 2022. https://doi.org/10.1101/2022.05.01.489089.
41. Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJ. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philos Trans R Soc Lond B Biol Sci* 2015; **370**: 20140236.
42. Jackson KJL, Kos JT, Lees W, *et al.* A BALB/c IGHV Reference Set, defined by haplotype analysis of long-read VDJ-C sequences from F1 (BALB/c x C57BL/6) mice. *Front Immunol* 2022; **13**: e888555.
43. Mullen Y. Development of the Nonobese Diabetic Mouse and contribution of animal models for understanding type 1 diabetes. *Pancreas* 2017; **46**: 455–466.
44. Bubier JA, Sproule TJ, Foreman O, *et al.* A critical role for IL-21 receptor signaling in the pathogenesis of systemic lupus erythematosus in BXSB-Yaa mice. *Proc Natl Acad Sci USA* 2009; **106**: 1518–1523.

45. Bygrave AE, Rose KL, Cortes-Hernandez J, *et al.* Spontaneous autoimmunity in 129 and C57BL/6 mice - implications for autoimmunity described in gene-targeted mice. *PLoS Biol* 2004; **2**: E243.
46. Henry RA, Kendall PL, Woodward EJ, Hulbert C, Thomas JW. Vkappa polymorphisms in NOD mice are spread throughout the entire immunoglobulin kappa locus and are shared by other autoimmune strains. *Immunogenetics* 2010; **62**: 507–520.
47. Johnston CM, Wood AL, Bolland DJ, Corcoran AE. Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. *J Immunol* 2006; **176**: 4221–4234.
48. Riblet R. Immunoglobulin heavy chain genes in the mouse. In: Honjo T, Alt FW, Neuberger M, eds. Molecular Biology of B cells. London: Elsevier Academic Press; 2004: 19–26.
49. Clements T, Rice TF, Vamvakas G, *et al.* Update on transplacental transfer of igg subclasses: impact of maternal and fetal factors. *Front Immunol* 2020; **11**: 1920.
50. Simister NE. Placental transport of immunoglobulin G. *Vaccine* 2003; **21**: 3365–3369.
51. Van de Perre P. Transfer of antibody via mother's milk. *Vaccine* 2003; **21**: 3374–3376.
52. de Haas M, Thurik FF, Koelewijn JM, van der Schoot CE. Haemolytic disease of the fetus and newborn. *Vox Sang* 2015; **109**: 99–113.
53. Pegoraro V, Urbinati D, Visser GHA, *et al.* Hemolytic disease of the fetus and newborn due to Rh(D) incompatibility: a preventable disease that still produces significant morbidity and mortality in children. *PLoS One* 2020; **15**: e0235807.
54. Hirayama F. Recent advances in laboratory assays for nonhemolytic transfusion reactions. *Transfusion (Paris)* 2010; **50**: 252–263.
55. Vlaar AP, Juffermans NP. Transfusion-related acute lung injury: a clinical review. *Lancet* 2013; **382**: 984–994.
56. Gouy A, Excoffier L. Polygenic patterns of adaptive introgression in modern humans are mainly shaped by response to pathogens. *Mol Biol Evol* 2020; **37**: 1420–1433.
57. Greenbaum G, Getz WM, Rosenberg NA, Feldman MW, Hovers E, Kolodny O. Disease transmission and introgression can explain the long-lasting contact zone of modern humans and Neanderthals. *Nat Commun* 2019; **10**: 5003.
58. Simonti CN, Vernot B, Bastarache L, *et al.* The phenotypic legacy of admixture between modern humans and Neandertals. *Science* 2016; **351**: 737–741.
59. Dannemann M, Andres AM, Kelso J. Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human Toll-like receptors. *Am J Hum Genet* 2016; **98**: 22–33.
60. Mendez FL, Watkins JC, Hammer MF. A haplotype at STAT2 introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am J Hum Genet* 2012; **91**: 265–274.
61. Valenti L, Griffini S, Lamorte G, *et al.* Chromosome 3 cluster rs11385942 variant links complement activation with severe COVID-19. *J Autoimmun* 2021; **117**: 102595.
62. Zeberg H, Paabo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 2020; **587**: 610–612.
63. Abi-Rached L, Jobin MJ, Kulkarni S, *et al.* The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 2011; **334**: 89–94.
64. Malmstrom M, Matschiner M, Torresen OK, *et al.* Evolution of the immune system influences speciation rates in teleost fishes. *Nat Genet* 2016; **48**: 1204–1210.
65. Karvonen A, Seehausen O. The role of parasitism in adaptive radiations - when might parasites promote and when might they constrain ecological speciation? *Int J Ecol* 2012; **2012**: e280169.
66. Bengten E, Quiniou S, Hikima J, *et al.* Structure of the catfish IGH locus: analysis of the region including the single functional IGHM gene. *Immunogenetics* 2006; **58**: 831–844.
67. Jiang N, Weinstein JA, Penland L, White RA 3rd, Fisher DS, Quake SR. Determinism and stochasticity during maturation of the zebrafish antibody repertoire. *Proc Natl Acad Sci USA* 2011; **108**: 5348–5353.
68. Magadan S, Krasnov A, Hadi-Saljoqi S, *et al.* Standardized IMGT nomenclature of salmonidae IGH genes, the paradigm of Atlantic salmon and rainbow trout: from genomics to repertoires. *Front Immunol* 2019; **10**: 2541.
69. Magadan S, Mondot S, Palti Y, Gao G, Lefranc MP, Boudinot P. Genomic analysis of a second rainbow trout line (Arlee) leads to an extended description of the IGH VDJ gene repertoire. *Dev Comp Immunol* 2021; **118**: 103998.
70. Macholan M, Baird SJ, Munclinger P, Dufkova P, Bimova B, Pialek J. Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? *BMC Evol Biol* 2008; **8**: 271.
71. DeKosky BJ, Lungu OI, Park D, *et al.* Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proc Natl Acad Sci USA* 2016; **113**: E2636–E2645.
72. Bomblies K, Weigel D. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat Rev Genet* 2007; **8**: 382–393.
73. Chae E, Bomblies K, Kim ST, *et al.* Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell* 2014; **159**: 1341–1351.
74. Yang H, Wang JR, Didion JP, *et al.* Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet* 2011; **43**: 648–655.
75. Wang JR, de Villena FP, McMillan L. Comparative analysis and visualization of multiple collinear genomes. *BMC Bioinformatics* 2012; **13**(Suppl 3): S13.